The Zebrafish Progranulin Gene Family

by

Benoît Cadieux

A thesis submitted to the Faculty of Graduate Studies
and Research in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Department of Medicine
Division of Experimental Medicine
McGill University
Montréal, Québec, Canada

Your file Votre référence
ISBN: 0-612-98219-X
Our file Notre référence
ISBN: 0-612-98219-X

ABSTRACT

Granulins are small cysteine-rich peptide growth factors that, in mammals, are derived from a common glycoprotein precursor (progranulin) containing one half and seven non-identical tandemly repeated granulin domains. While there is evidence for only one progranulin gene in mammalian genomes, work presented in this thesis demonstrates that granulins form an extended gene family in teleost fish. Two zebrafish genes that constitute likely co-orthologues to mammalian progranulin, progranulin-a and progranulin-b, encode precursors that encode 10 and 9 copies of the granulin motif, respectively. Two additional genes in zebrafish, designated progranulin-1 and progranulin-2, each give rise to precursors consisting of one and one-half granulin-like repeats only. The progranulin-1 and progranulin-2 genes are organized in tandem, and possess exonic complementarity to a single non-coding RNA gene transcribed in the antisense orientation from the complementary DNA strand. A cDNA encoding a chimeric precursor consisting of the amino-terminal progranulin-1 followed by the carboxyl-terminal region of progranulin-2 was characterized and is likely generated through the mechanism of splicing in *trans* between the two primary transcripts for progranulin-1 and progranulin-2. Chromosomal assignment of the zebrafish progranulin genes indicates that progranulin-a, but not progranulin-b, is located on a chromosome that displays syntenic correspondence to mammalian progranulin. Cumulatively, the data suggest that an ancestral progranulin gene was duplicated at the base of the vertebrate radiation to generate precursors with distinct architectures and that one was lost in the lineage leading to tetrapods after the split between sarcopterygians and actinopterygians. Like their mammalian counterpart, the expression of the zebrafish progranulins is widespread in adult tissues. During development, the maternal expression of progranulin-a and progranulin-b parallels that of the murine counterpart, while progranulins-1 and -2 are not detectable in the zygote, reflecting a possible devolution of function. Preliminary functional studies involving knockdown of progranulin-a indicate pleiotropic roles for this gene in vertebrate development. This includes axis formation, growth and regionalization of the central nervous system, cell

proliferation and protection against anoikis. Progranulin-a is also implicated in hematopoiesis, the establishment and integrity of the vasculature, in organogenesis, and in the formation of structures derived from pharyngeal endoderm such as the jaw.

ABRÉGÉ

Les granulines sont des facteurs de croissance peptidiques riches en cystéine qui, chez les mammifères, sont derivés d'une glycoprotéine commune (progranuline) comportant une demie et sept copies non-identiques du motif de la granuline organisées en tandem. Alors que les génomes des mammifères ne possèdent qu'un seul gène pour la progranuline, les recherches présentées dans cette thèse démontrent que les granulines forment une famille de gènes chez les poissons téléostéens. Deux gènes chez le poisson-zèbre, progranuline-a et progranuline-b, donnent lieu à des précurseurs qui comportent dix et neuf copies du motif de la granuline, respectivement, et qui constituent une relation de co-orthologues à la progranuline trouvée chez les mammifères. Deux autres gènes chez le poisson-zèbre, progranuline-1 et progranuline-2, encodent des précurseurs comportant seulement une copie et demie du motif de la granuline. Ces derniers sont organizés en tandem, et possèdent des regions exoniques qui sont complémentaires à un seul gène ne possédant pas de potentiel pour encoder un polypeptide, et dont la transcription provient du brin antisens de l'ADN génomique. Une ADNc qui encode une chimère composée des exons 1 et 2 de la progranuline-1, suivit des exons 3 à 5 de la progranuline-2, et qui semble provenir du méchanisme de l'épissage en *trans*, a été caractérisée. Progranulina-a, et non progranuline-b, est localizé sur le chromosome 3 dans une region qui partage une correspondance par synténie avec l'orthologue trouvé chez les mammifères. Une analyse comparative de la localization chromosomale des différents gènes de la progranuline trouvés chez le poisson-zebre suggère que la copie ancêtre de la progranuline a été dupliquée suite à l'origine des vertébrés pour donner lieu à deux types de précurseurs possédant des architectures distinctes, et qu'une copie a été perdue par la suite dans la lignée menant aux tétrapodes. Comme leur homologue chez les mammifères, les progranulines du poisson-zèbre sont exprimés dans une variété de tissus adultes. Dans le développement embryonaire, progranuline-a et progranuline-b sont exprimés de façon maternelle, alors que progranuline-1 et progranuline-2 ne sont pas détectés chez le zygote, ce qui réflète une devolution possible parmis ces gènes. Des études préliminaries portant sur la caractérisation fonctionnelle de progranuline-a utilisant la technique du "knockdown" indiquent des effets pléiotropiques pour ce gène

dans le development des vertébrés, notamment dans la formation des axes, la croissance et la regionalization du système nerveux central, ainsi que la proliferation et la survie cellulaire. Progranuline-a est également impliquée dans l'hématopoïèse, la formation et l'intégrité du système vasculaire, ainsi que dans la formation des organes viscéraux et des structures dérivées de l'endoderme des pharynges dont la mâchoire.

ACKNOWLEDGEMENTS

This project was seen to its completion because of the help and encouragements of the following individuals, whom I wish to thank:

Dr. Hugh Bennett, my mentor, for being honest and for caring. Thank you for your friendship, fatherly advice, and for your patience.

David Baranowski and Babykumari Perumal for their help with cloning, Western blot analyses, in situ hybridizations, RT-PCRs, sharing the task of keeping the fish alive and well, and, most importantly, for their friendship and moral support. You're like family, guys.

Susan James for her invaluable help with all the protein-related work, for teaching me about peptide purification, and for being a friend.

Drs. Marie-Andrée Akimenko and Marc Ekker of the University of Ottawa for welcoming me as a member of their lab and for introducing me to the zebrafish as model organism. Thank you for your inspiration, enormous support, and advice. My thanks also go their lab members and friends (present and past): Gary and Steve (for sharing plasmids), Lucille Joly for performing the LN54 panel chromosomal mapping experiments, Qiaoming Long for help screening the PAC library, Linda Laforest for teaching me the whole-mount *in situ* hybridization technique. Mireille, Noel and Amanda's help and encouragements are also greatly appreciated.

Dr. Roger Palfree and Dr. Fernando Congote for sharing their wealth of scientific and technical knowledge, constructive criticism, and especially for their human qualities.

Dr. Andrew Bateman for his honest criticism, numerous scientific discussions, pointing out weaknesses that need work in a constructive way, and for sharing data prior to publication and equipment. I also thank his lab members (present and past) for their technical advice, ideas, and friendship: Amin Ismail, Colin Ong, Gula Sadvakassova, Rachael Daniel and Ziheng He.

Barbara Gould of her friendship throughout the duration of this project and for sharing her expertise with RNA.

Dominique Besso, and the late Dr. Jerry Price, for their patience, smile, and for doing a wonderful job at caring for every student in the Division of Experimental Medicine.

Dr. Jean-Jacques Lebrun and his lab members (Hector, Johanne, Effie) for providing ideas and for sharing their imaging equipment.

The Team at the Sheldon Biotechnology Centre: Lenora Naimark, Jing Hu, Marcos Difalco, Nancy Qian and Sonia Bounardjian for their constant help through the years.

Members of the Fraser labs for sharing equipment (microspcopy and PCR), and, especially, their friendship.

Dr. Pierre Drapeau and his lab members (Guy, Shabana, and Dr. Joe McDearmid) for providing lab space and equipment for microinjection experiments, as well as for their kindness and encouragements.

The help of the following thesis committee members is also greatly appreciated: Dr. Janet Henderson, Dr. Roger Palfree, Dr. Fernando Congote, Dr. Andrew Bateman, and my academic advisor, Dr. Geoffrey Hendy. Thank you for caring and for challenging a mind that can sometimes be too obstinate.

Thank you all for making me grow as an individual.
This endeavor is as much yours as it is mine

TABLE OF CONTENTS

ABBREVIATIONS

AS	antisense
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CNS	central nervous system
DEL	deep layer
DEPC	diethylpyrocarbonate
DF	degenerate forward
DNA	deoxyribonucleic acid
DR	degenerate reverse
EGF	epidermal growth factor
EST	expressed sequence tag
EVL	enveloping layer
HPLC	high performance liquid chromatography
ICM	intermediate cell mass
IGF	insulin-like growth factor
FGF	fibroblast growth factor
HOX	homeobox
IRES	internal ribosome entry site
LINE	long interspersed element
MALDI-TOF	matrix-assisted laser desorption ionization time-of-flight

MHC	major histocompatibility complex
MITES	miniature inverted-repeat transposable elements
MM	mis-match
MO	morpholino
mRNA	messenger ribonucleic acid
MyHC	myosin heavy chain
ORF	open reading frame
PBS	phosphate buffered saline
PCDGF	PC cell-derived growth factor
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNAi	RNA interference
RTK	receptor tyrosine kinase
SDS-PAGE	sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SNP	single nucleotide polymorphism
TGF	transforming growth factor
UTR	untranslated region
VEGF	vascular endothelial growth factor
YSL	yolk syncytial layer

measures

bp	base-pair	mg	milligram
cR	centirad	min	minute
Ci	Curie	mM	millimolar
Da	dalton	mya	million years ago
dpf	days post-fertilization	ng	nanogram
g	gram	nl	nanolitre
hpf	hours post-fertilization	nM	nanomolar
kb	kilobase	sec	second
kDa	kilodalton	µg	microgram
M	molar	µl	microlitre
mCi	millicurie	°C	degree celcius

LIST OF TABLES

LIST OF FIGURES

There are no beginnings and there are no ends; there is only the revolving interplay between the pairs of opposites.

This work is dedicated to my family and friends, as well as to present and past teachers for sharing their enthusiasm and for caring.

To Claire, my mother, for showing me the joy of learning.
To Marc, my brother, for his constant support, and for being understanding.
To Michel, for being a father.

To friends at LaCité Apartments and at LaCité Gym (André, Eric, Etzel, Fadi, Franz, Francois, Gordon, Lee, Rob, Vincent, Sam, Thomas, Nicole, Stella, Robert, and many others). Your support and friendship made a world of difference.

I am particularly grateful to the following individuals for their continued friendship, advice and support through the years: Tom Jungen, Jean-Francois Lemieux, Justin Gauvin, Frédéric D'Amours, Nicholas Bombardier, Nathalie Brière, Patrick McDonnell, Fabienne Ranely-Dépré, Sandra-Michèle Bisson, Jonathan Shykofsky, and numerous others.

Philippe Calvé, for encouragements through the years, and for never losing faith. Couldn't have done it without you.

Luong Luu for her advice, honest criticism, and sense of humour. Thank you for your genuine love.

1. General Introduction

1.1 Granulins

1.1.1 General overview of granulins – Cellular proliferation and differentiation are processes normally kept under tight regulation by the interplay of a variety of growth factors and cell-cell (Fagotto and Gumbiner, 1996) or cell-matrix interactions (Ashkenas *et al.*, 1996). Aberrant stimulation by growth factors is often a hallmark of neoplastic cell growth, and may lead to abnormal embryogenesis and tissue development (Dickson and Salomon, 1998). As such, the characterization of candidate mitogens, and the elucidation of their modes of action, is critical to understanding various cellular processes ranging from the development of cell lineages to wound repair, as well as tissue aging and tumorigenesis. Granulins, also known as epithelins, are growth-modulating factors originally implicated in the control of epithelial cell proliferation *in vitro* (reviewed in Bateman and Bennett, 1998). Individual granulin peptides have an approximate molecular weight of 6 kDa, containing 12 cysteines arranged in a characteristic motif: $X_{2\text{-}3}\underline{C}X_{5\text{-}6}\underline{C}X_5\underline{CC}X_8\underline{CC}X_6\underline{CC}X_5\underline{CC}X_4\underline{C}X_{5\text{-}6}\underline{C}X_2$ (Bateman and Bennett, 1998). Carp granulin-1 has been used as a prototypic form to solve the NMR solution structure of this peptide family, and has been shown to consist of a parallel stack of β-hairpins held together by six disulfide bonds (Hrabal *et al.*, 1996).

The known roles and activities of granulins and of their common precursor, called progranulin, PC-cell-derived growth factor (PCDGF), or acrogranin, have expanded greatly since their discovery nearly 15 years ago. The current knowledge about progranulin biology has recently been reviewed (He and Bateman, 2003; Ong and Bateman, 2003; Serrero, 2003), and will not be covered at length here. Rather, in order to provide some background information, aspects concerning the structural organization of the mammalian granulins precursor, the involvement of this gene and its protein complement in the process of tumorigenesis, as well as its mode of action, are briefly overviewed below.

Granulins were first purified from human neutrophils and rat bone marrow (Bateman *et al.*, 1990), extracts of rat kidney (Shoyab *et al.*, 1990), and the hematopoietic organs of teleost fish (Belcourt *et al.*, 1993). In mammals, these peptides arise through the proteolytic processing of a common glycoprotein precursor, consisting of a signal sequence followed one amino-terminal half and seven tandemly repeated copies of the granulin consensus motif (Bhandari *et al.*, 1992; Plowman *et al.*, 1992; Bhandari *et al.*, 1993; Baba *et al.*, 1993a). The granulins precursor, progranulin, has been given different names depending on the context in which it has been identified. Secretion and action of intact progranulin (Baba *et al.*, 1993a; Zhou *et al.*, 1993; He and Bateman, 1999), as well as of its constituent granulin peptides (Bateman *et al.*, 1990; Shoyab *et al.*, 1990), have been reported. This implies that progranulin processing occurs differently in various tissues or cell types, and that individual peptides may have distinct biological activities. In agreement with this notion is the observation that granulin-A/epithelin-1 and granulin-B/epithelin-2 display pleiotropic growth factor-like activities on various cell lines that can apparently oppose each other. For example, while the growth of keratinocytes in culture is promoted by addition of granulin-A/epithelin-1, this stimulatory activity is antagonized by addition of granulin-B/epithelin-2 (Shoyab *et al.*, 1990).

Evidence has been accumulating for the role of progranulin itself in the process of tumorigenesis. Progranulin was purified as a candidate autocrine growth factor secreted from a highly tumorigenic murine teratoma cell line (Zhou *et al.*, 1993), where it is referred to as PC-cell-derived growth factor (PCDGF), and recently shown to modulate the growth of breast cancer cell lines *in vitro* (Lu and Serrero, 1999). Subsequently, it was demonstrated that inhibition of expression of PCDGF in the MDA-MB-468 human breast carcinoma cell line inhibited colony formation *in vitro* by up to 80%, while inhibiting tumour incidence and tumour weight by 90% following subcutaneous injection into athymic mice (Lu and Serrero, 2000). A corollary to these observations was provided by studies of SW13 epithelial cells engineered to over-express progranulin. These cells acquired phenotypic characteristics of transformed cells including evidence of anchorage-independent growth *in vitro*, and injection of

these cells in nude mice promoted tumour growth (He and Bateman, 1999). Together, these findings strongly suggest that the intrinsic level of expression of progranulin facilitates cellular proliferation and the propensity of tumour formation of epithelial cells (He and Bateman, 1999). Further support for these conclusions comes from the upregulation of progranulin gene expression in human neuroglial and other brain cancers relative to normal tissue (Liau *et al.*, 2000, Markert *et al.*, 2001).

In addition to its role in cell growth, progranulin expression is induced in skin fibroblasts and in endothelia subsequent to injury, suggesting a role for this factor in the wound healing response (He *et al.*, 2003). Exogenous addition of this factor at the wound site was shown to stimulate the appearance of inflammatory cells (macrophages and neutrophils), fibroblasts, as well as stimulation of angiogenesis (He *et al.*, 2003), confirming its involvement in this process. In addition, processing of progranulin by elastase, or its preservation in intact form through interaction with secretory leukocyte protease inhibitor, appear to lie at the interface between innate immunity and wound healing, respectively (Zhu *et al.*, 2002).

The mode of action of granulins has not yet been fully determined. Candidate high-affinity receptors for the peptides have been identified by cross-linking studies using the MDA-MB-486 breast tumour cell line (Culouscou *et al.*, 1993), and, for the precursor, on the CCL64 mink lung epithelial cells, murine PC cells, and mouse embryo fibroblast 3T3 cells (Xia and Serrero, 1998). Progranulin has been shown to be the only factor capable of stimulating the growth of 3T3 mouse embryo fibroblasts in which the IGF-I receptor has been deleted (R^- cells) (Xu *et al.*, 1998). Normally, these cells exhibit a cell cycle block for which no other known growth factor can bypass the requirement of IGF-I receptor signaling. This progranulin activity seems to be mediated by the stimulation of the mitogen-activated protein kinase and the phosphatidylinositol 3-kinase pathways in R^- cells, pathways known to be sufficient for IGF-I mediated mitogenesis (Zanocco-Marani *et al.*, 1999). Progranulin activity was also shown to protect cells against anoikis and to stimulate migration through the activation of focal adhesion kinase (He *et al.*, 2002). Interestingly, the protein core of a

heparan sulfate proteoglycan, namely perlecan, has been shown to interact with progranulin, and was suggested to modulate its growth-promoting activity (Gonzalez *et al.*, 2003).

1.1.2 Teleost granulins – Multiple granulin peptides have been isolated and characterized from extracts of the hematopoietic tissues of the carp (*Cyprinus carpio*) (Belcourt *et al.*, 1993). The major forms of granulin peptides found in spleen and head kidney differ in their relative abundance (Belcourt *et al.*, 1993), which is counter-intuitive to the notion that granulins all arise from a common precursor encoded by a single gene, as is the case in mammals. This suggests that granulins found in teleost fish may have different biosynthetic origins. Prior to the research presented in this thesis it was unclear whether these observations are a consequence of the presence of multiple copies of a similar gene structure, or if granulins arise from a multi-gene family. The data presented demonstrate that in teleost fish, granulins arise as members of an extended gene family of at least two architectures, and provide evidence that these structures arose as a result of a primordial chromosomal duplication event which was followed by the emergence of structural diversity.

1.2 Evolutionary origins of granulins

There is considerable interest in delineating the distribution of gene families and their constituent members across phyla through reading the evolution of the genomic text. Tracing back the origins of a gene family in extant species is now facilitated by the extensive deposition of expressed sequence tags (ESTs) from numerous model organisms at the NCBI. In turn, gene mapping and genome sequencing enable the establishment of paralogous and orthologous relationships between gene homologues found in different organisms. Aside from providing insight into the evolution of genomes and of gene families, this information permits a more direct comparison of the genetic and biochemical function of an equivalent gene product in different model organisms.

While the progranulin gene appears to be an exception, the majority of growth factors and cytokines - such as Wnts, BMPs, and interleukins - form extended gene families in mammals. In fact, data mining of genomic sequences and ESTs indicate that, whenever present, granulins are derived from a single gene in diploid representatives of both the protostome and deuterostome clades (**Figure 1**). The presence of a progranulin gene in the nematode worm *C. elegans*, but not in yeast or fungus, suggests that granulins evolved in the early stages of metazoan life. This is not surprising since unicellular organisms are known to express a limited number of secreted signaling factors involved in cell-cell communication. The antiquity of the granulin motif is further illustrated by its existence in some plants.

It is therefore remarkable that despite its ancient origin, the progranulin gene is represented only once in mammalian genomes, since genes present in invertebrates have usually undergone an expansion in terms of number at the base of the vertebrate radiation (Sidow, 1996; Pebusque *et al.*, 1998; Skrabanek and Wolfe, 1998). As a means of explaining this discrepancy, it could be argued that diversity of progranulin action is subserved by its unusual organization of tandemly repeated, non-identical granulin repeats, which potentially carries sufficient plasticity for generating biochemical diversity through a form of combinatorial biochemistry in combination with regulated processing. This would exclude the need for the generation and retention of duplicated gene pairs. Arguing against the action of this selective pressure is the presence of an extended gene family encoding the epidermal growth factor (EGF) motif. Commonly classified as the ErbB ligands, these comprise the EGF, TGFα, amphiregulin (AR), heparin-binding EGF-like growth factor (HB-EGF), betacellulin (β-CEL), epiregulin (Epi) and NDF/neuregulins α and β (Ben-Baruch *et al.*, 1998). Indeed, the canonical precursor for EGF, like progranulin, is made up of tandem repeats of non-identical EGF-like repeats (Gray *et al.*, 1983; Scott *et al.*, 1983; Derynek *et al.*, 1984; Bell *et al.*, 1986). Further, the duplication and retention of gene duplicates for EGF is manifest in arthropods, with three genes in *Drosophila melanogaster*: *spitz*, *vein*, and the TGF alpha-like ligand *gurken* (Stevens, 1998;Yarnitzky *et al.*, 1998).

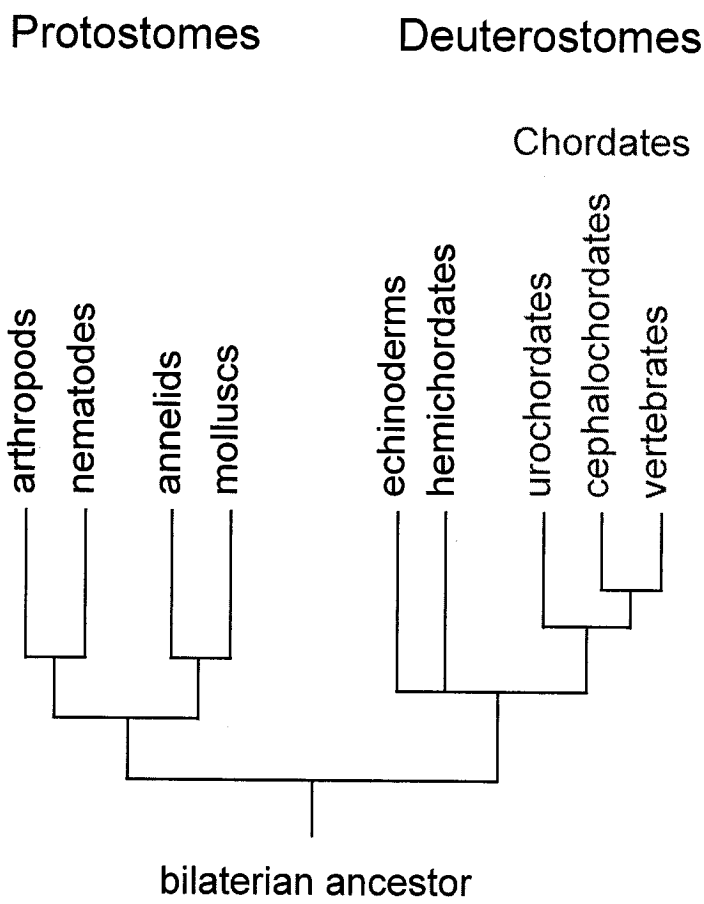


Figure 1. **Phylogeny of bilaterian animals**. Chordates, like echinoderms and hemichordates, are deuterostomes, a division of animals that are distinguished from other animals (protostomes) by the development of the embryonic blastopore into the anus rather than the mouth. Chordate features include a notochord and muscular tail, a sensory vesicle (brain) and a dorsal nerve chord. (Adapted from Dehal *et al.*, 2002)

Progranulin is not the only gene to occur as a singleton in the genome of mammals. Other examples include housekeeping genes that are ubiquitously expressed including manganese superoxide dismutase and the large subunits of RNA polymerases (Sidow, 1996). There are also genes that possess no clear invertebrate counterpart. For example, the hepatocyte growth factor (HGF), a potent stimulator of the growth, movement, and differentiation of epithelia and endothelia, together with macrophage stimulating protein (MSP), an effector of macrophage chemotaxis and phagocytosis, constitute a small growth factor family that share structural homology with the blood proteinase precursor plasminogen. These have evolved from a common ancestor consisting of an amino-terminal domain corresponding to plasminogen activation peptide, three kringle domains, followed by a serine proteinase domain. The absence of an equivalent composite gene architecture in invertebrates argues for its recent appearance within the vertebrate lineage.

Of course, it cannot be conclusively ruled out that there existed duplicated progranulin genes prior to, or at the base of, the vertebrate radiation with the subsequent loss of one copy within the lineage leading to mammals. Data from the genome of chordate ancestors is so far not indicative of this scenario: C*iona intestinalis*, a representative of the urochordate phylum whose genome offers insight into early chordate biology, appears to encode a single progranulin gene of similar architecture to that found in mammals (*Dehal et al.*, 2002, and http://ghost.zool.kyoto-u.ac.jpindexrl.html). Hence, unless demonstrated otherwise in the genome of a sister group to vertebrates such as that of the cephalochordate *amphioxus*, or in the agnathan *sea lamprey,* the sole occurrence trend favours the conclusion that the progranulin gene has never been duplicated, at least not outside the ray-finned (actinopterygian) fish lineage.

1.2.1 Emergence of the vertebrate genome: *progranulin* stands alone – From the viewpoint of genome evolution the absence of additional progranulin genes is quite important since the general dogma dictates that while protochordate ancestors present limited or no gene families, the emergence of gene families correlates with the radiation of vertebrata (Holland *et al.,* 1994; McLysaght *et al.*, 2002; Durand, 2002). This observation led Sushumo Ohno to propose that the excess genetic redundancy generated though gene duplications provided the necessary flexibility for morphological innovations (Ohno, 1970). Hence, single gene and entire genome duplication are considered to be two of the mechanisms responsible for the increased complexity of organisms in the evolution of life (Ohta, 1989). Simply stated, gene duplication promotes the formation of gene families and the functional diversification of genes, a process that tends to be translated into further genomic and phenotypic complexity (Iwabe *et al.*, 1996). This inference is derived largely from studies involving the phyletic distribution of Hox and paraHox gene clusters (Holland and Garcia-Fernandez, 1996; Bailey *et al.* 1997; see also review by Ferrier and Holland, 2001) as well as of other "developmentally important" genes (Wada *et al.* 1998) which provide a framework for evaluating both gene/genome duplications through evolution and their relationship with (i.e. whether causally related to) morphological complexity.

1.2.2 Vertebrate genome duplications – Individual genes (e.g. Hedgehog, Notch, MyoD, and integrin associated cytoplasmic proteins such as paxillin and FAK) or clusters (e.g. Hox, MHC), found as singletons in invertebrates and chordates, are often outnumbered by up to 4 copies in the vertebrate genome (**TABLE 1**; Comings, 1972; Atchley e*t al.*, 1994; Holland *et al.*, 1994; Spring, 1997; Rubin *et al.*, 2000). This has led to the creation of the one-two-four (1-2-4) rule to explain the evolution of the vertebrate genome from its deuterostome ancestors (Meyer and Schartl, 1999). Although not proven (Skrabanek and Wolfe, 1998), this hypothesis assumes that the single genome of an ancestral deuterostome underwent 2 consecutive rounds of duplication (2R), and derives support from the observation that genes from a given gene family are often part of linked clusters that maintain gene order on different chromosomes. Referred to as synteny, the location of 2 or more genes within a linkage

group is often conserved across large evolutionary distances, including humans, mice and fish (Postlethwait *et al.*, 2000). Syntenic conservation of groups or clusters of related genes cannot be easily reconciled with multiple independent individual gene duplication events. Whether syntenic correspondences arise through partial or whole genome duplications cannot be determined at present. This favours the phylogenetic approach of mapping gene families onto the deuterostome tree, rather than relying on counting the number of paralogues, the latter approach being unreliable because of unforeseeable independent processes of gene gain or loss.

In the case of genome duplications, while genes initially retain their position relative to one another, they now find themselves on the original chromosome as well as on its complement. This is in contrast to individual gene and segmental duplications that initially occur in tandem. On a global scale, genome duplication is thought to be brought about by polyploidization. In the event of tetraploidization or polyploidization, subsequent gradual reversion to the diploid state is coincident with extensive gene loss. Duplicated chromosomes, if retained, diverge in time to acquire new features enabling them to be fixed in the organism as non-identical but closely related chromosomes under a diploid state. Tetraploidization can either result from endoduplication (autopolyploidy) or involve hybridization to a closely related species (allopolyploidy). It is only in the former that the timing of the tetraploidization event can be estimated through phylogenetic analysis. The actual mechanism of tetraploidization and re-diploidization is poorly understood and beyond the scope of the present discussion. It should be noted, however, that aspects of the conceptual framework used for explaining the fate and retention of duplicate gene pairs (see later) could possibly be extrapolated to explain the appearance of duplications on a genome-wide scale. For the interested reader, an excellent review describing the current knowledge concerning tetraploidization events and their predicted consequences has recently been published (Wolfe, 2001).

Drosophila	Human	reference
insulin receptor (IR)	IR	Baserga, 1998
	IGF-1R	Wang and Gu, 2000
	IR-RELATED RECEPTOR (IRRR)	
aldolase	ALDO A	Wang and Gu, 2000
	ALDO B	
	ALDO C	
cyclin D	CYCLINS D1, D2, D3	Wang and Gu, 2000
Hedgehog (Hh)	SONIC Hh	Rubin *et al.*, 2000
	INDIAN Hh	
	Hh	
activin β	ACTIVIN β1–4	Rubin *et al.*, 2000 Wang and Gu, 2000
janus tyrosine kinase (JAK)	JAK1-4	Wang and Gu, 2000
EGF receptor	EGFR	Ben-Baruch *et al.*, 1998
	ERBB2	Wang and Gu, 2000
	ERBB3	
	ERBB4	
Notch	NOTCH 1-4	Rubin *et al.*, 2000
Hox cluster	HOX A-D	Wang and Gu, 2000
ParaHox cluster	ParaHOX A-D	Spring, 2002

TABLE 1. **Selected examples of invertebrate genes or clusters that are represented by up to four paralogues in the human genome.** Note that according to phylogenetic analyses, only the ACTIVIN , EGFR and JAK families conform to the predicted (AB) (CD) tree topology, consistent with the 2R genome duplication model (Wang and Gu, 2000). The HOX cluster was excluded from this analysis.

1.2.3 Evidence for 2R – It is advisable to regard the 1-2-4 hypothesis in genome evolution as a generality rather than being rigidly deterministic; it is well known that genes can be lost. It can be argued that looking at individual genes, rather than clusters, can present a biased view. In contrast, it is unknown whether the apparent clustering of genes provides a selective advantage to the organism, thus supporting their independent origin (Hughes, 1998). On what basis then can 2R be evaluated? The premise of two rounds of genome duplication is that it predicts a certain symmetry in gene topology in a given gene family tree, termed the (AB) (CD) model (**Figure 2**; Durand, 2003), which is not always supported through phylogenetic analyses (Hughes, 1999; Wang and Gu, 2000). The latter analyses should further reveal a concordant age-distribution for the birth of various gene families in support for autopolyploidy.

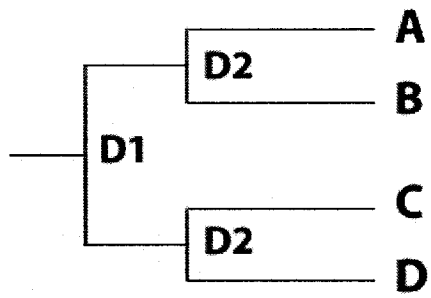


Figure 2. **The (AB) (CD) model of phylogenetic tree distribution.** According to the 2R hypothesis, individual genes generated via 2 consecutive rounds of genome duplication should display a characteristic topology through phylogenetic analysis.

Notwithstanding the contribution of molecular phylogenetics, difficulties are commonly encountered when attempts are made to determine whether homologous gene structures constitute paralogues that arose through either 2R or some other mechanism. For example, the tissue-type plasminogen activator (tPA), urokinase, blood coagulation factor XII (FXII), and hepatocyte growth factor activator (HGFA) family of serine proteases clearly defines a gene family based on conservation of coding sequence and exonic/intronic organization (Miyazawa *et al.*, 1998), but unequivocal support for or against two rounds of genome duplication cannot be provided unless their chromosomal localisation is resolved and analysed. Complicating matters is the observation that the serine proteases/trypsin gene family has presumably been further expanded through a more recent segmental duplication within the human genome (Bailey *et al.*, 2002). Another example in favour of this argument comes from the evolutionary analysis of the insulin/insulin-like growth factor (IGF) gene family. The discovery of a hybrid cDNA from the cephalochordate *Amphioxus* provides strong evidence that mammalian insulin and IGFs are derived from a common ancestral structure (Chan *et al.*, 1990). Further, the presence of three paralogues in humans (insulin, IGF-1, IGF-2) lends support to 2R, with the possibility that a fourth member was lost. However, gene mapping indicates that insulin and IGF-2 are closely linked on mouse chromosome 7, implicating individual gene duplication as the most likely mechanism for generating either insulin or IGF-2. A similar mechanism mediated by unequal crossing-over during meiosis has presumably given rise to other gene families for which additional members are found in tandem, such as rRNA genes and the globins.

Other cases are not easily reconcilable with the 1-2-4 hypothesis. This includes extended gene families that do not have an exact invertebrate counterpart, such as the TGF-β, α-inhibin, and Mullerian inhibiting substance (MIS) subfamilies (Rubin *et al.*, 2000). At the other extreme lies the extraordinary number of genes encoding the fibroblast growth factors (FGFs), with 23 representatives in the human genome.

At a first glance, it seems reasonable to infer that single and whole-genome duplications have both participated in the shaping of vertebrate genomes through the creation of gene families. Whole genome duplications should be rare events that are followed by extensive gene loss, whereas duplications and deletions on a smaller scale occur continuously (Gu *et al.*, 2002).

1.2.3.1 The revival of an hypothesis – Without the complementary analysis of chromosomal localisation, using the sole argument of gene number for establishing whether a given gene family provides support for, or against, two rounds of whole genome duplication in the vertebrate lineage can be misleading. Nevertheless, whether the fourfold complexity of the human genome was generated through individual gene or whole-genome duplications could not be conclusively determined through the analysis of the human genome alone (Lander *et al.*, 2001; Venter *et al.*, 2001). However, information derived from the comparative analysis of vertebrate gene maps has generated considerable support for 2R. The organization of the four Hox gene cluster paralogues, along with many animal-specific duplicated genes such as those encoding receptor tyrosine kinases (RTK) of the EGF receptor family or fibrillar collagens constitute benchmark observations favouring 2R (**Figure 3;** see commentary by Spring, 2002). The same has been observed for the paraHox clusters and their neighbours, namely the RTKs for platelet-derived growth factor (PDGF) and vascular endothelial growth factor (VEGF), and basement-membrane collagens (Miyata and Suga, 2001; Spring *et al.*, 2002). Other recent evidence for at least one round of genome duplication has come from the *en bloc* conservation of the four human MHC paralogous clusters and surrounding genes when compared to the corresponding region within the *Amphioxus* genome (Spring, 1997; Abi-Rached *et al.*, 2002).

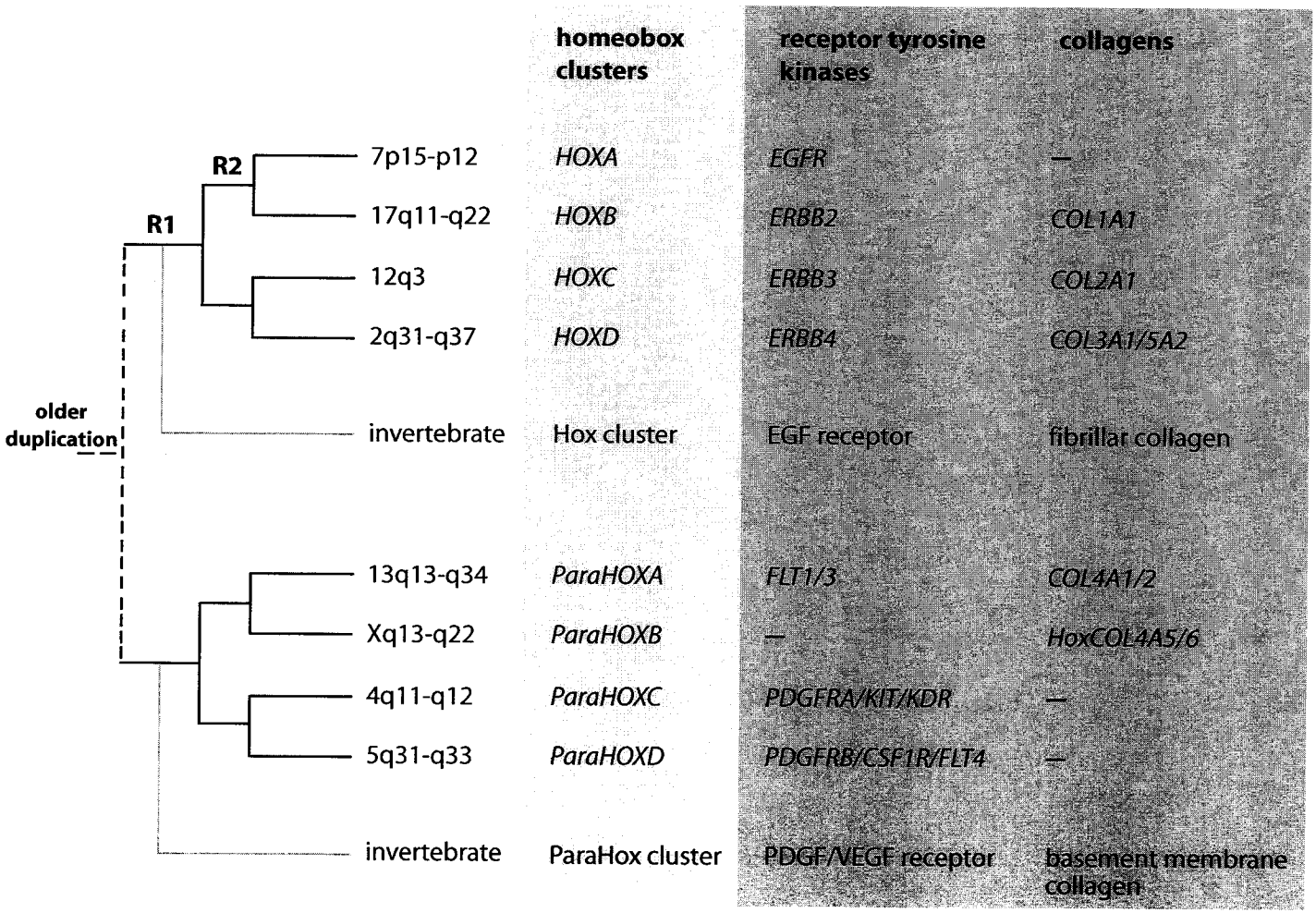


Figure 3. **Evidence for two rounds of genome duplications**. The respective *Hox* and *ParaHox* clusters of homeobox genes, with their linked receptor tyrosine kinases and collagens, are present as single copies in the genomes of invertebrates, while they number 4 paralogous groups in mammalian genomes, providing support for two rounds of polyploidization within the vertebrate lineage. (Adapted from Spring, 2002).

1.2.3.2 Timing of 2R – Although both large and small-scale duplications are required to explain the age-distribution of human gene families, the molecular phylogenetic analysis of large paralogous regions is consistent with duplications taking place within the vertebrate lineage (**Figure 4 and 5**; Gu *et al.*, 2002). Individual members of vertebrate gene families, compared to their invertebrate homologue used as an outgroup for the definition of an evolutionary timescale, led to the proposal that the fourfold increase in gene number predates the radiation of jawed vertebrates (**Figure 4**; Wang and Gu, 2000). More precisely, the first genome duplication probably occurred in a common ancestor of all agnathans and gnathostomes after its divergence from cephalochordates, ~ 594 mya (million years ago). The second round is presumed to have occurred ~ 488 mya, within the lineage leading to jawed vertebrates after the jawless line diverged, presumably before the split between cartilaginous and bony fish.

1.2.4 Vertebrate genes in early chordates – Gaining perspective on the early events in the evolution of vertebrate developmental genes is a major goal of modern biology. In particular, the discipline of evolution-development (evo-devo) aims at elucidating the reasons behind the presence of genes in early chordates whose function in vertebrates is associated with the formation of phylum-specific features, such as neural crest cells and the placodes. Does the expression of these genes presage the origin of vertebrate tissues? For example, *Ciona intestinalis* appears to contain three homologues of Noelin-1, a vertebrate olfactomedin family member whose protein product provides the neural tube with competence to generate neural crest cells (Bronner-Fraser, 2002). Other genes important for neural crest development, Hand1 and Hand2, also appear to have homologues in *Ciona* (Cañestro *et al.*, 2003). Also present in this urochordate are two homologues of the vertebrate Prox1 marker gene for lens, otic, ganglia and olfactory placodes (Rodiguez-Niedenfuhr *et al.*, 1999). The expression and activity of progranulin in the early stages of murine embryogenesis raises the question of the role of its counterpart in chordate ancestors, including *Ciona intestinalis* and *Amphioxus*.

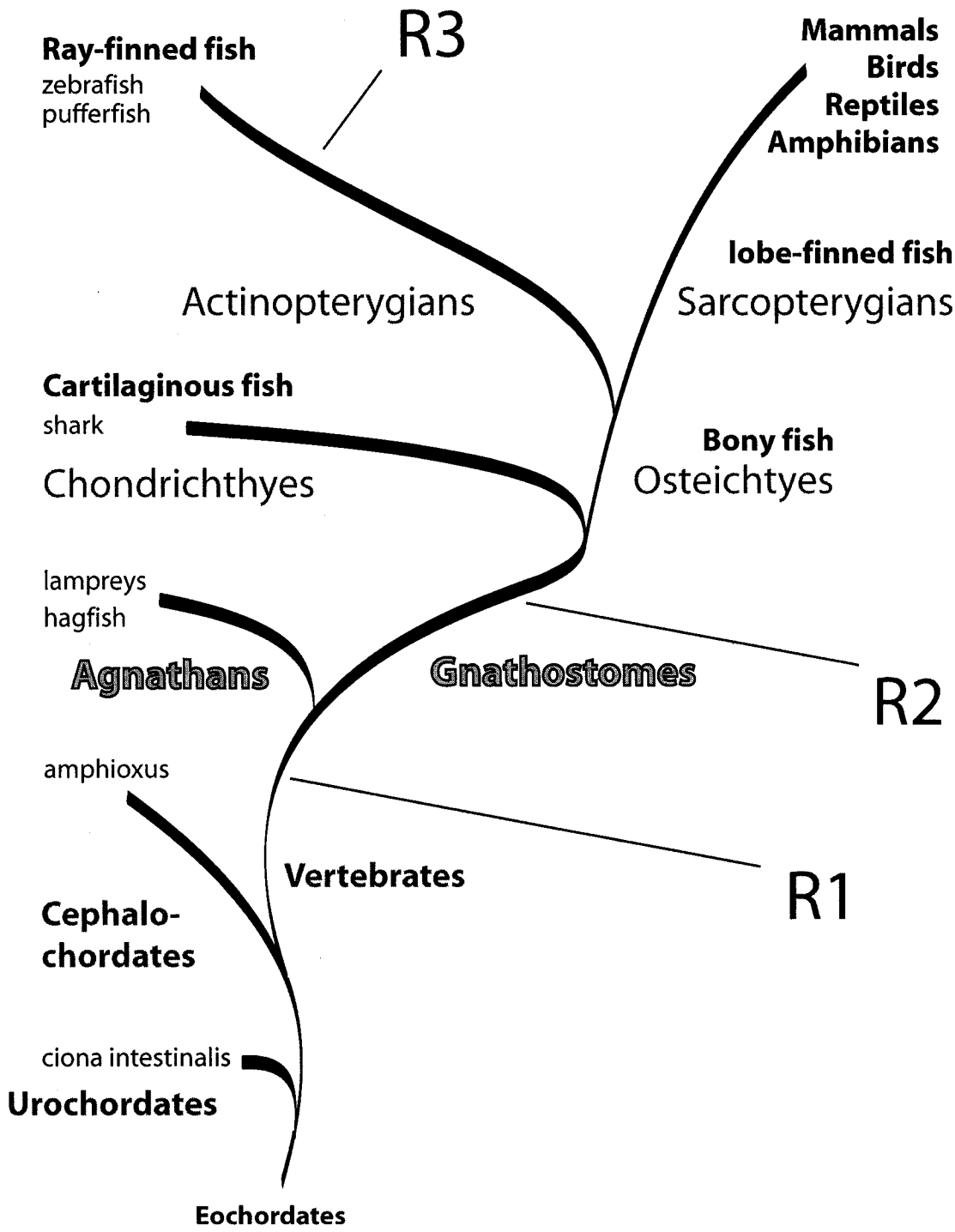


Figure 4. **Graphic representation of the vertebrate radiation from protochordate ancestors.** The estimated times for vertebrate-specific polyploidization events (R1 = 594 mya, R2 = 488 mya, R3= >100 mya) are derived from phylogenetic analyses. (Dates were taken from Wang and Gu, 2000)

1.2.4.1 Progranulin and vertebrate innovations – It is likely that at least one tetraploidization event took place in a vertebrate ancestor, culminating in the creation of gene families, thereby providing the necessary substrate for nature's experiments. It would appear then that progranulin did not contribute directly to the formation of vertebrate features through the mechanism proposed by Ohno, or that additional copies brought about by genome duplications were lost at some point in time within the vertebrate radiation. However, it should be remembered that novelties in morphology and speciation do not rely solely on extensive gene duplication, but also on the reiterative and differential use of already existing genetic information. The evolvability of progranulin or of any other gene product may have been used as a powerful source of functional novelty. In fact, developmental evolution can result from regulatory mutations of nonduplicated genes (Palopoli and Patel, 1996). For instance the mammalian pituitary hormone prolactin is a primary regulator of milk protein synthesis in mammals while playing an important osmoregulatory role in fresh water fish by preventing ion loss and promoting water uptake through the gill epithelium (Manzon, 2002). Interestingly, a role in osmoregulation may have been retained during the later stages of embryogenesis in vertebrates (Manzon, 2002).

Thus, arguments provided by evolutionary genomics cannot discount the possible participation of progranulin gene activity in the establishment of vertebrate innovations. Such features include the neural crest cell population and derived paired sensory organs, a highly segmented brain and complex nervous system, ectodermal placodes, cartilage and mineralized tissues, the elaborate cardiovascular system, and adaptive immunity (Kasahara *et al.*, 1997; Shimeld and Holland, 2000; Graham, 2000).

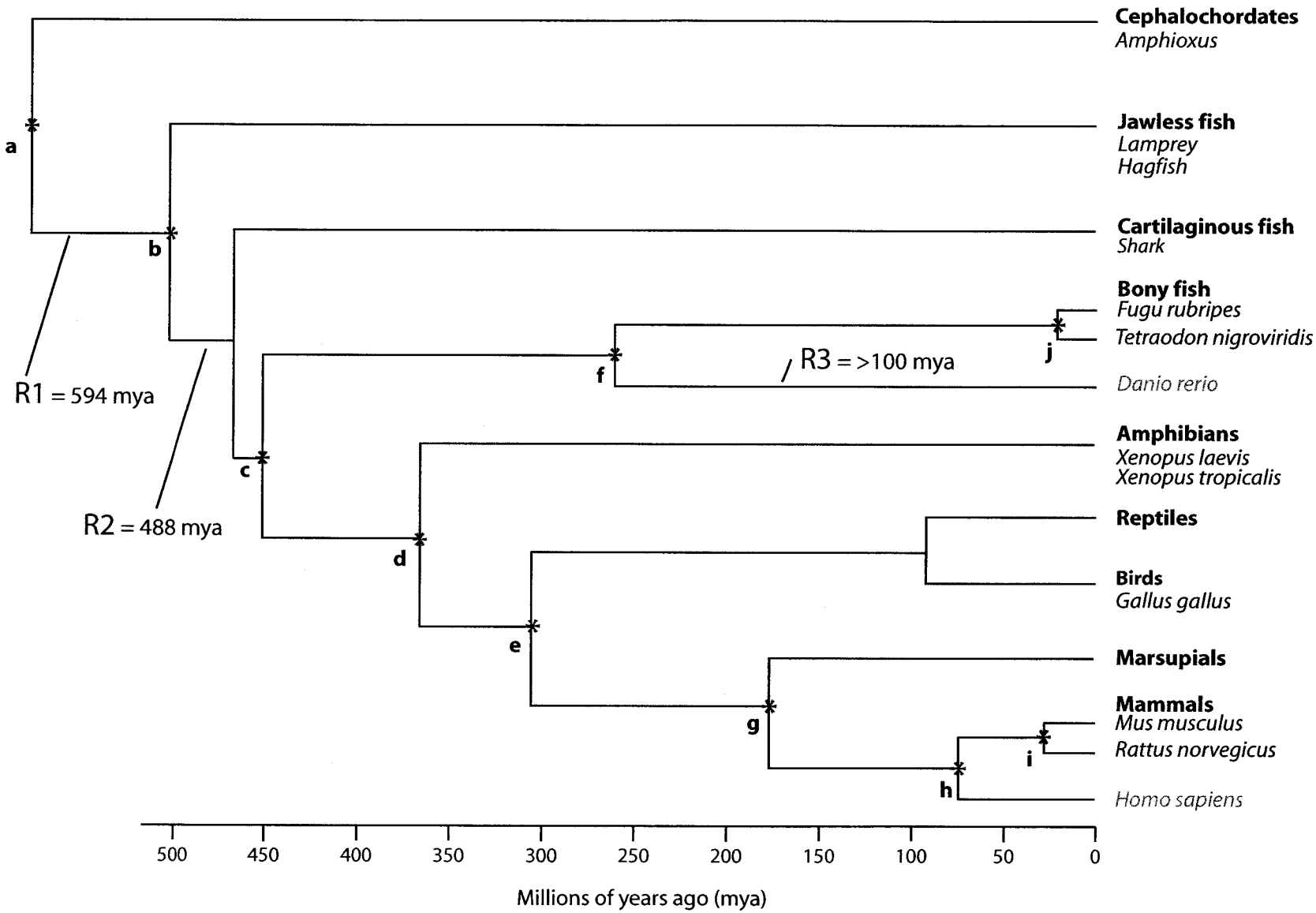

Cephalochordates
Amphioxus
Jawless fish
Lamprey
Hagfish
Cartilaginous fish
Shark
Bony fish
Fugu rubripes
Tetraodon nigroviridis
Danio rerio
Amphibians
Xenopus laevis
Xenopus tropicalis
Reptiles
Birds
Gallus gallus
Marsupials
Mammals
Mus musculus
Rattus norvegicus
Homo sapiens
a
b
c
d
e
f
g
h
i
j
R1 = 594 mya
R2 = 488 mya
R3 = >100 mya
500
450
400
350
300
250
200
150
100
50
0
Millions of years ago (mya)

Figure 5. **Evolutionary relationships and the timeline of the vertebrate radiation**. Current estimates of speciation events leading to the emergence of the various vertebrate clades from a common ancestor are based on the following: a: ~600 mya (Sidow, 1996); b: ~500 mya (Sidow, 1996); c: 415-485 mya (Kumar and Hedges, 1998); d: 356- 374 mya (Kumar and Hedges, 1998); e: 310 mya (Kumar and Hedges, 1998); f: 256-312 mya (Kumazawa *et al.*, 1999); g: 161-185 (Kumar and Hedges, 1998); h: 64-74 mya (Eizirik *et al.*, 2001); i: 14-41 mya (Jacobs and Pilbeam, 1980); j: 18-30 mya (Crnogorac-Jurcevic *et al.*, 1997). The emergence of man is estimated to have occurred 6 mya (Goodman *et al.*, 1998).

1.2.5 Carp granulin peptides: a conundrum – The common carp (*Cyprinus carpio*) has been shown to possess several granulin peptides, some of which are much less closely related to the mammalian sequences than others (Belcourt and Bennett, unpublished observations). This observation hinted at the possibility that there existed more genes encoding granulin peptides in bony fish. Differences in the relative abundance of granulins when extracted from various organs (granulin-1, -2 and -3), provided further evidence that these originated from distinct gene products (Belcourt *et al.*, 1993). In an era predating the undertaking of extensive characterization of ESTs and of the sequencing of mammalian genomes, this deduction led to the proposal that if the carp did indeed contain additional blueprints encoding granulins, then this finding was likely going to be extended to higher vertebrates. Similar approaches using fish were proven successful for the identification of novel members of extended gene families (Vaughan *et al.*, 1995). More importantly, the original characterization of regulatory peptides thought to be exclusive to teleost fish, but later demonstrated to possess orthologues with important roles in human physiology, has greatly contributed to the understanding of the evolution and biological roles of certain peptides (reviewed in Conlon, 2000). For example, urotensin-II and urotensin-I/urocortin, first discovered in the urophysis of fish, find orthologues in mammals with roles in regulating the cardiovascular system (Ames *et al.*, 1999; Nothacker *et al.*, 1999) and the stress response (Vale *et al.*, 1981; Fryer *et al.*, 1983; Vaughan *et al.* 1995; Spina *et al.*, 1996), respectively. The corpuscles of Stannius were found to be the site of origin of a major calcium and phosphate regulating hormone in teleosts, called stanniocalcin (Wagner *et al.*, 1991; Lu *et al.*, 1994), while cDNAs encoding glucagon-like peptides were first isolated from the Brockmann bodies of fish (Lund *et al.*, 1981; Lund *et al.*, 1982).

The possibility of uncovering additional granulin genes could have been dismissed easily as being a by-product of the tetraploid state of the carp genome (Larhammar and Risinger, 1994). Fortunately, the rise of the zebrafish, a closely related teleost also of the cypriniform clade but with a diploid chromosome set, heralded the possibility of avoiding this potential problem. Further, the zebrafish would allow for the investigation of the role(s) that granulins play in vertebrate development.

1.2.6 More genes: the "simple" fish complicates matters – Tetrapods, or descendents of the sarcopterygian lineage, which include coelacanth, lungfish, and all land vertebrates such as amphibians, reptiles, birds and mammals, have 4 *Hox* gene clusters (termed *Hox*A-D) in their genome (**Figure 4**). Convincing support for the notion that actinopterygians (ray-finned fish) possess more genes than sarcopterygians (lobe-finned fish) comes from work also concerning Hox gene clusters (Stellwag, 1999). It has been found that the zebrafish contains 7 such clusters, indicative of additional gene/genome duplications (Amores *et al.*, 1998). Based on sequence conservation and linkage analysis, mammalian Hox groups A, B, and C are represented twice in the zebrafish genome, whereas the Hox D cluster is represented by a single orthologue. Since that seminal finding, several papers provided evidence for the retention of duplicated zebrafish genes, now referred to as the "a" and "b" copies, each corresponding to a single mammalian orthologue (**Table 2**).

Human	Zebrafish co-orthologues	
BMP2	bmp2a bmp2b	*
DELTA (dll1)	delta a delta d	*
DLX2	dlx5 dlx2	*
ENGRAILED2	eng3 eng2	*
NODAL	squint nodal	Feldman *et al.*, 1998
PROGRANULIN	progranulin a progranulin b	Cadieux *et al.*, unpublished
SONIC HEDGEHOG	sonic Hh tiggy winkle Hh	*

TABLE 2. **Zebrafish co-orthologues of human genes resulting from a predicted 3R genome duplication within the actinopterygian lineage.** Orthology assignment was based on phylogenetic analysis and/or conserved synteny observed by gene mapping experiments (* Gates *et al.*, 1999). Note that none of the zebrafish gene duplicates are found in tandem within a given linkage group.

1.2.6.1 Another genome doubling in fish – It is important to point out that fish and mammals have similar numbers of chromosomes (e.g. the zebrafish has 50, medaka 48, human and mouse have 46), excluding the need to evoke differences in karyotype to explain large discrepancies in gene number. How then did the fish acquire additional genes? Two major mechanisms, which have found experimental support, are under suspicion. The first is an extension of the 1-2-4 rule (Meyer and Schartl, 1999), in which case the lineage leading to actinopterygians underwent an additional complete genome duplication that other vertebrates did not experience (**Figures 4 and 5**). When considering the outcome of such an event, it can be predicted that, on average, the fish genome bears gene families twice as large as those found in tetrapods. In congruence with this expected pattern is the aforementioned example of the duplicated fish *Hox* clusters (Amores *et al.*, 1998). Several rounds of independent gene duplications along the same lineage represent an alternative mechanism. This second hypothesis predicts that only selected genes would be present in larger number in fish compared to tetrapods, owing to the circumstantial duplication of only some genes, gene families or chromosomes.

Distinguishing between these two mechanisms can therefore be problematic if genome duplications occurred a long time ago, since eventual gene loss, chromosomal rearrangements (e.g. fusions and fissions) and lineage-specific gene amplifications can mask the true origins of certain gene families. Nevertheless, the advent of gene mapping and genome sequencing efforts has enabled the emergence of a powerful technique to evaluate the preponderance and contribution of these two mechanisms through the comparative analysis of gene order. As we will see, the combined use of several mapping panels for zebrafish initially developed to provide an infrastructure to facilitate the characterization of the >600 essential genes defined by mutations generated through genetic screens (Driever *et al.*, 1996; Haffter *et al.*, 1996), have been instrumental in providing insight into vertebrate chromosomal evolution.

Strong support for genome duplication as the mechanism underlying the appearance of the "a" and "b" copies in zebrafish has been derived from the analysis of

gene topologies within the genome. By scoring polymorphisms for genes and ESTs in the HS (heat shock) meiotic mapping panel, it was noticed that several zebrafish genes with synteny to human genes also have extra paralogues residing on duplicated chromosome segments. The zebrafish linkage groups carrying duplicated portions of conserved synteny include: LG3-LG12, LG5-LG21, LG7-LG25, LG11-LG23, and LG16-LG19 (Woods *et al.*, 2000).

Overall, this type of analysis of duplicated genes, as opposed to being solely based on phylogeny, assuming that the entire mapping information for each gene is complete, provides support for the 1-2-4-8 rule in fish, commonly referred to as 3R. The presence of paralogous copies finding residence on two chromosomes indicating synteny of each of the paralogous copies is in agreement with the scenario predicted from a complete genome duplication. Once more orthologous genes are mapped in zebrafish and humans (or once both genomes are sequenced and annotated), it will be interesting to establish whether they will still be mostly syntenic. Synteny is unlikely to have arisen independently several times. Further, a fish-specific genome duplication is suspected to have occurred during the Devonian period, since other diploid teleosts such as pufferfish (*Fugu rubripes*) and medaka (*Oryzias latipes*) also have extra genes (Wittbrodt *et al.*, 1998). This suggests that the duplication preceded the last common ancestor of these species >100 mya (million years ago) (Figures 3 and 4; Postlethwait *et al.*, 1998; Gates *et al.*, 1999; Santini and Tyler, 1999). Indeed this cannot be a recent event, otherwise the genome of these teleosts would be tetraploid like the clawed frog (*Xenopus laevis*) (Bisbee *et al.*, 1997) or pseudotetraploid like rainbow trout (*Onchorynchus mykiss*) (Young *et al.*, 1998). Surprisingly, fugu, medaka and stripped bass all contain 4 *Hox* clusters only (Holland, 1997; Aparicio *et al.*, 1997; Kurosawa *et al.*, 1999; Meyer and Malaga-Trillo, 1999; Snell *et al.*, 1999), indicating that the zebrafish is not a general representative of the ray-finned clade. It thus seems that the Hox repertoire likely numbered four at the base of the actinopterygian radiation, and that the additional genome duplication that occurred on the lineage leading to zebrafish took place after its divergence from the acanthopterygians perches, puffers, etc, and is restricted to members of the order Cypriniforme in which polyploidy is widespread

and includes the tetraploids goldfish *Carrasius auratus* and the carp *Cyprinus carpio* (Stellwag, 1999).

1.2.7 On the relatedness between the genomes of fish and man – Despite the evolutionary distance between humans and zebrafish, sharing a common ancestor ~ 450 mya, both genomes have chromosomes displaying portions with extensive syntenic correspondence. For example, a total of 167 conserved syntenies involving 2 or more candidate orthologous gene pairs have been found between *Homo sapiens* chromosome 6 (Hsa6) and zebrafish linkage groups 19 and 20 (LG19 and LG20), Hsa9-LG5, Hsa12-LG23, Hsa14-LG17-LG20, and Hsa17-LG3-LG12-LG15 (Woods *et al.*, 2000; Postlethwait *et al.*, 2000). This finding is a testament to the concept of an ancestral genome canvas that evolved differently subsequent to the splitting of ray-finned and lobe-finned fish. In addition, it provides the basis for a genome-wide understanding of sequence function and of the mechanisms governing genome evolution. A comparison of available gene maps of tetrapods to that of a vertebrate outgroup such as the zebrafish highlights the features unique to mammalian genomes and those that were inherited from a common ancestral state. In this respect, a frequent line of inquiry aims at defining the extent and length of conserved syntenies, and the preservation of gene order found therein. One such analysis revealed that the numerous blocks of conserved synteny between fish and man are usually large, but that gene orders are frequently transposed and inverted (Postlethwait *et al.*, 2000). An example of this arrangement is depicted in Figure 6, where most of human chromosome 17 (Hsa17) is orthologous to a portion of zebrafish linkage group 3, which in turn has duplicated loci found on LG12. Thus, although both mechanisms have certainly played an important role during vertebrate chromosome evolution, intrachromosomal rearrangements have been fixed more frequently than translocations, in disagreement with previous observations (Ehrlich *et al.*, 1997). If the reverse were true, it would be expected that gene order would be preserved within dispersed, relatively short conserved syntenies.

This finding is of fundamental importance for two reasons. First, it can help clarify the origins of the human genome, or at least shed light on the evolution of the vertebrate genome. Indeed, based on the extent to which zebrafish chromosomes are orthologous to several human chromosomes in a mosaic fashion, and the similar karyotype between these species, a question emerges: What was the genome organization of the last common ancestor of fish and mammals? Recall that humans have two less chromosomes in the haploid set than zebrafish. If tetraploidization is at the root of the multiple copies of zebrafish chromosome segments (Amores *et al.*, 1998; Postlethwait *et al.*, 1998) then zebrafish should have twice the number of chromosomes as humans in the absence of chromosomal rearrangements. Postlethwait and co-workers have proposed that the last common ancestor of zebrafish and humans had ~12 chromosomes, which fragmented in different places in different tetrapod lineages while doubling in the fish lineage, thereby establishing current karyotypes (**Figure 6**; Postlethwait *et al.*, 2000;). An alternative model, in support of an ancestral state of ~24 chromosomes that underwent excess chromosomal fusion within the fish lineage before or after 3R, resulting in the re-establishment of the original number, was not supported by the comparative mapping analysis.

The second contribution of these comparative approaches is the possibility of using the zebrafish for dissecting conserved genetic pathways involved in human disease and congenital malformation (Brownie *et al.*, 1998; Wang *et al.*, 1998; Donovan *et al.*, 2000; Feldman *et al.*, 2000; Parsons *et al.*, 2002)

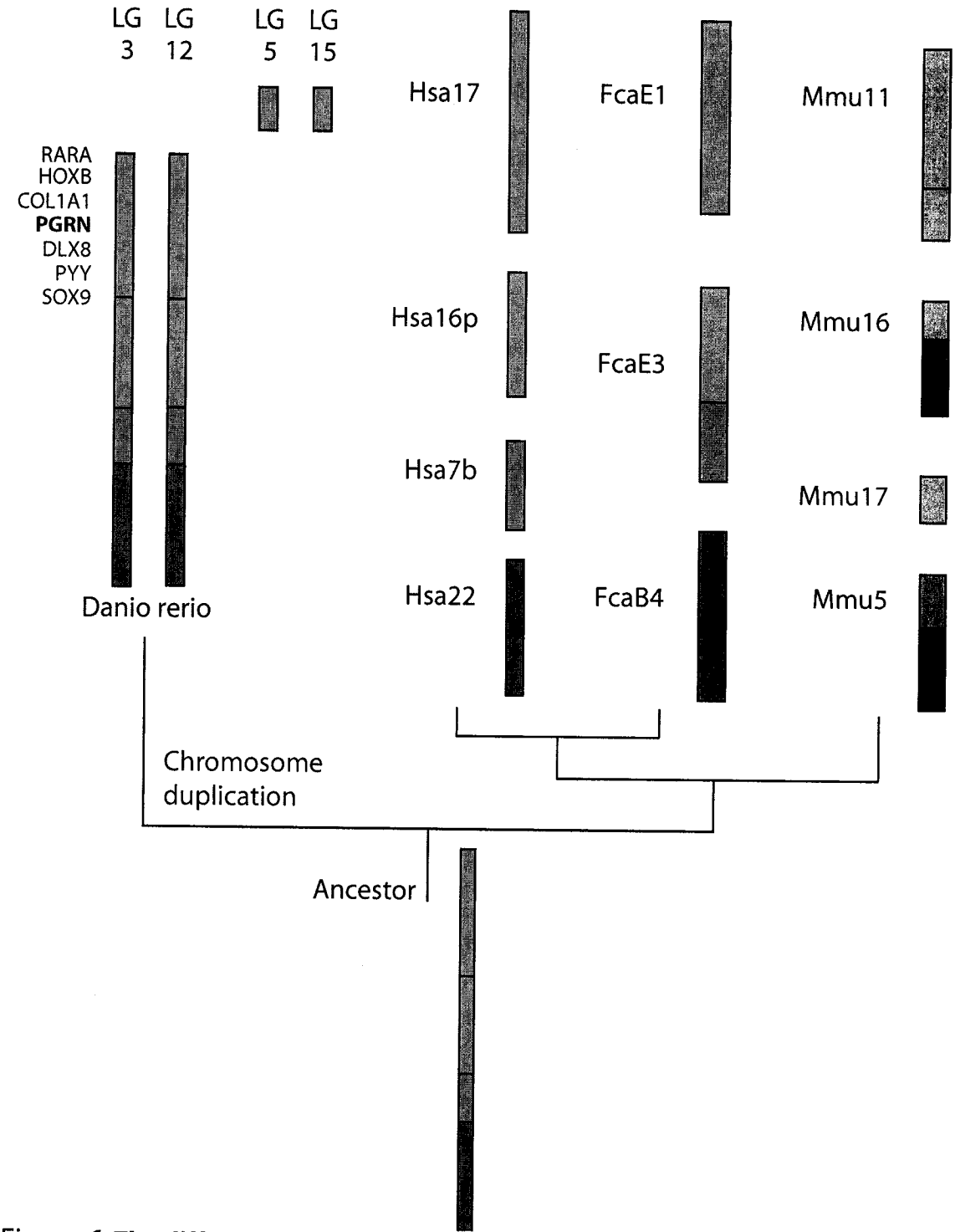


Figure 6. **The differential fragmentation of ancestral vertebrate chromosomes as a model for the establishment of current mammalian genome karyotypes.** A model for the origin of human chromosomes Hsa17, Hsa16p, Hsa7b, and Hsa22 is represented graphically. Analysis of conserved syntenies between orthologous segments in fish, mouse, cat, and human suggests that these are derived from a single ancestral vertebrate chromosome that underwent extensive fission, over fusion, in the tetrapod lineage. A similar pattern is observed for other human chromosomes. (Adapted from Postlethwait *et al.*, 2000)

1.2.8 More fishy tales about recent duplications in the human genome – Elucidating the prevalence of duplicated genomic loci has a far-reaching impact on the understanding of the evolution of vertebrate genomes, including that of humans. From the perspective of delineating the genetic etiology of disease, the contribution of comparative genomics is now recognized as a platform for establishing links from genotype to phenotype for conserved sequences across phyla. For this reason, and others, efforts have been made at determining the extent of presumed duplicated loci in the human genome that so far have remained elusive because of their high sequence identity. Indeed, the human genome carries dispersed segments for which the distinction between allelic variation (estimated at 1 base substitution per 1000 nucleotide) and highly similar duplicated sequences has proven to be a daunting task to resolve. Distinguishing between these two is of paramount importance in fully annotating the human genome, because unrecognised *bona fide* duplicated sequences inherently generate gaps in the final assembly.

Using a clone by clone approach to uncover sequences over-represented within available public whole-genome databases revealed the existence of 169 large, highly similar, paralogous segments, rather than allelic variations, corresponding to approximately one tenth of the genome (Bailey *et al.*, 2002). Mapping of these regions revealed a non-random chromosomal and genic distribution, thereby providing evidence for segmental duplication in expanding gene families. A well-documented set of segmental duplications can be found at http://humanparalogy.cwru.edu.

1.2.8.1 Chromosomal distribution of recent segmental duplications – Segmental duplications are distinct from those thought to have resulted from 2R, and appear to have occurred recently within the human genome, approximately 40 mya (Bailey *et al.* 2002). Recent segmental duplications within the human genome are non-uniform. The pericentromeric region of some chromosomes, such as 8p, 8q, Xp and Xq, are apparently lacking duplications between chromosomes. Chromosomes 7, 9, 15, 16, 17, 19, 22, as well as the male chromosome, known to be rich in gene content,

are surprisingly enriched in intra- and interchromosomal duplications relative to chromosomes 2, 3, 4, 5, 8, 14 and 20, which are considered to be the seven most gene poor. Although their location varies within different chromosomes, segmental duplications usually occur interchromosomally, within pericentromeric and subtelomeric regions (Eichler, 2001; Mefford and Trask, 2002), thus putting them under negative selective pressure because of their occurrence within gene poor regions. The increase in gene content for gene rich chromosomes is correlated with intrachromosomal duplications. Hence, a lack of negative correlation between gene density and duplication content for chromosomes challenges the prevalent view for the landscape distribution of segmental duplications.

1.2.8.2 What genes are found within the duplicated segments ? – A search for genes embedded within recent genomic duplications, while ensuring the presence of all exonic sequences that give rise to a defined transcript, suggested that the proteome is represented in a non-random fashion within segmental duplications. Particularly enriched in this set are genes involved in host defense and immunity (natural killer receptors, defensins, serine proteases, interferons and cytokines), membrane surface interactions (HLA, galectins, lipocalins, carcinoembryonic antigens), drug detoxification (cytochrome P450), and growth/development (somatotropins, chorionic gonadotropins) (Bailey *et al.*, 2002). Although unlikely, it is currently unknown whether an additional progranulin gene with allele-like characteristics is residing within a recently duplicated segment.

1.2.8.3 Genome annotation: the potential benefits derived from mere cataloguing – Identifying and assembling duplications are necessary for the proper completion and annotation of the human genome, even if descriptive in essence, is a worthwhile pursuit in itself, offering the promise of better understanding the evolutionary ramifications of our genome. Beyond this immediate contribution, and perhaps more importantly, lie two areas of genomic research directly impacted by these findings:

First, it can be envisioned that when duplications remain unrecognised, false assumptions can be made regarding the identification of single nucleotide polymorphisms (SNPs) within paralogous sequence variants. As expected, a higher density of SNPs has been demonstrated within duplicated segments compared with unique segments. This is especially important when considering the predisposition to disease for any given locus. It is currently estimated that ~100 000 paralogous sequence variants contaminate databases SNP.

The second field potentially benefiting from a proper description of genome organization is based on the observation that among the duplicated segments detected, several were previously associated with hot spots of genomic instability. It is now recognized that homologous recombination occurring between blocks of non-allelic duplicated sequences leads to inverted genomic segments, microdeletions and microduplications. When genes are located in the vicinity of these duplicated segments, these can be rearranged and lead to disease (Chen *et al.*, 1997; Christian *et al.*, 1999; Jenne *et al.*, 2001; Kuroda-Kawaguchi *et al.*, 2001). Such genomic instability commonly involve fairly long duplication segments (>10kb) with >95% similarity flanking 50kb to 10Mb of DNA (Stankiewicz and Lupski, 2002).

1.2.9 Vertebrate evolution revisited: role of duplications – In the end, it should be remembered that gene duplications *per se* do not directly result in an increase in the complexity of life. This is exemplified by comparing the diploid *Xenopus tropicalis* to the tetraploid *Xenopus laevis* (Amaya and Offield, 1998). On the other hand, since teleosts are known to be the most prolific vertebrates on the planet, with ~20 000 species, it is tempting to evoke 3R in the fish lineage as a contributing factor. The timing and type of duplication events occurring throughout evolution, and their coordinate effects on the repertoire of affected genes are perhaps more adequate avenues for deriving insight into the process of vertebrate evolution. In this respect, it is more fitting to address the relative importance of each mode of gene duplication. A recent study using 749 gene families suitable for reliable phylogenetic analysis, which approximates to one-quarter of human gene families, has shown that both small- and

large-scale duplications were instrumental in the shaping of the current vertebrate proteome (Gu et *al.*, 2002). More importantly, human paralogous genes have an age-distribution pattern that reveal three waves of duplication (waves I, II, III), the third being more ancient. The most recent events, constituting wave I, occurred in mammals and appears to be related to the recent segmental duplications and also comprises tandem duplications (Eichler, 2001). Characterized in this wave are immune-related molecules and several other large gene families (Nei *et al.*, 1997; Li *et al.*, 2001). Wave II duplication is in accordance with a whole-genome scale event having occurred in the early stages of vertebrate evolution, with the appearance of a large number of paralogues for developmentally important genes. This observation further lends credence for using zebrafish as model for functional genomics (Fishman, 2001). The oldest age-distributions have been encompassed within a third event that comprises extant duplications that took place approximately 830 mya, during or prior to metazoan evolution (Gu, 1998). Duplicates originating at the base of multicellular life are thought to be associated with the emergence of major signal transduction pathways (Iwabe *et al.*, 1996).

1.2.10 Evolutionary genomics: the breaking of the one into the manyfold – The field of evolutionary genomics has indeed come far from the original, and sometimes misleading, view that with major changes in morphological and physiological features across phyla are inevitably associated with equally profound genic alterations. The successful use of orthologues, separated through speciation events by over half a billion years, in rescue experiments testifies to the functional conservation of many gene products. Further, these genes often display similar expression patterns. This is generally regarded as an affirmation that a gene product fundamentally performs the same role, regardless of its incarnations in the various parallel universes of the tree of life. This, and the observation that phylum-specific gene classes are usually rare in a given clade, facilitates the use of an adequate model organism to address a given biological question. It is understandably unwise to overly generalize the concept of conserved gene function, since orthologous sequences do not necessarily share the same level of structural complexity. A simple observation, derived from the analysis of

the draft sequence of the human genome, is sufficient to dispel this line of reasoning. The unanticipated low complement of protein-coding genes in the human genome (~30,000 – 40,000), which is comparable to that of other vertebrates so far investigated, suggests that additional factors influence the complexity of life (Lander *et al.*, 2001; Venter *et al.*, 2001). A likely contribution to the enhanced sophistication encountered in humans is the incorporation of additional protein modules in selected transcription units, a process referred to as domain accretion, enabling higher order interactions within signalling complexes. In fact, based on genic content, mouse and man differ by only one percent (1%) (Mouse Genome Sequencing Consortium, 2002).

Notwithstanding the crucial differences between species, the era of comparative genomics is now paving the way for a more systematic approach to modelling genotype to phenotype and will certainly impact our understanding of human genetic diseases.

1.3 The zebrafish: a model – Since zebrafish *progranulin-a* is syntenic to mammalian *progranulin*, a gene knockdown approach for this gene in the developing zebrafish embryo was initiated (section 3.5). The well-document involvement of the mammalian progranulin gene and its biosynthetic products in the control of cell proliferation *in vitro* demanded its investigation in the context of embryonic development, where extensive morphogenetic movements (e.g. epiboly, convergent-extension migration of lateral mesoderm, etc.) is intermixed with the gradual acquisition of cellular identity (such as the formation of the primordial germ layers during gastrulation, and later organogenesis). These settings offer key paradigms to evaluate the enigmatic contribution of any growth factor during normal cellular proliferation and differentiation.

Although some studies aiming at establishing a role for the murine orthologue during early embryogenesis have been highly informative (see section 3.6.6) the use of the transparent nature of the developing zebrafish was attractive in exploiting some of its advantages over the murine model. For instance, unlike the mouse, zebrafish

embryos do not require uterine implantation for continued proper development. Also, the developing fish has the capacity to survive and develop even if its vasculogenic and angiogenic programmes are disrupted. Fish compromised in this way reach sufficiently advanced stages to allow for the study of other developed organ systems. These features should also be interpreted as a fundamental difference between these model organisms, and should be kept in mind when interpreting data derived from such comparative studies. Nonetheless, fundamental biochemical and physiological activities for the progranulin orthologues have presumably survived since they were last derived from a common ancestor.

It will be crucial to model the contributions provided by studies of the role of progranulin in vertebrate development using murine models, especially with the sophisticated repertoire of techniques now commonly used for transgenesis purposes. However, as will be highlighted during parts of the discussion, the zebrafish may offer additional benefits to uncover gene function during development that may not, in some cases, be as straightforward in the mouse, especially for genes that are represented by a limited number of paralogues in tetrapods.

Inherently, every model has its merits and limitations, from the simplest unicellular organism to the most complex metazoan. In light of the striking morphological differences encountered between fish and man, and the estimated 450 million years that have elapsed since both lineages split, it may be conjectured that the utility and feasibility of using zebrafish as a model for vertebrate development and human disease is somewhat far-fetched. Recent observations made in the zebrafish have greatly impacted our views and understanding of genes whose activities are better appreciated in mammals. For instance, a role for signal transducer and activator of transcription 3 (Stat3) in regulating the gastrulation movements was demonstrated for the zebrafish (Yamashita *et al.*, 2002). Specifically, it was shown that Stat3 is activated on the dorsal side of the embryo by the maternal Wnt/β-catenin pathway, and that a lack of Stat3 activity results in animals displaying abnormal cell movements: Stat3 is required cell autonomously for the anterior migration of dorsal mesendodermal cells

and non-cell autonomously for the convergence of neighbouring paraxial cells. Interestingly, this study also confirmed that cytokine signalling acting through Stat3 does not affect early cell fate specification (Yamashita *et al.*, 2002).

Further, gene disruption studies of the mouse p63 gene, a p53 gene family member, have indicated a requirement for this gene in epidermal growth and limb development (Mills *et al.*, 1999; Yang A *et al.*, 1999; Levrero *et al.*, 2000; Yang and McKeon, 2000; Irwin and Kaelin, 2001). The p63 gene contains two alternative promoters that give rise to protein products that have opposing effects: one form (TAp63) contains an amino-terminal transactivation domain and can activate p53-responsive promoters, whereas the other form (ΔNp63) lacks this domain (Yang A *et al.*, 1998). Because it is still capable of oligomerizing with the longer p63 protein, as well as with p53 or the related p73 protein, the ΔNp63 protein may act as a dominant-negative factor (Yang A *et al.*, 1998). Interestingly, the use of the knockdown approach for the zebrafish orthologue has resulted in similar observations, highlighting conserved biochemical activities in the respective organisms even if the structure of fish epidermis is very different from that of higher vertebrates (Lee and Kimelman, 2002). More importantly, however, this technique showed that selectively inhibiting translation from the second (alternate) promoter, while allowing for the full TAp63 form to be synthesized, leads to the skin defects, demonstrating that proper skin development, through the formation of epidermal stem cells, is predicated on the activity of the endogenous dominant-negative form of p63 (Lee and Kimelman, 2002).

Another example will suffice to illustrate the sometimes illuminating and surprising findings made by the use of alternative models to the mouse. In *Xenopus laevis*, a frog in which are employed similar strategies to those used with the zebrafish, abrogating IGF signalling using a secreted dominant-negative type I IGF receptor, and performing complementary overexpression experiments with the IGF ligands (Xenopus has three IGFs), led to the discovery that IGF signals appear to be both required and sufficient for anterior neural induction, and that this IGF activity is mediated in part through the potent inhibition of Wnt signalling at a step upstream of β-catenin (Pera *et al.*, 2001). This is in disagreement with mouse knockout studies that show that the IGF signalling pathway is a major regulator of body growth. Indeed, animals with disrupted IGF ligands, IGF type I receptor or insulin receptor substrates are proportionately reduced in size and possess smaller brains, but display no signs of preferential loss of head or forebrain structures (D'Ercole *et al.*, 1996; Efstratiadis, 1998). As was pointed out by Dr. De Robertis and co-workers, such a role for IGF signalling in neurogenesis may not be limited to the frog and was presumably never noticed in the mouse because of possible residual activity that may have remained using knockout approaches for individual receptor-ligand components (Pera *et al.*, 2001).

The future also looks bright for the use of zebrafish in the context of cancer genetics. Several labs are currently developing assays for the study of tumorigenesis and metastasis in adult zebrafish (Langenau *et al.*, 2003). Admittedly, this is beyond the scope of the present thesis, but may offer insight into the potential directions the research described in this thesis may be headed in the near future.

2. Materials and Methods

2.1 Materials – [γ-^{32}P]ATP (5000 Ci/mmol) and [α-^{32}P]dCTP (3000 Ci/mmol) were purchased from ICN Radiochemicals (Irvine, CA) or Amersham Biosciences (Baie D'Urfé, Québec, Canada). [α-^{32}P]UTP (400 Ci/mmol) was purchased from Amersham Biosciences (Baie D'Urfé, Québec, Canada). Random prime labelling was performed with a kit based on the method of Feinberg and Vogelstein (Feinberg and Vogelstein, 1983; Feinberg and Vogelstein, 1984) purchased from Amersham Biosciences (Baie D'Urfé, Québec, Canada). Probe purification was performed using either Sephadex G-50 columns (Amersham Biosciences) or NucAway spin columns (Ambion, Austin, Texas). Agarose LE was obtained from Roche Diagnostics (Laval, Quebec, Canada) and DNA ladders (100 bp and 1 kb) were purchased from MBI Fermentas Inc. (Flamborough, Ontario, Canada). *Taq* DNA polymerase was purchased from Amersham Biosciences or from MBI Fermentas Inc. Agarose gel extraction of PCR products was achieved using either a DNA extraction kit (MBI Fermentas) or a MinElute gel extraction kit based on the modified silica beads protocol (Vogelstein and Gillespie, 1979) (Qiagen, Mississauga, Ontario). TOPO TA cloning kits (pCR2.1 and pCRII) were obtained from Invitrogen (Carlsbad, CA) and pbluescript II phagemid vector was purchased from Stratagene (LaJolla, CA). *E.coli* cells chemically-competent for transformation were either prepared (XL1-blue) according to published protocols (Mandel and Higa, 1970; Cohen *et al.*, 1972; Tang *et al.*, 1994), or purchased (DH5α) from Bio S&T (Montreal, Quebec, Canada); electro-competent (Top10 F') cells were obtained from Invitrogen (Carlsbad, CA) or prepared according to the "Preparation of electrocomp^{TM} cells" protocol, version 2.0, from Invitrogen. T4 polynucleotide kinase, shrimp intestinal alkaline phosphatase and T4 DNA ligase were purchased from MBI Fermentas Inc (Flamborough, Ontario, Canada). Plasmid DNA isolation was initially achieved according to a modified protocol (Birnboim and Doly, 1979), and subsequently using the High Pure Plasmid Isolation kit, the Geno Pure plasmid midi and maxi kits (Roche Diagnostics, Laval, Quebec, Canada), as well as the HiSpeed plasmid midi kit (Qiagen). All oligonucleotide primers were obtained from the Sheldon Biotechnology Centre of McGill University.

2.2 DNA reactions with modifying and restriction enzymes – All restriction enzymes and modifying enzymes were purchased from Amersham Biosciences (Baie D'Urfé, Québec, Canada), MBI Fermentas Inc. (Flamborough, Ontario, Canada), or Roche Diagnostics (Laval, Quebec, Canada). Restriction enzyme digests were carried out at 37°C for 2 hours to overnight, in buffers supplied by the manufacturers. DNA modifications, which include ligations and end-labelling reactions, were performed as outlined in Molecular Cloning, a Laboratory Manual (Sambrook *et al.*, 1989), or according to the manufacturer's instructions.

2.3 DNA sequencing – All sequencing was performed at the Sheldon Biotechnology Centre of McGill University, using either an ALF Express DNA sequencer (Amersham Biosciences, Baie D'Urfé, Québec, Canada), an ABI 373 DNA sequencer (Applied Biosystems, Foster City, CA) or a MegaBACE 500 capillary electrophoresis sequencer (Amersham Biosciences) using primer sequencing or cycle sequencing methods.

2.4 Tissue extraction and peptide purification – Carp (*Cyprinus carpio)* were purchased live at a local fish market (Waldman Plus, Montréal, Canada). The spleens from two fish were extracted and subjected to reversed-phase enrichment using C_{18} Sep-Pak cartridges (Waters, Milford, MA) following an established protocol described previously (Bennett *et al.*, 1981). The granulins were purified by reversed-phase high-performance liquid chromatography (RP-HPLC) on a C_{18} μBondapak column (Waters, Milford, MA) using a gradient of increasing acetonitrile concentration containing 0.1% (v/v) trifluoroacetic acid throughout as previously described (Belcourt *et al.*, 1993). Column fractions were screened for cysteine content by amino acid analysis using a Beckman 6300 autoanalyzer (Beckman, Palo Alto, CA) as described previously (Bateman *et al.*, 1990). Granulin-1 immunoreactivity in these fractions was also measured by radioimmunoassay, using an antibody directed against carp granulin-1 (Belcourt *et al.*, 1993). Fractions containing cysteine and granulin-1 immunoreactivity were further purified by RP-HPLC using solvents containing 0.13% (v/v)

heptafluorobutyric acid throughout, and subsequently purified to apparent homogeneity using the original solvent system containing trifluoroacetic acid . Mass determinations of the purified material was undertaken using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry on a Voyager mass spectrometer (Applied Biosystems, Framingham, MA) located at the Sheldon Biotechnology Centre of McGill University.

2.5 Microsequencing of carp granulin-A – Putative carp granulin-A was pyridylethylated according to a previously published protocol (Belcourt *et al.*, 1993). The alkylated derivative of carp granulin-A was purified by RP-HPLC and used for all subsequent sequencing procedures. Approximately 20μg of the alkylated carp granulin-A was digested with sequencing grade chymotrypsin (Roche Diagnostics) according to the manufacturer's directions, and the resulting fragments separated by RP-HPLC on a C_{18} μBondapak column. Amino-terminal sequence analysis of carp granulin-A and its chymotryptic fragments was undertaken using a Procise sequencer (Applied Biosystems, Foster City, CA) at the Sheldon Biotechnology Centre of McGill University.

2.6 Extraction of genomic DNA from tissues – Carp and zebrafish tissues were finely chopped, washed in 1X phosphate buffered saline (PBS) (10X PBS is composed of 0.01M KH_2PO_4, 1.37M NaCl, 0.027M KCl, pH 7.0), and resuspended in 2 ml of a 10mM Tris-HCl pH 7.6, 50mM NaCl, 0.2% SDS solution. Proteinase K (200 μg/ml) was added and the reaction mixed gently. Following incubation at 42°C for approximately 16 hours with gentle agitation to digest the proteins, the DNA was extracted with water-saturated phenol-chloroform-isoamyl alcohol (24:24:1). Extractions were performed by gently rocking the DNA sample in the aqueous phase with an equal volume of the organic solvent for 10 min to obtain a complete emulsion, centrifuging at 600Xg (using a Beckman JS-4.2 rotor in a Beckman J6-Bcentrifuge; Beckman instruments Inc, Fullerton, CA) for 3 min, and recovering the aqueous phase. The extraction was repeated once with phenol-chloroform-isoamyl alcohol (24:24:1) and twice with chloroform-isoamyl alcohol (24:1). The DNA was precipitated from the

aqueous phase by adding 0.02 volume of 3M NaCl and 1 volume of isopropanol. The DNA was then removed with a glass rod, washed 3X with 70% ethanol and resuspended in 1mM Tris-HCl pH 7.6, 0.1mM EDTA, at 4°C until the DNA was fully dissolved. The concentration of the DNA was determined by measuring the absorbance at 260nm (Hitachi U-2000 Spectrophotometer, Hitachi Instruments, Danbury, CT).

2.7 Bacteriophage lamba cDNA library screening – Initially, a carp spleen cDNA library constructed with bacteriophage lambda gt11 (obtained from the Wageningen Agricultural University, The Netherlands) was transferred to nitrocellulose membranes (Xymotech Biosystems Inc., Mt. Royal, Québec, Canada) and screened with degenerate oligonucleotides as probe whose sequences were based on carp granulin-1 peptide. End-labelling of oligo probes with ^{32}P was performed using T4 polynucleotide kinase (Amersham Biosciences). Approximately $3.5x10^6$ independent clones were screened without success. Duplicate nitrocellulose filters (Schleicher and Schuell, Keen, NH) were prehybridized in 6X SSC (20X SSC is 3M NaCl, 0.3M sodium citrate, pH 7.0), 5x Denhardt's blocking reagent (50X Denhardt's is 1% polyvinylpyrrolidone, 1% Ficoll, 1% bovine serum albumin), 0.2% SDS, at 42°C for 3 hours. Hybridization was carried in the same solution, prepared fresh for 14-16 hours at temperatures empirically determined according to the oligos used. Filters were washed once at room temperature in 2X SSC, 0.1% SDS for 30 min, twice with the same solution for 30 minutes each at 37°C, and exposed at –70°C to Kodak X-OMAT film (Eastman Kodak Co., Rochester, NY) with an intensifying screen. No positive colony was isolated.

2.8 Bacteriophage Lambda DNA isolation – DNA from bacteriophage lambda was isolated using the lambda mini kit (Qiagen, Chatsworth, CA) as described by the manufacturer's instructions. Lysates of bacteriophage lambda clones were prepared as described in Molecular Cloning (Sambrooke *et al.*, 1989; Preparation of plate lysates stocks: Protocol I, p2.65), with the following modification: agarose was used instead of agar. Chloroform was added to the plates at a final concentration of 5%, vortexed, and the debris removed by centrifugation at 5000xg (JA-20 rotor; Beckman J2-21

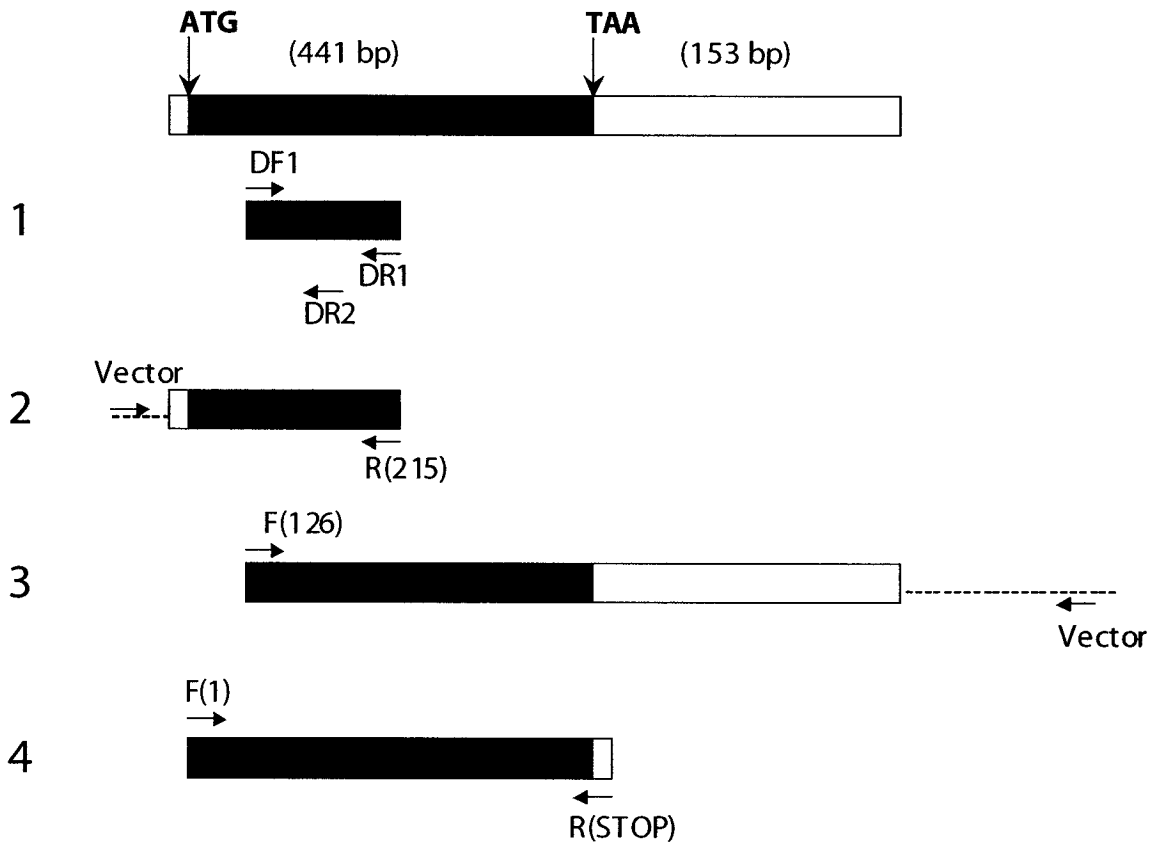
centrifuge) for 10 min at 4°C. RNase A (200μg/ml) and DNase1 (10μg/ml) were added and the reaction was left to proceed for 2 hours at 37°C. Polyethylene glycol (PEG) 8000 and NaCl were then added to a final concentration of 10% and 1.25 M, respectively, and the sample was incubated on ice or at 4°C for nearly 20 hours. Lambda bacteriophage particles were recovered from the solution by centrifugation at 12000Xg for 15 min, and resuspended in diluent (100mM NaCl, 8mM MgCl2, 50mM Tris-HCl pH 7.5, 0.01% gelatin). EDTA, SDS and proteinase K were added to a final concentration of 12.5mM, 0.5% SDS and 200μg/ml, respectively. After incubating at 65°C for 30 min, the sample was extracted three times with phenol, then twice with chloroform, and the DNA was precipitated with 2.5V 100% ethanol. The DNA was removed with a glass rod, washed with 70% ethanol, resuspended in H_2O and the concentration determined by measuring the absorbance at 260 nm (Hitachi U-2000 Spectrophotometer).

2.9 Amplification and cloning of zebrafish progranulin-1 and progranulin-2 cDNAs, and chimeric progranulin – Initially, the isolation of cDNA sequences encoding carp granulin-1 were attempted via conventional library screening procedures. Briefly, a carp spleen cDNA library constructed in λgt11 bacteriophage was screened by the plaque lift approach using degenerate oligonucleotide probes of approximately 18-mer based on the carp peptide sequence, without success. A zebrafish cDNA encoding a deduced precursor for zebrafish progranulin-1 was achieved using a PCR strategy (**Figure 7**). The amino acid sequence of carp (*Cyprinus carpio*) granulin-1 was used to design degenerate forward DF1 (5'- GTI ATY CAY TGY GAY GC – 3') and reverse DR1 (5'- CAR CAR TGR ATI CCR TC – 3') and DR2 (5'- TCR CAR TGR TAI CCR TG – 3') primers for use in the polymerase chain reaction (PCR). The sequences of these and other oligonucleotides used in the amplification of other cDNAs are listed in **TABLE 3**.

PCR amplifications were performed with *Thermus acquaticus* (Taq) DNA polymerase, unless specified otherwise, using a Hybaid thermal cycler from Bio/Can Scientific Inc. (Mississauga, Ontario, Canada). Amplified products were isolated by agarose gel electrophoresis, purified with the QIAquick Gel Extraction Kit (Qiagen Inc., Chatswoth, CA), and sequenced after cloning into TOPO pCR2.1. The template for the initial amplification reaction was a 5'-STRETCH plus cDNA library cloned in lambda gt10 vector (Clontech, Palo Alto, CA). cDNA for this library was prepared from 1-month-old zebrafish by a combination of oligo-dT and random priming. 0.25 µl of library (approximately 10^8 pfu/ml) was used in a final reaction volume of 100 µl for each new amplification attempt. The annealing temperature was varied empirically to maximize yield of the product. An initial reaction using the DF1 and DR1 primer pair yielded several products. 5 µl of this reaction was subjected to re-amplification using DF1 primer in combination with the nested (anchored) DR2 primer, which revealed a product of 126-bp encoding a partial sequence for granulin-1 (**Figure 7, step 1**). New progranulin1 primers F126 (5'– ACTGTGTGTCCAGACGG – 3') and R215 (5'– CCATCCCTGCAACACTG – 3') were then designed based on this sequence and were used, respectively, in combination with flanking gt10 primers in order to obtain the 5'- and 3'-untranslated region (UTR) cDNA sequences (**Figure 7, Steps 2 and 3**). Finally, the entire ORF was amplified with Pwo DNA polymerase (Roche Diagnostics, Laval, Québec, Canada), using forward F1 (5' – ATGTTCCCAGTGTTGATG – 3') and reverse R(STOP) (5' – GCTTA CAACTCCAACCCG – 3') primers (**Figure 7, Step 4**). This PCR was performed in a final volume of 100 µl, containing 0.25 µl of library, 50mM KCl, 10mM Tris-HCl, (pH 8.8), 1.5mM $MgCl_2$, 0.1% Triton X-100, 0.2mM concentration of each dNTP, 0.5 unit of Pwo DNA polymerase, and 100pmol of each primer. An initial denaturation step was carried out at 94° C for 3 min. Annealing temperatures of 54° C, 56° C and 58° C were used sequentially for 10 cycles each. Typical denaturation, annealing, and amplification reactions were carried out at 94° C for 30 sec, 54° C for 1 min, and 68° C for 1min, respectively. A final extension step of 10 min at 72°C was carried out after adding 0.25 unit Taq DNA polymerase. An amplification product specific for granulin-1 was sequenced on both strands. The 5'-

UTR, and a portion of the 3'-UTR for progranulin-1, were amplified using progranulin-1-specific primers in conjunction with a lambda gt10 primer. Sequences encoding progranulin-2 and a chimeric progranulin were uncovered through this approach. Each transcript was confirmed through sequencing of multiple independent amplification reactions using template cDNA derived from various sources, including adult organs and staged embryos (data not shown).

A



B

```
1    MFPVLMLLMAALVAADEPLLDLSIPVETVDTSAS  34
35   VIHCDAQTVCPDGTTCCLSPYGIWSCCPYSMGQCCRDGIHCCQHGYRCDSTSTRCLR  91
92   GWLTLPSSFQKATRTFQKDQTHAE  115
116  TVQCEGNFYCPAEKFCCKTRTGQWGCCSGLEL*  147
```

Figure 7. **Cloning Strategy for zebrafish cDNA encoding the precursor for granulin-1, and deduced amino acid of the open reading frame.** *Panel A*: The full-length cDNA for progranulin-1 is represented at the top of the diagram. Black rectangles represent the ORF, and open rectangles represent the respective 5'- and 3'-UTRs. The dashed lines represent lambda phage (vector) sequences. Bold numbers on the left represent the sequential order of PCRs. See text for details. *Panel B*: Deduced amino acid sequence for the precursor encoding granulin-1, consisting of one and one-half repeats of the granulin consensus motif. Characteristic cysteines are in bold and underlined. Predicted leader sequence is shown in italics. The full granulin-1 peptide sequence (35 to 91) is separated from the N-terminal half-granulin sequence (116-147) by an intervening sequence. Stop codon is represented by *. Numbers represent amino acid position.

Granulin-1 primers:

DF1	5′- GTI ATY CAY TGY GAY GC – 3′
DR1	5′- CAR CAR TGR ATI CCR TC – 3′
DR2	5′- TCR CAR TGR TAI CCR TG – 3′
F1	5′– ATG TTC CCA GTG TTG ATG – 3′
F126	5′– ACT GTG TGT CCA GAC GG – 3′
R215	5′– CCA TCC CTG CAA CAC TG – 3′
R(STOP)	5′- GCT TAC AAC TCC AAC CCG – 3′

Granulin-A primers:

F1	5′– ATG TTG AGA CTG ACA GTC TGC – 3′
R486	5′– CCA TAT GAT GTA GAC ATC AGC – 3′

Vector primers:

5′-gt10	5′– GAG CAA GTT CAG CCT GG – 3′
3′-gt10	5′– GAG TAT TTC TTC CAG GG – 3′

TABLE 3. **List of primers used for the cloning of zebrafish cDNAs encoding precursors for granulin-1 and granulin-A, respectively, as well as those used to clone the progranulin-1 gene**. The abbreviations used include: D, degenerate; F, forward; and R, reverse. Numbers in primer designations indicate nucleotide positions, with the "A" in the cDNA translation initiation codon designated "+1". IUPAC codes are used to refer to the bases in primer sequences.

2.10 Cloning and structural analysis of the zebrafish progranulin-1 gene – For all screening procedures, the progranulin-1 cDNA ORF was radioactively labelled with [α-^{32}P] dCTP by random priming performed with the Oligolabeling kit (Amersham Biosciences) for use as probe. Labelled probe was purified using a Sephadex G-50 column (Amersham Biosciences). Kodak X-OMAT AR film was used for autoradiography (Fisher Scientific Ltd, Whitby, Ontario, Canada). A zebrafish genomic library constructed in P1 artificial chromosome (PAC) (Amemiya and Zon, 1999) and represented on filters at high-density, was obtained from RZPD GmbH (Berlin, Germany) and screened for the presence of the progranulin-1 gene using standard procedures. Three positive clones (706K2254Q, BUSMP706K14116Q2, 706F20133Q2) were detected by autoradiography. Clones 1 and 2 were confirmed to carry at least part of the progranulin-1 gene by PCR analysis using the F1 (5'– ATGTTCCCAGTGTTGATG – 3') and R215 (5'– CCATCCCTGCAACACTG – 3') primer pair. DNA from a positive clone (706K2254Q) was purified with the QIAGEN Plasmid Midi Kit. 1.5 µg of this DNA was subjected to restriction digest with EcoRI to generate fragments suitable for cloning into pBluescript II KS (Stratagene), and was followed by transformation in TOP 10F' electrocompetent cells (Invitrogen). Screening of colonies transferred onto nitrocellulose membranes (Xymotech), employing the same probe used for the original library screening, was performed in the following prehybridization and hybridization conditions: 2x SSC, 0.5% SDS, 0.05% Na Pyrophosphate at 65° C. Membranes were washed twice in 1x SSC, 0.1% SDS, 0.05 % Na Pyrophosphate at 60° C for 15 min, followed by two washes in 0.1% SSC, 0.1% SDS, 0.05% Na Pyrophosphate at 60° C for 10 min. Plasmid DNA from a positive clone was purified using the high pure plasmid isolation kit (Roche Diagnostics). An insert of ~9-kb was fully sequenced and revealed the presence of the promoter region and approximately half of the progranulin-1 gene. The remaining gene sequence was found in a ~6-kb insert clone isolated by re-screening the colony lifts with ^{32}P-labeled reverse R(STOP) oligonucleotide as probe. A PCR was performed using primers flanking this EcoRI site, and sequenced to confirm that the isolated 9 kb and 6 kb clones represent continuous sequences within the original PAC clone.

2.11 Isolation of an antisense transcript to progranulin-1 and progranulin-2 (AS progranulin-1/-2) – While in the process of analysing progranulin-1 genomic sequences through BLAST searches, an EST harbouring sequences for unspliced progranulin-1 but in the reverse complement orientation was noticed. This clone (accession number AW777232) was purchased (RZPD, Berlin, clone ID: DKFZp717B091Q2) and further analysed through sequencing. Sequencing of the genomic region containing the genes for progranulin-1 and progranulin-2 (GenBank accession number AC124903) confirmed the authenticity of the progranulin antisense gene and its associated spliced transcript, sharing complementary exonic sequences to exons 2 and 3 of both progranulins-1 and -2 genes (see results).

2.12 Amplification and cloning of zebrafish progranulin-a and progranulin-b cDNAs – A similar PCR strategy to that described for progranulin-1, using the same zebrafish cDNA library as source of template, was employed to amplify the precursor encoding granulin-A, using primers based on the sequence represented in GenBank EST clone AW174591. A cDNA encoding a partial ORF for zebrafish progranulin-a was amplified using the granulin-A F1 primer (5'– ATGTTGAGACTGACAGTCTGC – 3') in conjunction with vector primer 3'-gt10 (5'– GAGTATTTCTTCCAGGG – 3'), which did not overlap with the 3'end read for this EST (accession number AW233473). From this sequence, a primer was designed (5'-CTCAACTCACTCACATCC – 3') to amplify the remaining part of the ORF with a primer whose sequence was derived from EST AW233473 (5'-GTTTATAGAGTTAGGGCTCG – 3'). The 5'-UTR was amplified using granulin-A R486 (5'– CCATATGATGTAGACATCAGC – 3') primer with the same vector primer. Both PCRs were performed using 0.5 unit *Taq* polymerase in NH_2SO_4-containing buffer with 1mM $MgSO_4$ (MBI Fermentas), under the following annealing conditions for 10 cycles each, sequentially: 58° C, 60° C, 62° C. All products obtained were cloned into TOPO pCR2.1, and colonies were subsequently transferred to nitrocellulose film. Positive clones were isolated after screening with a [γ-^{32}P] ATP end-labelled progranulin-a -specific primer as probe, sequenced, and assembled. A final assembly of the full-length cDNAs encoding both zebrafish progranulins-A and -

B was created by sequencing two ESTs deposited at GenBank (accession numbers AW174591 and AW184435, respectively) that were purchased from RZPD (Berlin, Germany), corresponding to clones UCDMp574E2318Q2 and UCDMp574I0223Q2, respectively. After sequencing each clone on both strands in full, these were compared to corresponding genomic sequences whenever available through the Sanger Center as part of the Zebrafish Genome Sequencing Project (Vogel, 2000).

2.13 Chromosomal mapping by radiation hybrid (RH) analysis – Primers used for linkage group (chromosome) assignment are in **TABLE 4**. The LN 54 radiation-reduced zebrafish/mouse hybrid panel (Hukriede *et al.*, 1999) was used to map the chromosomal location of progranulin-1 and progranulin-2 by PCR using a forward (5' – ACTGTGTGTCCAGACGG – 3') primer in combination with either reverse1 (5' – CCATCCCTGCAACACTG – 3') or reverse2 (5' – TCTGGTGGAGGCGAAATT – 3') primers, which do not discriminate between these two genes, using a Tetrad DNA engine (model PTC225) by MJ Research. Each reaction was carried in a final volume of 20 µl containing 100ng "hybrid DNA", 500 mM KCl, 100 mM Tris-HCl (pH 8.3), 15 mM $MgCl_2$, 0.2 mM each dNTP, 1 unit *Taq* DNA polymerase, and 5 pmol of each oligo. Denaturation, annealing and amplification were performed at 94° C for 30 sec, 55° C for 30 sec, and 72° C for 30 sec, respectively, followed by an extension step of 7 min at 72° C. The PCR results were submitted online (http://www.zfish.wustl.edu) to locate the radiation hybrid map position of progranulins-1 and -2 using the RH MAPPER program. The map positions of progranulin-a and progranulin-b were resolved employing the same approach, using an annealing temperature of 60° C with the following primers:

progranulin-a map forward (5' – ATGTTGTGCAGTGTGAAGGAC – 3') and map reverse (5' – CATCAGCAGACAGGAGAACTC – 3'); progranulin-b map forward (5' – AACGCATGCACTGCTCTGATC – 3') and map reverse (5' – AGTTCCCTGATAACAGCAGTG – 3'). The synteny analysis was performed by using the consolidated zebrafish maps available from ZFIN (http://zfin.org/ZFIN/) and data available from LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink/).

Progranulin-1 and Progranulin-2

map forward	5′ – ACT GTG TGT CCA GAC GG – 3′
map reverse1	5′ – CCA TCC CTG CAA CAC TG – 3′
map reverse2	5′ – TCT GGT GGA GGC GAA ATT – 3′

Progranulin-a

map forward	5′ – ATG TTG TGC AGT GTG AAG GAC – 3′
map reverse	5′ – CAT CAG CAG ACA GGA GAA CTC – 3′

Progranulin-b

map forward	5′ – AAC GCA TGC ACT GCT CTG ATC – 3′
map reverse	5′ – AGT TCC CTG ATA ACA GCA GTG – 3′

TABLE 4. **Primers used in the linkage group assignment of the zebrafish progranulin genes.** Before use on the LN 54 mapping panel, conditions for each primer combination were optimized by PCR with the use of zebrafish genomic DNA derived from the AB wild type strain. The authenticity and specificity of each amplicon was verified by sequencing after cloning into the pCRII plasmid (Invitrogen). For the assignment to zebrafish linkage groups, each PCR amplification experiment was performed at least twice.

2.14 Fish maintenance – Wild type zebrafish were purchased from Scientific Hatcheries (San Diego, CA) and maintained on a 14h/10h light/dark cycle at 28.5°C in an Allentown Aquaneering tank system (New Jersey), fed twice daily, and bred as described elsewhere (Mullins *et al.*, 1994). Embryos for developmental studies were collected from tanks and staged according to Kimmel *et al.* (Kimmel *et al.* 1995).

2.15 RNA extraction – Adult zebrafish were anesthetized in aquarium water containing 90 mg/L 3-aminobenzoic acid ethyl ester (A-5040, Sigma), then were patted dry with tissue paper and killed by cervical transection. Individual organs were dissected from 5-10 adult zebrafish under a MS5 dissecting microscope (Leica Microsystems Inc, Wollowdale, Ontario, Canada) and frozen in liquid nitrogen. Organ-specific total RNA was isolated using Trizol LS reagent (Gibco BRL) according to the manufacturer's instructions. Similarly, total RNA was isolated from approximately 50 embryos staged according to Kimmel (Kimmel *et al.* 1995) after freezing in liquid nitrogen and subsequent homogenization by passing through a 28G1/2 gauge syringe (Becton Dickson, Oakville, Ontario, Canada) using either Trizol reagent or a modified SDS/phenol protocol (Stachel *et al.*, 1993). The latter procedure involved initial homogenization in 0.2M LiCl, 50mM Tris-HCl, pH 8.0, 10mM EDTA and 1% SDS (1 ml per 100 embryos). Subsequently, an equal volume of water-saturated phenol was added to the homogenate and incubated at 65°C for 10 min. The aqueous phase was re-extracted once with phenol and twice with chloroform/isoamyl-alcohol (24/1). Total RNA was precipitated by adding one-quarter volume of 10M LiCl at –20°C for 1 hour and collected by centrifugation. The pellet was washed with 75% ethanol, dried, resuspended in DEPC-treated H_2O and stored at –80°C. Amounts of total RNA were assessed by measurement of absorbance at 260 nm and visualization of rRNA bands through TAE-agarose gel electrophoresis next to an 0.24-9.5 Kb RNA ladder (Gibco BRL).

2.16 Cloning of 5' and 3' untranslated regions – The rapid amplification of cDNA ends (RACE) was performed with the GeneRacer kit (Invitrogen) using total RNA isolated from adult zebrafish intestine. For progranulin-1 and progranulin-2 transcripts, the 5' – CCATCCCTGCAACACTGACCC – 3' reverse primer, corresponding to nucleotides 195-215 of each ORF, was used to perform the 5' RACE, while the 5' – ATGTTCCCAGTGTTGATGTTAC – 3' forward primer, corresponding to nucleotides 1-22 of each ORF, was used to perform the RACE in the 3' direction. Similarly a 5' UTR sequence for progranulin-a was obtained using the 5'– CATCAGCAGACAGGAGAACTC – 3' map reverse primer.

2.17 Adult tissue and developmental RT-PCR – Prior to cDNA synthesis, total RNA was treated with DNase I according to the manufacturer's instructions (MBI Fermentas Inc. (Flamborough, Ontario, Canada). Oligo dT-primed template cDNA for RT-PCR analysis was synthesized from approximately 5 µg of total RNA using the RevertAid H Minus First Strand Synthesis Kit (K-1632, MBI Fermentas Inc., Flamborough, Ontario, Canada). To control for genomic DNA contamination, the samples were also incubated in the absence of reverse transcriptase. The resulting cDNA reaction mixture was resuspended in a final volume of 200µl in dH_20, from which 10 µl aliquots were used in each subsequent amplification reaction. Primer sets, along with the size of amplicons are found in **TABLE 5.** Individual polymerase chain reactions were performed as follows on a ThermoHyBaid (model HBSPO2110) thermal cycler (Ashford, UK) using Taq (Amersham Biosciences): initial denaturing at 94°C for 2 min, followed by 30 cycles of 94°C for 45 sec, annealing at 60°C for 1 min and extension at 72°C for 1 min. A final extension step at 72°C was carried for 7 min. The authenticity of the amplicons was confirmed by either Southern analysis with an internal oligo end-labelled with T4 polynucleotide kinase (Amersham Biosciences) used as probe or through sequencing. Hybridizations were carried out overnight at 45°C in 6x SSC, 5x Denhardt's, 0.5% SDS, with 3 washes at 45°C in 2x SSC and 0.1% SDS for 20 min each.

To facilitate discrimination between the transcripts for progranulin-1, progranulin-2 and the progranulin hybrid, the PCR was performed using the $(NH_4)_2SO_4$ buffer system with $MgCl_2$ at a final concentration of 1mM (MBI Fermentas Inc. (Flamborough, Ontario, Canada). In control amplification experiments, 50ng of template cDNA cloned in pbluescript was used. For progranulin-1-specific amplification, the progranulin-1 forward and reverse1 or reverse2 primer combinations were used at an annealing temperature of 60°C. Progranulin-2 amplification was achieved using progranulin-2 forward with either progranulin-2 reverse1 or reverse2 primers at 63°C, while the transpliced transcript was detected using the progranulin-1 forward primer with either of the two progranulin-2 reverse primers, also at 63°C. All PCR products were resolved on 2% agarose gels, ethidium bromide stained and visualized on a Terochem Scientific UV projector (Toronto, Ontario, Canada). Pictures were taken using Polaroid 667 Film, and analysed using adobe PhotoShop 7.0 software after scanning.

A

gene	primers	size
progranulin-a	FOR: 5′- CTCAACTCACTCACATCCGC - 3′ REV: 5′- GTTTATAGAGTTAGGGCTCG - 3′	301 bp
progranulin-b	FOR: 5′- ACAATGGTGCGTGCAGCTTTC - 3′ REV: 5′- GCACACGGCATTTTTCATAGG - 3′	381 bp
generic progranulin-1 and -2	FOR: 5′- ACAGACAGCAGCAGAAGCATC - 3′ REV: 5′- TTCACATTGGACAGTCTCAGC - 3′	409 bp
AS progranulin	FOR: 5′- TTAACTTGTCGAGCGTTTCCC - 3′ REV: 5′- GCAGACCACTAATACTCTCCT - 3′	556 bp
actin	FOR: 5′- ATGGATGATGAAATTGCCGC - 3′ REV: 5′- TGTCATCTTTTCCCTGTTGG - 3′	253 bp

B

progranulin-1	FOR: 5′- G**G**TGGAGACTG**T**AGAC**AC** - 3′	
progranulin-2	FOR: 5′- G**A**TGGAGACTG**A**AGAC**GT** - 3′	
progranulin-1	REV1: 5′- TC**G**GGTGGAGG**T**G**G**AATC - 3′	190 bp
progranulin-2	REV1: 5′- TC**T**GGTGGAGG**C**G**A**AAT**T** - 3′	
progranulin-1	REV2: 5′- GCTTACAACTCCAAC**C**C**G** - 3′	372 bp
progranulin-2	REV2: 5 - GCTTACAACTCCAAC**T**C**A** - 3′	

TABLE 5. **Primers for RT-PCR.** *Panel A*: Primer combinations used for the amplification of individual progranulin genes, and for the antisense transcript. Note that "generic" primers denote that the primers do not discriminate between *progranulin-1* and *progranulin-2*. *Panel B*: Primers used for the specific amplification of *progranulin-1*, *progranulin-2* and the transpliced *hybrid progranulin*. Nucleotide substitutions are in bold. Amplification conditions are described in the accompanying methods section. In all cases, primers were designed on consecutive exons. Expected size of amplicons is indicated.

2.18 Northern blot analysis – For Northern analysis, 10 µg of total RNA was dissolved in 10mM $NaPO_4$ buffer pH 7.4 and 2x glyoxal, denatured for 1 hour at 55°C, chilled on ice, supplemented with 1/5 volume of loading dye (30% Ficoll, 0.1%, bromophenol blue, and 0.1% xylene cyanol), loaded on a gel (1% agarose, 10 mM $NaPO_4$) and run in 10mM $NaPO_4$ buffer with an exchange pump. After transfer onto a positively charged nylon membrane (Roche Diagnostics; Laval, Quebec, Canada) in 10X SSC, RNA was cross-linked to the filter by baking for 2 hours at 80°C. Visualization of rRNA bands was monitored after transfer by staining with 0.04% methylene blue according to Sambrooke *et al.*, 1989 (method 2, section 7.51). Hybridizations to detect progranulin transcripts were carried out overnight using α-UTP-labelled single-stranded complementary RNA probes (see *in situ* hybridizations for plasmid construct preparations) prepared with the Strip-EZ™ RNA probe synthesis and removal kit (Ambion, Texas). Blots were prehybridized for 2 hours, and hybridized overnight at 70°C in ULTRAhyb solution (Ambion, Texas). Subsequently, membranes were washed twice in 2x SSC, 0.1% SDS at 70°C for 15 min, followed by three washes for 15 min in 0.1x SSC, 0.1% SDS at 70°C, and then exposed to Kodak BIOMAX MS supersensitive x-ray film at -80°C for 2-5 days. Transcript sizes were calculated from the mobilities of RNA standards (Gibco BRL). RNA concentrations were estimated from measurement of absorbance at 260 nm, and the 18S rRNA band intensities were used as loading controls.

2.19 Whole-mount *in situ* hybridization – In order to prevent the appearance of melanin pigmentation, embryos at approximately 18-24hpf were grown in egg water supplemented with 0.003% of the tyrosinase inhibitor 1-phenyl-2-thiourea (phenythiocarbamide P-5272, Sigma). Staged embryos were manually dechorionated and fixed for 2 hours at room temperature or overnight at 4°C in 4% paraformaldehyde/PBS. After several washes in PBS, embryos were stored in 100% methanol until needed. Whole-mount *in situ* hybridization with digoxygenin-labeled RNA probes and antibody staining were essentially done according to the method of Schulte-Merker *et al.* (1992) or Thisse *et al.* (1993) at a hybridization temperature of 70°C. An extended description of the procedure is appended at the end of this section.

In some cases, polyvinyl alcohol was added to the staining solution in order to minimize the occurrence of background, especially when the reaction was required to proceed for several days (DeBlock and Debrouwer, 1993). The description of anatomical orientations for zebrafish embryos follows suggested conventions (Moorman, 2001). Stained whole-mount and sectioned embryos were mounted in glycerol and visualized under a Leica MZFLIII stereomicroscope. Pictures were taken with a Leica DC350F camera and processed with Adobe Photoshop 7.0 software.

2.20 Preparation of constructs for cRNA riboprobe synthesis – For the following riboprobe synthesis reactions, RNA polymerases (T3, T7 and Sp6), ribonuclease inhibitor and DNase1 were purchased from MBI Fermentas Inc (Flamborough, Ontario, Canada). Digoxigenin-11-UTP and the anti-digoxigenin-AP (alkaline phosphatase) Fab fragments, for probe preparation and detection, respectively, were purchased from Roche Diagnotics. The NBT/BCIP chromagen mix was used for alkaline phosphatase staining. NBT (nitroblue tetrazolium) and BCIP (bromochloroindolyl phosphate) were purchased from MBI Fermentas Inc.

Individual insert fragments were cloned into the EcoRI site of pbluescript: For progranulin-1, a 576 bp fragment was subcloned and riboprobes corresponding to it were transcribed using T3 polymerase with SmaI-linearized template (antisense) and T7 polymerase with HindIII-linearized template (sense). A 595 bp fragment corresponding to progranulin-2 was transcribed to generate antisense and sense riboprobes using T7 polymerase with EcoRV-linearized, and T3 polymerase with SmaI-linearized templates, respectively. Riboprobes were also synthesized using T3 polymerase with SmaI-linearized template (antisense), and T7 polymerase with EcoRV-linearized template (sense) corresponding to a 487 bp subcloned fragment of the hybrid progranulin cDNA. In addition to the full-length clone for the antisense transcript, two fragments were subcloned into pbluescript for riboprobe synthesis. The first, named AS exons 1-3, corresponds to the first three exons portions of the antisense gene (i.e. antisense to progranulin-1 and progranulin-2 genes), and is 1014 bp in size. Riboprobes corresponding to it were transcribed using T7 polymerase with

EcoRV-linearized template (antisense) and T3 polymerase with SmaI-linearized template (sense, i.e. antisense to progranulin-1 and -2). The second fragment, named AS tzf, is 886bp and corresponds to part of exon 4 of the antisense transcript. For this fragment, riboprobes were transcribed using T3 with SmaI-linearized (antisense) and T7 polymerase with EcoRV-linearized template (sense). Control riboprobe synthesis transcribed from the pbluescript backbone for subsequent quality assessment of *in situ* hybridization experiments used a ~2.5kb fragment corresponding to zebrafish sonic hedgehog (*Shh*) subcloned into pbluescript (gift from the Akimenko lab, University of Ottawa). *Shh* riboprobes were transcribed using T7 polymerase with EcoRI-linearized template (antisense) and T3 polymerase with XhoI-linearized template (sense).

The remaining riboprobes were synthesized from fragments subcloned into the NotI and SalI sites of the vector pSPORT. Inserts corresponding to full-length *progranulin-a* (3649 bp), *progranulin-b* (2820 bp), the full antisense transcript (*AS progranulin-1/-2*) (1989 bp), and the transcript for *P2X3* (~4kb) as control (gift from the Drapeau lab, McGill University), were used to transcribe riboprobes using Sp6 polymerase with SmaI-linearized template (antisense) and T7 polymerase with NotI-linearized template (sense). In all stages examined, both sense and antisense probes were analysed for all transcripts.

2.21 Production of progranulin-A and progranulin-B polyclonal antisera – Rabbit polyclonal antisera were generated against synthetic peptides corresponding to residues 242 to 256 of zebrafish progranulin-A (RAEWEDHKQKKPETQC), and residues 153 to 173 of zebrafish progranulin-B (CGSSPFLRKFAARRRKPLEKNA), respectively. The cysteine residues were introduced into the sequences and used for thiol coupling to keyhole limpet hemocyanin (KLH). Peptide synthesis and production of antibodies were generated through the Sheldon Biotechnology Centre of McGill University. The production of a rabbit polyclonal antibody directed against the carp granulin-1 peptide has been reported elsewhere (Belcourt *et al.*, 1993). Polyclonal antiserum was purified using the Montage Antibody Purification$_{\text{PROSEP-A}}$ Kit (P36486, Millipore, Nepean, Canada). IgG was quantified according to published values where

an O.D. of 1.32 at 280nm equals 1.0 mg/ml of IgG (Leslie and Clem, 1969). After desalting and concentrating, the antibodies were resuspended in 1% BSA/PBS.

2.22 Whole-mount immunohistochemistry – For immunodetection of zebrafish embryos, staged embryos were fixed for 90 min in 4% PFA/PBS at room temperature, washed in PBS, manually dechorionated, and then dehydrated in methanol. Rehydration of embryos was performed for 5 min each in successive solutions of MeOH / PBST (25%/75%, 50%/50%, 75%/25%), then 3 times in PBST. Embryos were then permeabilized with proteinase K diluted in PBST at a final concentration of 10 µg/ml. Duration of treatment increased with age as follows: 9 min for 24hpf embryos, 20 min for 48hpf embryos, 25 min for 72hpf embryos. Post fixation was then carried out in 4% PFA/PBS for 20 min at room temperature, and embryos were rinsed 3 times in PBST. Preincubation in blocking solution was carried out in PBST / 2% BSA / 5% Goat or Calf serum / 3% DMSO / 0.5% Triton X-100 at room temperature for 2-5 hours. Primary antibody was then added at a dilution of 1/5000 or 1/10000 and incubated at 4°C overnight. On the following day, embryos were rinsed 6 times in PBST for 15 min. Detection of the primary antiserum was performed using the ABC Staining System (sc-2018, Santa Cruz Biotechnology, Santa Cruz, CA) according to the manufacturer's instructions. Briefly, embryos were incubated with pre-adsorbed secondary antibody for 2-3h at a dilution of 1/200, and then rinsed 6 times in PBST. Staining was performed with diaminobenzidine (DAB) reacting with the horseradish peroxidase-conjugated secondary antibody. Alternatively, immunodetection was carried out with NBT/BCIP using a goat anti-rabbit IgG secondary antibody conjugated to alkaline phosphatase (#A-3937, Sigma, St-Louis, Missouri, USA).

After staining, embryos were rinsed 3 times in PBST, postfixed in 4% PFA/PBS, and stored a 4°C in PBST. Alternatively, embryos were transferred in increasing concentrations of glycerol as follows: 25%, 50%, 75%, 100% glycerol for 30 minutes each, rinsed in glycerol twice and store at 4°C. The glycerol treatment renders the embryos with enhanced optical clarity.

Preadsorption of antibodies was performed on the first day. 50-1000 spare embryos (usually 24-48hpf embryos) were quickly rehydrated in PBST, and rinsed 3 times in PBST. Incubation of the antibody in the blocking solution (PBST / 2% BSA / 5% Calf serum) containing the spare embryos was carried out at room temperature for several hours and then stored at 4°C until used. Preadsorption was done with the primary antibody at a concentration 5-10 times higher than the final concentration needed for the incubation step during protein detection. Antibodies were diluted at working concentrations prior to use (1/1000 for anti-progranulin-A; 1/1000 for anti-progranulin-B; 1/400 for anti-progranulin-1; 1:200 for anti-acetylated alpha-tubulin used as positive control (mouse monoclonal; T-6793, Sigma).

2.23 Morpholino injections – Morpholine-based oligonucleotides were designed according to published criteria (Summerton and Weller, 1997) and purchased from GeneTools, LLC (Philomath, Oregon, USA). Negative controls include individual morpholinos incorporating 4 or 5 nucleotides mis-matches, and a standard morpholino of scrambled sequence. A morpholino complementary to the zebrafish *chordin* transcript was used as positive control to monitor individual morpholino dose and efficacy of injections. For injections, morpholinos were diluted in 1x Danieau buffer (Danieau buffer is 58 mM NaCl, 0.7mM KCl, 0.4mM $MgSO_4$, 0.6mM Ca $(NO_3)_2$, 5 mM HEPES, pH 7.6) with 0.1% phenol red according to a published protocol (Nasevicius and Ekker, 2000), and injected into the yolk cell adjacent to the blastomeres of 1- or 2-cell stage embryos using a PLI-100 pico-injector (Harvard Apparatus, Holliston, Massachusetts,USA). Typically, 1 to 5 nl (nanoliter) of diluted morpholino were injected per embryo for a final amount of up to a maximum amount of 10 ng (nanogram) of morpholino.

2.24 SDS-PAGE and Western blot analysis – 15 morpholino-injected or wild-type embryos were collected at the 72hpf stage, frozen in liquid nitrogen, and stored at -80°C until needed. Embryos were homogenized in 30 µl 2 x Laemmli buffer (10 x Laemmli is 0.25 M Tris, 1.92 M Glycine and 1 % SDS in aqueous solution; Laemmli, 1970) by gentle pipetting according to a published protocol (Schulte-Merker *et al.*,

1994). To this, 1/100 volume of reducing agent mercaptoethanol was added. A BroadRange high molecular weight (M.W.) protein standard (Bio-Rad, Hercules, CA, USA) used as migrating markers was similarly prepared using 5µl of M.W. marker stock solution. The samples were boiled for 5 min at 95°C to facilitate protein denaturation, chilled on ice for 30 seconds and centrifuged at 13 000 rpm for 10 min at 4°C. The samples were then loaded and resolved through migration on an 8% acrylamide gel (Sambrook *et al.*, 1989, section 18.47) in running buffer (5x stock running buffer is 1.51% Tris-base, 7.21% Glycine, 0.5% SDS in ddH_2O) for 90 min at 150 volts. Electrophoresis was monitored using pre-stained M.W. markers (Bio-Rad).

For western analysis, the gel was washed twice in ddH_2O for 2 min, soaked in transfer buffer (3X stock is 0.9% Tris-base, 1.32% Glycine, 0.6%MeOh, pH 8.3) for 10 min (Sambrook *et al.*, 1989, section 18.60), and then immersed in 100% MeOH and rinsed in transfer buffer for 10 min. After transferring the gel onto a nitrocellulose membrane (Amersham Biosciences) for at least 2 hours, the membrane was washed twice in ddH_2O, and blocked overnight with 1% BSA, in TBST (10X TBST is 1.5 M NaCl, 100 mM Tris-HCl, 1% Tween-20, pH 7.6) with agitation. For the detection of zebrafish progranulin-A translated products, the anti-progranulin-A antibody was diluted at 1:5000. Detection was achieved using an anti-rabbit IgG antibody conjugated to HRP as secondary antibody (Amersham Biosciences), using the ECL detection kit (Amersham Biosciences). Control immunodetection was performed using an anti-mouse beta-catenin monoclonal antibody (#C-7202, Sigma) as primary antiserum, diluted at 1:5000, which cross-reacts with the zebrafish homologue. Membranes were exposed to Kodak X-OMAT film for signal development, and images were analysed using Adobe PhotoShop 7.0 software after scanning.

2.25 Histochemical staining for granulocytes and hemoglobin – Myeloperoxidase histochemical staining was based on the method of Kaplow (Kaplow, 1965) and was performed on whole zebrafish embryos for the detection of granulocytes essentially as described (Lieschke *et al.*, 2001). Briefly, fixation was carried in 1 in 10 dilution of 37% formaldehyde stock in 100% ethanol for 60 seconds, followed by washing in

water for 30 seconds. Excess water was carefully removed. Fixed embryos were transferred in freshly prepared incubation mixture, made by mixing a pre-prepared stock solution (30% ethanol with 3 mg/ml benzidine dihydrochloride [B-3383; Sigma], 1.32 mM $ZnSO_4$, 0.123 M sodium acetate, and 0.0146 M sodium hydroxide) with hydrogen peroxide (20 vol [6%]) in a ratio of 25 ml stock solution to 87.5µl H_2O_2. The stain reaction was allowed to proceed under direct visualization for 1 to 10 minutes until focal staining of cells was evident, and stopped by removing embryos and washing them repeatedly in water. Myeloperoxidase-positive cells were characterized by a blue-black precipitate immediately after staining, but the colour changed to brown within the first 24 hours of storage. Leaching of the precipitate was also observed during further storage in 4% paraformaldehyde in PBS. For histochemical staining of hemoglobin, embryos were placed in freshly prepared *o*-dianisidine stain solution (40% ethanol with 0.01 M sodium acetate, 0.65% H_2O_2, and 0.6 mg/ml *o*-dianisidine [D-9143; Sigma]) for 15 minutes and then were washed in water.

2.26 Analysis of knockdown phenotypes – To monitor the phenotypic consequences of zebrafish progranulin-a knockdown, several marker gene expression and histochemical techniques are used at appropriate developmental timepoints. Cell proliferation is monitored using the mouse monoclonal anti-phospho-histone H3 antibody (06-570; Upstate, Charlottesville, VA) and proliferating cell nuclear antigen (PCNA) mRNA expression. Apoptosis is monitored using the *In Situ* Cell Death Detection Kit (Roche Diagnostics). The following riboprobes have been constructed by amplifying a segment of the full-length mRNA using Taq DNA polymerase, and cloned into pCR2.1 (Invitrogen). After sequencing of individual amplicons for confirmation of their authenticity and verification of their orientation, corresponding antisense riboprobes were synthesized using conventional methods (see above) using either the Sp6 or T7 RNA polymerases. A PCNA riboprobe (1089 nucleotides) corresponding to nucleotides 21-1110 of the PCNA mRNA (accession BC049535); a flk-1 riboprobe (1209 nucleotides) corresponding to nucleotides 17-1226 of the flk-1 mRNA (accession AF487829); a angiopoietin receptor (Tie-2) riboprobe (1072 nucleotides) corresponding to 2471-3543 of the Tie2 mRNA (accession AF053632); a

GATA-2 riboprobe (1159 nucleotides) corresponding to nucleotides 3613-4772 of the GATA-2 mRNA (accession AF001220); a ikaros riboprobe (1109 nucleotides) corresponds to nucleotides 1171-2280 of the ikaros mRNA (accession AF092175); a rag1 riboprobe (1062 corresponds) corresponding to nucleotides 71-1133 of the rag1 mRNA (accession U71093); a sox17 riboprobe (1102 nucleotides) corresponding to nucleotides 11-1113 the sox17 mRNA (accession AF168614); a preproinsulin riboprobe (437 nucleotides) corresponding to nucleotides 1-437 of the preproinsulin mRNA (accession AF036326); a pdx-1 riboprobe (914 nucleotides) corresponding to nucleotides 1-914 of the pdx-1 mRNA (accession AF036325); a trypsin riboprobe (767 nucleotides) corresponding to nucleotides 34-801 of the trypsin precursor mRNA (AF541952); an E-cadherin (cdh1) riboprobe (1093 nucleotides) corresponding to nucleotides 2111-3204 of the cdh1 mRNA (accession AF364811); a N-cadherin (cdh2) riboprobe (1157 nucleotides) corresponds to nucleotides 18-1175 of the cdh2 mRNA (accession AF418565); a neural adhesion molecule L1.1 N-cadherin riboprobe (1240 nucleotides) corresponding to nucleotides 241-1481 of the L1.1 mRNA (accession AY376855); a neural adhesion molecule L1.2 N-cadherin riboprobe (1258 nucleotides) corresponding to nucleotides 86-1344 of the L1.2 mRNA (accession AY376856); a zash1 riboprobe (1855 nucleotides) corresponding to nucleotides 83-1938 of the zash1 (neurod) mRNA (accession NM130978).

The generosity of the following colleagues for their gift of riboprobe constructs is greatly appreciated: Jennifer Rhodes (The A. T. Look lab, Harvard University, Boston) for cebpα, Pu.1, mpo, l-plastin, scl, gata-1, alpha globin, and flk-1; Didier YR Stainier (UCSF, San Francisco) for fkd7 (foxA1) and fkd2 (foxA3); Hitoshi Okamoto and Hiroshi Segawa (RIKEN Brain Science Institute, Japan) for islet-1 and islet-2; Alain Ghysen (Université Montpellier II, France) for the lateral liner markers CB403 and CB701. All other markers (e.g. neurogenin, crestin) were purchased from RZPD (Germany).

2.27 Preparation of digoxygenin-labelled RNA probes –

Linearize 10 µg template DNA,
Extract with phenol/chloroform,
Precipitate with NaOAc + EtOH (sodium acetate and ethanol)
Resuspend in 10 µl RNase-free H_2O (RF H_2O)

Reaction mix:

1 µg linearized template DNA	-	µl
* NTP labelling mix	2	µl
5x transcription buffer	4	µl
RNAse inhibitor (20 u/µl)	1	µl
DTT (0.75 M)	0.3	µl
RNA polymerase (20 u/µl)	1	µl
RF ddH_2O	f inal 20	µl

- incubate at 37°C for 1 hr.
- add 1 µl of RNA polymerase.
- incubate at 37°C for 1 hr.
- add 2 µl DNase I, RNase-free (RF) (20 units/µl) and incubate for 15 min.
- Stop the synthesis reaction and DNase I digestion with 2 µl 0.2M EDTA pH 8.

- precipitate with 2.5 µl 4 M LiCl and 75 µl 100% ethanol. Mix well and keep at –70°C for 30 min.
- spin 30 min in microfuge, wash pellet with 50 µl 70% ethanol, spin 3 min, dry, and resuspend in 50 µl RF H_2O (30 min at 37°C).
- run 1-2 µl on agarose gel to verify probe quality.
- this reaction should yield about 10 µg of RNA. (100-200 ng/µl)

note: In some cases, incubation at 30°C for the transcription reaction of longer transcripts (probes) can help (see Sambrook *et al.*, 1989, section 10.37)

* NTP-labelling mix:

100 mM ATP	2.5µl
100 mM CTP	2.5 µl
100 mM GTP	2.5 µl
100 mM UTP	1.62 µl
10 nM DIG-UTP	8.75 µl
RF H_2O	7.13 µl

2.28 Whole-mount mRNA *in situ* hybridization of fish embryos and larvae (complete protocol) –

Use RNase-free solutions. Unless otherwise stated, the procedures are carried out with embryos in baskets and solutions in 6 well cell culture plates (Corning Inc, Corning, NY) (4ml/well).

1- Fixation I:

Fix embryos at room temperature (RT) for 2 hours or overnight (O/N) at 4°C in 4% paraformaldehyde (PFA) in PBS X1.

Wash 2 x 5 min in PBS

2- Dehydration:

2 x 5 min in 100% methanol (MeOH) at RT

Store embryos in fresh 100% MeOH at -20°C for at least 60 min, or longer (days-months) if storage is required.

3- Rehydration:

1 x 5 min in 75% MeOH / 25% PBS
1 x 5 min in 50% MeOH / 50% PBS
1 x 5 min in 25% MeOH / 75% PBS
3 x 5 min in PBST
(1x PBS with 0.1% tween-20 from 20% stock)
(250 µl 20% tween in 50 ml PBS 1X)

4- Proteinase K digestion:
Stock 20mg/ml; Required Concentration 10µg/ml

2µl of 20 mg/ml Proteinase K in 4ml PBST

blastula & gastrula	30 sec
early somitogenesis	1 min
late somitogenesis	5 min
24hpf embryos	15 min
36hpf-48hpf	30 min
72hpf	45 min
5 dpf	60 min

1 quick wash in PBST
1x 5' in PBST

5- Fixation II:

Fix for 20-60 min in 4% PFA
wash 2 x 5 min PBST

Can stop here. If you chose to proceed, set oven or waterbath at 70°C and defrost deionised formamide (stored at -20 °C).

6- Acetylation: To reduce endogenous phosphatase activity

1x 10 min in acetylation mix (made immediately prior to use)

125 µl Triethanolamine (in solvent cabinet)
27 µl acetic anhydride (in fridge)
10 ml H_2O

Wash 2 x 10 min PBST

Transfer embryos into 1.5 ml centrifuge tubes with wide bore pipette tip. Division by groups is done at this stage.

7- Preadsorption of antibody:

Preadsorb 1/6 total embryos with serum:

Remove the majority of PBST from centrifuge tube, leaving 50 µl

Add 20 µl calf serum (freezer)
20 µl of 100mg/ml BSA (freezer)
2 µl anti-DIG Antibody (fridge)
960 µl PBST

Leave at RT for 2-4 hours, then store in fridge O/N. Take out first thing in the morning to let it reach RT.

8- Prehybridization:

5/6 total embryos at least 3 hours in 200 µl HYB mix at 70°C.

HYB MIX: (Note: mix for washes does not contain tRNA)

50% deionised formamide	5 ml
5x SSC	2.5 ml of 20x SSC
0.1% Tween-20	50 µl of 20% Tween-20
50µg/ml heparin	100 µl of 5 mg/ml stock
citric acid	92 µl of 1M (pH 6.0)
H_2O	add volume to 10 ml

100 µg/ml yeast tRNA stock is 10 mg/ml (kept at -20 °C)

Prepare mix without tRNA, then add tRNA only to make volume needed for pre-hybridization and hybridization.

9- Hybridization: Add 1 µl of probe (100-200 ng)
Hybridize O/N at 70°C

10- Washes: No need to be RNase-free from here onward.
Carried out in six-well plates.

- 10 min in 75% HYB mix / 25% 2x SSC at 70°C
- 10 min in 50% HYB mix / 50% 2x SSC at 70°C
- 10 min in 25% HYB mix / 75% 2x SSC at 70°C
- 10 min in 100% 2xSSC at 70°C
- 2 x 30 min in 0.2xSSC at 65°C

- 5 min in 75% 0.2x SSC / 25% PBST at RT
- 5 min in 50% 0.2x SSC / 50% PBST at RT
- 5 min in 25% 0.2x SSC / 75% PBST at RT
- 5 min in PBST at RT

11- Preincubation:

Incubate the hybridized embryos for 1-4 hours (usually 3 hours), with gentle shaking, in:

PBST	3.2 ml
10% calf serum	400 µl
40 mg/ml BSA from 100 mg/ml stock	400 µl

12- Incubation with anti-DIG antibody (Ab):

Take 800 µl of preadsorbed Ab (step 7) and complete to 4 ml with PBST. Incubate the hybridized embryos for 2-4 hours at RT with gentle shaking, or O/N at 4°C.

13- Washes:

2 quick washes in PBST
(can store embryos overnight or weekend at this stage)

6 x 15 min in PBST at RT
1 x 5 min in STAINING solution at RT:

100 mM	Tris-HCl pH 9.5	1.5 ml of 1 M stock
50 mM	$MgCl_2$	0.75 ml of 1 M stock
100 mM	NaCl	0.3 ml of 5 M stock
0.1 %	Tween-20	0.075 ml of 20% stock
1 mM	levamisol	(fresh solution of 100 mM stock = 24mg/ml)
10 %	polyvinyl alcohol (optional)	

Make up to 15 ml with H_2O

14- Staining:

Stain in the same solution (4 ml), but add the following in this order and mix between each addition:

- 5-Bromo-4-Choloro-3-Indolyl Phosphate (**BCIP**) 14 µl
- Nitro Blue Tetrazolium (**NBT**) 27 µl

Wrap in aluminum foil and stain for 15 min to 2 hours.
Change stain solution every 3 hours if required
(up to 5 days for some transcript targets).

Wash 2 or 3 x for 5-15 min in PBST at RT.

Put embryos in methanol for 30 min. This converts the signal to a true purple colour, and washes away non-specific staining.

Wash again 3 x 5 min in PBS

15- Post-fixation:

4% PFA for 2h at RT
wash in PBS, keep in PBS + 5mM NaAzide at 4°C.

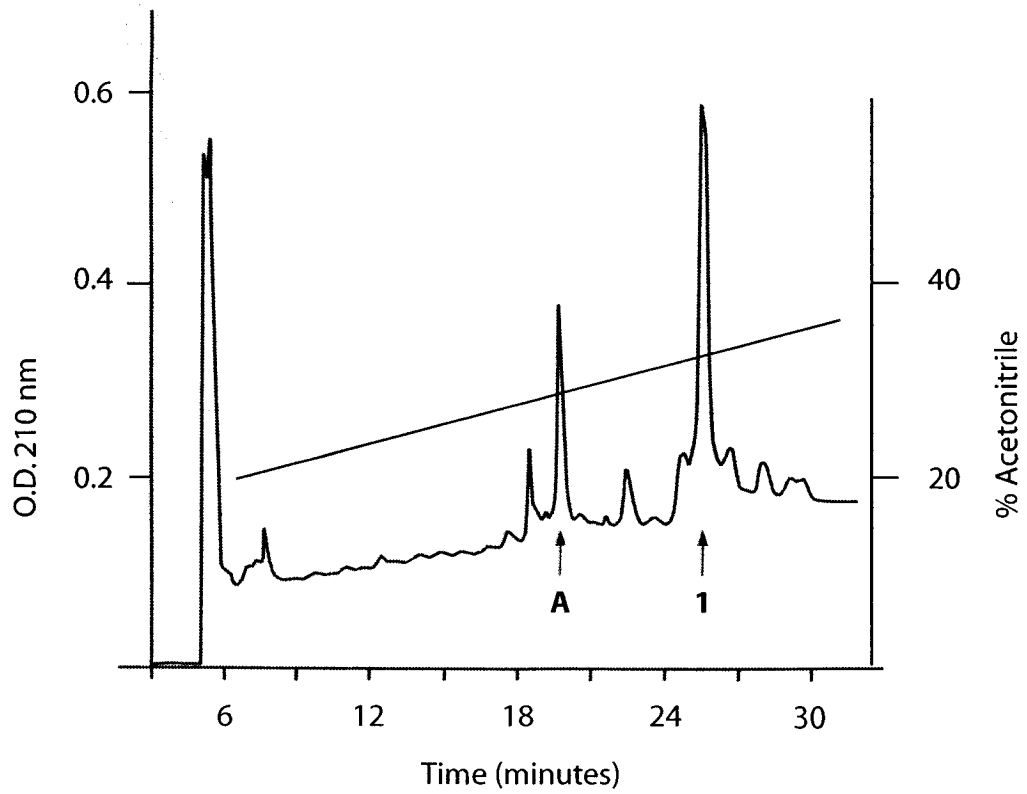
Optional - Glycerol series:
30 min each in 30%, 50%, 70%, 80%, 90%, 95%, 100%.
Glycerol gives pinkish color to yolk. Better for older embryos.

3. Results and Discussion

3.1 Characterization of the zebrafish progranulin gene family

3.1.1 Identification of granulin-1 and granulin-A in extracts of carp spleen - Extracts of carp spleen were subjected to reversed-phase HPLC as described previously using trifluoroacetic acid as counterion (Belcourt *et al.*, 1993). Column fractions were monitored for the presence of granulin-1 immunoreactivity using a previously characterized polyclonal antiserum (Belcourt *et al.*, 1993) in combination with a screen for cysteine content by amino acid analysis. The subsequent purification step using heptafluorobutyric acid as counterion gave rise to two major peaks of apparent homogeneity (**Figure 8, panel A**). Characterization of the material contained within these two peaks was achieved by peptide sequencing and MALDI-TOF mass spectrometry. The complete amino acid sequence of granulin-A was derived by amino-terminal sequencing of the alkylated derivative and of an overlapping chymotryptic fragment (**Figure 8, panel B**). The sequence was confirmed by comparing the observed (5857.9 kDa) and calculated masses (5859.6 kDa). The identity of granulin-1 was resolved by mass spectrometry only, since the signature mass of granulin-1 (6277.1 kDa), identical to the predicted value previously reported (Belcourt *et al.*, 1993), was obtained. The sequences of these carp peptides are compared to the structures of their zebrafish counterparts (**Figure 8, panel B**), which are deduced from the cDNA sequences reported later in this thesis.

A



B

granulin-A

```
CARP    1 DVPCDSTKSCPDGSTCCKTKEGGWACCPLPQAVCCDDFIHCCPHGTTCNVAAGSCDE 57
ZF      1 ....ND.AA.A..T......D.........E....E.........KK.D.......D 57
```

granulin-1

```
CARP    1 VIHCDAATICPDGTTCCLSPYGVWYCCPFSMGQCCRDGIHCCRHGYHCDSTSTHCLR 57
ZF      1 ......Q.V.............I.S...Y.............Q...R......R... 57
```

Figure 8. **Reversed-Phase HPLC purification of granulin-A and granulin-1 peptides from extracts of two carp spleens, and sequence comparison with deduced zebrafish homologues**. *Panel A*: Resolution of HPLC column fractions enriched in cystine and granulin-1 immunoreactivity derived from an extract of two carp spleens into two major components, granulin-A (peak A) and granulin-1 (peak 1). Peaks are indicated by arrows. *Panel B*: Sequence comparison of carp granulin peptides with their respective zebrafish candidate orthologues deduced from cloned cDNA sequences. Numbers correspond to amino acid position. Characteristic cysteines are in bold.

3.1.2 Isolation and characterization of the cDNA encoding a precursor for zebrafish granulin-1 – The carp granulin-1 amino acid sequence was used to design degenerate oligonucleotide primers to make possible a PCR-based approach for the cloning of a cDNA encoding the precursor for the orthologous zebrafish peptide (**Figure 7, panel A**, methods section 2.9). When the PCR was performed using the degenerate DF1 and DR1 primer pair with zebrafish cDNA library as template, several products in the vicinity of the expected size were obtained. In order to increase the specificity of the desired product, an anchor primer (DR2) was used in combination with DF1 to re-amplify an aliquot of the original PCR reaction to yield a product of 126-bp. The sequence of this amplified product was used to design zebrafish granulin-1–specific primers that were used with anchor primers to perform 5'- and 3'-rapid amplification of cDNA ends (RACE) using the zebrafish cDNA library. The zebrafish granulin-1 cDNA amplified by this approach includes a 441-bp ORF that encodes a 147-amino acid protein with a calculated molecular mass of 14.5 kDa after deducting the putative signal peptide sequence (Nielsen *et al.*, 1997). No potential N-glycosylation site (N-X-S/T) is found within this polypeptide (Gupta *et al.*, 2002). The deduced architecture of the granulin-1 precursor consists of one and one-half granulin-like repeats, with the half granulin domain corresponding to the amino terminal moiety of a full granulin peptide (**Figure 7, panel B**). For this reason, this precursor is referred to as progranulin-1. This is in contrast to mammalian progranulin, which encodes one amino-terminal half followed by seven repeats of the granulin motif. The 5'-UTR and 3'-UTR of progranulin-1 are 41- and 153 nucleotides in length, respectively, according to the 5'- and 3'- RACEs performed using cDNA library as template (see above). The putative translation initiation codon deduced from the progranulin-1 cDNA (AAGAUGU) differs by only one residue at position +4 (U instead of G) from the consensus sequence proposed by Kozak (A/GNNAUGG) (Kozak, 1987; Peri and Pandey, 2001). Neither a poly-A tract nor a polyadenylation signal (AAUAAA) was detected using this cloning strategy.

Simultaneously, a zebrafish cDNA encoding a granulin precursor identical to progranulin-1 both in size and in the architecture of its translated product, but

harbouring deduced sequences homologous to carp granulin-2, was noticed in the process of sequencing several independent clones obtained through our cloning strategy. This precursor is regarded as progranulin-2. The nucleotide sequence identity between the progranulin-1 and progranulin-2 ORFs is 92.3%, and substitutions are distributed evenly throughout their entire sequence (data not shown). The deduced polypeptide sequence has a calculated molecular mass of 14.8 kDa after deducting the putative signal peptide sequence (Nielsen *et al.*, 1997). Initially, the authenticity of each clone was further suggested by 1: re-amplifying and sequencing the respective deduced ORFs from independent sources - cDNA prepared from total RNA isolated from staged embryos (a kind gift from Marc Ekker, University of Ottawa) and from whole adult zebrafish); 2: amplifying and sequencing corresponding genomic segments that are bisected by at least one intron and; 3: comparison to deposited zebrafish ESTs at GenBank. The nucleotide sequences for progranulin-1 and progranulin-2, respectively, along with their deduced polypeptide product, are depicted in **Figure 9**. The respective 5'UTR for these cDNAs was also verified by independent 5'RACE (**Figure 10**). The deduced translated sequences for these precursors have been aligned for comparison (**Figure 11**). Interestingly, progranulin-1 and progranulin-2 share an identical leader sequence, a feature sometimes observed for other small peptide precursors, such as members of the defensin gene family (Maxwell *et al.*, 2003). The conserved signal peptide sequence for progranulin-1 and progranulin-2 and the overall high sequence identity may reflect a recent duplication event that gave rise to the two respective precursors. The possibility of a common biosynthetic origin was refuted through isolation of the respective genes (see section 3.1.3). Functionally, it can be speculated that selective pressure may be operating on the signal sequence because of exerted translation initiation control on these precursors in a context-dependent manner such as the innate immune response, as was conjectured for the antimicrobial defensin precursors (Maxwell *et al.*, 2003). The conserved signal peptide sequence for progranulins-1 and -2 may ensure targeting to appropriate vesicular compartments during transit through the endoplasmic reticulum, especially since granulin peptides were inferred to be residing within granules of human neutrophils, hence the derivation of their name (Bateman *et al.*, 1990).

Progranulin-1

5'- AGAACGACACAGACAGCAGCAGAAGCATCCCGAACAGGAGGAAACCCAGCCAGCAAG
ATG TTC CCA GTG TTG ATG TTA CTC ATG GCG GCT CTG GTG GCT GCA
M F P V L M L L M A A L V A A
GAT GAG CCT CTG CTT GAT CTG TCT ATC CCG GTG GAG ACT GTA GAC
D E P L L D L S I P V E T V D
ACC TCT GCC TCT GTG ATT CAC TGC GAT GCT CAA ACT GTG TGT CCA
T S A S **V I H C D A Q T V C P**
GAC GGA ACA ACA TGC TGT CTG AGT CCA TAT GGC ATA TGG TCA TGT
D G T T C C L S P Y G I W S C
TGT CCT TAC TCA ATG GGT CAG TGT TGC AGG GAT GGG ATT CAC TGT
C P Y S M G Q C C R D G I H C
TGT CAA CAT GGA TAT CGC TGC GAT TCC ACC TCC ACC CGA TGC CTG
C Q H G Y R C D S T S T R C L
CGG GGC TGG CTG ACG CTG CCA TCG TCT TTC CAG AAG GCC ACC AGA
R G W L T L P S S F Q K A T R
ACC TTC CAG AAA GAT CAG ACC CAC GCT GAG ACT GTC CAA TGT GAA
T F Q K D Q T H A E T V Q C E
GGA AAT TTC TAC TGC CCG GCT GAG AAG TTT TGC TGC AAG ACT AGA
G N F Y C P A E K F C C K T R
ACT GGC CAG TGG GGC TGC TGC AGC GGG TTG GAG TTG **TAA**GCAAACAA
T G Q W G C C S G L E L .
CCACTGTTACACTGGTTCACCTCTTACTGAAGACTGTCAAAATGCAAGAGGCTCCAAAC
TTCATTCATGTTTCAGGCTATATTCTTACATCCAGGCTGAATTAGTGTGAAATAAAAAC
TAATAGAATTACTGTCAAAATGCACCGTCAAGGG -**3'**

Progranulin-2

5'- AGAATGACACAGACAGCAGCAGAAGCATCCCGAACAGAAGGAAAACCAGCCAGAAAG
ATG TTC CCA GTG TTG ATG TTA CTC ATG GCG GCT CTG GTG GCT GCA
M F P V L M L L M A A L V A A
GAT GAG CCT CTG CTG GAT CTG TCC ATT CCG ATG GAG ACT GAA GAC
D E P L L D L S I P M E T E D
GTC TCT GCC TCT GTG ATT CAC TGC GAT GCT CGA ACT GTG TGT CCA
V S A S **V I H C D A R T V C P**
GAT AGA ACA ACA TGC TGT CGG ACT CCA TAT GGC AAA TGG ACA TGT
D R T T C C R T P Y G K W T C
TGT CCC TTC CCA ATG GGT CAG TGT TGC AGG GAT GGG ATT CAC TGC
C P F P M G Q C C R D G I H C
TGT CGA CAC GGT TAT CGC TGC AAT TTC GCC TCC ACC AGA TGC CTG
C R H G Y R C N F A S T R C L
CGA GGC TGG CTG TCA CTG CCA TCG TCT TTC CAG GAG GCC ACC AGA
R G W L S L P S S F Q E A T R
ACC TTC GAG AAA GAT CAG ACC CAG GCT GAG ACT GTC CAA TGT GAA
T F E K D Q T Q A E T V Q C E
GGA AAT TTC TAC TGC CCG GCT GAG AAG TTT TGC TGC AAG ACT GGA
G N F Y C P A E K F C C K T G
ACT GGC CAG TGG GGC TGC TGC AGT GAG TTG GAG TTG **TAA**GCAAACAA
T G Q W G C C S E L E L .
CCACTGTTACACTGGTTCACCTCTTACTGAAGACTGTCAAAATGCAAGAGGCTCCAAAC
TTCATTCATGTTTCAGGCTATAGTTTACTCTTACTGAATTAGTGTGAAATAAAACTAAA
AGAATTACTGTCAAAATGCAC -**3'**

Figure 9. **Complete nucleotide sequence and deduced translated sequence for zebrafish progranulin-1 and progranulin-2.** Predicted translation initiation (ATG) and termination (TAA) codons are underlined and in bold. Deduced granulin-1 and granulin-2 peptide sequences are in bold. Both cDNAs encode a 441 nucleotide ORF that encodes a 147-amino acid polypeptide with a calculated molecular mass of approximately 14.5 kDa and 14.8 kDa, respectively, after deducing the putative signal peptide corresponding to residues 1- 16 (italics).

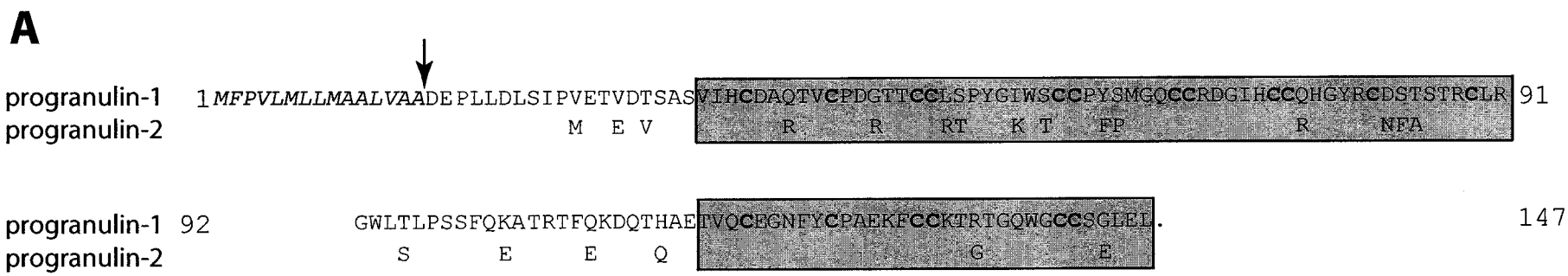
```
prograulin-1  5'-AGAACGACACAGACAGCAGCAGAAGCATCCCGAACAGGAGGAAACCCAGCCAGCAAGATG
prograulin-2         T                                    A      A       A

prograulin-1     TTCCCAGTGTTGATGTTACTCATGGCGGCTCTGGTGGCTGCAGATGAGCCTCTGCTTGAT
prograulin-2                                                              G

prograulin-1     CTGTCTATCCCGGTGGAGACTGTAGACACCTCTGCCTCTGTGATTCACTGCGATGCTCAA
prograulin-2          C  T   A         A    GT                            G

prograulin-1     ACTGTGTGTCCAGACGGAACAACATGCTGTCTGAGTCCATATGGCATATGGTCATGTTGT
prograulin-2                  TA               G  C           A    A
```

Figure 10. **Characterization of the 5′ untranslated sequences of zebrafish progranulin-1 and progranulin-2**. The rapid amplification of cDNA ends (RACE) technique was performed to characterize the 5′ untranslated regions and transcription start sites corresponding to the progranulin-1 and progranulin-2 transcripts. No amplicons were obtained suggesting the occurrence of these sequences on a common transcript, suggesting that hybrid progranulin may result through a mechanism other than alternative splicing of a larger bicistronic pre-mRNA. Similar conclusions were obtained for the 3′ RACE (data not shown). Translation initiation codon is underlined and in bold.



B

granulin-1

carp 1 VIHCDAATICPDGTTCCLSPYGVWYCCPFSMGQCCRDGIHCCRHGYHCDSTSTHCLR 57
zebrafish Q V I S Q R R

granulin-2

carp 1 VVYCNARTTCPSRTTCCRSPFGVWYCCPFLMGQCCRDGRHCCRHGYRCDSTSTLCLR 57
zebrafish IH D V D T Y K T P I NFA R

Figure 11. **Sequence alignment of deduced translated sequences for zebrafish precursors encoding granulin-1 and granulin-2.** *Panel A*: 2 additional cDNAs from zebrafish, each with a predicted 441 nucleotide-long open reading frame (ORF), respectively encode a deduced precursor protein carrying one and one amino-terminal half granulin repeats (boxed) whose sequences are distinct from those contained within the larger co-orthologues of mammalian progranulin. These two precursors, called *progranulin-1* and *progranulin-2*, share 92.3% identity at the nucleotide level, and 85.7% identity in their overall translated sequences. Both carry an identical signal peptide (italics) (von Heijne, 1986; Nielsen *et al.*, 1997), whose predicted cleavage site is indicated by an arrow. *Panel B*: Sequence comparison of translated sequences indicates the deduced zebrafish granulin peptides are likely orthologues of carp granulin-1 and granulin-2, respectively. Conserved arginine residues for granulin-2 are shown in red.

3.1.3 Cloning of the zebrafish progranulin-1 gene – A probe corresponding to the progranulin-1 ORF was used to screen high density filters representing a zebrafish genomic DNA library. One clone (706K2254Q) was further fragmented and re-screened twice with the same probe and an additional oligonucleotide probe, respectively, for the isolation of two non-overlapping fragments containing the entire cDNA ORF. The two fragments were used to determine the zebrafish progranulin-1 gene structure and sequence (**Figure 12**). The zebrafish progranulin-1 gene is encoded on 5 exons, and spans 4.2 kb in length. There is one nucleotide substitution between the zebrafish gene and the respective cDNA sequences: the substitution of T at nucleotide position 297 in the cDNA sequence to C in the corresponding gene sequence (data not shown). This substitution makes no change in the corresponding amino acid sequence. The size of exons and introns are represented in **Figure 13** (see below). Intron phase classes for this gene and for others are represented in **TABLE 6**. All exon/intron boundaries are consistent with the consensus GT/AG sequence at the donor and acceptor sites for RNA splicing (Breathnach and Chambon, 1981; Mount, 1982). The translation initiator codon, ATG, is encoded in the second exon (exon 2). A canonical TATA(A/T)A box sequence is located in the promoter region of the gene, at nucleotide position –784. A putative poly(A) tail of the zebrafish granulin-1 precursor cDNA can be added 21 nucleotides downstream of a consensus signal sequence (AATAAA) located in exon 5. Extension of the 5'- and 3'-UTRs relative to the cloned cDNA was made possible using the sequences from GenBank EST clones BG884011 and AW232436, respectively.

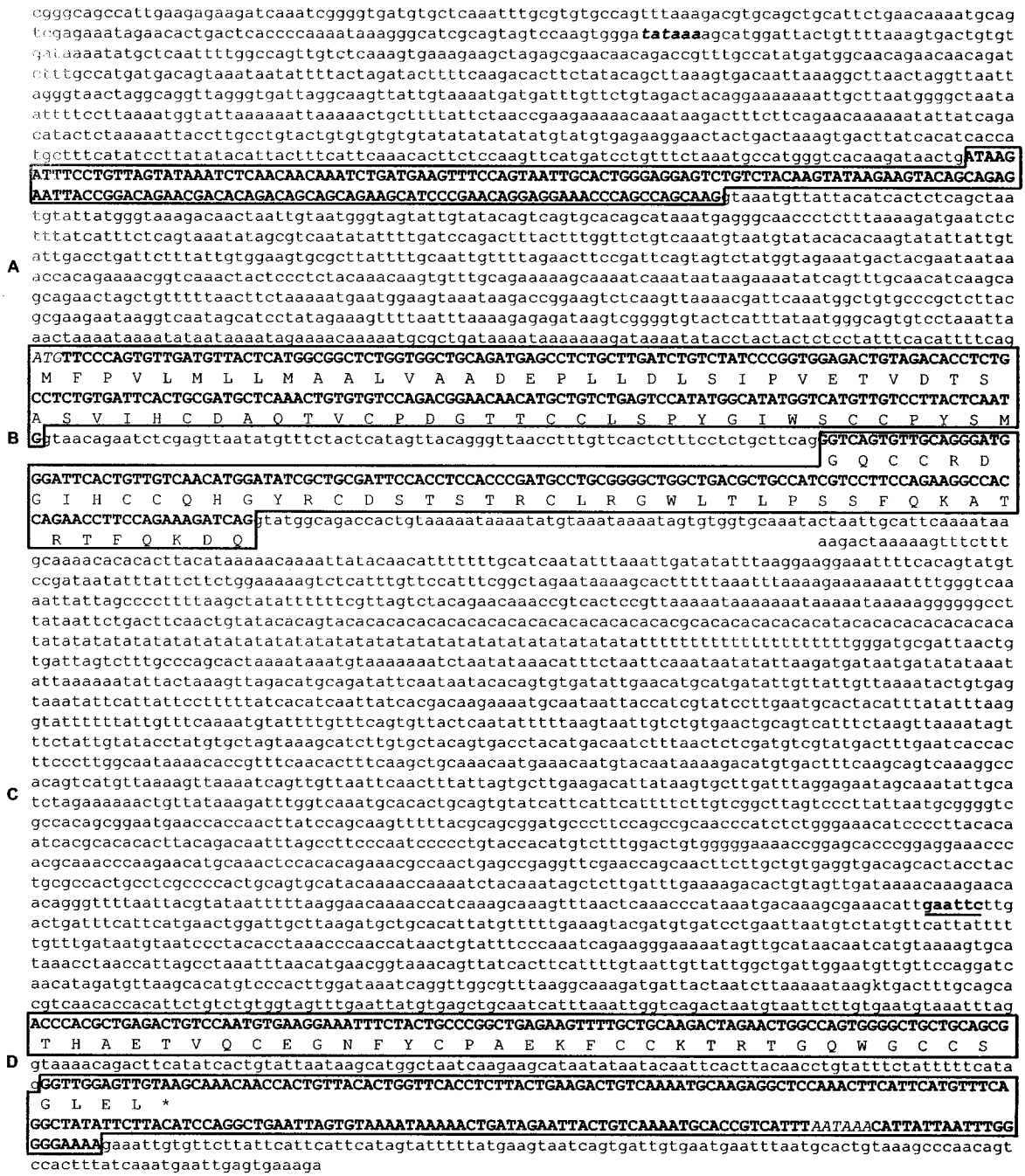


Figure 12. **Genomic architecture of the zebrafish progranulin-1 gene.** Exons (1-5) and deduced amino acid sequences are shown in uppercase letters and are boxed, while intron (A-D) and flanking sequences are shown in lowercase. 5' extension of exon 1 is based on a deposited EST sequence (accession number BG884011). A potential TATA box in the promoter region is shown in bold and italics. The translation initiation codon (ATG) and polyadenylation signal sequence (AATAAA) are in italics. The EcoRI restriction site (gaattc) used for cloning the fragments containing this gene is located in intron C, and is underlined and bold (see methods).

Close inspection of the genomic organization of the progranulin-1 gene reveals that the 12-cysteine motif of the granulin-1 peptide is encoded on two separate exons in a matter identical to that seen in the mammalian progranulin gene (Bhandari and Bateman, 1992; Baba *et al.*, 1993b) (**Figure 13**). This implies the joining of two peptide moieties bearing six cysteines, corresponding to the amino- and carboxyl-terminal halves of the full peptide, respectively. Specifically, these sequences arise through a conserved splicing of the amino-terminal fragment occurring 4 amino acids after cysteine number 6 (exon 2), and two amino acids before cysteine 7 (exon 3).

Several palindromic sequences, inverted and direct repeats are localized upstream and within intronic sequences of the progranulin-1 gene (data not shown); sharing significant homologies with portions of zebrafish ESTs deposited at GenBank. Some of these are reminiscent of the miniature inverted-repeat transposable elements (MITEs), which are frequently found in proximity to genes (Izsvak *et al.*, 1999). A characteristic of MITES is the potential to form stem-loop or cruciform structures *in vitro* (Izsvak *et al.*, 1999). There are indications that the distribution of stem-loop structures throughout eukaryotic genomes may facilitate homology search and pairing of homologous chromosomes during meiosis (Forsdyke, 1995). The presence of cruciform structures may alleviate torsional stress during transcription thus providing the basis for the conservation of MITES in the genomes of zebrafish and other eukaryotes (Izsvak *et al.*, 1997). Interestingly, some of the sequences surrounding the progranulin-1 gene exonic sequences sharing high identity with zebrafish ESTs of unknown identity are also detected within the zebrafish Apolipoprotein E, Bone Morphogenetic Protein 4 (BMP4), and GATA-1 genes (data not shown). The significance of this observation is not known.

A candidate long interspersed element (LINE) retroposon is also found approximately 3.5 kb upstream of the progranulin-1 gene, which shares similarity to a retrovirus-related pol polyprotein (reverse transcriptase) encoded by a zebrafish EST (accession number: AI545754), and to a Drosophila retrotransposon (accession number: DBU133521) (data not shown). A non-redundant BLAST search

(http://www.ncbi.nlm.nih.gov/BLAST/) for genomic sequences deposited at NCBI using the translated sequence of this candidate retroposon also revealed extensive identity to several areas repeated within the immunoglobulin light chain gCL5 gene cluster of Gadus morhua (accession number: AF104899). Finally, a candidate member of the DANA family of zebrafish short interspersed elements was found within intron C (Izsvak *et al.*, 1996) (**Figure 13**).

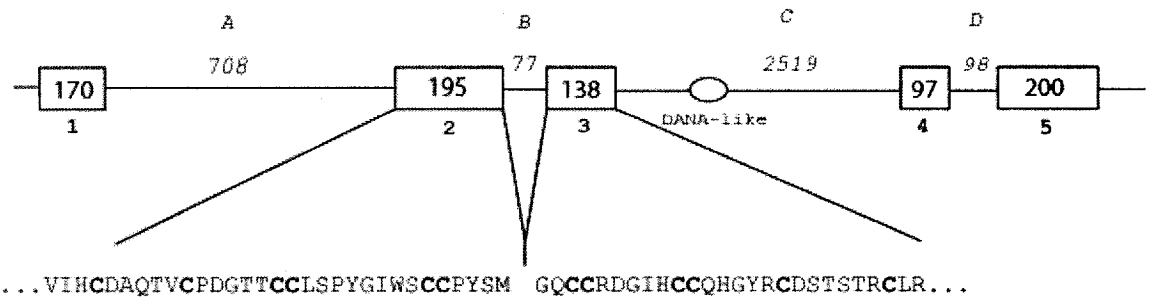


Figure 13. **Conserved exonic organization of zebrafish progranulin-1 and progranulin-2 genes relative to mammalian progranulin.** Like mammalian granulin repeat units, the nucleotide sequences encoding zebrafish granulin-1 and granulin-2 are derived from the joining of two spliced exons with phase 0 boundaries. The characteristic splice positions occur respectively at 4 residues after cysteine 6 and 2 residues before cysteine 7, of the translated granulin peptide sequence (see Bhandari and Bateman, 1992; Baba *et al.*, 1993b).

3.1.4 Cloning of the zebrafish progranulin-2 gene – While in the process of cloning the gene for zebrafish progranulin-2 using a strategy identical to that used for progranulin-1, a genomic clone deposited at GenBank (accession number AC124903) was noticed as part of the zebrafish genome sequencing initiative. This sequence was found to harbour the full exon-intron complement to the *progranulin-2* gene, as well as that of *progranulin-1*. The gene for *progranulin-2* possesses an exonic organization identical to that seen for *progranulin-1*. The splice junctions for the *progranulin-2* gene, as well as the phase of the introns, are depicted in **TABLE 6**.

5' donor	consensus splice site sequences	3' acceptor
	(C/A)**AG / gt** (a,g)agt.....**INTRON**..... .(c,t)**ag / G**	
		splice phase
	progranulin-1	
	(A)AG / gt (a) aat.....intron A.....(c)ag / A	0
	(A)TG / gt (a) aca.....intron B.....(c)ag / G	0
	(C)AG / gt (a) tgg.....intron C.....(t)ag / A	0
	(G)CG / gt (a) aaa.....intron D.....(t)ag / G	1
	progranulin-2	
	(A)AG / gt (a) aat.....intron A.....(c)ag / A	0
	(A)TG / gt (a) aga.....intron B .....(c)ag / G	0
	(C)AG / gt (a) tgg.....intron C.....(t)ag / A	0
	(G)CG / gt (a) aaa.....intron D.....(t)ag / A	1
	antisense progranulin	
	(G)TG / gt (a) aga.... ...intron A.....(c)ag / T	N/A
	(T)AG / gt (a) t(a/g)t.... intron B.....(c)ag / T	
	(T)AG / gt (a) tgt.... .... intron C.....(c)ag / T	

TABLE 6. **Splice junctions of the zebrafish progranulin-1, progranulin-2, and the non-coding RNA antisense gene.** The consensus sequence for splice donor and acceptor sites is shown on the top line (Breathnach and Chambon, 1981). The nucleotide sequences surrounding the sites for introns A-D for the *progranulin-1* and *progranulin-2* genes, respectively, as well as for introns A-C of the antisense gene, are shown. The sequences of exons (uppercase) and introns (lowercase) are indicated. The phase of the introns interrupting open reading frames are shown.

3.1.5 Cloning of a transcript harbouring a hybrid structure corresponding to portions of progranulin-1 and progranulin-2 – The first clone fully sequenced among a series of cDNA sequences encoding precursors for zebrafish peptides found to be closely related to carp granulin-1 (**Figure 8, panel A, step 4**) revealed a structure whose deduced architecture was later shown to be a candidate hybrid between progranulin-1 and progranulin-2. The obtained nucleotide sequence for this hybrid cDNA is compared to two allelic variants for *progranulin-1* (grn1a and grn1b) and to *progranulin-2*, the latter three sequences derived from cloned and sequenced cDNA as well as genomic sources to minimize the intrusion of sequencing errors (**Figure 14, panel A**). Careful inspection of the alignments of available sequences indicates that the only nucleotide substitution (nucleotide 429, T) found within the hybrid sequence that does not match either the corresponding sequence (C) within the *progranulin-1* or the *progranulin-2* transcripts (located at the end of exon 4 for these genes). Interestingly, all other substitutions for the hybrid find a corresponding match exclusively in *progranulin-1* (exon 2), or *progranulin-2* (exons 3 and 4). This is in contrast to the polymorphic character of the substitution pattern observed between *progranulin-1a* and *1b* (**Figure 14, panel A**).

The deduced translated sequence of the progranulin hybrid is compared to those for progranulin-1 and progranulin-2, and highlights the architecture of this hybrid progranulin, which consists of the amino-terminal half derived from progranulin-1, coupled to the carboxyl-terminal half derived from *progranulin-2* (**Figure 14, panel B**). A schematic representation of this cDNA is depicted in **Figure 14, panel C**.

A

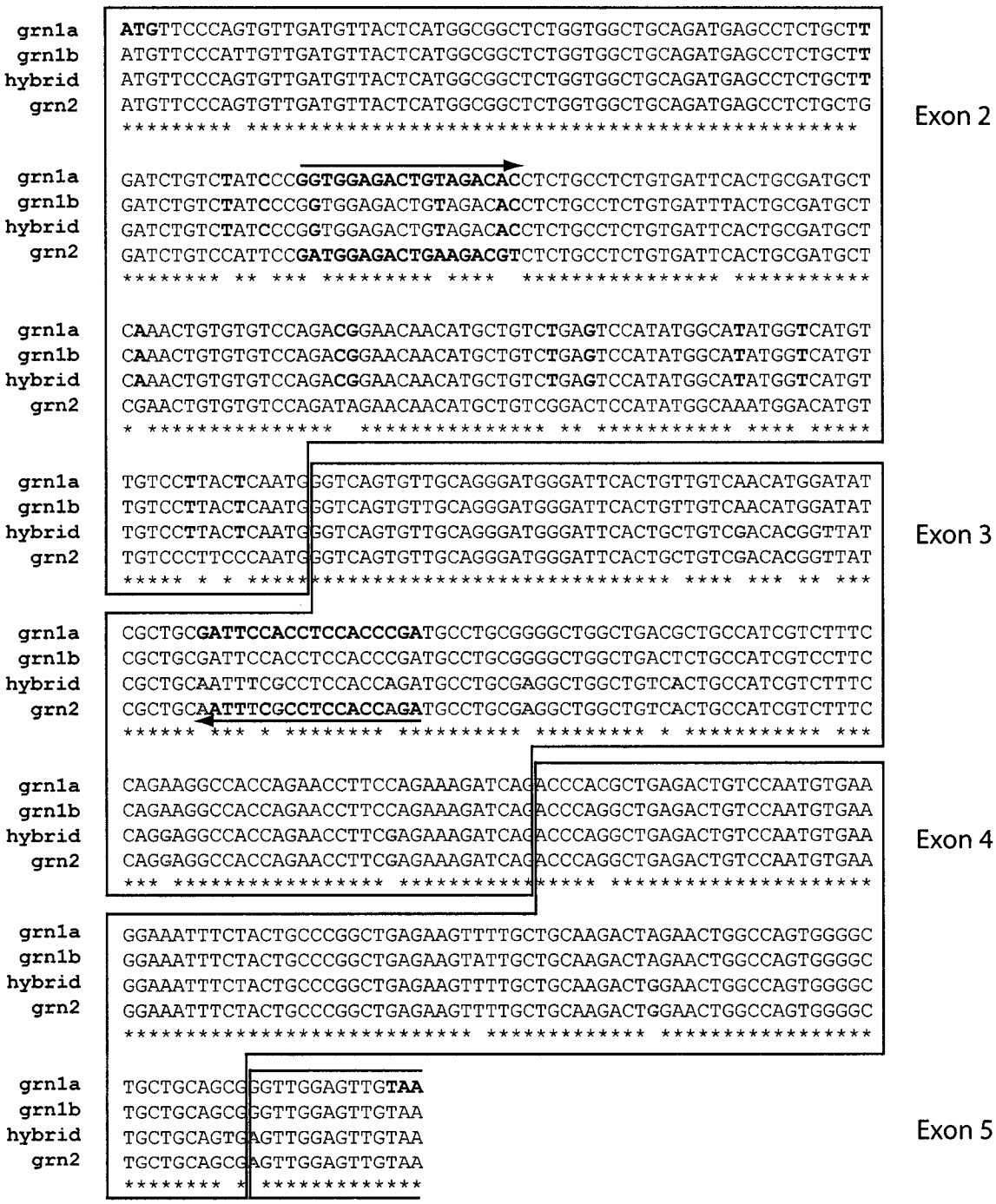


Figure 14

B

```
                intron A
progranulin-1   ▽ MFPVLMLLMAALVAADE PLLDLSIPVETVDTSAS
hybrid
progranulin-2                               M  E V

                                                  intron B
progranulin-1   VIHCDAQTVCPDGTTCCLSPYGIWSCCPYSM ▽ GQCCRDGIHCCQHGYRCDSTSTRCLR
hybrid                                                        R     NFA
progranulin-2         R      R   RT    K T   FP               R     NFA

                                     intron C                               intron D
progranulin-1   GWLTLPSSFQKATRTFQKDQ ▽ THAETVQCEGNFYCPAEKFCCKTRTGQWGCCS ▽ GLEL.
hybrid             S      E     E      Q                      G        E
progranulin-2      S      E     E      Q                      G        E
```

C

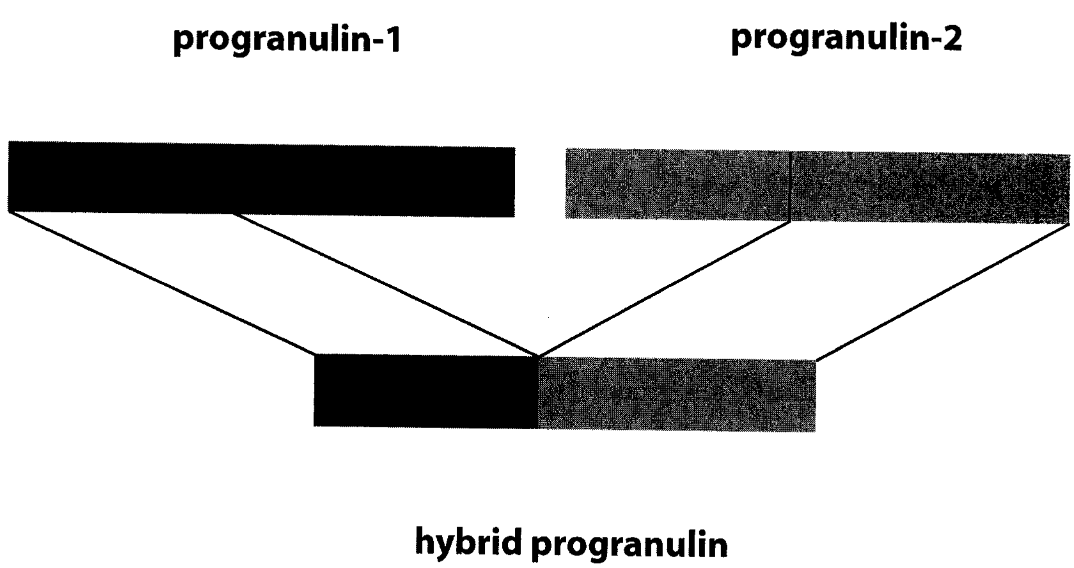


Figure 14

Figure 14. **Cloning of a candidate chimeric progranulin cDNA.** *Panel A*: Nucleotide sequence alignment of the ORFs for progranulin-1 allelic variants (1a and 1b), a chimeric progranulin, and progranulin-2. The cDNAs possess an identical 3'UTR (not shown). The cDNA nucleotide sequences for progranulin-1 and progranulin-2 were verified by comparing their respective exonic segments (boxed). Note that the nucleotide substitutions for the hybrid transcript are distributed in a non-random manner among the corresponding exonic sequences for the progranulin-1 and progranulin-2 genes (highlighted in blue for progranulin-1, exon 2; and in orange for progranulin-2, exons 3-5), suggesting that its origin is other than though polymorphism. Progranulin-1 and progranulin-2 are 92.3% identical, and both respectively share 95.6% nucleotide identity with the hybrid. The translation initiation (ATG) and termination (TAA) codons are in bold. Arrows indicate primers used for RT-PCR analyses. *Panel B*: Sequence alignments of deduced translated sequences for progranulin-1, progranulin-2 and hybrid progranulin. Introns (A-D) located in the respective progranulin-1 and progranulin-2 genes are indicated by arrowheads (see later). *Panel C*: Schematic representation of the hybrid progranulin cDNA. The candidate chimeric transcript consists of the joining of the N-terminal portion of progranulin-1 (exons 1 and 2) and of the C-terminal portion of progranulin-2 (exons 3 to 5).

3.1.6 Cloning of a cDNA with complementarity to zebrafish progranulin-1 and progranulin-2 – During the cloning of the progranulin-1 gene, an EST deposited at GenBank (accession AW777232) was noticed whose sequence was an exact match to a portion of the *progranulin-1* gene, but apparently in the incorrect orientation. Specifically, part of the 5'read of this incompletely sequenced EST encompassed exons 2 and 3 of the *progranulin-1* gene, as well as the intervening intron B (**Figure 15**).

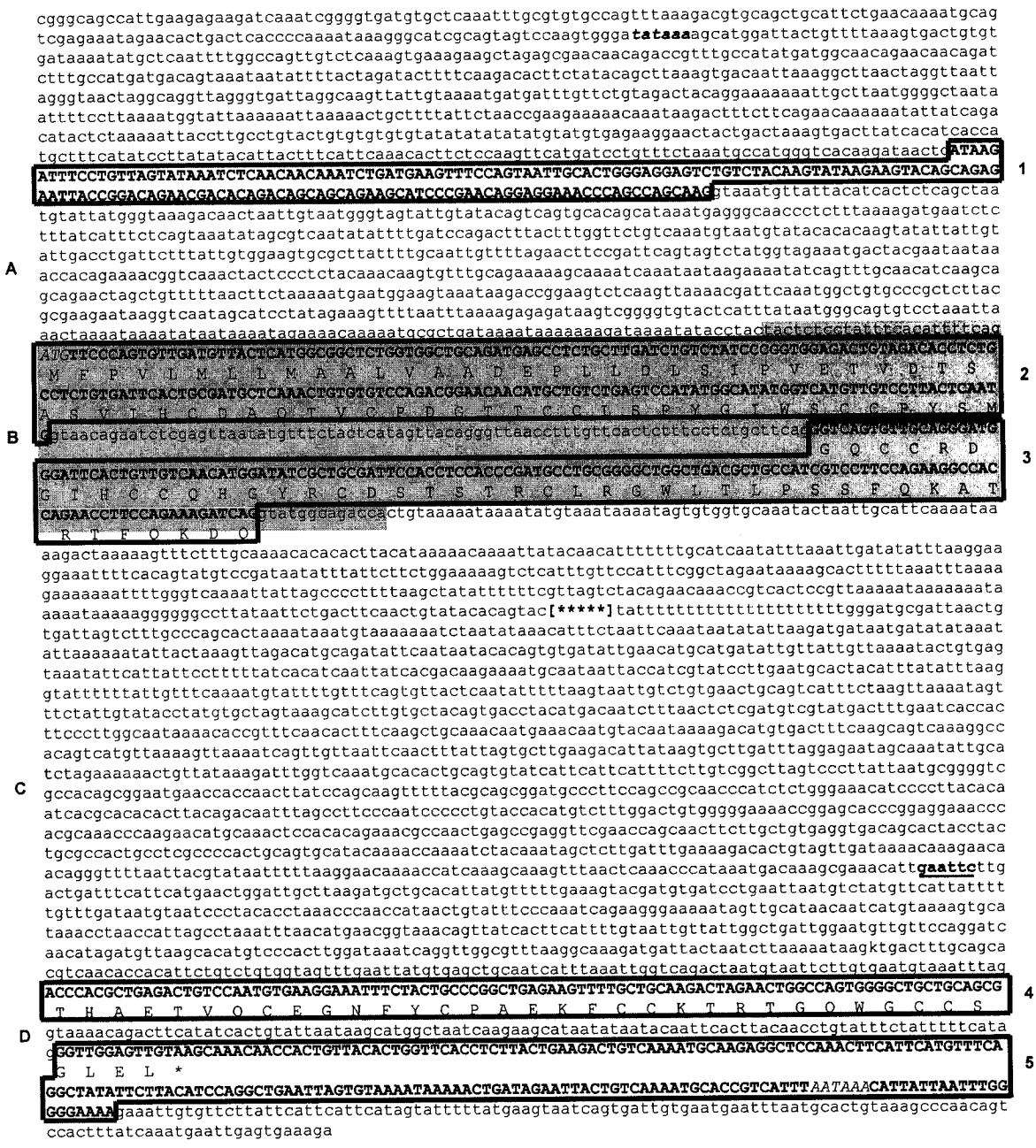


Figure 15. **Detection of an EST whose sequence is complementary to parts of the progranulin-1 gene.** Part of the EST (accession AW777232) is the exact reverse complement sequence (pink) of exons 2 and 3, as well as part of the flanking introns, constituting the progranulin-1 gene. An identical pattern of antisense complementarity is observed between exon 2 of the *antisense* gene and the corresponding region of the *progranulin-2* gene (not shown).

This EST was obtained from RZPD (see methods) and sequenced in full on both strands (**Figure 16**). This cDNA is 1989-nucleotides long, includes a polyadenylated tail, and is encoded on four exons derived from a gene that spans >19.8 kb (presented later). Overall, this transcript does not appear to give rise to a predictable ORF in view of the presence of several termination codons in all three possible reading frames. Although it remains possible that this RNA gives rise to undiscovered small peptides resulting from the use of internal ribosome entry sites (IRES) (data not shown), this is improbable. This conclusion was reached because of a lack of observed sequence conservation while performing *tblastn* searches at GenBank for translated sequences using these conceptual peptides. For this reason, this gene transcript is considered to be non-protein coding.

Further inspection of this RNA reveals other interesting and unconventional features, which are enumerated below (**Figure 16**). First, part of the non-coding RNA is complementary to genomic portions of both the progranulin-1 and progranulin-2 genes, thereby establishing this non-coding RNA as a naturally occurring antisense (AS) transcript. Because of its transcriptional overlap and partial complementarity with the progranulin-1 and progranulin-2 genes, the AS gene is referred to as the AS progranulin-1/-2 gene. As was presented above (**Figure 15**), a complementary sequence to part of the progranulin-1 gene encompassing exons 2 and 3 as well as flanking intronic sequences is reiterated immediately downstream of this sequence, but is complementary to the identical portion of the progranulin-2 gene. This demonstrates that the progranulin-2 gene is located upstream of the progranulin-1 gene and are arranged in a head-to-tail organization. Further, this appears to represent the first example of a single spliced transcript with antisense complementarity to two tandemly organized paralogous protein-coding genes. It having been established that the middle portion of the antisense RNA is complementary to portions of the progranulin-1 and progranulin-2 genes, the remaining 5' and 3' sequences of this transcript were examined for additional features. While the most 5' portion of the AS RNA is not similar to any known sequence, the remaining sequence downstream of the reverse complement progranulin-2 gene segment is a mutated (i.e. defective) transposase gene

belonging to the tzf subclass of the *Tc1/mariner* superfamily of class II DNA transposable elements (Lam *et al.*, 1996) (**Figures 17 and 18**; see section 3.2.5 for more details). It is important to consider that the tzf element is, like the progranulin gene segments, in reverse complement orientation relative to the AS transcript; thus, the AS progranulin-1/-2 gene has never possessed the ability to encode a translatable transposase protein.

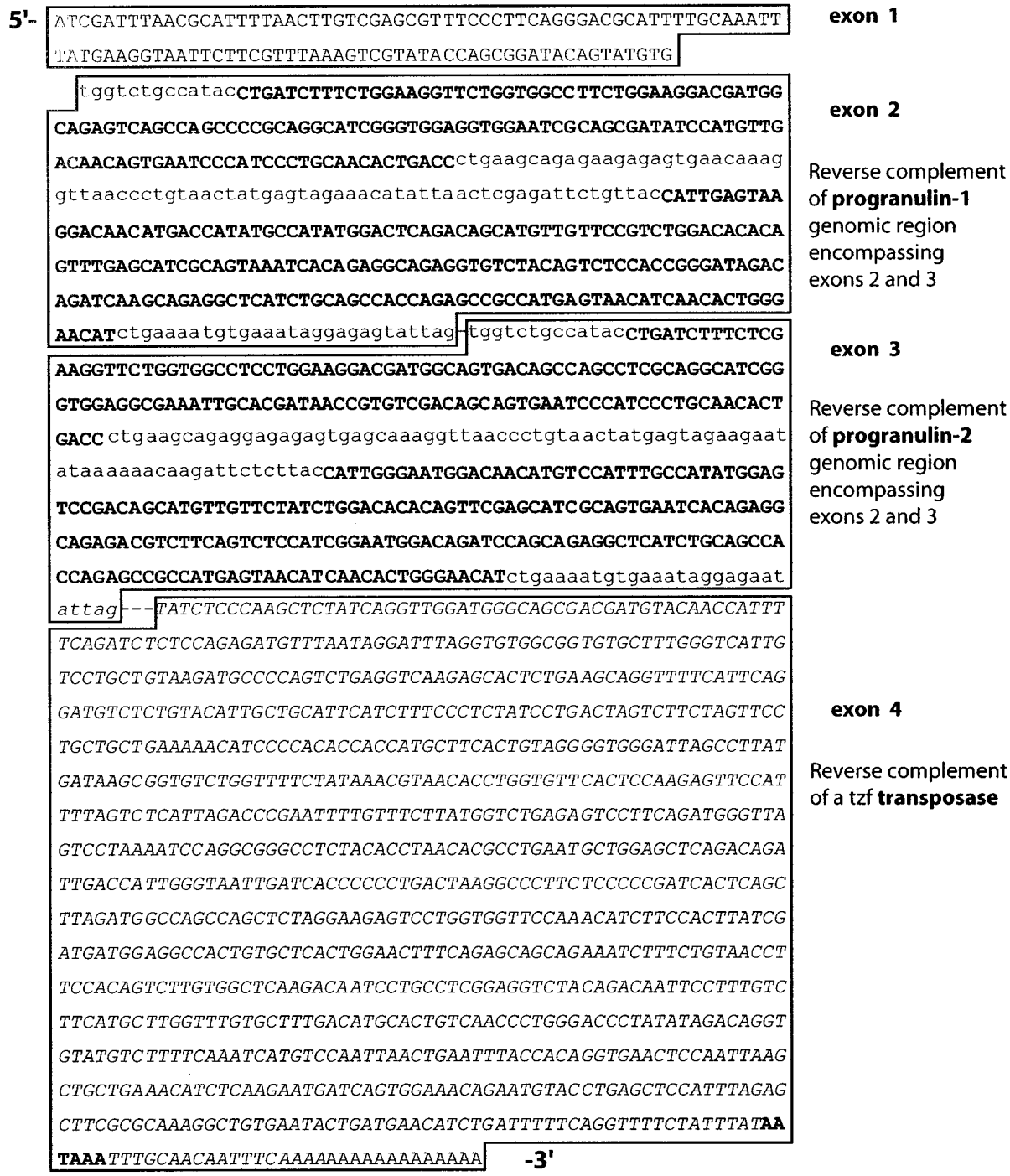


Figure 16. **Nucleotide sequence of the transcript antisense to the progranulin-1 and -2 genes.** The antisense RNA is 1989 nucleotides in length, is polyadenylated due to the presence of polyadenylation signal (AATAAA), and is encoded on four exons (boxed). Sequences derived from exons 2 and 3 are complementary to regions encompassing the second and third exons (uppercase and bold) and intronic sequences (lowercase) of the progranulin-1 and progranulin-2 genes, respectively. The nucleotide sequence of exon4 corresponds to a mutated transposase gene of the *tzf* transposon sub-class of the *Tc1/mariner* superfamily, but in the reverse complement orientation. The antisense transcript appears to be a non-coding RNA due to the absence of a predictable open reading frame.

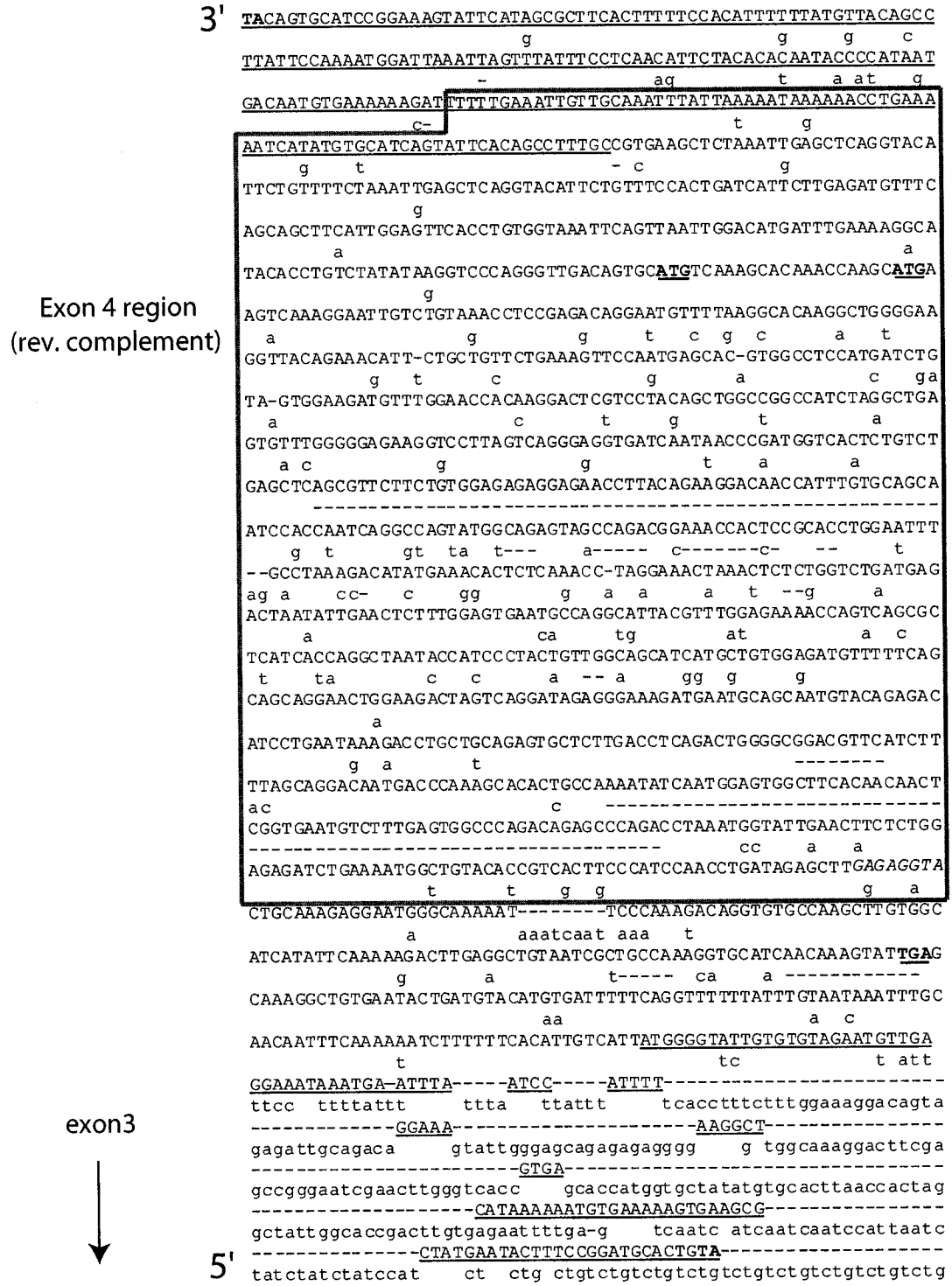
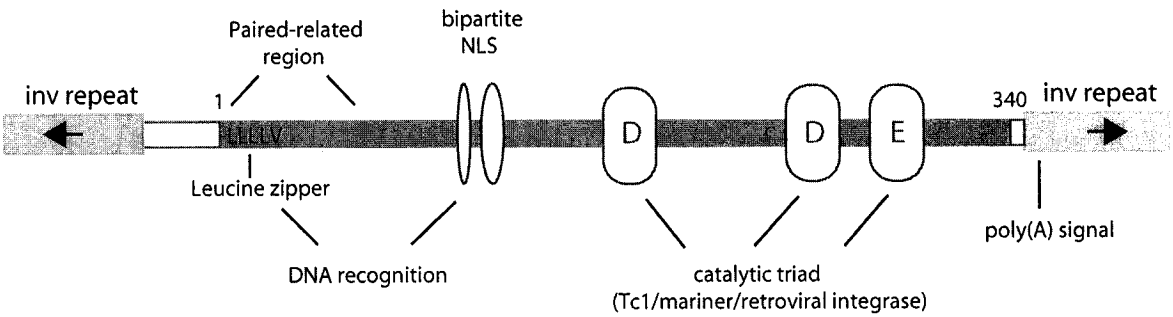


Figure 17

Figure 17. **The genomic region encompassing exon 4 of the antisense gene encodes a transposon of the tzf family, in the reverse orientation.** Reverse complement sequence of the antisense gene (i.e. sense orientation to progranulin-1 and progranulin-2) in the region encompassing exon 4 (boxed) is shown. A consensus 1621 nucleotide sequence for the *tzf* transposon (top, uppercase) was deduced by the majority rule of aligned sequences obtained from Genbank EST entries (U51226-U51230) and published data (Lam *et al.*, 1996). Terminal dinucleotides (TA) used for integration are indicated in bold. Flanking 200 nucleotides inverted repeats are underlined. Two in frame translation initiation (ATG) codons and a termination (TGA) codon are indicated in bold and underlined. Conceptual translation from the first ATG gives rise to the transposase protein product. Aligned sequences corresponding to the antisense genomic region are shown underneath the transposon sequence, with mismatches indicated in lowercase. The dashed lines within the genomic sequence (underneath) represent deletions relative to the consensus tzf (above) sequence. Conversely, the presence of these dashed lines within the tzf (above) sequence, denote inserted sequences within the corresponding region of the genomic sequence (underneath). Note that intronic sequences flanking exon 4 indicate that the remaining *tzf* sequences, including inverted repeats, are preserved, confirming that this region belongs specifically to the *tzf* transposon family.

A



B

```
Tzf     MSKHKPSMKTKELSVDLRDRIVSRHKAGEGYRNISAALKVPMSTATSIIRKWKMFGTTRT
AS grn  MSKHKPSMKTKELSVDLRGRIVLSHKTVEGYRKISAALKVPVSTVASIIDKWKMFGTTRT
        ****************** ***  **  **** ********:**.:*** **********

Tzf     RPRAGRPSKLSVRGRRSLVREVINNPMVTLSELQRSSVE-R-GEPYRRTTICAAIHQS--
AS grn  LPRAGWPSKLSDRGRRALVRGVINYPMVNLSELQHSGVEGRCLEPPG-LFLELAGHLS.V
         **** ***** **** *** *** *** ***** * ** *  **         * * *

Tzf     -GLYGR------VARR----KATPHLEFAKR--------HLKDSQNIRNKILWSDETKIE
AS grn  IGGEGP.SGG.SITQWSICLSSSIQAC.V.RPAWILGLTHLKDSQTIRNKIRV..D.-NG
         *  *                          *        ****** *****

Tzf     LFGVNARRYVWRKPGSAHHQANTIPTVKHGGGSIMLWGCFSAAGTGRLVRIEGKMNAAMY
AS grn  TLGVNTRCYVYRKPDTAYHKANPTPTVKHGG.SMVVWGCFSAAGTRRLVRIEGKMNAAMY
          *** * ** ***  * * ** ******** *   ************************

Tzf     RDILNKDLLQSALDLRLGRRFIFQQDNDPKHTAKISMEWLHNNSVNVLEWPSQSPDLNPI
AS grn  RDILNENLLQSALDLRLGDWGILQQDNDPKHTAECS.PQT-GASYSRTMTQSTPPHLNPI
        *****  ***********  ***********          *       *  * ****

Tzf     EHLWRDLKMAVHRRFPSNLIELERYCKEEWAKIPKDRCAKLVASYSKRLEAVIAAKGASTKY
AS grn  KHLWRDLKMVVHRRCPSNLIELGRYCTSSLP-IQPDR-----AW-----EILIFSY--FTF
        ******** **** ******* ***         **     *      *  *      *
```

Figure 18

Figure 18. **The reverse complement sequence of the 3'end region of the antisense RNA corresponds to a mutated tzf transposase.** Panel A: Schematic representation of the structural features of *Tc1*-like transposable elements (Ivics, 1996). N-terminal region of the transposase contains a DNA-binding paired domain similar to the Pax family of transcription factors (Walter *et al.*, 1991). Some elements also contain an amphipatic leucine zipper motif that overlaps with the paired-related region (L11 L18 L25 L32 V39). A bipartite-type nuclear localization sequence (NLS) (two basic amino acids followed by a 10 amino acid spacer and a cluster of five amino acids, of which three are basic (Dingwall and Laskey, 1991)) is also present. Putative casein kinase II sites (S/TXXD/E) are found in the vicinity of the NLS (Rihs *et al.*, 1991). A catalytic triad consisting of the DDE box motif is also shown. An endogenous polyadenylation signal is often found downstream of the transposase gene and is located within the inverted repeat region (Adapted from Ivics *et al.*, 1996). Panel B: Sequence alignment between a consensus sequence of the tzf transposase amino acid sequence (top) (Lam *et al.*, 1996) and corresponding sequences derived from the AS progranulin gene (bottom)(see Figure 21). The latter is non-functional due to several insertions, deletions (---), and point mutations resulting in premature stop codons (.). Stars represent conserved sequences. Residues corresponding to the DDE motif are indicated with arrowheads (Ivics *et al.*, 1996; Haren *et al.*, 1999; Leaver *et al.*, 2001).

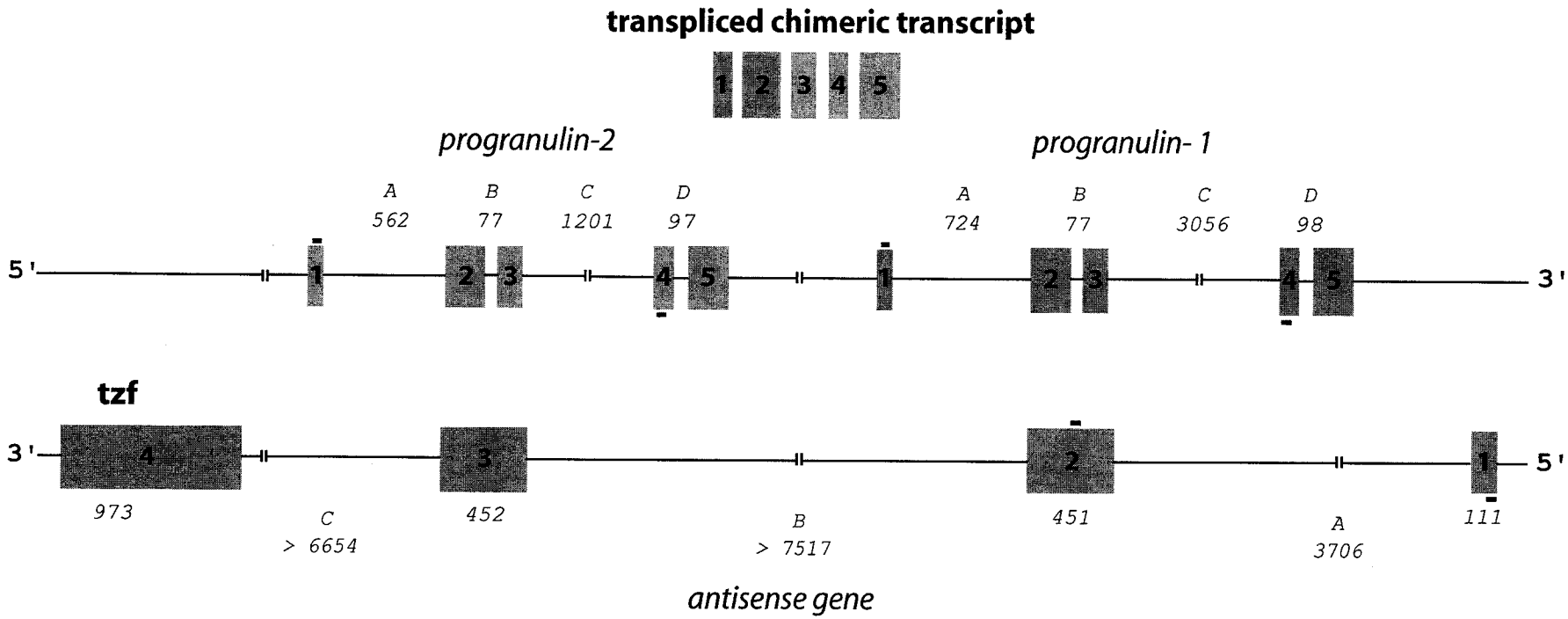


Figure 19. **Genomic organization of zebrafish progranulin-1, progranulin-2 and their complementary antisense gene**. The zebrafish progranulin-1 and -2 genes are organized in tandem in a head-to-tail orientation and share an identical exonic organization (exons 1-5, orange and blue boxes, respectively), but differ in their intronic size (introns A, C and D). A spliced and polyadenylated RNA with no apparent open reading frame is derived from four exons (pink boxes) located on the complementary DNA strand that spread over at least 19 864 kb of genomic DNA. Sequence analysis of this antisense RNA confirmed the gene topology between progranulin-1 and progranulin-2. Interestingly, the reverse complement sequence of exon 4 of this antisense gene (i.e. sense relative to the progranulins) encodes a defective transposase of the *tzf* family of transposons. Exon 2 of the antisense gene is complementary to the progranulin-1 gene in a region that encompasses exons 2 and 3, as well as flanking intronic sequences. Similarly, the third exon of the antisense gene is complementary to the same corresponding region of the progranulin-2 gene. Shown on top is a schematic representation of the *trans*-spliced chimeric progranulin mRNA depicting exonic contributions from the progranulin-1 and -2 genes. Bars (-) indicate regions used for the design of primers for RT-PCR. (See text for details)

3.1.7 Genomic organization of progranulin-1, progranulin-2, and of their antisense gene – The determination of the organization of the gene topology and the estimated intergenic distance between progranulin-1 and progranulin-2 was made possible by the finding of all exonic segments giving rise to the antisense transcript within the deposited AC124903 clone (see above). A schematic representation of the progranulin-1 gene locus is presented in **Figure 19**. Note that the authenticity of the AS progranulin-1/-2 transcript is validated by its full resolution onto four distinct exonic segments. Note also that progranulin-1 and progranulin-2 have an identical exonic organization, and that progranulin-1 is located downstream of progranulin-2. The topology of the progranulin-2 and progranulin-1 genes inferred by the antisense transcript has major implications for the interpretation of the genesis of the progranulin hybrid transcript (see section 3.2.4 for discussion).

3.1.8 Cloning and characterization of a cDNA for zebrafish progranulin-a – A search performed at GenBank for zebrafish ESTs encoding granulins revealed the presence of transcripts with leader sequences distinct from that of progranulin-1, implying that additional granulin genes existed. A similar approach to that used for cloning the granulin-1 precursor was used to isolate the ORF encoding a deduced granulin-A peptide. The incomplete cloned ORF results matched sequences found in public databases (EST clone AW174591). This clone was obtained from RZPD and sequenced in full on both strands. This cDNA encodes a 2619-nucleotides ORF (**Figure 20**) that gives rise to a deduced polypeptide of 873 amino acids with several potential N-glycosylation sites (not shown). This polypeptide, which is referred to as progranulin-A, encodes 10 full, tandemly repeated but non-identical, granulin peptides. The organization of these granulin peptides within the precursor is presented later. Like progranulin-1, the translation initiation codon for zebrafish progranulin-a (AAAAUGU) conforms to Kozak's rule for efficient translation initiation (Kozak, 1987). A predicted signal peptide sequence (residues 1-16) for progranulin-A is different from the one observed for progranulin-1 (and progranulin-2) (Nielsen *et al.*, 1997). After deducting the signal peptide, the predicted molecular mass of progranulin-A is 91.3 kDa.

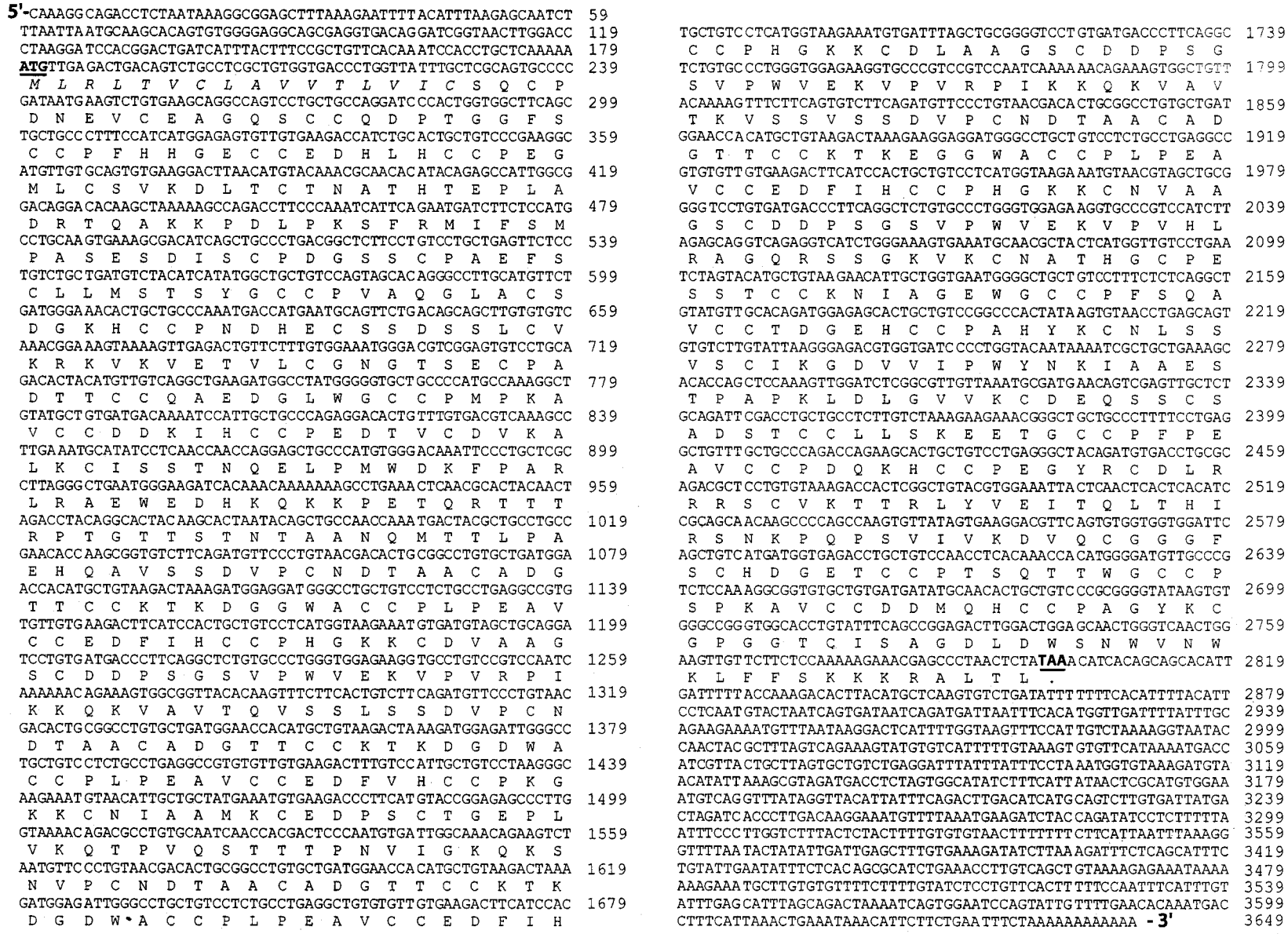

5'-CAAAGGCAGACCTCTAATAAAGGCGGAGCTTTAAAGAATTTTACATTTAAGAGCAATCT 59
TTAATTAATGCAAGCACAGTGTGGGGAGGCAGCGAGGTGACAGGATCGGTAACTTGGACC 119
CTAAGGATCCACGGACTGATCATTTACTTTCCGCTGTTCACAAATCCACCTGCTCAAAAA 179
ATGTTGAGACTGACAGTCTGCCTCGCTGTGGTGACCCTGGTTATTTGCTCGCAGTGCCCC 239
M L R L T V C L A V V T L V I C S Q C P
GATAATGAAGTCTGTGAAGCAGGCCAGTCCTGCTGCCAGGATCCCACTGGTGGCTTCAGC 299
D N E V C E A G Q S C C Q D P T G G F S
TGCTGCCCTTTCCATCATGGAGAGTGTTGTGAAGACCATCTGCACTGCTGTCCCGAAGGC 359
C C P F H H G E C C E D H L H C C P E G
ATGTTGTGCAGTGTGAAGGACTTAACATGTACAAACGCAACACATACAGAGCCATTGGCG 419
M L C S V K D L T C T N A T H T E P L A
GACAGGACACAAGCTAAAAAGCCAGACCTTCCCAAATCATTCAGAATGATCTTCTCCATG 479
D R T Q A K K P D L P K S F R M I F S M
CCTGCAAGTGAAAGCGACATCAGCTGCCCTGACGGCTCTTCCTGTCCTGCTGAGTTCTCC 539
P A S E S D I S C P D G S S C P A E F S
TGTCTGCTGATGTCTACATCATATGGCTGCTGTCCAGTAGCACAGGGCCTTGCATGTTCT 599
C L L M S T S Y G C C P V A Q G L A C S
GATGGGAAACACTGCTGCCCAAATGACCATGAATGCAGTTCTGACAGCAGCTTGTGTGTC 659
D G K H C C P N D H E C S S D S S L C V
AAACGGAAAGTAAAAGTTGAGACTGTTCTTTGTGGAAATGGGACGTCGGAGTGTCCTGCA 719
K R K V K V E T V L C G N G T S E C P A
GACACTACATGTTGTCAGGCTGAAGATGGCCTATGGGGGTGCTGCCCCATGCCAAAGGCT 779
D T T C C Q A E D G L W G C C P M P K A
GTATGCTGTGATGACAAAATCCATTGCTGCCCAGAGGACACTGTTTGTGACGTCAAAGCC 839
V C C D D K I H C C P E D T V C D V K A
TTGAAATGCATATCCTCAACCAACCAGGAGCTGCCCATGTGGGACAAATTCCCTGCTCGC 899
L K C I S S T N Q E L P M W D K F P A R
CTTAGGGCTGAATGGGAAGATCACAAACAAAAAAAGCCTGAAACTCAACGCACTACAACT 959
L R A E W E D H K Q K K P E T Q R T T T
AGACCTACAGGCACTACAAGCACTAATACAGCTGCCAACCAAATGACTACGCTGCCTGCC 1019
R P T G T T S T N T A A N Q M T T L P A
GAACACCAAGCGGTGTCTTCAGATGTTCCCTGTAACGACACTGCGGCCTGTGCTGATGGA 1079
E H Q A V S S D V P C N D T A A C A D G
ACCACATGCTGTAAGACTAAAGATGGAGGATGGGCCTGCTGTCCTCTGCCTGAGGCCGTG 1139
T T C C K T K D G G W A C C P L P E A V
TGTTGTGAAGACTTCATCCACTGCTGTCCTCATGGTAAGAAATGTGATGTAGCTGCAGGA 1199
C C E D F I H C C P H G K K C D V A A G
TCCTGTGATGACCCTTCAGGCTCTGTGCCCTGGGTGGAGAAGGTGCCTGTCCGTCCAATC 1259
S C D D P S G S V P W V E K V P V R P I
AAAAAACAGAAAGTGGCGGTTACACAAGTTTCTTCACTGTCTTCAGATGTTCCCTGTAAC 1319
K K Q K V A V T Q V S S L S S D V P C N
GACACTGCGGCCTGTGCTGATGGAACCACATGCTGTAAGACTAAAGATGGAGATTGGGCC 1379
D T A A C A D G T T C C K T K D G D W A
TGCTGTCCTCTGCCTGAGGCCGTGTGTTGTGAAGACTTTGTCCATTGCTGTCCTAAGGGC 1439
C C P L P E A V C C E D F V H C C P K G
AAGAAATGTAACATTGCTGCTATGAAATGTGAAGACCCTTCATGTACCGGAGAGCCCTTG 1499
K K C N I A A M K C E D P S C T G E P L
GTAAAACAGACGCCTGTGCAATCAACCACGACTCCCAATGTGATTGGCAAACAGAAGTCT 1559
V K Q T P V Q S T T T P N V I G K Q K S
AATGTTCCCTGTAACGACACTGCGGCCTGTGCTGATGGAACCACATGCTGTAAGACTAAA 1619
N V P C N D T A A C A D G T T C C K T K
GATGGAGATTGGGCCTGCTGTCCTCTGCCTGAGGCTGTGTGTTGTGAAGACTTCATCCAC 1679
D G D W A C C P L P E A V C C E D F I H
TGCTGTCCTCATGGTAAGAAATGTGATTTAGCTGCGGGGTCCTGTGATGACCCTTCAGGC 1739
C C P H G K K C D L A A G S C D D P S G
TCTGTGCCCTGGGTGGAGAAGGTGCCCGTCCGTCCAATCAAAAAACAGAAAGTGGCTGTT 1799
S V P W V E K V P V R P I K K Q K V A V
ACAAAAGTTTCTTCAGTGTCTTCAGATGTTCCCTGTAACGACACTGCGGCCTGTGCTGAT 1859
T K V S S V S S D V P C N D T A A C A D
GGAACCACATGCTGTAAGACTAAAGAAGGAGGATGGGCCTGCTGTCCTCTGCCTGAGGCC 1919
G T T C C K T K E G G W A C C P L P E A
GTGTGTTGTGAAGACTTCATCCACTGCTGTCCTCATGGTAAGAAATGTAACGTAGCTGCG 1979
V C C E D F I H C C P H G K K C N V A A
GGGTCCTGTGATGACCCTTCAGGCTCTGTGCCCTGGGTGGAGAAGGTGCCCGTCCATCTT 2039
G S C D D P S G S V P W V E K V P V H L
AGAGCAGGTCAGAGGTCATCTGGGAAAGTGAAATGCAACGCTACTCATGGTTGTCCTGAA 2099
R A G Q R S S G K V K C N A T H G C P E
TCTAGTACATGCTGTAAGAACATTGCTGGTGAATGGGGCTGCTGTCCTTTCTCTCAGGCT 2159
S S T C C K N I A G E W G C C P F S Q A
GTATGTTGCACAGATGGAGAGCACTGCTGTCCGGCCCACTATAAGTGTAACCTGAGCAGT 2219
V C C T D G E H C C P A H Y K C N L S S
GTGTCTTGTATTAAGGGAGACGTGGTGATCCCCTGGTACAATAAAATCGCTGCTGAAAGC 2279
V S C I K G D V V I P W Y N K I A A E S
ACACCAGCTCCAAAGTTGGATCTCGGCGTTGTTAAATGCGATGAACAGTCGAGTTGCTCT 2339
T P A P K L D L G V V K C D E Q S S C S
GCAGATTCGACCTGCTGCCTCTTGTCTAAAGAAGAAACGGGCTGCTGCCCTTTTCCTGAG 2399
A D S T C C L L S K E E T G C C P F P E
GCTGTTTGCTGCCCAGACCAGAAGCACTGCTGTCCTGAGGGCTACAGATGTGACCTGCGC 2459
A V C C P D Q K H C C P E G Y R C D L R
AGACGCTCCTGTGTAAAGACCACTCGGCTGTACGTGGAAATTACTCAACTCACTCACATC 2519
R R S C V K T T R L Y V E I T Q L T H I
CGCAGCAACAAGCCCCAGCCAAGTGTTATAGTGAAGGACGTTCAGTGTGGTGGTGGATTC 2579
R S N K P Q P S V I V K D V Q C G G G F
AGCTGTCATGATGGTGAGACCTGCTGTCCAACCTCACAAACCACATGGGGATGTTGCCCG 2639
S C H D G E T C C P T S Q T T W G C C P
TCTCCAAAGGCGGTGTGCTGTGATGATATGCAACACTGCTGTCCCGCGGGGTATAAGTGT 2699
S P K A V C C D D M Q H C C P A G Y K C
GGGCCGGGGTGGCACCTGTATTTCAGCCGGAGACTTGGACTGGAGCAACTGGGTCAACTGG 2759
G P G G T C I S A G D L D W S N W V N W
AAGTTGTTCTTCTCCAAAAAGAAACGAGCCCTAACTCTATAACATCACAGCAGCACATT 2819
K L F F S K K K R A L T L .
GATTTTTACCAAAGACACTTACATGCTCAAGTGTCTGATATTTTTTTCACATTTTACATT 2879
CCTCAATGTACTAATCAGTGATAATCAGATGATTAATTTCACATGGTTGATTTTATTTGC 2939
AGAAGAAAATGTTTAATAAGGACTCATTTTGGTAAGTTTCCATTGTCTAAAAGGTAATAC 2999
CAACTACGCTTTAGTCAGAAAGTATGTGTCATTTTTGTAAAGTGTGTTCATAAAATGACC 3059
ATCGTTACTGCTTAGTGCTGTCTGAGGATTTATTTATTTCCTAAATGGTGTAAAGATGTA 3119
ACATATTAAAGCGTAGATGACCTCTAGTGGCATATCTTTCATTATAACTCGCATGTGGAA 3179
ATGTCAGGTTTATAGGTTACATTATTTCAGACTTGACATCATGCAGTCTTGTGATTATGA 3239
CTAGATCACCCTTGACAAGGAAATGTTTTAAATGAAGATCTACCAGATATCCTCTTTTTA 3299
ATTTCCCTTGGTCTTTACTCTACTTTTGTGTGTAACTTTTTTTCTTCATTAATTTAAAGG 3359
GTTTTAATACTATATTGATTGAGCTTTGTGAAAGATATCTTAAAGATTTCTCAGCATTTC 3419
TGTATTGAATATTTCTCACAGCGCATCTGAAACCTTGTCAGCTGTAAAAGAGAAATAAAA 3479
AAAGAAATGCTTGTGTGTTTTCTTTTGTATCTCCTGTTCACTTTTTCCAATTTCATTTGT 3539
ATTTGAGCATTTAGCAGACTAAAATCAGTGGAATCCAGTATTGTTTTGAACACAAATGAC 3599
CTTTCATTAAACTGAAATAAACATTCTTCTGAATTTCTAAAAAAAAAAAA - 3' 3649

Figure 20

Figure 20. **Full nucleotide and deduced translated sequence of zebrafish progranulin-a**. A zebrafish cDNA with a predicted 2619 nucleotide-long ORF encodes a deduced polypeptide of 873 amino acids carrying 10 repeats of the granulin motif. Of the 3649 nucleotides comprising the complete sequence of this cDNA, 179 and 851 residues correspond to the 5'UTR and 3'UTR, respectively. Translation initiation (ATG) and termination (TAA) codons are in bold and underlined. The calculated mass of the polypeptide after deducting the predicted signals sequence (italics, residues 1-16) (von Heijne, 1986; Nielsen *et al.*, 1997) is approximately 91.3 kDa.

3.1.9 Cloning and characterization of a cDNA for zebrafish progranulin-b – We isolated a second zebrafish precursor related to progranulin-a, which we sequenced in full (EST clone AW184435; see methods). Progranulin-b encodes a 729 amino acid-long polypeptide deduced from a cDNA with a predicted ORF 2820 nucleotides in length (**Figure 21**). This polypeptide also contains sites potential for N-glycosylation (not shown), and has a predicted molecular mass of 76.5 kDa after removing the mass of the predicted signal peptide sequence (residues 1-19) (Nielsen *et al.*, 1997). The translation initiation codon for progranulin-B (ACAAUGG) is found within a perfect Kozak consensus (Kozak, 1987). Progranulin-B is thus slightly smaller than progranulin-A and encodes 9 tandemly repeated granulin-like peptides (see section 3.1.10).

3.1.10 Comparison of progranulin-A, progranulin-B, and human progranulin – The zebrafish granulins precursor repertoire is thus not limited to small gene structures. Although it cannot conclusively be determined whether the cloned cDNAs encoding progranulin-A and -B represent full-length or alternatively spliced structures derived from their respective progranulin genes, it is safe to assume that because of their organization, progranulin-A and progranulin-B are more closely related to mammalian progranulin than are progranulin-1 and progranulin-2. The presence of peptides related to mammalian granulin-A within the longer zebrafish progranulins also supports this argument.

As expected, sequence alignment between the full zebrafish progranulin-a, progranulin-b and human progranulin protein products reveals extensive homologies, especially within the granulin domains (**Figure 22**). However, when zebrafish granulin peptides deduced from progranulin-a are individually compared to mammalian progranulin using the ClustalW program (http://www.ch.embnet.org/software/ClustalW.html), the granulin motifs found within the zebrafish precursor do not always show congruent co-linear homology assignment on the basis of the nomenclature attributed to mammalian granulin peptides (i.e., in the order of granulin-G, -F, -B, -A, -C, -D, -E; Bhandari *et al.*, 1992; Plowman *et al.*, 1992; Baba *et al.*, 1993a; Bhandari *et al.*, 1993). This discrepancy likely results from the close sequence conservation among granulin peptides (**Figure 22**). The actinopterygian (zebrafish) and tetrapod (human) granulins precursors have therefore evolved differently since they diverged from the last common ancestor (**Figure 4**; see also section 3.2.2 for discussion).

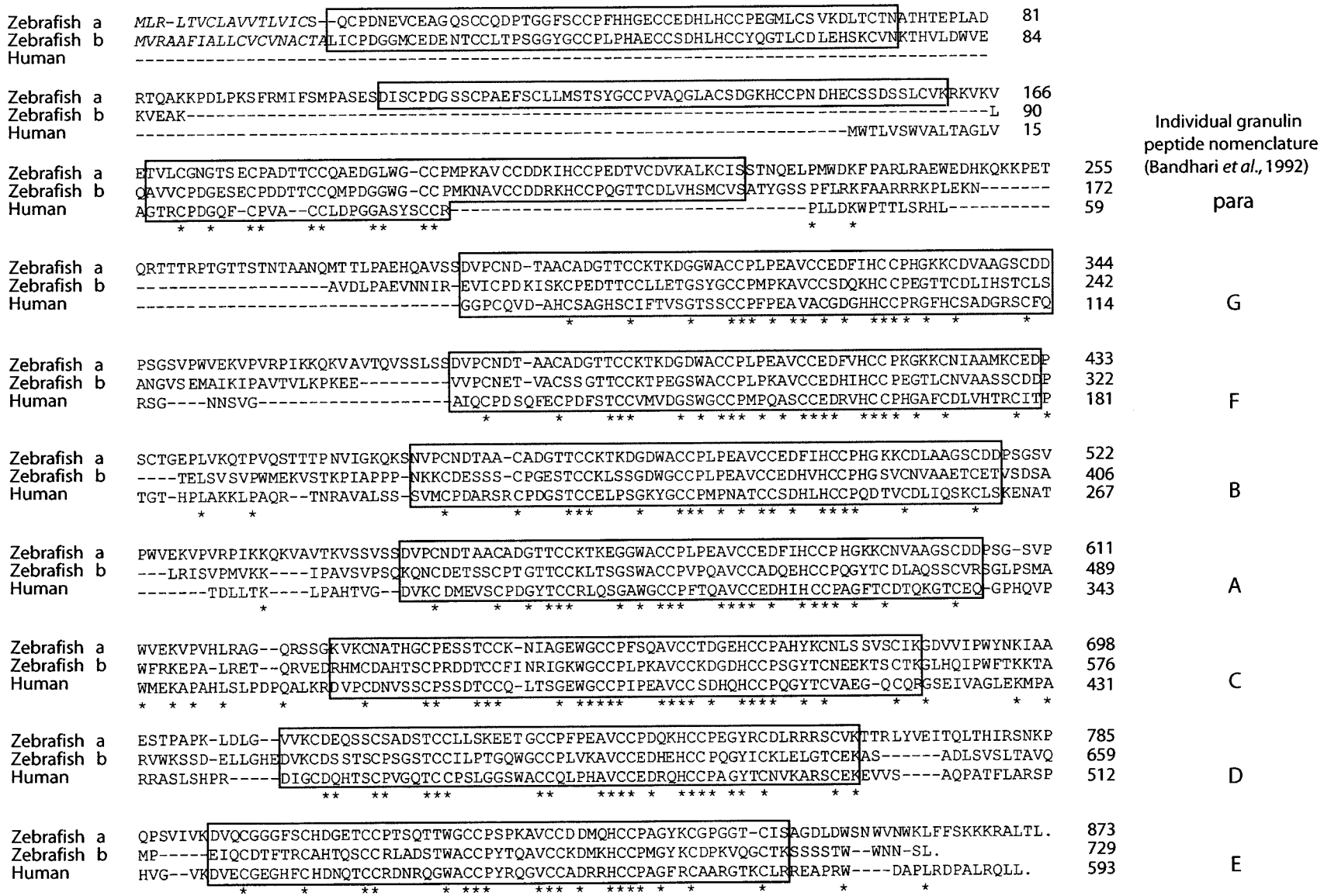
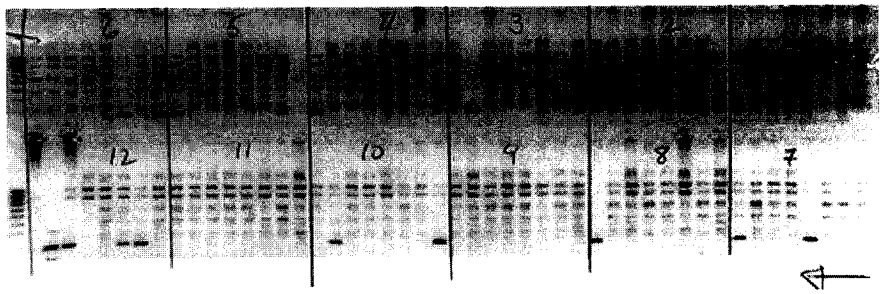

Zebrafish a
Zebrafish b
Human
Individual granulin peptide nomenclature (Bandhari et al., 1992)
para
G
F
B
A
C
D
E

Figure 22. **Sequence alignment between zebrafish progranulins-a and -b, and human progranulin**. Zebrafish possesses two orthologues of mammalian progranulin, referred to as progranulin-a and progranulin-b, respectively. Each carries 10 and 9 copies of the granulin motif (boxed), respectively. This is in contrast to mammalian progranulin, which is constituted of one N-terminal granulin half motif and seven full domains. Colinear assignment of the mammalian nomenclature to deduced granulin motifs of the fish precursors is based on ClustalW sequence alignments (Bandhari *et al.*, 1992).

3.1.11 Chromosomal location of the respective zebrafish progranulin genes – Primers used for the amplification of genomic segments corresponding to individual zebrafish progranulin genes are listed in **TABLE 4**. An example of the raw data derived from each mapping experiment and used for software analysis (RH MAPPER program, see methods) is provided in **Figure 23**. Radiation hybrid mapping indicates that zebrafish *progranulin-1* is located on linkage group 19 (LG19 = chromosome 19) with a lod score of 12.7 (an estimate of confidence and accuracy: the higher the value, the greater the significance), 5.98 centiRays (cR) from EST clone fb47h01 located next to the simple sequence length polymorphism SSLP marker z6661 (where 1 centiRay = 148 kilobases, the estimated average breakpoint frequency for the LN54 RH panel) (Hukriede *et al.*, 1999). Since a primer pair used for the mapping of *progranulin-1* does not discriminate between *progranulin-1* and *progranulin-2*, and in view of the close physical proximity between these two genes, it was not surprising that the two genes localized to the same chromosomal location. Primers specific for *progranulin-2*, but that might cross-react with *progranulin-1* at the annealing temperature used (see methods) localized this gene 8.66 cR from fb47h01 with a LOD value of 11.3. *Progranulin-a* is located on LG3 with a LOD value of 12.0, 9.92 cR from SSLP marker z22516 and is found 14.6 cR from EST fb77c07. *Progranulin-b* is located on LG24 9.76 cR from EST fc18g06 with a LOD value of 20.8. A schematic representation of the linkage assignment for the cloned zebrafish progranulin genes, with the exclusion of *progranulin-b*, is represented in **Figure 24**. Syntenic relationships are considered in the discussion (section 3.2.2.3).



1	2	4	5	6	8	9	11	1
13	(16)	18	(19)	23	25	27	29	2
30	33	36	39	40	41	47	48	3
49	50	52	54	57	59	63	65	4
66	(70)	73	74	79	80	83	84	5
85	86	87	88	89	91	92	96	6
97	98	101	104	105	106	108	109	7
114	117	119	121	123	125	132	135	8
136	137	138	150	151	152	153	167	9
169	174	175	176	178	182	183	184	10
190	300	301	302	303	304	305	306	11
308	309	310	311	312	MIX	AB9	B78	12

Figure 23. **PCR for the progranulin-1 gene on the LN54 mapping panel.** Typical PCR data for the mapping of individual progranulin genes, as demonstrated here for the progranulin-1 gene (an identical result was obtained for the progranulin-2 gene). Each positive (highlighted in yellow) is scored for analysis using the RH mapper program (http://www.zfish.wustl.edu).

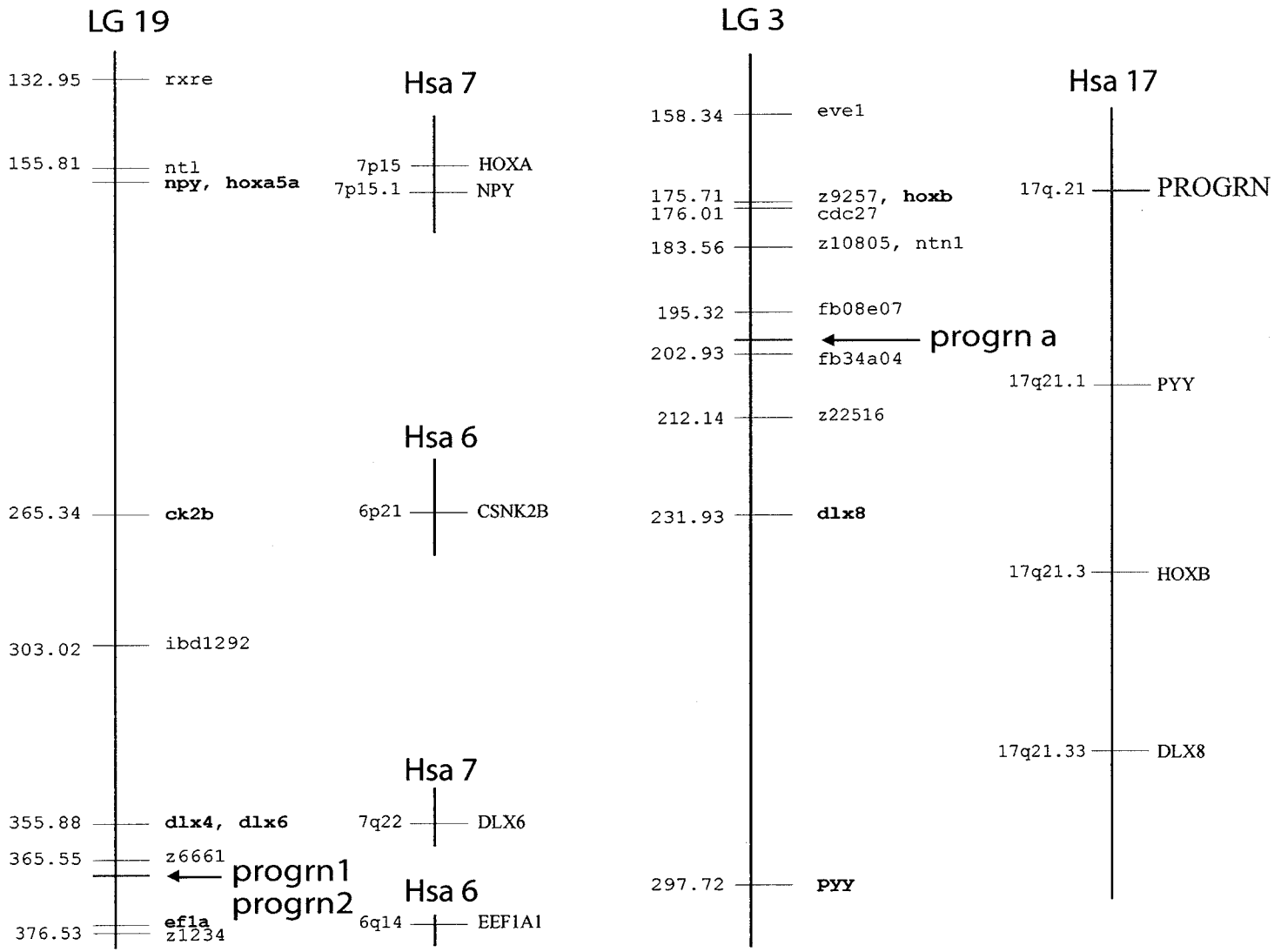

LG 19
132.95
rxre
155.81
ntl
npy, hoxa5a
265.34
ck2b
303.02
ibd1292
355.88
dlx4, dlx6
365.55
z6661
progrn1
progrn2
376.53
ef1a
z1234
Hsa 7
7p15
HOXA
7p15.1
NPY
Hsa 6
6p21
CSNK2B
Hsa 7
7q22
DLX6
Hsa 6
6q14
EEF1A1
LG 3
158.34
eve1
175.71
z9257, hoxb
176.01
cdc27
183.56
z10805, ntn1
195.32
fb08e07
progrn a
202.93
fb34a04
212.14
z22516
231.93
dlx8
297.72
PYY
Hsa 17
17q.21
PROGRN
17q21.1
PYY
17q21.3
HOXB
17q21.33
DLX8

Figure 24

Figure 24. **Chromosomal assignment of zebrafish progranulin genes, and syntenic relationship to human progranulin.** Zebrafish progranulin-a is located next to genes (*HoxB* cluster, *dlx8*, *pyy*) that form an extensive bloc of conserved synteny to human chromosome 17 (Hsa17), indicating an orthologous relationship to human progranulin. Zebrafish progranulin-b maps to LG24 in a region devoid of syntenic correspondence to Hsa17, rather than being located on LG12 as predicted from 3R (see Figure 6) (data not shown). Co-orthology for progranulin-b is based on sequence conservation, as well as 3R within the actinopterygian lineage and the absence of additional progranulins within tetrapods. Progranulin-1 and progranulin-2 are both linked on the same chromosome (LG19), which displays syntenic correspondences mainly fragmented onto two human chromosomes (Hsa6 and Hsa7). The fortuitous linkage of these genes to the *HoxA* cluster, *npy* and *dlx8* genes, suggests that progranulin-1 or -2 may have originated in concert with the mechanism leading to the emergence of duplicated Hox clusters at the base of the vertebrate radiation (R1 or R2)(see Figure 3). Presumably, this gene was subsequently lost at the base of the tetrapod lineage, since there is no evidence of a progranulin-1 counterpart in mammals, birds or amphibians in public databases.

3.2 Discussion of the zebrafish progranulin gene family

3.2.1 Introductory remarks – The first part of this section discusses the contrasting presence of a progranulin gene family in zebrafish relative to observations made in genomes of other animals so far investigated. The retention of duplicated progranulins in the genome of some teleosts offers the opportunity to discuss the usual fate of duplicons and their consequences upon organismal evolution, a topic that is presented below. In this sense this part can be viewed as a continuation of the introduction (Sections 1.2).

In light of the cloning of a hybrid progranulin, an account of the recent findings pertaining to the formation of chimeric transcripts, as well as the types and extent of these unusual protein-coding mRNAs is overviewed. In the same vein, the discovery of the AS progranulin-1/-2 gene and of its unique attributes are discussed in relation to the phenomenal presence of mobile genetic elements in eukaryotic genomes, including their consequences. This is followed by an account of what is generally known about the emerging field of antisense transcription and of the nature of non-coding RNAs. These excursions constitute the second half of this discussion and have been made in order to provide a system of correspondences upon which to relate the observations presented in this thesis. Similarities to the granulin story will sometimes come to mind, but differences will be more often encountered - these are meant to be informative and to provide perspective.

3.2.2 General discussion of the zebrafish progranulin gene family

3.2.2.1 Progranulin gene number: the zebrafish breaks the rule – Having established that the fish has undergone a third round of genome duplication (3R), a question arises as to the prevalence of mammalian genes that will find two co-orthologues in zebrafish. Using a combination of genetic mapping with the HS meiotic panel and phylogenetic analyses, an initial survey of >300 zebrafish genes estimated the retention of duplicate gene pairs to be at least 20 % (Postlethwait *et al.*, 2000). In

other words, one in every 5 mammalian genes will be represented by 2 co-orthologues in the zebrafish. The high percentage of duplicate gene pairs in zebrafish has prompted the creation of Wanda, a public database cataloguing co-orthologuous fish genes (Van de Peer *et al.*, 2002; http://www.evolutionsbiologie.uni-konstanz.de/Wanda/). It is thus not surprising to find two progranulin genes (progranulin-a and progranulin-b) in *Danio rerio* (**TABLE 2**; **Figures 20-22**).

What is unexpected, however, is the presence of two additional blueprints, named progranulin-1 and progranulin-2, that share 92.3% identity in their respective nucleotide sequence, and that each carries an uninterrupted open reading frame (ORF) encoding one full and one amino-terminal half granulin sequence only (**Figure 9**). Another distinctive feature of these small proteins is the reduced sequence conservation between their encoded granulin peptides relative to those found within the larger progranulins, i.e. the granulin-1 and granulin-2 peptide sequences have diverged to the extent where their correspondence to granulin peptides residing within progranulins-a or –b cannot be assigned with confidence. However, close inspection of the translated sequences of zebrafish progranulins-1 and -2 cDNAs indicates that these correspond to likely orthologues of carp granulin-1 and -2 (**Figure 11**), and that they do not fall in the category of pseudogenes.

Surprisingly, an additional transcript with 95.6% identity with either progranulin-1 or progranulin-2 was also uncovered using our cloning strategy (see Methods) (**Figure 14**). Independent amplification and sequencing of this cDNA from distinct sources used as template indicates that this product is unlikely to be an artifact of the PCR. A caveat in this interpretation is the high rate of polymorphism encountered in some fish, such as fugu with an estimated change occurring every 0.4% of the nucleotides, corresponding to four times the frequency estimated for humans (Aparicio *et al.*, 2002). However, it was noticed that the pattern of nucleotide change is non-randomly distributed over the entire length of the amplicon. In fact, its structure represents a chimera consisting, approximately, of the N-terminal portion of progranulin-1 followed by the C-terminal portion of progranulin2, thereby discrediting

the likelihood of polymorphism or allelic variation as a simple explanation for the existence of this precursor (**Figure 14**).

The authenticity of these transcripts was confirmed through characterization of the gene structure(s) from which they originated. Possible scenarios envisioned were the generation of progranulin-1 and -2 via tandemly duplicated genes organized in a topology inferred by the chimeric transcript, suggesting further that the latter was derived from the alternative processing of a larger common primary transcript. Although suspected to be rare, the tandem organization of some factors on a bi-cistronic transcript has been documented, and is illustrated by the precursor encoding zebrafish Wnt8a and Wnt8b (Levken *et al.*, 2001). Nevertheless, long transcription across neighbouring genes that normally act as independent transcription units has been documented in several cases (Magrangeas *et al.*, 1998; Finta and Zaphiropoulos, 2000; Communi *et al.*, 2001).

Northern blot analyses of total RNA isolated from various organs of carp and adult zebrafish (data not shown), in combination with the rapid amplification of cDNA ends (RACE) performed in both directions according to the cloning strategy and subsequently independently, provided evidence that progranulin-1 and progranulin-2 transcripts do not arise from alternative splicing of a common gene or are part of a bicistronic transcript (**Figures 7 and 10**).

In the event that the coding portion of a gene is uninterrupted by intronic sequences, the usual inference is that it is a pseudogene (Mighell *et al.*, 2000). This hallmark is not sufficient, however, since several examples of intronless genes exist, and include some G-protein-coupled-receptors (GPCRs) (Dearry *et al.*, 1990; Corness *et al.*, 1993; Glusman *et al.*, 2000) and the ferritin gene (Levi *et al.*, 2001). Determining the genic exon/intron architecture of the smaller zebrafish granulin precursors, although eventually achieved via conventional strategies, was initially assayed with a primer set non-discriminating between progranulin-1 and -2, for use in the PCR amplification of genomic DNA. These attempts served two purposes. First, they would allow for the use of primers in chromosomal localization experiments.

Second, it was believed that due to the smaller size of the transcripts, the desired gene(s) could potentially be cloned using a PCR strategy. In fact, granulin-coding gene segments are not invariably bisected by an intron, as is demonstrated by the structure of *C. elegans* progranulin. Partial products of identical size, as judged on ethidium-stained agarose gel electrophoresis, were sequenced and shown to carry distinct intronic sequences corresponding to both progranulin-1 and -2, respectively, but no amplicon, as anticipated, was uncovered encoding the chimeric structure (data not shown).

The use of this primer set in the somatic radiation hybrid mapping panel localized the sequences encoding progranulin-1 and/or progranulin-2 next to a single SSLP marker (z6661) on linkage group 19 (**Figures 23 and 24**). This observation suggested that *progranulin-1* and *progranulin-2* were genetically linked. However, the lack of discrimination between *progranulin-1* and *-2* amplicons generated with these primers prevented this definitive conclusion. Also, the LN54 mapping panel covers at least 88% of the zebrafish genome, an estimation that was calculated based on the placement of unbiased markers (Hukriede *et al.*, 1999). It is therefore possible that other progranulin-1-like genes could be residing within the genomic regions not represented in the panel. However, Southern blot analysis of zebrafish genomic DNA supports the conclusion that there exists no progranulin genes other than those presented in this thesis (data not shown).

3.2.2.2 Progranulin-1 and progranulin-2 are distinct genes – Confirmation that progranulin-1 and progranulin-2 do not share a common origin came from the characterization of the gene encoding progranulin-1 (**Figure 12**). While in the process of isolating the progranulin-2 gene using a strategy identical to the one used for progranulin-1, genomic sequences encoding progranulin-2 were uncovered within whole genome shotgun sequences deposited at NCBI (accession number AC124903). Analysis of these sequences confirmed that progranulin-1 and progranulin-2 share an identical exonic structure.

Contained within this assembly were also the entire sequences for progranulin-1, corroborating its suspected physical proximity to progranulin-2. Comparison of nucleotide sequences derived from cDNA to those of genomic origin revealed no discrepancy in the cloned cDNAs. This sequence conservation permitted a more direct validation of the authenticity of the chimeric transcript. Specifically, it consists of sequences derived from exons 1 and 2 of the progranulin-1 gene, joined to those from exons 3-5 of the progranulin-2 gene (**Figures 14 and 19**). Therefore, noting that the substitutions encountered within the chimeric transcript all occur specifically within predicted exons from progranulin-1 or progranulin-2, the notion that this hybrid cDNA does not result from polymorphism was reinforced. Unfortunately, sequencing errors within the AC124903 clone due to presence of repetitive elements prevented the complete assembly of this genomic segment, and thus did not allow for the confirmation of the topology between the progranulin-1 and -2 genes. The hybrid progranulin will be discussed further in another section.

3.2.2.3 Origin of the progranulin gene family – In fish, the 3R hypothesis leads to the prediction that, on average, gene families will be larger than those encountered in mammals. However, this is not always the case: again, genes can be lost. The observation that granulins form an extended gene family in some teleosts appears to be an elaborate exception to the rule. Orthology and paralogy assignment has been reinforced by the comparative genomic mapping approach (**Figure 24**). As was alluded to earlier in the results section, there appears to be no progranulin-1 structural counterpart within the tetrapod lineage. There are additional examples of zebrafish genes apparently without a human orthologue. For example, genes of the EVX family of transcription factors have been found lying adjacent to the 5' ends of HOX A and HOX D clusters in both zebrafish and humans. In the vicinity of the HOX B cluster of zebrafish, however, an additional EVX paralogue called eve1 is present, which was presumably lost in the mammalian lineage (Amores *et al.* 1998). Hoxc1 and Hoxc3 are also present in fish (Aparicio *et al.* 1997; Amores *et al.*, 1998; Naruse *et al.* 2000) but absent from mammals, showing that the common ancestor of actinopterygians and sarcopterygians had a greater Hox repertoire than extant

tetrapods. Apparently, the ependymin gene represents another example (Sterrer et al., 1990). Unlike the case of progranulin-1 and progranulin-2, the examples above are predicated on an *a priori* knowledge of the expected location of the additional (zebrafish) or missing (mammalian) orthologous gene. Hence, the progranulin gene of the one and one-half structural motif may be a unique example of convergent evolution in the teleost lineage, as opposed to it arising within ancestral gnathostomes and followed by selective loss in tetrapods.

The likely mechanism responsible for generating the "b" copy of zebrafish progranulin has been extensively discussed (see introduction). Progranulin-1 and progranulin-2, on the other hand, provide an example of individual gene duplication that, as expected, generates a tandem array on an individual chromosome. This leaves us with the daunting task of tracing back the origin of the one and one-half gene motif. To reiterate, there is no evidence for a progranulin-1-like sequence in mammals, whether through deposited EST or genome sequences. The conserved exonic organization observed for progranulin-1 (and -2), relative to mammalian progranulin, argues against their having originated independently from an ancestral progranulin gene (**Figure 13**). This observation is significant, since not all orthologous (and hence paralogous) genes share the same exon-intron organization between different species. An extreme case in point was made through comparison of the genomes of *Fugu* and humans: 327 human orthologues of intronless *Fugu* genes contained multiple introns and 317 *Fugu* orthologues of human intronless genes contained multiple introns (Aparicio et al., 2002). It is presently impossible to determine whether a progranulin-1-like architecture existed in a common ancestor of mammals and the zebrafish. However, conserved synteny is observed for the linkage between progranulin, pyy, dlx8 and HoxB gene in humans and zebrafish (progranulin-a) (**Figure 24**). Even if progranulin-b is localized to a linkage group 24 that does not display this syntenic correspondence, its co-orthology assignment is predicted by the apparent sequence conservation and preservation of the longer architecture. A surprising finding revealed through the mapping studies is the localization of progranulin-1 and -2 genes on linkage group 19 where paralogues of these linked genes (npy, dlx4 and dlx6, HoxA

cluster) also reside, the latter zebrafish chromosome being syntenic, at least when considering these genes, to human chromosome 7 (Postlethwait *et al.*, 1998; Barbazuk *et al.*, 2000; Postlethwait *et al.*, 2000). Interestingly, its was recently shown that human chromosomes 7 and 17 possess several segments of shared sequences and are thus presumed duplicates from a common "ancestral" *Amphioxus* chromosome (Panopoulou *et al.*, 2003). Combined, these observations suggest that the smaller progranulin architecture existed at the base of the vertebrate radiation, but that it underwent selective loss in the lineage leading to tetrapods after the split between sarcopterygians and actinopterygians (Ahlberg and Milner, 1994).

Thus, a possible scenario for the emergence of the smaller progranulin gene is that an ancestral form underwent recombination during meiosis via non-allelic chromosome pairing. This mechanism is difficult to reconcile with two observations, since it requires that a non-replicative mechanism be at play. The first is the acquisition of a promoter and associated cis-regulatory elements, coupled to an in frame stop codon, enabling the newborn gene to be fixed within the genome, without disrupting the coding potential of the ancestral gene. In fact, if the exons were derived from either extremity (5'end: promoter, 3'end: stop codon) of the donor gene, this would have undoubtedly resulted in an inactive ancestral progranulin gene. An alternative process suggests that progranulin-1-like sequences could still be found within the ancestral form, which is difficult to evaluate because of subsequent sequence drift. This also raises the question of whether partial (endoduplication) instead of whole-gene duplication has taken place. In this sense, it could be informative to look at the relationship between the EGF precursor gene and its smaller paralogous structure, TGF-α. Was the EGF precursor the ancestral gene that gave rise to TGF-α, through whole-gene duplication with subsequent truncation to reduce it to its current smaller form? Conversely, did an ancestral TGF-α undergo duplication and subsequently expansion to form a tandem array of closely related motifs found in the structure of the EGF precursor? Answers to these questions might lie in the relatedness between the EGF precursor and TGF-α, at the nucleotide level. This analysis is clearly hampered by the sequence drift inherent with time and the differential selective pressure imposed

upon duplicate gene pairs (see later). Furthermore, granulin peptide motifs vary in number and in sequence among different granulin precursors deduced from various species (e.g., *C. elegans* numbers 3; *Ciona intestinalis* has at least 7; zebrafish progranulins-1 and -2 encode 1 and 1/2, while progranulins-a and –b harbour 10 and 9, respectively; and mammalian progranulin has 1/2 and 7), demonstrating that this gene evolves rapidly across phyla. The notable absence of progranulin from the genome of *Drosophila melanogaster*, suggests that this gene is dispensable with respect to the development or normal physiology of bilaterians. It is currently unknown if this is due to the divergence of the characteristic cysteine motif in arthropods, or if this presents a rare case of non-orthologous gene displacement. The purification of a granulin-like peptide from the brain of the locust (*Locusta migratorias*) suggests loss within the arthropod lineage leading to the fruitfly (Nakakura *et al.*, 1991).

The second consideration is based on the timing of the appearance of progranulin-1, and supposes that a replicative mechanism was at play. If progranulin-1 finds origin before the duplication of the ancestral form leading to progranulin-a and -b, then neither progranulin-a nor -b would be expected to carry granulin-1-related sequences, as is the case. This would suggest that the one and one-half copy was also subjected to 3R, generating an additional copy elsewhere in the genome, irrespective of the duplicated copy in tandem. There is no indication so far for additional granulin-like sequences in the zebrafish coming from either deposited EST or high throughput genome sequencing efforts. This, combined with results obtained from chromosome assignment using a radiation hybrid panel, indicates that if they once existed, additional granulin genes drifted beyond recognition or were deleted. Interestingly, *Fugu* also appears to possess additional progranulin cDNAs that differ in length and sequence conservation relative to the mammalian form, and also diverge from the one and one-half progranulin architecture presented in this thesis (data not shown). This observation will necessitate a proper annotation of *Fugu* sequences in order to draw conclusions, but may suggest that newly arisen granulin genes might have emerged at the base of the teleost radiation, and that these evolved rather rapidly.

3.2.3 Fate of the duals: why are gene duplicates preserved ? – The occurrence of gene duplicates is extensive in extant eukaryotic genomes. In fact, it has been estimated that approximately one quarter of yeast genes exist as duplicates (Gu *et al.*, 2003), and that half of all duplicated vertebrate genes have been retained (Nadeau and Sankoff, 1997). Thus, the high occurrence of gene pairs arising from the latest duplication event in the lineage leading to zebrafish is not a phenomenon unique to teleost fish. At least 15% of human genes exist as duplicates (Li *et al.*, 2001), with segmental duplications covering 5.2% of the genome (Bailey *et al.*, 2002). In contrast, only 1.2% of the mouse genome has apparently been subjected to recent segmental duplications (Cheung *et al.*, 2003). This opens up a reflection upon the usual fate and consequences of retained gene pairs. The next section overviews the proposed mechanisms underlying the retention of duplicated genes, as well as their functional consequence, through a discussion of several well-characterized examples derived from various model organisms.

3.2.3.1 The race between gene loss and fixation – Since not all genes evolve at the same rate, the relative evolutionary rates for individual transcription units must be considered when evaluating the contribution of consecutive polyploidization events underlying the expansion of gene families (see introduction, section 1.2.2). For instance, several mammalian protein hormones have undergone rapid bursts of apparent positive selection even if during most of evolutionary time there has been an underlying basal constant rate of evolutionary drift (substitutions/amino acid site/yr). Well-documented examples include growth hormone and prolactin (Wallis, 2001), and possibly serine protease inhibitors (Hill and Hastie, 1987).

The rate of silencing is surprisingly lower than what was originally anticipated through the study of gene duplicates of isozymes arising from polyploidization within tetraploid fish (reviewed in Li, 1980). For example, the catostomids, presumed to have originated from a tetraploidization event 50 mya (Uyeno and smith, 1972) have, on average, lost the function of a gene duplicate in only slightly greater than 50% of cases (Ferris and Whitt, 1977).

In this regard, the differential integration and use of duplicated elements may be at the origin of genomic incompatibility, conceivably influencing the creation of speciation barriers, at least in theory (Ryu *et al.*, 1998; Lynch and Conery, 2000; Lynch and Force, 2000; Lynch *et al.*, 2001). It can be anticipated that work aiming at uncovering the genetic basis underlying the differences between the limnetic and benthic sticklebacks (distinct morphs that live sympatrically) may shed light on this important issue (Peichel *et al.*, 2001).

3.2.3.2 The redundant duplicate – In its initial stage of life, assuming faithful replication of all elements constituting the transcriptional unit, the duplicate can simply be regarded as a backup copy, essentially performing all the functions of the ancestral gene. This scenario does not adversely affect the fitness of the organism when gene dosage is not critical.

In the classical view, the new copy is usually regarded as being free to acquire nucleotide substitutions at a high rate due to reduced selective constraints, while the former copy retains the original function through purifying selection. The outcome is the accumulation of degenerative mutations leading the duplicate to become non-functional (i.e. to become a pseudogene). Although considered a rare event, the chance accumulation of a favourable mutation over time within the coding or cis-regulating portions can predispose the new gene to evolve a beneficial mutation, ensuring its retention in the population. In this respect, the impact of the process of neo-functionalization of a gene duplicate on organismal evolution can be regarded as no less important than that of the creation of new genes as a source of genetic innovation. This model, known as mutation during non-functionality (MDN), suggests that the redundant duplicate is freed from selective pressure and can thus accumulate random mutations predisposing it to the chance acquisition of a new function or to drift, can no longer be regarded as being valid.

The MDN is not supported by the observation that both duplicate genes within the tetraploid organism *Xenopus laevis* undergo less amino acid change substitution

(non-synonymous rate of nucleotide substitution) than expected (Hughes and Hughes, 1993). Both copies are under functional constraint (i.e. purifying selection acting to eliminate amino acid differences) to a similar extent to their mammalian counterpart (Hughes and Hughes, 1993).

It has been postulated that new proteins arise through the recombination of pre-existing functions rather than by evolution of a new function (Doolittle, 1985). In some cases where domain accretion and exon shuffling is not the driving force behind the innovation of gene function, it can be anticipated that gene duplication is required, otherwise the function of the ancestral gene may be disrupted.

3.2.3.3 Relatedness of gene duplicates to their ancestral copy – Although the MDN model cannot account for the fates of all gene duplicates, it has not been fully discredited. Support for the notion that gene duplicates are subject to different selective pressures has been recently obtained from the comparison of the four human MHC paralogous cluster regions and their neighbouring genes to that of the single orthologous region in *Amphioxus* (Abi-Rached *et al.*, 2002). Reconstruction of the putative pre-duplicated region indicates that the ancestral organization is maintained in one of the duplicated regions, while the remaining three diverged furthermore. This plesiomorphy (i.e. closeness to the ancestral state) was also observed at the level of individual genes within each cluster, each displaying the same substitution pattern.

3.2.3.4 The problem of redundancy: consequences of duplicate gene retention – When considering the modes through which gene duplicates are generated (polyploidization, chromosomal duplication or the more common duplication in tandem) and preserved, redundancy could be viewed as an inevitable result of genome evolution. The recognition that pervasive redundancy is likely to be encountered in several information networks must therefore be accounted for in experiments designed to uncover a more complete genetic characterization of cellular and developmental processes.

It is thus not surprising that, in recent years, genetic and biochemical analyses of gene function have repeatedly revealed the complexity of regulatory pathways. A testament to this is the lack of an observable phenotype for loss-of-function mutations generated for many genes through use of the knockout approach. This often masks their essential function and highlights the phenomenon of compensation in biological systems. Furthermore, cross-talk between, or convergence onto, various signal transduction pathways, along with feedback loops in information networks provide a sophisticated type of redundancy that enable pathway activity to be maintained even when a component of the network is inactivated through mutation or gene loss. A simple explanation for the perdurance of duplicated genes is to promote the establishment of compensatory mechanisms.

A classical example illustrating the problem with redundancy has been the difficulty in reconciling the conflicting results obtained between the *in vivo* and *in vitro* analyses of the functions of the basic helix-loop-helix transcription factors Myf-5 and MyoD. Either one of these paralogues is able to induce transfected fibroblast cells to differentiate into skeletal muscle cells, as judged morphologically and by the activation of the marker gene myogenin, a downstream myogenic factor also of the helix-loop-helix transcription factor family (Braun *et al.*, 1989; Edmonston and Olson, 1989). Despite this activity, gene disruption of either MyoD or Myf-5 in the mouse does not lead to apparent skeletal muscle abnormalities (Braun *et al.*, 1992; Rudnicki *et al.*, 1992). Rather, the expected absence of myofibers in the developing mouse is only revealed through the combination of both mutations, indicating that Myf-5 and MyoD are redundant for the function of one another in this process (Rudnicki *et al.*, 1993). Cross-regulation between these paralogues, as exemplified by the three fold increase in Myf-5 transcript levels in MyoD null mice, is presumed to be responsible, at least in part, for this redundancy (Rudnicki *et al.*, 1992; reviewed in Weintraub, 1993). Further, although Myf-5 expression is normally repressed by MyoD activity, Myf-5 normally exerts positive regulation upon MyoD transcription. The downstream regulator myogenin provides an additional influence on the network by acting positively on the transcriptional regulation of MyoD in a positive feedback manner

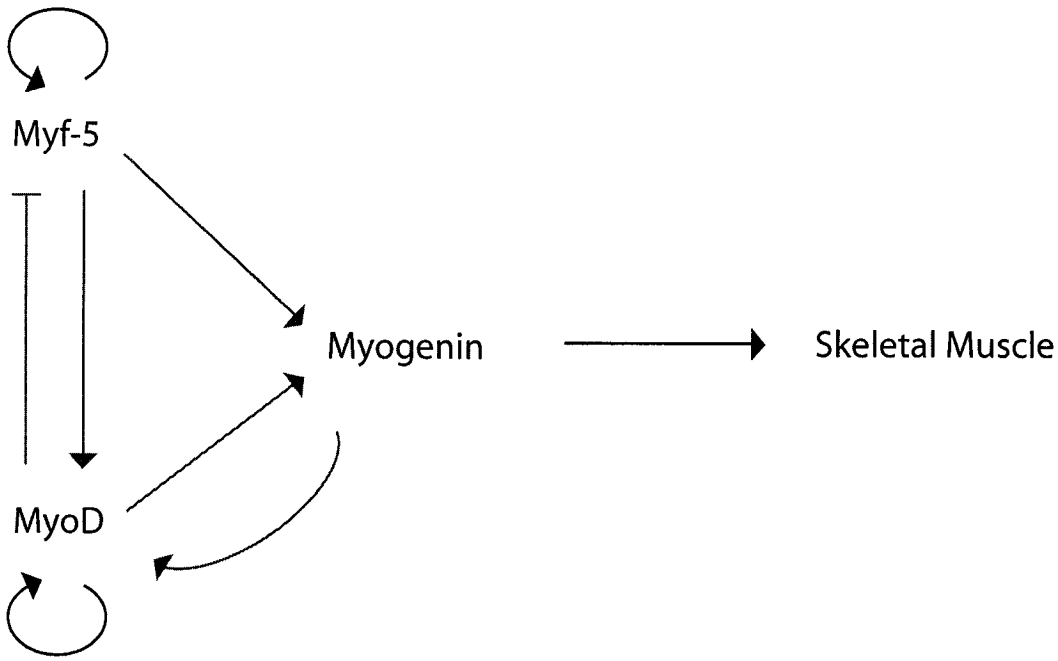


Figure 25. **Simplified representation of redundant gene function leading to skeletal myogenesis.** (Adapted from Weintraub, 1993) The duplicated paralogues *Myf-5* and *MyoD* are partially redundant for establishing a genetic programme leading to the expression of the transcription factor *myogenin* and for subsequent skeletal muscle formation. Genetic and biochemical analyses have revealed complex cross-regulation between these genes. Green arrows indicate positive influence on transcription, while the red line indicates transcriptional repression. Loops indicate positive autoregulation.

(**Figure 25**) (Rudnicki *et al.*, 1992; reviewed in Weintraub, 1993). These observations hint at the possibility that positive and negative cis-regulatory elements were present in the common ancestral MyoD-like gene in invertebrates, which were subsequently partitioned between the duplicated vertebrate paralogues MyoD and Myf-5, allowing for the evolution of a complex autoregulatory network necessary for the formation of skeletal muscle.

Gene duplication is not invariable synonymous with redundancy and may give rise to other modes of paralogous gene interactions with possible benefits to organismal evolution. This is exemplified by the observed antagonistic activities of duplicated Wnts genes (Topol *et al.*, 2003; Westfall *et al.*, 2003, reviewed in Weidinger and Moon, 2003).

3.2.3.5 DDC: the duplication-degeneration-complementation model – The full spectrum of redundant gene activity ranges from complete redundancy seen among duplicated housekeeping genes, to those that have evolved completely independent functions through the acquisition of novel activities. These extreme cases contrast with the more commonly encountered duplicates that share redundant functions in a subset of cellular or developmental processes, but have distinct roles in other processes. Such observations hint at a link between redundancy and pleiotropy of gene activity.

The partitioning or devolution of the pleiotropic activity of an ancestral gene onto duplicate pairs led to the creation of the duplication-degeneration-complementation (DDC) model to explain the high incidence of duplicate gene pair retention (Force *et al.*, 1999). The DDC model incorporates the concept of sub-functionalization, which proposes that the two copies arising subsequent to gene duplication undergo loss-of-function mutations in complementary but independent sub-functions. Such changes, often observed as partially non-overlapping expression patterns and biochemical function, are viewed as governing elements predisposing duplicates to fixation in the genome through their mutual requirement to reproduce the full complement of functions of the ancestral gene.

3.2.3.6 Examples of functional devolution – The idea of functional devolution originally came from observations made for the duplicated lens *crystallins* genes (described below) (Piatigorsky and Wistow, 1991). Some crystallins, however, have derived a novel activity via an alternate route independent of, and prior to, gene duplication. For example, the metabolic enzymes lactate dehydrogenase-B/ε-crystallin and α-enolase/τ-crystallin are encoded by single genes in the duck. They have been recruited to evolve a structural role with respect to the refractive properties of the lens through a change in gene expression, while preserving their catalytic function in non-lens tissues. This phenomenon, whereby two (or more) protein phenotypes result from the same transcriptional unit, has been called gene sharing (Piatigorski, 1988; Hughes, 1994), and may be a contributing factor to the pleiotropic activities of certain gene products.

Gene sharing is not unique to the crystallins. Other examples include the enzyme phosphohexose isomerase and the neurotrophic factor neuroleukin (Chaput *et al.*, 1988; Faik *et al.*, 1988), as well as a remarkable protein with four known functions (protein disulphide isomerase, thyroid hormone binding protein, the β-subunit of prolyl hydroxylase, and the glycosylation site binding component of oliogosaccharyl transferase) (Geetha-Habib *et al.*, 1988). Translational frameshifting is another sophisticated mechanism used by certain retroviruses as well as by the intracisternal A particles (IAP) elements long-terminal-repeat (LTR) retrotransposon: these can express protease and polymerase proteins from the same transcription unit (Fehrmann *et al.*, 1997). A spectacular example of gene sharing is offered by the tumour suppressor ink4a/ARF locus (Lowe and Scherr, 2003). Transcripts derived from this gene are translated via two alternative reading frames to encode either $p16^{ink4a}$ or $p19^{Arf}$ (Kamijo *et al.*, 1997), proteins that are involved in suppressing the cell cycle by stimulating the activities of the *retinoblastoma* and *p53* transcription factors, respectively (Bates *et al.*, 1998).

As indicated previously, some crystallins have been subjected to partitioning of function through the generation of gene duplicates, presumably subsequent to gene sharing. This is illustrated by the contrasting properties of the tandemly duplicated chicken argininosuccinate lyase (ASL)/δ-crystallin genes. While δ1-crystallin produces >95% of both transcripts in the lens, the δ2-crystallin appears to encode the enzymatically active ASL and is the predominant form in non-lens tissues, even if its expression level is still higher in the lens (Thomas, 1990).

A recent and convincing demonstration of the devolution of function following gene duplication comes from the existence of duplicated RNase genes which allows for one enzyme to specialize in the degradation of single-stranded RNA at low pH in the small intestine, leaving the breakdown of double-stranded RNA to the other gene product (Zhang *et al.*, 2002).

Hox paralogues have been used as a paradigm to evaluate further the functional contribution of duplicates, and the extent to which they display redundancy. Despite the largely overlapping developmental expression domains of three Hox paralogues in group 3, namely Hoxa3, Hoxb3 and Hoxd3 (Hox genes come in 13 groups), unique functions were attributed to Hoxa3 and Hoxd3. For example, inactivation of Hoxa3 by homologous recombination results in postnatal lethality due to improper formation of the neural-crest derived pharyngeal tissues (Chisaka and Capecchi, 1991). Removal of Hoxd3, in contrast, leads to viable mice with skeletal abnormalities (Condie and Capecchi, 1993). Disparity in phenotypic consequences for these genes was originally interpreted as being the result of differing biochemical properties of each encoded protein product, since these display less than 50% sequence conservation. The combined removal of these Hox genes resulted in the worsening of the vertebral defects normally observed for $Hoxd3^{-/-}$ mice, further suggesting a quantitative contribution from these duplicated paralogues (Condie and Capecchi, 1994; Manley and Capecchi, 1997; Manley and Capecchi, 1998). The extent to which both qualitative and quantitative traits contributed to the observed phenotypes was investigated further by replacing either of the respective coding portion of these two Hox genes by that of

its paralogue, while preserving the regulatory sequences intact. Mice generated through these elegant gene knock-in strategies were indistinguishable from wild type siblings, indicating that expression levels are key determinants of the overall activity of at least some paralogous Hox genes (Greer *et al.*, 2000).

A similar paradox was noticed when *Junb* was introduced into the *Jun* locus using the knock-in technique (Passegué *et al.*, 2002), where JunB protein rescued the liver defects and embryonic lethality normally encountered in *Jun*-deficient mice (Eferl *et al.*, 1999). This result was unanticipated, since the individual knockout for these paralogues displays different phenotypes (Eferl *et al.*, 1999; Schorpp-Kistner *et al.*,1999), and their respective protein product are known to have opposing effects upon proliferation and gene expression *in vitro*. Jun acts positively on proliferation (Pfarr *et al.*, 1994) through stimulating cell cycle regulators such as cyclin D1 expression (Bakiri *et al.*, 2000) and downregulating p53 and its target p21 (Schreiber *et al.*,1999). In contrast, JunB antagonizes Jun activity through stimulating the expression of p16INK^{4a}, a cyclin-dependent kinase inhibitor (Bakiri et al., 2000; Passegué and Wagner 2000)

In summary, the retention of some classical paralogues can be under the influence of altered gene expression. An implication for this mechanism to be operative is the functional equivalence of the respective protein products. Prevailing mechanisms are the partitioning of gene dosage within the same spatio-temporal context or through non-overlapping expression patterns. A variation on this theme is the apparent evolution of the regulatory regions of duplicated yeast genes (Papp *et al.*, 2003)

3.2.3.7 The retention of paralogues defining co-orthology to a single gene – While the previous examples provide insight about some of the mechanisms behind the retention of paralogous duplicate genes, strong support for the DDC model proper comes from the functional analysis of zebrafish genes defining co-orthologous relationship to a mammalian counterpart. In each case, although much functional overlap exists, distinct functional attributes account for the retention of both duplicate pairs. In some cases, the presence of zebrafish duplicates allows for the analysis of a homozygous mutant condition otherwise not possible in other vertebrates. This condition often permits researchers to unravel sub-functions for the mouse counterpart. In order to highlight known variations on the mechanisms of devolution of function, some of which may be applicable to zebrafish granulins, four pertinent examples are presented below.

3.2.3.7.1 Example 1: *Nodal* – In mouse, a single Nodal gene of the TGF-β superfamily exists that displays pleiotropic activities during development (Zhou *et al.*, 1993; Conlon *et al.*, 1994; Jones *et al.*, 1995; Varlet *et al.*, 1997; Brennan, *et al.* 2002; reviewed in Whitman, 2001). The presence of functional duplicates of this gene in zebrafish uncovered as the mutations *squint* - also called nodal-related1 - and *cyclops*, has facilitated the study of Nodal activity during embryogenesis. Although *squint* is maternally expressed and *cyclops* transcripts occur zygotically, the activities of both are partially redundant with respect to the establishment of the zebrafish shield (organizer) and in the formation of the endoderm and mesoderm germ layers (Feldman *et al.*, 1998; Dougan *et al.*, 2003). A late requirement for *cyclops* in the patterning of the neural plate has also been revealed through the presence of early *squint* activity, which allows for the embryo to develop at this stage (Hatta *et al.*, 1991; Sampath *et al.* 1998). The latter activity for Nodal in the mouse is masked due to early lethality.

Interestingly, *squint* and *cyclops* also differ in the way they exert their influence on the developing embryo. While *squint* is the first vertebrate factor demonstrated to fulfill all the requirements to act as a morphogen, *cyclops* does not share those properties (Chen and Schier, 2001).

3.2.3.7.2 Example 2: *Sox9* – Sox9 provides another example in which the understanding of a mammalian gene has been extended through the analysis of zebrafish co-orthologues. Mutations in the human SOX9 gene have been shown to be the cause of a complex condition characterized by sex reversal and extensive cartilage phenotypes referred to as campomelic dysplasia (Foster *et al.*, 1994; Kwok *et al.*, 1995, Cameron *et al.*, 1996). Again, early lethality observed for this gene in the homozygous mouse has prevented the detailed study of SOX9 in cartilage formation (Bi *et al.*, 1999; Bi *et al.*, 2001; Kist *et al.*, 2002).

A duplicated zebrafish SOX9 gene, *Sox9a*, is mutated in the *jellyfish* mutation (Yan Y-L, *et al.*, 2002). Despite being a recessive-lethal mutation, *jellyfish* embryos develop to larval stages, allowing for a more extensive analysis of cartilage formation in this genetic background. Interestingly, it has been observed that SOX9 function is not required for the initial specification and migration of neural crest cells fated to give rise to cartilage, but rather is necessary for proper cartilage morphogenesis and differentiation (Yan *et al.*, 2002).

3.2.3.7.3 Example 3: *Mitf* – Zebrafish has duplicated microphthalmia-associated transcription (*Mitf*) genes, which are required for the formation of pigment cells. In humans, different mutations affecting the single *Mitf* gene lead to syndromes that affect sensory systems and pigmentation, namely the Waardenburg syndrome type 2a (Tassabehji *et al.*, 1994) or Tietz syndrome (Smith *et al.*, 2000), respectively. In mouse, the *Mitf* gene undergoes extensive alternative splicing. When mutated, this gene leads to loss of both pigmented melanocytes derived from neural crest and retinal pigment epithelium (Hodgkinson *et al.*, 1993). Crest-derived melanocytes are missing in zebrafish for which the *mitfa* gene is mutated (the *nacre* mutant) (Lister *et al.*, 1999). The intact retinal epithelium of *nacre* fish is brought about in part by the co-expression of *mitfa* and *mitfb* in this tissue (Lister *et al.*, 2001). Interestingly, *mitfa* and *mitfb* genes are homologous to distinct isoforms of the mammalian *Mitf* gene (Lister *et al.*, 2001). In fact, the duplicated *mitf* genes do not differ significantly in their sequence; rather *mitfb* utilizes an alternative 5'exon allowing *mitfa* and *mitfb* to

recapitulate the expression pattern and actions of the isoforms of the mammalian orthologue (Altschmied *et al.*, 2002).

Interestingly, the *nacre* mutant cannot be rescued to the same extent by the two *Mitf* genes, indicating that the two tissue-specific isoforms generated by alternative splicing of the ancestral gene were converted into two genes with slightly differing properties.

The previous two examples indicate that DDC mechanisms in the teleost lineage, giving rise to the devolution of ancestral function onto duplicated genes, can provide insight into the pleiotropic roles of human genes with disease.

3.2.3.7.4 Example 4: *Hoxb1* – Another striking example of devolution of function generated through the subdivision of ancestral gene expression is provided by the analysis of the zebrafish *hoxb1a* and *hoxb1b* duplicates.

Transient expression of *hoxb1b* is initially observed in rhombomere4 up to 10 hpf in the developing zebrafish embryo, then gradually retreats towards the posterior. *Hoxb1a*, in contrast, has a later onset of expression (10hpf) in this region of the hindbrain and is stably maintained because of autoregulation. Combined, these expression patterns recapitulate that of the mouse *hoxb1* orthologue, which is in accordance with the DDC model. Hence, the *hoxb1b* shares the early expression pattern of the mouse hoxb1 gene in the hindbrain of gastrulating embryos, whereas *hoxb1a* mimics the later expression pattern in a single segment of the hindbrain (r4) at the neurulation stage. Complementary degenerative loss in the cis-regulatory elements of the gene duplicates has been recognized (McClintock *et al.*, 2002), and is presumed to be sufficient for allowing the fixation of both genes in the teleost genome.

The partitioning of the expression pattern of the ancestral *hoxb1* gene onto duplicates predicts that zebrafish *hoxb1a* and *hoxb1b* genes will functionally phenocopy the mouse *hoxb1* activity when combined. However, this is not strictly the

case. Homozygous null mice for hoxb1 display a change in neuronal identity due to a failure of facial neurons derived from rhombomere4 to undergo their normal migration towards the posterior (Goddard *et al.*, 1996; Studer *et al.*, 1996; Gaufo *et al.*, 2000). Knockdown of zebrafish *hoxb1a* reveals that this duplicate is similarly involved in the correct migration of facial neurons (McClintock *et al.*, 2002). Unexpectedly, *hoxb1b* is instead required for the proper segmental organization of the hindbrain (McClintock *et al.*, 2002). The latter activity is associated with *Hoxa1* in the mouse (Carpenter *et al.*, 1993). Functional shifting of a HoxA gene to a HoxB gene probably occurred because of the extensive redundancy encountered among Hox paralogue group 1 genes, allowing function shuffling to occur. This is distinct from non-orthologous gene displacement, since the zebrafish has a *Hoxa1* gene (see Koonin *et al.*, 2000).

3.2.4 Tracking down the origins of the chimeric granulin transcript – Chimeric transcripts usually result through one of the following mechanisms: chromosome translocations, alternative splicing in *trans* (described below, section 3.2.3.4.2), or transcription through neighbouring genes as a single transcription unit. In all cases, joining of exons is predicted to occur through the recognition of canonical splice acceptor and donor sites.

A hybrid granulin structure has previously been reported through cloning of cDNA sequences in the rat (Plowman *et al.*, 1992). Specifically, a structural variant of the full sequence was retrieved and predicted to encode a granulin domain consisting of the amino-terminal domain of granulin-C fused to the carboxyl-terminus of granulin-D. As was interpreted by Bateman and co-workers, this smaller precursor harbouring a hybrid form was consistent with the removal of an exon from the deduced larger primary transcript (Bhandari et al., 1993), based on the genomic structure of human progranulin (Bhandari and Bateman, 1992). The hybrid progranulin reported in this thesis originates through a mechanism other than alternative splicing from a larger primary transcript.

Explaining the origin of the transcript encoding a hybrid progranulin through gene conversion between progranulin-1 and -2 cannot be excluded, but is unlikely since it would require partial exchange between these genes. Unravelling the contribution of gene conversion in this case would require the sequencing of genomic sequences from different tissues or animals. The inability to isolate a chimeric transcript of the opposite character, namely consisting of N-terminal progranulin-2 joined to C-terminal progranulin-1, also argued against the alternative splicing of a larger bicistronic primary transcript consisting of progranulin-1 followed by progranulin-2, unless the creation of the documented hybrid architecture is favoured by some other mechanism. It could also be possible that two exons encoding progranulin-1 are located upstream of, and in the same orientation to, the progranulin-2 gene as a consequence of a partial gene duplication event. However, we find no evidence for this by analyzing genomic sequences.

The most parsimonious explanation for the existence of a hybrid progranulin, is the process of trans-splicing, similar to that originally documented in Trypannosomatids. The discovery of a single transcript in antisense orientation to, and sharing exonic complementarity with, the progranulin-1 and -2 genes provides indirect support for *trans*-splicing (the nature and consequence of this transcript are discussed in detail in a later section). In fact, full sequencing and annotation of the antisense transcript at the genomic level confirms that the progranulin-1 gene is in tandem and downstream to progranulin-2, contrary to that suggested by the chimeric transcript (**Figure 19**). This, along with the absence of intronless genomic sequences corresponding to hybrid progranulin, suggests that this chimeric mRNA is not generated through conventional splicing, or splicing in *cis*.

3.2.4.1 *Trans*-splicing: a retrospective – The production of chimeric RNA molecules through the recombination of distinct pre-mRNA transcripts, involving the splicing reaction in *trans*, has been documented to a varying extent in all phyla.

Trans-splicing was originally observed in (unicellular) trypanosomatids through the discovery that several transcripts are discontinuous in their 5'end sequences (Campbell *et al.*, 1984; Kooter *et al.*, 1984). A common 5'end exon observed in many if not all transcripts was later shown to be derived from a small RNA expressed from various locations within the genome (Milhausen *et al.*, 1984), and that this was somehow joined to other RNAs expressed from independent loci. This type of splicing, whereby the 5' region of a pre-mRNA is displaced by the 5' end of another small spliced leader RNA, is referred to as *SL RNA trans*-splicing (**Figure 26**).

Trans-splicing involving a *SL RNA* has been documented in metazoa, specifically in two phyla of the protostome clade: the *C. elegans* nematode (Kraus and Hirsch, 1987) and the platyhelminthe (trematode) *Schistosoma mansoni* (Rajkovic *et al.*, 1990). More recently, at least 7 independent mRNAs with a 5' end sequence

derived from a common 96-nucleotide SL RNA were discovered in *Ciona intestinalis* (Vandenberghe *et al.*, 2000), a deuterostome chordate thought to be related to vertebrate ancestors (Katz, 1983). Although cultured mammalian cells are capable of supporting the *C. elegans* SL *trans*-splicing reaction (Bruzik and Maniatis, 1992), no example of endogenous 5' SL *trans*-splicing has been reported to date in vertebrates.

Several functions have been ascribed to the phenomenon of SL *trans*-splicing. These include the conversion of polycistronic pre-mRNA (RNAs organized in tandem or overlapping the coding regions of other RNAs) into mature monogenic mRNAs (Johnson *et al.*, 1987; LeBowitz *et al.*, 1993; Blumenthal, 1995), influences on mRNA stability and translatability by providing identical 5' terminal cap structures (Perry *et al.*, 1987; Maroney *et al.*, 1995), and the potential creation of protein-coding mRNAs from RNA polymerase I-generated primary transcripts (Lee and Van der Ploeg, 1997).

3.2.4.2 Non SL-mediated transplicing – There is increasing evidence that mature eukaryotic RNAs do not always follow a linear combination of exonic sequences. Examples of likely transplicing between two RNA molecules transcribed from a single gene resulting in exon repetition include the rat carnitine octanoyltransferase (Caudevilla *et al.*, 1998), sodium channel SNS (Akopian *et al.*, 1999), hypertension-related SA (Frantz *et al.*, 1999), and human Sp1 genes (Takahara *et al.*, 2000).

Although rare, scrambled or intergenic RNA molecules consisting of exons originating from distinct genes through a trans-splicing reaction are also possible in eukaryotes (Sullivan *et al.*, 1991; Vellard *et al.*, 1991; Finta and Zaphiropoulos, 2000; Breen and Ashcroft, 1997; Kim *et al.*, 2000). For instance, the acyl-CoA:cholesterol acyltransferase-1 (*ACAT-1*) hybrid mRNA occurs naturally in humans (Li *et al.*, 1999). Even more surprising is the finding of the *trans*-splicing of the voltage-gated sodium channel in response to nerve growth factor stimulation, indicating that this alternative mode of splicing can be a regulated process (Akopian 1999).

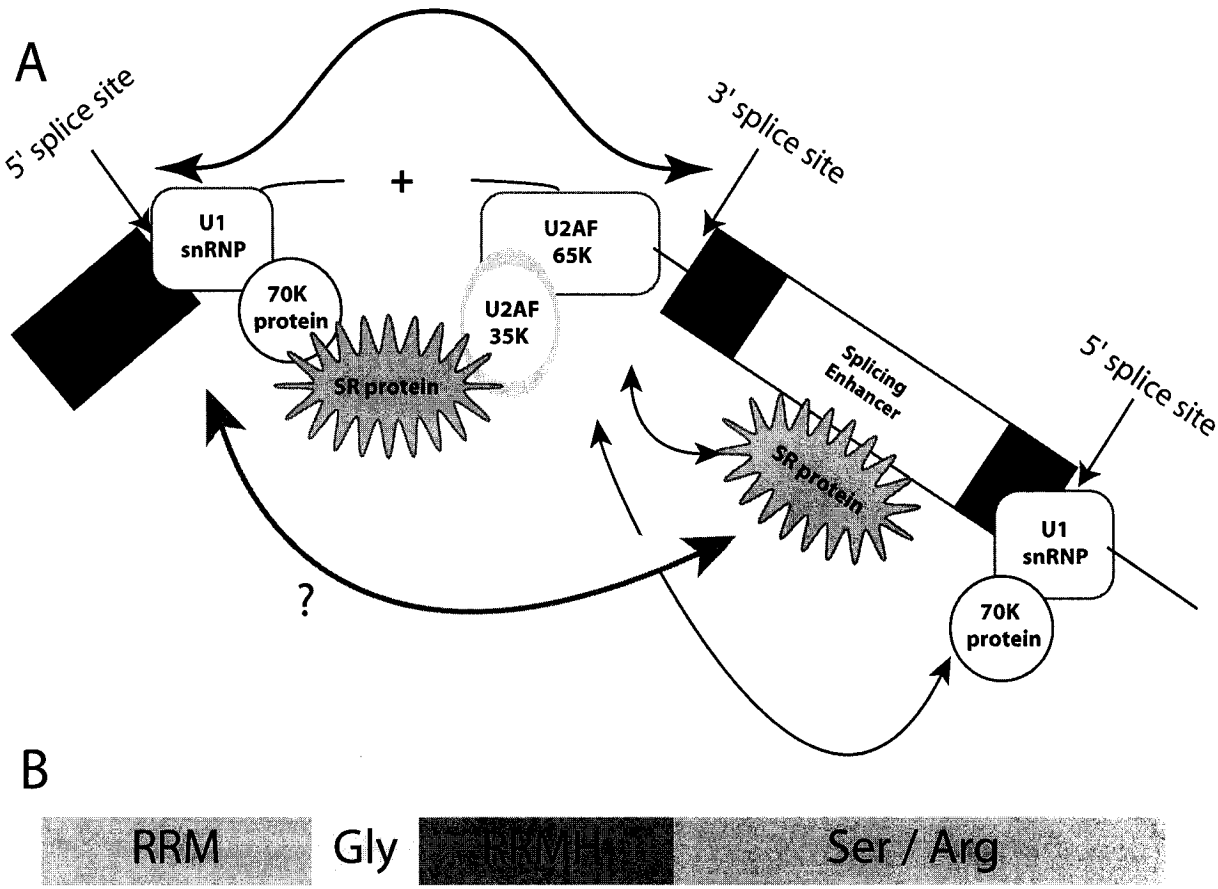


Figure 27. **The involvement of SR proteins in *cis*- and *trans*-splicing of heterologous precursor RNAs.** Panel A: Similar to *trans*-splicing involving a *SL RNA*, distinct eukaryotic pre-mRNAs originating from different loci can be spliced together to yield a mature chimeric mRNA. This event appears to be favoured by the presence of splicing enhancer sequences that bind to various splicing factors of the Ser-Arg family, which in turn are proposed to provide a bridge between 5' and 3' splice site-binding components of the spliceosome in *cis* or in *trans* (Bruzik, 1996). Specifically, the mechanism can be envisioned as follows (taken as is from Bruzik, 1996): "If the U1 snRNP interacts with the 5' splice site via base pairing with the U1 snRNA (Zhuang and Weiner, 1986) and the U1 70 kDa protein (containing an SR domain) is an integral component of the U1 snRNP (Query *et al.*, 1989), then these factors serve to mark the 5'splice site. At the other end of the intron, the U2AF 65 kDa protein specifically recognizes the region upstream of the 3' splice site (the polypyrimidine tract) (Zamore and Green, 1989) and then tightly interacts with its heterodimeric partner, U2AF 35 kDa (Zhang *et al.*, 1992). With these two complexes demarcating the important regions of the pre-mRNA transcript, other SR proteins, namely SC35 and/or SF2/ASF can serve as a bridge between the 5' and 3' splice sites by simultaneously interacting with both U1 70 kDa and U2AF 35kDa (Wu and Maniatis, 1993)". Panel B: Ser-Arg proteins are generally constituted of the following domains from the amino- to carboxyl-terminus (all domains are not invariably present in all SR proteins): the RNA recognition motif (RRM), including RNP2 and RNP1; a Glycine-rich region; a second degenerate RRM ~30% homologous to the RRM termed the RRM homologous region (RRMH); and the Ser-Arg (SR) domain (Bruzik, 1996).

The expression level of trans-spliced transcripts is usually low relative to transcripts normally spliced in *cis*, making their detection often problematic. Nevertheless, sporadic cases have been documented for nearly 20 years (Konarska *et al.*, 1985 Solnick, 1985). In one of these original cases, it has been suggested that base-pairing between the substrate primary transcripts, is necessary for splicing in *trans* to proceed (Solnick, 1986; Sharp and Konarska, 1986). The requirement for sequence complementarity in trans-splicing is estimated to be unlikely due to the sense-sense transcript joining of unrelated genes, but has not been systematically ruled out.

3.2.4.3 Mechanism of splicing in *trans* – The mechanism of *trans*-splicing is similar to conventional splicing in *cis* (**Figure 27**). In each case, the recognition of the same splice donor and acceptor sites, nucleotide sequences involved in the process of exon and intron definition, is achieved by a macromolecular complex termed the spliceosome (Lamond, 1993; Berget, 1995). Assembly of the spliceosome involves a cocktail of RNAs collectively termed the U1, U2, U4/U6 and U5 small nuclear RNAs that complex with proteins to form the small nuclear ribonucleoprotein particles (snRNPs) (Wise 1993; Nilsen, 1994). Other protein constituents, called splicing factors, are also required for the recognition and removal of introns (Wu and Maniatis, 1993; Fu, 1995). Families of protein sequence motifs involved in pre-mRNA interaction and removal of introns include the RGG (Arg-Gly-Gly), RRM (RNA recognition motif), and SR (Ser-Arg) domains (Lamm and Lamond, 1993; Burd and Dreyfuss, 1994; Fu, 1995).

Trans-splicing shares several features with *cis*-splicing. For example, the structure of the 3' splice acceptor site consists of an invariant dinucleotide AG, an upstream polypyrimidine tract of variable length, and internal adenosine residues for branch point structure formation (the lariat structure) (Huang and Van der Ploeg, 1991; Hummel *et al.*, 2000). Thus, there is no evidence suggesting that splicing in trans is favoured by the use of a second class of spliceosomal machinery, such as the one involving the U4atac, U6atac, U11 and U12 snRNA, that apparently specialize in the

recognition of introns demarcated by non-consensus AT-AC splice sites (Tarn and Steitz, 1996; Sharp and Burge, 1997).

3.2.4.4 Possible involvement of the SR protein splicing factors – How are the 5' and 3' splice sites initially juxtaposed in the trans-splicing reaction ? As mentioned previously, whether sequence complementarity is required between pre-mRNAs to allow for *trans*-splicing is at present speculative (Bruzik and Maniatis, 1992). This is unlikely, since many chimeric RNAs have no obvious nucleotide sequence complementarity. Nevertheless, RNA-RNA interactions are typically stronger than other combinations of nucleic acids; it is therefore possible that only small stretches of complementarity are necessary for brief juxtaposition of pre-mRNAs.

The coincidence of sequence complementarity between each region encompassing exons 2 and 3 of progranulin-1 and progranulin-2 pre-mRNAs (which include the 5' and 3' splice sites involved in *trans*-splicing) to exons 2 and 3, respectively, of the antisense transcripts is striking (**Figure 16**). This accident of circumstance implicating antisense transcript-mediated base-pairing for enhancing proximity between the trans-spliced primary transcripts may facilitate the genesis of the hybrid progranulin, irrespective of potential RNA interference (RNAi) elicited responses (RNAi is described later, section 3.2.6). From this viewpoint, the hybrid progranulin may in such a case be a fortuitous residual by-product of these interactions. Whether this mechanism is tenable and the chimeric RNA is functionally beneficial to the organism remain to be demonstrated.

Joining of the 5' and 3' splice sites in *cis* (Horowitz and Krainer, 1993; Staknis and Reed, 1994), and in *trans* (Bruzik and Maniatis, 1995), is mediated in part by the SR proteins, which interact with complexes formed at 5' and 3' splice sites. SR domain-containing proteins are able to associate directly with pre-mRNA and splicing factors located at either end of the intron. In addition, SR proteins undergo protein-

protein interactions potentially capable of forming bridges between 5' and 3' splice sites (**Figure 27**) (Wu and Maniatis, 1993).

The reaction in *trans* may be facilitated through the recognition of regulatory elements called splicing enhancers that require binding of SR proteins for activity (Tian and Maniatis, 1993; Ramchatesingh *et al.*, 1995; Bruzik and Maniatis, 1995). In fact, the presence of enhancers can dramatically stimulate the utilization of weak splice sites during conventional splicing (Tian and Maniatis, 1993; Lavigeur *et al.*, 1993). Similarly, the *in vitro* splicing reaction normally proceeding in *cis* for pre-mRNA substrates carrying both weak and strong 3' splice site sequences can be efficiently directed in *trans* in the presence of SR proteins when the 3' exon contains a splicing enhancer (Bruzik and Maniatis, 1995).

Thus, splicing enhancers and SR proteins promote and mediate splicing in *trans*, at least in HeLa cell nuclear extracts. From this observation, it has been proposed that stable spliceosome complex formation at 5' and 3' splice sites located on separate pre-mRNAs can functionally interact and that SR proteins play a critical role in this interaction (Bruzik and Maniatis, 1995; Chiara and Reed, 1995).

Although the documented examples are so far scarce, if *trans*-splicing is required for the production of physiologically relevant mRNAs, then splicing enhancers and SR proteins are likely to be involved in determining splice-site selections. As indicated previously, several protein motifs are involved in RNA recognition during the splicing event. The ability of SR proteins to undergo homotypic interactions and to recognize discrete sequences in RNA (Tacke and Manley, 1995) may be discriminating factors contributing to decision making between *cis*- and *trans*-splicing. As an example, exonic sequences functioning as splicing enhancers have been uncovered and include 5' splice site-like sequences and motifs containing the element (G/C)GAC(G/C) (Boukis and Bruzik, 1999). Whether a necessary enhancer sequence can be recognized downstream of the 3' *trans*-splicing acceptor site and within exon 3 of the *progranulin-2* pre-mRNA, and is responsible for promoting unidirectional

splicing in *trans* of the *progranulin-1* and *progranulin-2* pre-mRNAs, deserves further investigation.

3.2.4.5 The unsuspected common occurance of chimeric RNAs – Various examples of apparently naturally occurring *trans*-spliced vertebrate mRNAs have been enumerated previously. A large-scale screening effort within the HGI and RefSeq databanks for chimeric sequences that originate either through translocation events or from *bona fide* multi-locus transcription was recently conducted (Romani *et al.*, 2003). Surprisingly, 21742 chimeric sequences, representing approximately 1% of database transcripts, have been identified. Of this number, however, only 1% has been considered to be of non-artefactual nature, as judged by the apparent junction of two independent pre-mRNAs in a sense-sense orientation at canonical splice sites. The remainder appear disjointed at their predicted donor-acceptor exon-bordering sites, thereby likely representing artefacts of the PCR or incorrect ligations and abnormal reverse-transcription in the creation of cDNA libraries (Brakenhoff *et al.*, 1991; Chetverina *et al.*, 1999; Chang and Taylor, 2002). Some chimeras also possess segments in antisense orientation (Lehner *et al.*, 2002; Romani *et al.*, 2003). It should be noted that hybrid transcripts with incongruent splice donor and acceptor sites might in some situations represent naturally occurring RNAs. The finding that the rat Leukocyte Common Antigen-Related (*LAR*) tyrosine phosphatase receptor contains an alternative 3' UTR (Zhang *et al.*, 2003), supports the notion that not all chimeric RNAs invariably result from cloning artefacts, but may rather be the result of non-conventional or atypical spliceosome-mediated splicing in *trans*.

3.2.4.6 Physiologically relevant naturally occurring chimeric RNAs ? – Chimeric transcripts among the CYP2C group of cytochrome P450s have been isolated from human liver and epidermis (Zaphiropoulos, 1999; Finta and Zaphiropoulos, 2000). The discovery of *trans*-splicing for P450 cytochromes genes has been extended to members belonging to the CYP3A family, demonstrating that *trans*-splicing may be more prevalent and widespread than originally presumed. Specifically, CYP3A paralogues are heme-containing monooxygenases cytochrome P450s involved in the

hydroxylation of endogenous steroids and of xenobiotics that collectively form a cluster of four genes (CYP3A43, CYP3A4, CYP3A7, and CYP3A5) on human chromosome 7q21-22.1. Using various combinations of exon-specific primers, hybrid RNAs consisting of CYPA43 exon1 joined to distinct sets of exons derived from either CYP3A4 or CYP3A5 have been uncovered (Finta and Zaphiropoulos, 2002). Interestingly, no hybrid RNA of other character was amplified, suggesting directionality in the formation of these intergenic mRNAs. The key observation in the interpretation of the mechanism underlying the origin of these chimeric transcripts is the head-to-head topology of the CYP3A43 gene relative to those of CYP3A4 and CYP3A5.

Of course, cataloguing authentic chimeric transcripts will be of little significance unless meaningful physiological contributions can be ascribed to their encoded products. A plausible example is found among the different CYP3A chimeric mRNAs, namely one composed of CYP3A43 exon 1 joined to exons 2-13 of CYP3A4 which encodes a protein that is capable of hydroxylating testosterone *in vitro* (Finta and Zaphiropoulos, 2002). In this specific case, the amino-terminal exon switching may be physiologically important by either modifying the *in vivo* metabolising properties of this enzyme, or in modulating its targeting, insertion or retention in membranes of appropriate cellular compartments (Neve and Ingelman-Sundberg, 2000; Finta and Zaphiropoulos, 2002). The amino-terminal portion of the respective granulin-1 and granulin-2 peptides differ in hydrophobicity, due to the selective presence of four arginine residues within granulin-2, giving the latter a more basic character (**Figure 11 and Figure 14, panel B**). Unfortunately, little is known about the biochemical function of individual granulins and the associated structural attributes that enable them to interact with other cellular components (e.g. binding proteins, receptor(s)). Once the structure-function relationships are better defined for granulin-1 and granulin-2, specific hypotheses aimed at determining to what extent, if at all, hybrid progranulin is advantageous to the organism may become apparent.

3.2.4.7 *Trans*-splicing diversifies the proteome – The next example convincingly demonstrates that under special circumstances, *trans*-splicing can be essential to the organism. In *Drosophila*, functional diversity is generated by the mechanism of *trans*-splicing at the *modifier of mdg4* locus. Expresion of the *mod(mdg4)* gene illustrates an extraordinary level of complexity through the joining of its 4 exons to one of 26 independently transcribed 3'ends encoding alternative C-terminal exons of the protein (Dorn *et al.*, 2001; Buchner *et al.*, 2000). Surprisingly, several of these 3' exons are transcribed from the opposite DNA strand (Labrador *et al.*, 2001; Dorn *et al.*, 2001).

Although they share the first 4 exons, providing a common BTB/POZ domain, most, if not all, resultant protein isoforms possess specific functions within the nucleus (Dorn and Krauss, 2003). This is apparently due to the different DNA-interacting properties of the alternative 3' exons, allowing the isoforms to bind to multiple and different sites on polythene chromosomes (Buchner *et al.*, 2000).

As would be expected, homozygous mutations encountered within any of the *mod(mdg4)* exons disrupt gene function in all product variants. However, *trans*-heterozygotes for mutations of this gene display interallelic complementation. Specifically, independent mutations located within the 5' exons (i.e. exons1-4) on one allele of the gene can complement mutations in an alternative 3' end exon on another allele (Mongelard *et al*, 2002). *Trans*-splicing between the two mutant pre-mRNAs in this case is able to restore wild-type levels of functional protein.

3.2.4.8 *Trans*-splicing and the clinic – Interestingly, the mechanisms of *trans*-splicing is currently being exploited as a tool for therapeutic intervention in gene therapy (Garcia-Blanco *et al.*, 2000; Sullenger *et al.*, 2002; Garcia-Blanco, 2003). For instance, a spliceosome-mediated RNA *trans*-splicing (SMaRT) strategy has been used in cell culture and in animals for correcting the function of mutant proteins such as CFTR (Puttaraju *et al.*, 1999; Mansfield *et al.*, 2000; Puttaraju *et al.*, 2001; Liu *et al.*, 2002). Further, SMaRT has recently been successfully applied to the correction of

haemophilia A mice *in vivo* (Chaos *et al.*, 2003), while another clinical *trans*-splicing strategy employing a ribozyme approach is now being optimised for use *in vivo* (Rogers *et al.*, 2002).

3.2.4.9 Chimeric progranulin summary: a miniature Pandora's box – In summary, *hybrid progranulin* is a full-length mRNA, and is not derived from the transcriptional termination by-pass of two tandemly duplicated genes. All exons are spliced at canonical sites, indicating that the progranulin1-2 hybrid is a result of a pre-mRNA splicing process. Directionality is favoured in the formation of the progranulin chimeric RNA, as no product encoding the opposite set of exons was amplified and sequenced. This suggests that sequence differences between the progranulin-1 and progranulin-2 pre-mRNAs such as the relative presence of a splicing enhancer or absence of a splicing repressor, may predispose the *trans*-splicing reaction to occur in one direction, but not the other. Unfortunately, the limited number of nucleotide sequence divergence observed between progranulin-1 and progranulin-2 makes it technically challenging to select primers whose sequences discriminate between the different isoforms without ambiguity in the PCR. This is not to say that a solution to this problem is not achievable. This also prevents the semi-quantitative analysis of each transcript concerned using the real-time 5'exonuclease (TaqMan) PCR approach (Heid *et al.*, 1996).

Quantitation via the RNase protection assay is similarly hampered for the following reason: the protected fragments for all transcripts are of identical size, preventing cross-reactive hybridization (protection) from being noticed during electrophoresis.

Nevertheless, it is predicted that the joining of progranulin-1 exon2 to progranulin-2 exon3 (*hybrid progranulin*) (**Figure 14, panel A; Figure 19**) is a much more rare event than the canonical joining of progranulin-1 exon2 to progranulin-1 exon3.

Finally, during the synthesis of mature mRNAs *in vivo*, transcription, capping, splicing and polyadenylation are mutually regulated (Hirose and Manley, 2000). However, during the formation of chimeric RNAs, transcription and *trans*-splicing are likely to occur apart from each other, both spatially and in time, especially for pre-mRNAs originating from distant loci and whose expression are under different regulation. This is consistent with the finding that *in vivo* trans-splicing and polyadenylation may occur independently (Graessmann *et al.*, 1996). The detection of an amplicon corresponding to this chimeric RNA from cDNA prepared from oligo-dT-primed RNA template suggests that *hybrid progranulin,* like progranulin-1 and progranulin-2, is polyadenylated (data not shown).

3.2.5 The antisense progranulin-1/-2 gene

3.2.5.1 Overview of the antisense progranulin transcript – The 1989 nucleotides-long spliced and polyadenylated transcript in antisense orientation to the *progranulin-1* and *-2* genes is encoded on four exons, and spans at least 19.8 kb of genomic DNA (**Figures 16 and 19**). The second and third exons of this transcript, respectively, display full complementarity to regions encompassing exons 2 and 3 of *progranulin-1* and *progranulin-2*. Put differently, a genomic portion spanning exons 2 and 3 of the *progranulin-1* gene is complementary to the full sequence of *AS progranulin-1/2* exon 2. An identical portion within the *progranulin-2* gene is complementary to exon 3 of that same antisense gene (**Figures 16 and 19**).

Some mobile elements, including *tc1/mariner* transposons (section 3.2.5.2) and *Alu* repeats, possess their own polyadenylation signal downstream of the transposase gene, located within or in the vicinity of the inverted repeat region (Ivics Z *et al.*, 1996; Boeke, 1997). Whether or not the AS progranulin-1/-2 transcript derives its poly(A) signal from the tzf segment is not certain (the tzf is in the reverse complement orientation, thus requiring a poly(A) signal in each flanking repeat region (**Figure 18, panel A**). Nevertheless, a canonical polyadenylation sequence has been located within exon 4 (**Figure 16**).

AS progranulin-1/-2 possesses no apparent ORF besides a transposase sequence, belonging to the tzf transposon of the *tc1/mariner* superfamily of mobile genetic elements (section 3.2.5.3), but in the wrong orientation (**Figures 16 to 18**). This entire sequence conservation is restricted to exon 4, and has several insertions and deletions and stop codons in the transposase gene, suggesting that the encoded enzyme is no longer functional and was, like the majority of transposase genes, "dead-on-arrival" or shortly thereafter (Comeron, 2001).

BLAST searches aimed at uncovering homologous translated sequences within the remaining portion (exons 1-3) failed to detect any submitted sequence at *GenBank*,

even when considering possible internal ribosome entry sites (IRES) for the generation of small translated peptides.

3.2.5.2 Part of the AS progranulin-1/-2 gene harbours a *Tc1/mariner*-type transposon – The partial mobile element sequence found within the mature AS progranulin transcript is in the reverse complement orientation (**Figure 17**), precluding its potential role as a conventional transposon *per se*. Regardless of this, it is pertinent to overview briefly the general architecture of DNA transposons, along with aspects of the consequences of transposable element activity within the eukaryotic genome. This may help to interpret how the tzf transposon located within the AS progranulin-1/-2 transcript may have contributed to the formation of the smaller progranulin genes in zebrafish or to their exclusion outside the teleost lineage.

Mariner and *Tc1* from *Drosophila mauritania* and *Caenorhabditis elegans*, respectively, are canonical representatives of the *Tc1/mariner* superfamily of DNA transposons (Plasterk *et al.*, 1999). These mobile elements are universally present in eukarya and, like other transposons, are usually characterized by the presence of a single gene whose product, termed the transposase, is commonly involved in the recognition, excision, and insertion of cognate elements (**Figure 18, panel A**). The exclusive activity of each transposase on members of its own subfamily is predicated on the recognition of specific structural features concealed within flanking inverted repeat sequences that are unique to each subgroup (Lohe *et al.*, 1995; Plasterk *et al.*, 1999). This feature may contribute to the mechanisms of "overproduction inhibition" and dominant-negative complementation (Lohe and Hartl, 1996), which are partly responsible for the vertical inactivation of some elements (Lohe *et al.* 1996; Hartl *et al.*, 1997), and presumably underlies the differential representation of certain subfamilies among various genomes. This is in contrast to stochastic loss resulting from random genetic drift. The governing influences on the fitness of a particular transposon class over another in a given phylum, clade, or between closely related species such as in Drosophilae (Brunet *et al.*, 1994; Moschetti *et al.*, 1998), are at present not understood.

3.2.5.3 Attributes of the tzf transposon – A deduced active structure consisting of 1621 nucleotides for the *tzf* element has been assembled and shown to resemble that of other zebrafish transposons such as the *Tdr1*, but with only 38% amino acid sequence conservation overall (Lam *et al.*, 1996). Further, the *tzf* and *tdr1* end repeat sequences differ. Specifically, the *tzf* transposase gene encodes a predicted product of 339 amino acids, and is flanked by 200 bp of inverted repeats (Lam *et al.*, 1996; **Figure 18**). The presence of TA dinucleotides at both ends of the repeats suggests that, like *mariner* and *tc1*, the *tzf* integrates at the sequence TA (Ketting *et al.*, 1997; Plasterk *et al.*, 1999). The amino-terminal region of *Tc1* element transposases harbour sequences responsible for the interaction with DNA composed of two helix-turn-helix (HTH) motifs bearing similarity to the paired DNA-binding domain of transcription factors and the homeodomain DNA-binding domain of recombinases (Ivics *et al.*, 1996; Plasterk *et al.*, 1999). This is usually followed by a nuclear localization sequence (NLS) enabling receptor-mediated transport into the nucleus. The NLS is flanked by phosphorylation target sites for casein kinase II, which may act as a functional checkpoint in the regulation of transposition. DNA cleavage and joining reactions of transposition are mediated in part by three characteristically positioned residues collectively termed the DDE motif (Bukrinsky *et al.*, 1993; Doak *et al.*, 1994; Leaver *et al.*, 2001). The presence of these aspartic and glutamic acid residues is required for all catalytic activities (van Luenen *et al.*, 1994; Haren *et al.*, 1999).

A consensus sequence of 1621 nucleotides was derived from the alignment of five Tzf sequences (U51226-U51230) using the majority rule, and was used to compare the genomic portion of the AS progranulin-1/-2 gene (exon 4 area) (**Figure 17**).

Although *tzf* elements have suffered many disruptions in their transposase gene, these are very abundant in the zebrafish and are apparently still actively mobilizing within the genome, at an estimated 8 events per generation (Lam *et al.*, 1996). Interestingly, if the haploid genome of zebrafish is estimated to contain 50 000 genes, then, assuming every insertion into a transcription unit is disruptive, the mutation rate

would be expected to be 8 $X10^{-5}$ per individual (i.e. per generation), or 0.01% (Lam *et al.*, 1996). As pointed out by the authors, this number is similar to the calculated average background mutation rate for four pigmentation genes in zebrafish (Mullins *et al.*, 1994). This suggests that if a mobilization of the *tzf* elements located upstream to progranulin-2 and within the AS progranulin-1/-2 gene took place to generate the current organization of the smaller zebrafish progranulin genes, then this event may not necessarily be ancient. This reasoning assumes a causal relationship between the co-occurrence of the sense-antisense gene pairs concerned (progranulin-1 and/or progranulin-2 vs AS progranulin-1/-2), which has not been demonstrated. This is discussed more in section 3.2.5.9.

Active transposition with consequent gene disruption has been reported in several organisms. Interestingly, both *de novo* insertion of the *Tol2* DNA transposon in and excision from the *tyrosinase* gene of the medaka fish *Oryzias latipes* resulting in albinism, have been observed (Koga *et al.*, 1996). Autonomy of *Tol2* was later demonstrated using the zebrafish as propagating host (Kawakami *et al.*, 1998). The *Tc1* elements of *C. elegans* and the *P*-elements of the *Drosophila* are also a source of mutations in these animals (Robertson, 1993; Robertson and Lampe, 1995). Interchromosomal mobilization of *tc1* elements has been observed in mouse embryonic stem cells, suggesting that transposons have the potential to contribute to the background mutation rate in mammals (Luo *et al.*, 1998). In fact gene disruption events due to the insertion of mobile elements have been documented in humans (reviewed in Kazazian, 1999; Prak and Kazazian, 2000).

3.2.5.4 Genome architects: mobile parasitic elements providing structure to life ? – Transposable elements are coeval with the shaping of genomes (Doolittle and Sapienza, 1980; Comeron, 2001; Hurst and Werren, 2001), and are usually classified into two categories depending on their mechanism of transposition: Class I transposons mobilize via an RNA intermediate, while class II members transpose directly via a cut-and-paste mechanism. *Tc1/mariner* elements belong to the second category. Irrespective of their mode of mobilization, transposing DNA elements are

often considered to be selfish in nature. This trait is attributed to their apparent sole purpose, namely that of self-propagation in a random manner into a wide selection of host genomes, with little direct detectable functional contribution to the organism. This reputation is illustrated by several examples, including the spectacular colonization of the human genome by mobile elements, where they constitute approximately 42% of euchromatic DNA (Smit *et al.*, 1999). On the other hand, bona fide class II DNA transposon sequences represent roughly 2-3% of the human genome (Smit, 1999; Hattori *et al.*, 2000). Of these, m*ariner* elements distantly related to the *Tc1-mariner* group, comprise less than 3% of the bulk of DNA transposon copies (Smit, 1999), in keeping with the low copy number for these elements in previous reported estimates (Smit and Riggs, 1996).

The type and copy number of mobile DNA elements can vary greatly from one organism to the next: the distribution of a particular transposon subclass might be sporadic among closely related species, but be detected in the genome of an unrelated organism (Mayurama and Hartl, 1991; Robertson, 1993; Leaver, 2001). The vertical discontinuity in the distribution of mobile elements along the tree of life has led to the proposal that transposons are possibly subject to lateral (horizontal) transfer (Lohe *et al.*, 1995; Leaver, 2001). Possible vectors for the exchange of transposons, especially for aquatic species, are the parasitic helminth cousins of *C. elegans*, which have a complex life cycle normally involving multiple host species. There is also evidence suggesting that some *Drosophila* genes have been transferred through a similar vector mechanism (Houck *et al.*, 1991). Several observations point to the possible lateral gene transfer between organisms, ranging from bacteria to vertebrates (Genereux and Logsdon, 2003). For example, analyses of genic content in the human genome have identified many transcriptional units of potential direct bacterial origin (Ponting, 2001).

3.2.5.5 Mobile elements and disease – The disruptive consequence of mobile element integration within a transcription unit leading to human disease was first demonstrated for the inactivation of the factor VIII gene, in which long- interspersed-element-1 (LINE-1) sequence insertion resulted in haemophilia A (Kazazian *et al.*, 1988). Since then, several examples of the contribution of mobile element to human disease have been reported. Cases of Duchenne muscular dystrophy, β-thalassemia, and retinitis pigmentosa have each been reported to result from the inadvertent insertion of mobile elements in the respective dystrophin (Narita et al., 1993), β-globin (Divorsky *et al.*, 1996) and X-linked retinitis pigmentosa 2 genes (Schwann *et al.*, 1998). Further, gene-disruptive LINE-1 integration has been linked to chronic granulomatous disease (Meishi *et al.*, 2000), and as a contributing factor in the etiology of colon cancer by mutating the APC gene (Miki *et al.*, 1992). Transposon integration does not always produce null alleles of gene activity, however. In fact, depending on the location and nature of integration, chimeric transcripts may be produced (Holmes et al., 1994). It should be pointed out that human genetic disease resulting from L1 and other the insertion of other mobile elements is estimated to be infrequent (Kazazian, 1999).

Mobile elements can be substrates for homology-dependent recombination, thereby potentially providing an alternative mechanism for their implication in disease (Burwinkel and Kilimann, 1998). This is apparently the case for two *mariner*-like sequences distanced by about 1.5 Mb on human chromosome 17p11.2-p12 (Hartl, 1996). Analysis of the *CMT1A-REP* repeats has implicated the *mariners* as the hotspot of recombination (Kiyosawa and Chance, 1996; Reiter *et al.*, 1996). The *mariners* are part of a larger duplication of approximately 30 kb called the *CMT1A-REP* repeats. During male meiosis, the *CMT1A-REP* repeats are prone to undergo misalignment and unequal crossing-over, the outcome being recombinant chromosomes that either carry duplications of the 1.5 Mb region, or its complementary deletion. Importantly, the gene for peripheral myelin protein 22 (PMP22) is located within the 1.5 Mb region, and the duplication and deletion products result in distinct hereditary neurological syndromes,

namely, Charcot-Marie-Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP), respectively.

3.2.5.6 Benefits derived from mobile elements – The parasitic nature of mobile elements is not without beneficial contributions to the host: the random mobilization of transposons has often resulted in the provision of enhancer sequences for endogenous genes (Britten, 1997; Deininger and Batzer, 1999) and can affect gene expression both qualitatively and quantitatively (Girard and Freeling, 1999), and are key players in the genomic shuffling of exons (Ejima and Yang, 2003). For instance, the expression of the apolipoprotein-A gene is regulated by an enhancer that is located within a nearby LINE-1 element (Yang *et al.*, 1998). Similarly, the antisense promoter of L1 elements is utilized by many human genes (Nigumann *et al.*, 2002). Finally, some Alu elements may provide nearby transcription units with functional retinoic acid response elements (Vasant and Reynolds, 1995).

Genes found in transposons and retroelements may have been co-opted for other purposes in host organisms. For instance, the gene encoding the protein responsible for maintaining the length and structure of chromosome ends, telomerase, is similar to the reverse transcriptase enzyme of non-long-terminal-repeat (non-LTR) retrotransposons (Nakamura *et al.*, 1997). In Drosophila, there are two telomeric transposable elements, *HeT-A* and *TART*, which possess the remarkable activity of transposing to chromosome ends only (Pardue *et al.*, 1996). Further, each of these elements has a distinct pattern of transcription; a feature that is conserved between different *Drosophila* species despite sequence divergence of the elements (Danilevskaya *et al.*, 1999). Whether telomerase evolved from a non-LTR retrotransposon reverse-transcriptase, or vice versa, is a matter of debate (Eickbush, 1997; Nakamura and Cech, 1998).

Vertebrate immunity may also have derived a boost from the parasitic nature of transposons. The mechanism of transposition involves a series of DNA hydrolysis and transesterification reactions, and is similar to V(D)J recombination (Oettinger et al.,

1990). Interestingly, the recombinase-activating gene 1 (RAG1), one of the enzymes responsible for immunoglobulin gene rearrangement, may have one or more DDE motifs (Plasterk *et al.*, 1999; Haren *et al.*, 1999; Fugmann et al., 2000). Unfortunately, oncogenic translocations can potentially result from RAG-mediated DNA transposition (Hiorn *et al.*, 1998).

Finally, LINE-1 retrotransposons can insert at double-stranded DNA breaks (DSB), especially in cells made deficient in non-homologous end joining through mutations or deletions (Morrish *et al.*, 2002). This suggests that LINEs either can participate in DSB repair, or take advantage of the available insertion site (Eikbush, 2002)

3.2.5.7 Is the movement of transposable elements a regulated process ?: The influence of DNA methylation – What regulates the expression and transposition of mobile elements? Not all transposons mobilize randomly. This is best exemplified by the classic finding that transposition of the *P* element of *Drosophila* is tissue-specific and regulated at the level of RNA splicing (Laski *et al.*, 1986). Little is known about host factors influencing the rate of transposition, but screening efforts in *C. elegans* have identified 27 genes whose products lead to transposon silencing in these animals (Vastenhouw *et al.*, 2003).

A reciprocal relationship between the methylation status of endogenous genes and its resident transposable elements has been reported in a chordate genome: when genes are methylated, the transposable elements are not in *Ciona intestinalis* (Simmen *et al.*, 1999). Surprisingly, mice deficient for DNA methyltransferase 1 show activation of transposable elements in all tissues, and this finding is highlighted by the complete relaxation of the normally transcriptionally silenced mouse intracisternal A particle (IAP) retrovirus, the most aggressive parasitic sequence of the murine genome which is normally constrained by DNA methylation (Walsh *et al.*, 1998).

A link between CpG island methylation, imprinting and transposable elements may exist since long-interspersed-elements (LINEs) have been proposed to participate in X-chromosome inactivation (Bailey *et al.*, 2000). More indirect evidence for this link is provided by the regulation of LINE retrotransposition by members of the SRY gene family in humans (Tchenio et al., 2000). Another suggestive observation in this line of reasoning is the coincidental presence of a transcriptional silencer of the Wilms tumour gene WT1 located within an Alu repeat (Alu elements are short-interspersed-elements, or SINEs) (Hewitt *et al.*, 1995). Clearly, more work needs to be done to provide clear evidence for the possible interplay between mobile elements and the DNA methylation status of an organism. For instance, it is not clear whether patterned DNA methylation during invertebrate development serves the purpose of regulating development through influencing gene activation/repression, or is utilized to protect the genome against the inadvertent consequences of random transposition events (Regev *et al.*, 1998).

3.2.5.8 The expression of transposable elements – Due to their usual unregulated nature, the expression of most families of mobile elements is usually ubiquitous. It is therefore conceptually intriguing to consider that transposable elements can sometimes display a restricted spatiotemporal expression pattern. For instance, in mouse testes, the express of LINE-1 elements is limited to a subset of cell types during development (Branciforte and Martin, 1994). Another family of retrotransposons is also developmentally regulated during *Xenopus* development (Shim et al, 2000). Specifically, these elements, called *Xretpos*, are restricted to ventro-posterior specific regions and are induced by UV-irradiation and BMP-4 overexpression in a cycloheximide-sensitive manner.

The regulated expression of class II DNA transposons has been less well characterized in vertebrates due to difficulties in characterizing endogenously active copies (Ivics *et al.*, 1997). As pointed out earlier, *P* element expression in *Drosophila* is tissue-specific, indicating that expression of diverse mobile elements can be regulated, whether quantitatively or spatio-temporally (Laski *et al.*, 1986).

Surprisingly, not only is the expression of certain mobile elements regulated, transcriptional activation itself is apparently sufficient to alter the expression of nearby endogenous genes (Kashkush *et al.*, 2003).

Riboprobes corresponding to the full-length sequence of the AS progranulin-1/-2 transcript, or to a region located within the fourth exon only (AS tzf - see methods), detect abundant transcription in a fairly ubiquitous pattern during the development of the zebrafish (data presented later, **Figure 41**). It would thus appear that endogenous *tzf* transposons are not developmentally regulated. However, it is possible that the *tzf* transposase sequence is sufficiently related to that of other closely related elements, allowing for riboprobe cross-hybridization. Independent confirmation that *tzf* expression is unrestricted would require the use of sequence-specific inverted-repeat sequences as probe.

3.2.5.9 A role for AS progranulin-1/-2? – The presence of a transcript antisense to *progranulin-1* and *progranulin-2* raises several interesting questions. Clearly, although the last exon of the AS progranulin-1/2 gene harbours a transposase sequence, neither the pre-mRNA nor the mature transcript gives rise to a mobile element. This is due to the following reason: the integration of the tzf transposon is in the reverse orientation and has suffered substantive genetic drift, especially in one of the flanking inverted repeats, precluding its recognition and mobilization by yet another endogenously active tzf transposase. This is not to say that the sequences encoded by the fourth exon cannot base-pair with other cognate tzf sequences (**Figure17**).

The genomic *tzf* transposon in sense orientation to and upstream of the progranulin-2 gene may thus be regarded as a molecular fossil, perhaps reflecting an ancient event, such as in the creation of either the progranulin-1 or progranulin-2 gene from an ancestral progranulin by unequal crossing-over during non-allelic homologous recombination (Hey, 1998). Likewise, a similar mechanism may have been utilized in a later event in the creation of the tandem duplicate, such as in the case of the CMT1-

REP repeats (see discussion above). This line of reasoning does not take into account the antisense transcription of this element as part of the AS progranulin-1/-2 gene.

It is unlikely that an active *tzf* transposon participated in the creation of a small progranulin gene through the mechanism of transposition. Although the tzf transposase gene is intronless, like that of many class II transposable elements including the *Alu* elements in humans (Boeke, 1997), it mobilizes directly via a cut-and-paste mechanism. Even if this *tzf* mobilized with the indirect help of a retrotransposon, which generates an RNA intermediate that is reverse-transcribed prior to integration, this would have presumably resulted in the formation of a processed progranulin pseudogene (Maestre *et al.*, 1995; Esnault *et al.*, 2000). In fact, the presence of a gene encoding a reverse-transcriptase was detected upstream of and in the same orientation to the progranulin-1 gene (data not shown). Interestingly, it was recently shown that Alu transposons (class II elements like tzf) are sometimes hitchhiking with mobilizing LINE-1 retrotransposons (Schmid, 2003; Dewannieux *et al.*, 2003). Exon shuffling of granulin sequences through the mechanism of retrotransposon-mediated 3' transduction is also conceivable through the action of this upstream element (Moran *et al.*, 1999; Boeke and Pickeral, 1999).

Notwithstanding whether or not this DNA element has a causative relationship with the original emergence of a progranulin-1-type gene, it is remarkable to consider that there is no indication of existence of the tzf element outside the teleost radiation (Lam *et al.*, 1996). Furthermore, like progranulin-1 and -2, transposons of the *tzf* class appear to be restricted to the teleost order cyprinidae, which includes goldfish, carp, and zebrafish. If this is the case, then it is interesting to consider that the restriction of this element to the teleost radiation has an influence on the apparently similarly restricted distribution of the progranulin-1 gene in fish. This genomic trait may represent a crucial factor underlying key morphological differences between the anatomy of (some teleost) fish and mammals, which should find support in the expression patterns of these genes during development. This does not invalidate the study of these smaller progranulin genes in their own right, as they may offer insight

into the physiological function of the mammalian equivalent in a context-dependent fashion through the performance of conserved biochemical functions. It should also be remembered that this feature is distinct from the often encountered gene function partitioning onto two structurally similar duplicated paralogues, since we know that mammalian progranulin is represented by two zebrafish co-orthologues.

It would be equally interesting to determine whether AS progranulin-1/2 transcription is regulated by cytosine methylation and acts as an agent of epigenetic silencing favouring evolution by gene duplication (Rodin and Riggs, 2003). In fact, atypical non-coding RNAs are often imprinted (Sleutels *et al.*, 2000). Whether the zebrafish utilize the methylation and deacetylation in the process of imprinting remains unknown (McGowan and Martin, 1997).

Of course, it may be that this co-occurrence is purely coincidental, and persists as an accident frozen in time. In this case, neither progranulin-1 nor progranulin-2 would require antisense regulation. AS progranulin-1/-2 could function in juxtaposing the progranulins-1 and -2 pre-mRNAs for the creation of a *trans*-spliced product. This would require the combined spatio-temporal presence of the primary transcripts originating from the three promoters in sense and antisense orientation at the progranulin-1 gene locus.

Did the ancestral progranulin gene also possess an antisense transcript that was carried over with the creation of the smaller genes? It is also possible to consider that antisense transcription is unique to *progranulin-1* and *progranulin-2*. If either the progranulin-a or -b genes were demonstrated to possess convergent transcription units, then AS progranulin-1/-2 could be explained as an emergent transcription unit simultaneous to progranulin-1 gene formation. Evidence for antisense progranulin transcription is weak, since preliminary searches of NCBI databanks have detected only a few reportedly unidirectionally cloned ESTs corresponding to murine and chick progranulin (data not shown), as well as to zebrafish progranulin-a (accession numbers CD585963, CD586001, CD588938), but in the reverse complement orientation.

Although clear evidence of antisense transcription to mammalian progranulin is lacking, the next section discusses the current views regarding eukaryotic antisense transcription. This is intended for the interested reader. Some of the examples provided will be applicable to AS progranulin-1/-2 transcription.

3.2.6 Antisense transcription: an intersecting antiparallel world – The last decade has seen the sequencing of various eukaryotic genomes, providing insight into the genetic constituents of life (Rubin *et al.* 2000; Lander *et al.*, 2001; Venter *et al.*, 2001; Mouse Genome Sequencing Consortium, 2002; Dehal *et al.*, 2002). An interesting and unanticipated contribution from this international enterprise has been the eventual discovery that although individual transcriptional units are preferentially residing in regions of euchromatin, the complementary overlap of genes at the same locus is not that infrequent. In fact, since the early description of a few sporadic examples of transcriptional units sharing partial complementarity, including those of the basic FGF gene and the insulin-like growth factor 2 loci nearly 15 years ago (discussed later), recent reports have indicated that the phenomenon of antisense strand transcription is common in higher vertebrates including mice and humans. For instance, of the estimated 40 000 human gene number, greater than 1600 transcribed sense-antisense gene pairs have been identified, representing at least 8% of the human transcriptome (Yelin *et al.*, 2003). Similar numbers have been obtained for the murine genome (Kiyosawa *et al.*, 2003). These numbers were presaged by the earlier unsuspected discovery of transcription originating from the complementary DNA strand of several human genes (Fahey *et al.*, 2002; Lehner *et al.*, 2002) including such genes as those associated with endocrine function, namely pro-melanin-concentrating hormone (Miller *et al.*, 1998), preproenkephalin (Weisinger *et al.*, 1998), a novel human gonadotropin releasing hormone (GnRH) receptor (Millar *et al.*, 1999) and urocortin (Shi *et al.*, 2000). Future efforts will undoubtedly aim at determining the extent of sense-antisense gene pair conservation among vertebrate genomes, their functional significance, along with their mechanism of action.

Vertebrate antisense transcription is therefore more prevalent than once presumed, and understanding its overall function may benefit from knowledge gained through well-documented use of antisense transcription by prokaryotes and viruses (Gerhart *et al.*, 2002). For example, antisense RNA-mediated regulation of gene expression in prokaryotes usually leads to reduced protein synthesis of such targets as bacteriophage targets sense transcript by interference with the assembly of the

translational machinery at the 5'end or by regulating the transposition rates of insertion elements (Inouye, 1988; Simons, 1988). Alternatively, double-stranded RNA (dsRNA) generated by virtually all types of viruses enables the infecting agent to cope with host defense mechanisms, and vice-versa (Ratcliff *et al.*, 1997; Kasschau and Carrington, 1998; Kumar and Carmichael, 1998; Diebold *et al.*, 2003). Although of paramount importance in plants, antisense-mediated transcription occurring in the vegetal kingdom will not be discussed at length here.

In the majority of the cases, the consequences of antisense transcription in vertebrates is expected to be similar to those observed in lower organisms, namely reduced mRNA stability via the mechanism of RNA interference which results in post-transcriptional gene silencing (Montgomery and Fire, 1998; Hunter, 2000; Ketting and Plasterk, 2000; Hammond *et al.*, 2001; Zamore *et al.*, 2002). Although it was initially discovered as a method of inactivating genes in the germ line of *C. elegans* (Guo and Kemphues, 1995), the actual mechanism of RNAi action is allegedly to suppress the mobilization and propagation of transposons (Ketting *et al.*, 1999; Tabara *et al.*, 1999).

The process of RNA interference (RNAi) initially involves the recognition of double-stranded RNA (dsRNA) precursors by the multidomain ribonuclease III protein called Dicer (Hutvagner *et al.*, 2001; Nicholson and Nicholson, 2002). The dsRNA is then cleaved into segments of approximately 21-26 nucleotides long referred to as small interfering RNAs (siRNA). These are the active agents of the RNAi effect, namely the effective degradation of complementary RNAs (Hamilton and Baulcombe, 1999; Hammond *et al.*, 2000; Zamore *et al.*, 2000; Elbashir *et al.*, 2001; Sharp, 2001; Carthew, 2001). This process is distinct from the processing of dsRNA from specialized enzymes such as dsRNA-specific nucleases (Wu, 1998) and dsRNA adenosine deaminase (Bass, 1997).

Because of its evolutionary conservation, post-transcriptional gene silencing via RNA interference is now commonly used in loss of function studies aimed at elucidating the role of candidate genes in the *C. elegans* and *Drosophila melanogaster*

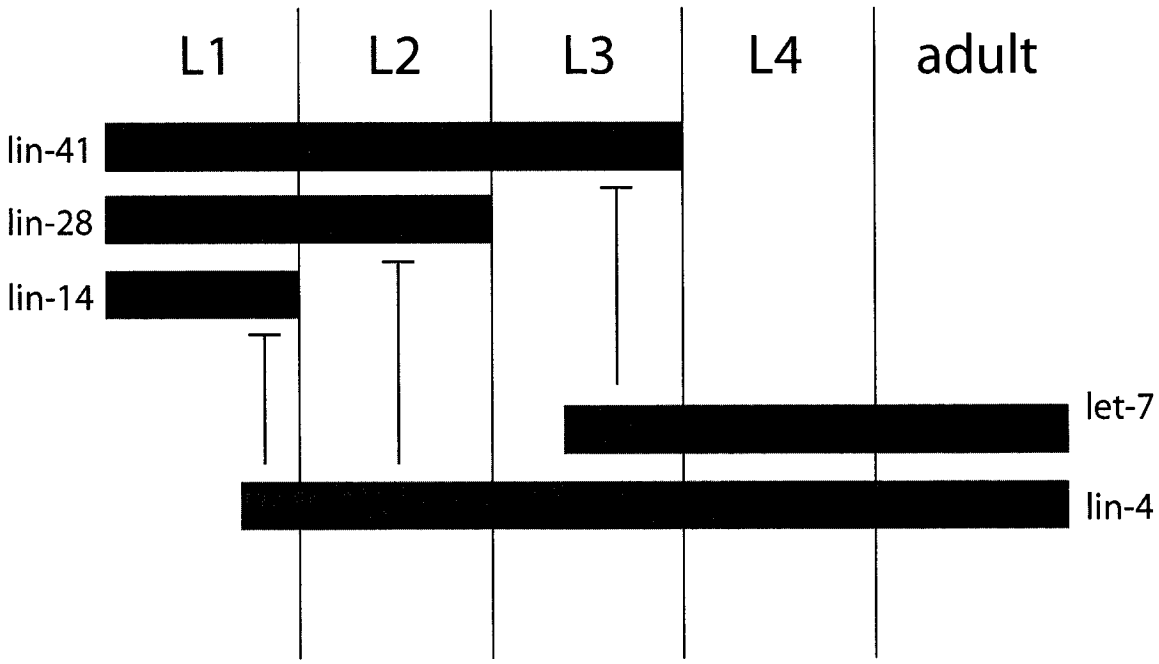


Figure 28. **Heterochronic gene function during *C. elegans* development.** The four larval stages (L1-L4) of *C. elegans* development, along with the adult stage, are listed at the top. Shown in green are the approximate times of protein expression for LIN-41, LIN-28, and LIN-14. In red is the timing of RNA expression for the small heterochronic antisense transcripts let-7 and lin-4. lin-4 encodes two small RNA transcripts that inhibit the translation of the lin-14 (see text) and lin-28 transcripts through heteroduplex formation, and the small let-7 antisense transcript operates similarly through interfering with the sense lin-41 transcript (Slack and Ruvkun, 1997; Ambros, 2000). (Adapted from Thummel, 2001)

invertebrate models of development (Tabara *et al.*, 1998; Fire *et al.*, 1998; Bass, 2000; Kennerdell and Carthew, 2000; Tavernarakis *et al.*, 2000), in murine embryos (Wianny and Zernicka-Goetz, 2000) and in cell culture (Clemens *et al.*, 2000; Sharp, 2001; Kiger *et al.*, 2003). For reasons that are at present not understood, this technique is considered unreliable for use in the zebrafish due to documented non-specific effects (Oates *et al.*, 2000), notwithstanding some reported successful attempts (Wargelius *et al.*, 1999; Li *et al.*, 2000).

3.2.6.1 Imperfect, small antisense transcription – Antisense-mediated transcription has been documented for a few eukaryotic genes, and is usually classified, depending on the length and complementarity of the RNA duplex formed, into two categories. Constituting the first group are short (less than 100 bp) RNAs that are usually imperfect complements to a sense transcript and that originate from a different genomic locus. Such interactions include pre-mRNA splicing substrates and the small nuclear ribonucleoprotein (snRNP) particle RNAs, as well as the involvement of small nucleolar RNA in RNA maturation.

Perhaps the most intriguing and better studied example of small imperfect complementary (antisense) transcripts comes from the lin-4 and lin-7 heterochronic genes involved in regulating the normal larval transitions of the developing *C. elegans* (**Figure 28**). Lin-14 mRNAs are normally present throughout all stages of development, but the translated product is not. The lin-14 transcript possesses two small antisense transcripts sharing partial complementarity to its 3'UTR, each with a respective length of 22 (processed) and 61 (precursor RNA) nucleotides, and that originate from the lin-4 gene located at a distinct locus. Failure to express "antisense" lin-4 transcripts leads to a retarded phenotype characterized by the inability to undergo a transition between the L1 to L2 stages (mutant animals reiterate the L1 larval stage rather than progress to later stages of development) (Chalfie *et al.*, 1981). While the molecular mechanisms remain to be elucidated, the role of the antisense transcripts in this context is to exert translational repression of the target lin-14 mRNA (Wightman *et al.*, 1993). Interestingly, the imperfect complementarity of the lin-4 transcripts

appears to be required for its activity; bulges formed by the unpaired nucleotides are crucial for translational repression of its target (Lee *et al.*, 1993; Wightman *et al.*, 1993).

Similarly, the let-7 transcript, which is 21 nucleotides in length, regulates developmental timing through translational repression of another distantly transcribed target, the lin-41 mRNA (Reinhart *et al.*, 2000). At least some mechanisms of developmental timing appear to be conserved in vertebrates as suggested by the functional characterization of the lin-41/let-7 gene pair, but not of the lin-14/lin-4 gene pair, in zebrafish and humans (Pasquinelli *et al.*, 2000). In fact, the let-7 RNA is almost 100% conserved and expressed as a small 21-nucleotide RNA in all bilaterally symmetrical animals, but not in jellyfish (cnidarians) or sponges (poriferans). Finally, it is interesting to note that the size of the processed let-4 and let-7 microRNAs is identical to that of siRNAs (Moss, 2000).

Another recent example falling in this category is the *bantam* microRNA which simultaneously stimulates proliferation and prevents apoptosis during *Drosophila* development (Brennecke *et al.*, 2003). Interestingly, there are more than 200 known microRNA genes in vertebrates (Huttenhofer et al., 2001; Lagos-Quintana *et al.*, 2003; Lim *et al.*, 2003), for which the target RNA and respective role in development remain to be determined. Once this is established, it will also be worthwhile to investigate the hormonal cues that drive the correct spatiotemporal expression of so-called heterochronic (e.g. lin-4 and let-7) and other microRNA genes in order to achieve regulated co-occurring transcription with the respective "sense " targets.

3.2.6.2 Perfect, long antisense transcription – The second category of antisense transcription involves longer mRNAs originating from the complementary DNA strand of the sense gene. In cases where transcription at a given locus is bi-directional, the complementarity is perfect and can be limited to a portion of the exonic sequence of the sense transcript or extend through its entire length. The transcript antisense to the zebrafish progranulins 1 and 2 genes belongs to this category.

In mammals, convergent transcription was first described for two processed mRNAs, Surf-2 and Surf-4, which display a 133 bp overlap at their 3' untranslated termini (Williams and Fried, 1986). This report was followed by a small wave of papers demonstrating the presence of other convergent or antisense transcriptional units in vertebrates, some of which have been shown to play regulatory roles. In order to provide evidence for the biological significance of antisense transcription, selected examples of this phenomenon arising in the animal kingdom are discussed in greater detail below.

3.2.6.3 Examples of antisense transcription

3.2.6.3.1 Myosin Heavy Chain – The myosin heavy chain (MyHC) is the molecular motor driving muscle contraction. The heart expresses two isoforms, α-MyHC and β-MyHC, which arise from two closely linked paralogous genes. Regulated expression of these isoforms during development and in the adult has been related to the different requirements for their distinct intrinsic ATPase activity (Lompre *et al.*, 1984; Krenz *et al.*, 2003).

Naturally occurring antisense transcripts to each of the α- and β-MyHC isoforms were identified in the heart of newborn rats (Luther *et al.*, 1998). As revealed through Northern analysis, each of these complementary transcripts is of similar size to its corresponding sense mRNA, although these differ in their respective extent of overlap (Luther *et al.*, 1998).

The common occurrence of antisense transcription for these tandemly organized paralogues may have important evolutionary implications. Whether antisense transcription occurs in the ancestral pre-duplicated myosin heavy chain gene has not been investigated, but is arguably expected. This may impact gene dosage through post-transcriptional regulation of either MyHCs or the developmental regulation of expression of the distinct isoforms (Lompré *et al.*, 1984), both scenarios possibly influencing duplicate gene retention.

Significantly, antisense β-MyHC transcription leads to attenuated sense primary transcript processing into functional mRNA (Haddad *et al.*, 2003). In addition, sense α-MyHC and antisense β-MyHC are both negatively co-regulated in response to hypothyroidism and in diabetes (Haddad *et al.*, 2003). It is thus possible that the absence of antisense β-MyHC transcription under these pathophysiological conditions is responsible for the diagnostic upregulation in β-MyHC mRNA levels and concomitant decrease in expression of the α-isoform (Haddad *et al.*, 1997).

In this regard, it is noteworthy to consider the regulated antithetical shift in expression of these MyHC genes during development (Lompré et al., 1984; Chizzonite and Zak, 1984). During the initial three weeks post-partum, α-MyHC levels gradually rise to complement the dramatic loss in β-isoform levels from >90% to trace levels. The coinciding surge in thyroid hormone levels in this period of development hints at the possibility that induced expression of the antisense β-MyHC transcript facilitates or directs the coordinated β-α antithetical gene switching (Haddad *et al.*, 2003). As pointed out previously, this mechanism may be relevant in pathological situations such as in cardiac myopathy and failing hearts where the normally high expression levels of the α-MyHC isoform are severely downregulated (Nakao *et al.*, 1997; Reiser *et al.*, 2001).

3.2.6.3.2 basic Fibroblast Growth Factor – The basic fibroblast growth factor (bFGF) gene locus also gives rise to transcripts derived from each complementary DNA strand. While originally cloned from *Xenopus laevis* (Volk *et al.*, 1989), antisense FGF, now referred to as the *gFg* gene, is conserved in mammals (Borja *et al.*, 1993; Knee *et al.*, 1994; Murphy and Knee, 1994).

Early expression studies in the developing chick, has revealed an inversely proportional relationship for this sense-antisense pair, suggesting a possible regulatory role for the antisense transcript (Borja *et al.*, 1993). In the adult, however, the relative

abundance of each transcript varies in a tissue-dependent context. For instance, the human sense transcript is generally more abundant than the antisense transcript in the majority of tissues examined, but the latter transcript predominates in testes, skeletal muscle, heart and liver (Knee et al., 1994). In the developing brain, sense–antisense RNAs display a clear reciprocal relationship, where bFGF is almost exclusively observed (Knee *et al.*, 1994; Li *et al.*, 1996; Knee *et al.*, 1997). Albeit low, the residual levels of antisense transcripts in the CNS regulate bFGF mRNA translation (Li and Murphy, 2000).

The interesting co-occurrence in some tissues for the bFGF and its antisense transcripts suggests the potential for a non-antisense function for the gFg gene. This is supported by the demonstration that the 1.5kb polyadenylated antisense RNA has an intact open reading frame encoding a protein with homology to the prokaryotic MuT family of nucleotide hydrolases (Volk *et al.*, 1989, Li *et al.*, 1996). Interestingly, this protein is broadly expressed in tissues outside the central nervous system (CNS) in the postnatal rat, in agreement with the low levels of *gFg* RNA transcription found in the brain (Knee *et al.*, 1997).

In summary, the FGF/gFg pair highlights two key points. First, not all transcripts derived from complementary DNA strands are invariably expressed in a reciprocal fashion. This notion is also supported by the developmentally regulated but non-complementary expression pattern of the chromogranin *Nesp* gene and its associated antisense *Nespas* gene, notwithstanding their oppositely imprinted pattern of expression (Ball *et al.*, 2001). Second, although expected to be rare, some antisense transcripts have preserved the capacity to encode a translated protein product.

3.2.6.3.3 Thyroid Hormone Receptor c-Erbα - Some reports hint at the possibility that antisense transcription may modulate splicing of the sense primary transcript. A candidate for such a function is the *Rev-ErbAα* transcript. In rat, a thyroid hormone receptor (c-erbAα) gene gives rise to two transcripts, *c-erbAα1* and *c-erbAα2*, which differ through the alternate incorporation of 3'-terminal exons. Like *c-*

erbAα1, *c-erbAα2* encodes a thyroid hormone receptor, but with an altered ligand binding domain. An mRNA originally thought to encode an additional member of the steroid/thyroid hormone receptor, but found to be transcribed from the opposite strand of c-ErbAα, has been shown to overlap by 269 bp with c-erbAα2, but not with *c-erbAα1* (Lazar *et al.*, 1989). This gene, referred to as *Rev-ErbAα*, is also present in humans (Lazar *et al.*, 1990).

Because it base pairs with *c-erbAα2*, the *rev-erbAα* transcript is predicted to influence the differential levels of the *c-erbAα1* and *c-erbAα2* mRNAs through negative regulation (Munroe and Lazar, 1991). Indeed, *rev-erbAα* inhibits splicing of *c-erbAα2*, demonstrating a role for an antisense transcript in regulating alternative processing of transcripts (Munroe and Lazar, 1991). This is further supported by the strong correlation observed between the ratio of *c-erbAα1* to *c-erbAα2* mRNA and expression levels of the antisense *rev-erbAα* transcript in cell lines representing various progressive stages of differentiating B-lymphocytes (Hastings *et al.*, 1997). Further, recent findings have provided a direct role for the antisense transcript in regulating thyroid hormone receptor expression postranscriptionally (Hastings *et al.*, 2000).

3.2.6.3.4 N-myc – Another example providing evidence for the involvement of an antisense transcript in the regulation of sense pre-mRNA processing is ascribed to the N-myc antisense gene. In humans, bi-directional overlapping transcription units characterize the N-myc gene. Several antisense transcripts are initiated from various locations within the first intron of N-myc pre-mRNA, some polyadenylated, others not (Krystal *et al.*, 1990). The 5'ends of the sense N-myc mRNA and of the polyadenylated antisense transcripts are complementary, and these form heteroduplexes in the cytoplasm. dsRNA occurs only with some forms of the multiple N-myc mRNA, indicating that selectivity is dictated by the transcriptional start site of each RNA (Krystal *et al.*, 1990). Interestingly, the RNA duplexes are comprised of sequences for both exon1 and intron A of N-myc, suggesting that dsRNA formation modulates RNA processing through inhibiting intron A splicing (Krystal *et al.*, 1990).

3.2.6.3.5 HoxA11 – Transcripts antisense to the 5'end of the gene for mammalian Hoxa11 have been uncovered (Hsieh-Li *et al*; 1995; Potter and Branford, 1998) and their expression patterns have been investigated in three contexts with interesting results. These are briefly reviewed below.

First, sense and antisense *HoxA11* transcripts display a striking complementary pattern of expression in the developing limb. Specifically, in the early (E9.5) limb bud, the sense transcript is abundant throughout, and the antisense transcript is absent. One day later, at E10.5, antisense RNAs are very abundant in the distal bud, immediately flanking and somewhat overlapping the expression domain of the now more proximally restricted *HoxA11* sense RNA (Hsieh-Li et al., 1995).

Second, complementarity of expression is not observed in the kidney. Rather, opposite-strand transcripts, as opposed to truly antisense, are present. These are alternately processed to exclude any sequence overlap to sense *HoxA11*, arguing for an additional non-antisense function for this transcript (Potter and Branford, 1998).

Third, sense and antisense HoxA11 transcripts have a similar cellular tissue distribution in the stroma of endometrium (Chau *et al.*, 2002). During the menstrual cycle, the respective level of these transcripts varies inversely, with peak antisense levels occurring in the midproliferative phase when the endometrium is not receptive to blastocyst implantation, while those of *HoxA11* are at their lowest (Chau *et al.*, 2002). In contrast, there is evidence that the upregulated expression of sense *HoxA11* normally seen during the luteal (secretory) phase is, at least in part, resulting from the downregulating effect of progesterone on *HoxA11* antisense transcription. Thus, antisense *HoxA11* appears to exert its negative influence on the presence of sense transcripts via transcriptional interference, rather than through heteroduplex formation with the sense transcript with consequent transcript degradation.

3.2.6.4 Antisense transcription and mammalian imprinting: The IGF system – The phenomenon of genomic imprinting is characterized by the inherited pattern of transcriptional activation/repression of a gene on either the maternal or paternal chromosome (Tilghman, 1999). Key elements governing allele-specific imprinting and transcriptional repression are methylation at CpG islands (Cedar *et al.*, 1988; Li *et al.*, 1993) and complementary histone deacetylation (Tazi and Bird, 1990) (see section 3.2.6.4.1 for overview). Interestingly, genomic imprinting is often associated with antisense transcription. The functional contribution of antisense transcription in the process of imprinting is obscured by the apparent absence of imprinting outside the eutherian mammalian lineage (Sleutels *et al.*, 2000). Clear evidence for imprinting in the zebrafish is currently lacking, but cannot be conclusively dismissed (McGowan and Martin, 1997). Notwithstanding this discrepancy, the following briefly discusses two examples of imprinted loci in which antisense non-coding RNA transcription has been documented in conjunction with coding sense RNA, namely those of the IGF type 2 receptor and the IGF2 ligand (**Figure 29**). In mammals, the insulin-like growth factor 2 receptor gene (IGF2R) is not expressed from the paternal chromosome (Barlow *et al.*, 1991). Structurally, the IGF2R product is unrelated to the IGF type I and insulin receptors. Rather, it is identical to the cation-independent mannose-6-phosphate receptor (Morgan *et al.*, 1987; Kiess *et al.*, 1988) in which incarnation it subserves the function of directing lysosomal enzymes from mainstream traffic to lysosome compartment within the secretory pathway (Kornfeld and Mellman, 1988). In this case, gene sharing is achieved through interactions with mannose-6-phosphate or IGF2 on distinct binding sites located on the mannose-6-phosphate receptor/IGF2R (Kiess *et al.*, 1988). The structural basis of the IGF2R suggests that it does not participate in signal transduction of the IGF2 ligand. This prediction was compounded by the consequence of IGF2R gene disruption in the mouse (Barlow et al., 1991), which does not phenocopy that of the IGF2 knockout (DeChiara *et al.*, 1990; DeChiara *et al.*, 1991). For this reason, the IGF2R has been considered to act as a sink, effectively sequestering freely available ligand (Kiess *et al.*, 1987; Haig and Graham, 1991).

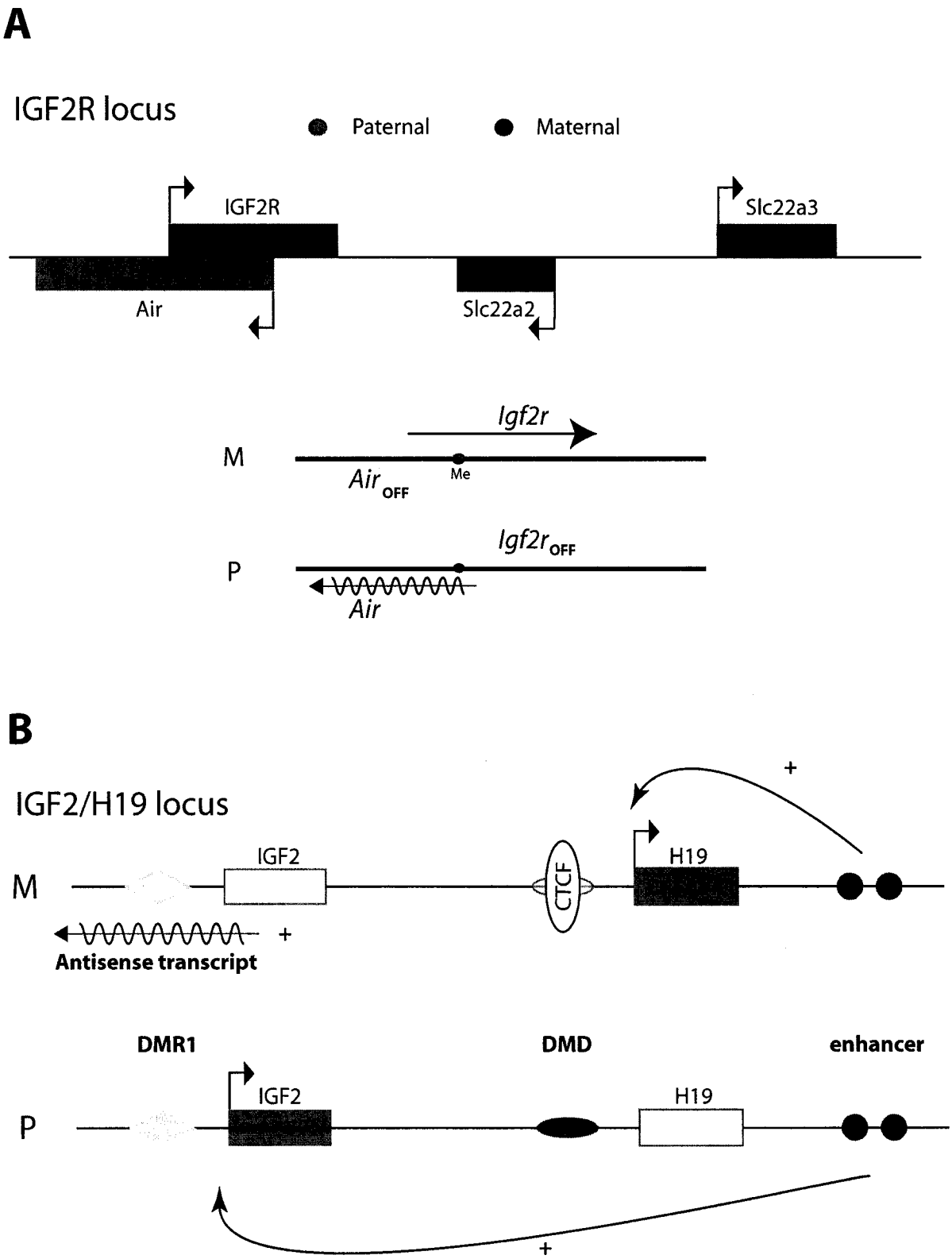

A
IGF2R locus
Paternal
Maternal
IGF2R
Slc22a3
Air
Slc22a2
Igf2r
M
Me
Air OFF
Igf2r OFF
P
Air
B
IGF2/H19 locus
+
IGF2
CTCF
H19
M
+
Antisense transcript
DMR1
DMD
enhancer
IGF2
H19
P
+

Figure 29

Figure 29. **Imprinting and antisense transcription at the IGF2R and IGF2/H19 genomic loci.** Panel A: Top – The paternally imprinted *igf2r*, and *Slc22a2* and *Slc22a3* genes are regulated by the reciprocally imprinted expression of the antisense *Air* gene. Bottom - Regulated allelic expression is under the influence of methylated CpG islands in the vicinity of the respective gene promoters. Panel B: The *igf2* and *H19* genes are reciprocally imprinted, and this mechanism is partly governed by an enhancer boundary mediated by the binding of the CTCF transcription factor on a differentially methylated domain (DMD) in red. Differential methylation of another region (DMR1) in yellow, as well as a CpG region located within intron 2 of the igf2 gene (DMR2) (not shown) also regulate antisense transcription to *igf2*. (See text for details) (Adapted from Rougeulle and Heard, 2002 and Arney, 2003, respectively)

The differential expressivity of *igf2r* from the maternal chromosome is sufficient to explain the discrepancy observed in mice heterozygotes for a disrupted *igf2r* allele. Specifically, under the situation where the paternal copy is removed, embryonic development proceeds normally, in contrast to the disrupted maternal allele which leads to early lethality (Barlow *et al.*, 1991).

Differential methylation of a 2 kb CpG island referred to as region 2, located within the second intron of the IGF2R gene was thought to underlie its maternal expression (Stojer *et al.*, 1993; Wutz *et al.*, 1997) (**Figure 29, panel A**). However, when region 2 is unmethylated, as on the paternal chromosome, antisense (*Air*) transcription is observed (Wutz *et al.*, 1997). Absence of methylation of this region is required for antisense transcription since it is not observed from the maternal chromosome where reciprocal imprinting of the antisense transcription is observed. This indicates that methylation of CpG residues, located in the vicinity of either promoters for *igf2r* or *Air*, effectively abrogates transcriptional activity.

Surprisingly, the imprint signal in the CNS is not maintained for the sense IGF2R transcript, where expression occurs biallelically (Hu *et al.*, 1998; Hu *et al.*, 1999). Methylation of CpG residues within region 1 located in the vicinity of the IGF2R gene promoter of the paternal allele occurs only in the periphery, preventing its transcription from this allele (Hu *et al.*, 1998). The corollary does not apply to the antisense transcript, in that the maternally imprinted DMR region 2 CpG methylation is preserved in both the periphery and the CNS (Stoger *et al.*, 1993; Hu *et al.*, 1999). Thus, since Air transcription is necessary to imprint sense igf2r expression, its role in the brain remains enigmatic. Regardless, it may be possible that the *Air* transcript facilitates imprinting of IGF2R in a manner similar to the *Tsix* transcript in antisense orientation to *Xist* (X(inactive)-specific transcript), the latter directing mammalian X chromosome inactivation (Lee *et al.*, 2000; Sado *et al.*, 2001; Migeon *et al.*, 2002). Indeed, the *Air* transcript is not only required for igf2r imprinting, but also for the transcriptional silencing of the non-overlapping *Slc22a2 and Slc22a3* genes (Sleutels *et al.*, 2002; Rougeulle and Heard, 2002).

Another example of complex imprinting in relation to antisense transcription occurs at the insulin-like growth factor 2 (IGF2)/H19 locus itself (Moore *et al.*, 1997) (**Figure 29, panel B**). During murine fetal development, the IGF2 gene expression is silenced on the maternal allele. Located upstream of the IGF2 gene is a region containing a tandem repeat, from which originate several antisense and sense transcripts. Methylation of regions flanking the tandem repeat, on the paternal chromosome, as well as the production of both sense and antisense transcripts, are required for imprinting of igf2. Further, these two phenomena necessitate the presence of a functional H19 gene, since imprinting of the IGF2 gene is lost in H19 null mice (Moore et al., 1997).

Consistent with the maternally inherited imprint is the phenotypic distribution following disruption of the IGF2 gene disruption studies: hemizygous animals that carry an intact maternal allele are reduced in size but remain proportionate, whereas those that inherit a functional paternal locus display no defect (DeChiara *et al.*, 1990; DeChiara *et al.*, 1991).

This striking opposite pattern of imprinting for a ligand (IGF2) and its receptor (IGF2R) has been interpreted as a mechanism of natural selection in the context of the cost of intrauterine growth to the mother in eutherian mammals (Haig and Westoby, 1989; DeChiara *et al.*, 1991; Haig and Graham, 1991; Moore and Haig, 1991).

The field is now confronted with the daunting task of elucidating the series of events leading to an established imprinting signal, and its dissimulation. For instance, it will be interesting to determine whether DNA methylation or histone acetylation levels come first to establish the imprinting mark, and their conjoined integration of signal transduction events. Also, it is pertinent to assess the extent of regulated antisense transcription and its role in directing silencing (Matzke *et al.* 2001). In a likely scenario, efforts will undoubtedly be aimed at solving the relationship between

decreased antisense RNA levels and *de novo* methylation of intronic CpG islands of the Igf2R gene in the context of aging mice (Yang, 2000).

As alluded to earlier, a role for DNA methylation in either the initiation or maintenance of imprinting has been corroborated from studies in which the DNA methyltransferase 1 (Dnmt1) gene has been disrupted in mice. In this case, normal imprinting of igf2r, igf2 and H19 is lost (Li *et al.* 1993). Unfortunately, the level of antisense transcription was not assessed in these animals. Thus, if the purpose of regulated DNA methylation is to repress transcriptional activity, it can be envisioned that the default state is active antisense transcription in these animals. Furthermore, in the IGF2 and H19 alleles pair of reciprocally imprinted coding and non-coding RNA genes, the expressed alleles are associated with more abundantly acetylated histones than the imprinted alleles (Pedone *et al.*, 1999). The same applies to the sense and antisense transcripts of the IGF2R locus (Hu *et al.*, 2000).

3.2.6.4.1 The link between histone acetylation and CpG methylation in imprinting and in AS transcription – Core histone acetylation is known to modulate the expression of several genes (Jenuwein and Allis, 2001). Generally, acetylated histones are associated with increased gene expression, whereas decreases in histone acetylation favours the transcriptionally repressed chromatin state. The molecular mechanism underlying the negative influence exerted by hypoacetylated histones on gene transcription has been linked to positively charged lysine residues of histones, enabling tight interactions with the DNA. This presumably prevents the recruitment of positively acting cis-regulatory transcription factors and stable formation of the transcriptional apparatus on the promoters of the affected genes. Thus, addition of an acetyl group to histone protein constituents, usually histone4, neutralizes the positive charges, disrupting the electrostatic interactions between the histone core and DNA.

Hypoacetylated histone cores, along with a concomitant increase in histone H1, are known to associate with DNA carrying a high content of methylated CpG islands (Tazi and Bird, 1990), suggesting that DNA methylation and histone acetylation are

commonly involved in the process of imprinting. Concurrently, the expression of genes whose promoter regions are heavily methylated is usually decreased (as we have seen for either sense or antisense IGF2).

An intimate interrelationship at the molecular level between histone acetylation and DNA methylation has been provided by the demonstration of complex formation between histone deacetylase and the methyl-CpG-binding protein MeCP2 involving the corepressor mSin3A protein (Jones *et al.*, 1998; Nan *et al.*, 1998). Further, Dnmt1, the enzyme responsible for CpG methylation, interacts with histone deacetylase 1, suggesting that DNA and histone modifications are co-regulated (Fuks *et al.*, 2000).

3.2.6.5 Antisense transcription in the normal control of proliferation and differentiation – During development, DNA methylation is critical for determining cellular differentiation (Riggs and Pfeifer, 1992). Further, although embryonic stem cells can survive if DNA methylation is decreased, these die subsequent to the onset of differentiation. This is in keeping with the lethal phenotype observed during gestation of murine embryos with a disrupted Dnmt1 gene (Li *et al.*, 1993). Whether linked to imprinting and DNA methylation or not, antisense transcription may also participate indirectly in the process of cellular differentiation. As noted earlier, there is a correlation between the levels of Rev-erbAα transcription and B-lymphocyte maturation (Hastings *et al.*, 1997). More suggestive evidence, however, comes from studies of antisense-mediated gene regulation of p53 activity during embryogenesis. For instance, tissue differentiation during embryonic development of mouse and chick is accompanied by a reduction of stable p53 mRNA, while the transcription rate remains unchanged (Louis *et al.*, 1988). A similar phenomenon is observed when mouse F9 embryonal carcinoma stem cells differentiate after being subjected to retinoic acid treatment and dibutyryl cyclic AMP (Dony *et al.*, 1985). Again, an induced RNA transcript in the reverse orientation appears to control p53 at the post-transcriptional level (Khochbin *et al.*, 1988). Two polyadenylated transcripts, one 1.3kb in length that can anneal to the 5'end part of the first intron of the p53 gene as a result of a B1 repetitive element (Khochbin and Lawrence, 1989), the other larger and

apparently complementary to the entire p53 gene (Khochbin *et al.*, 1992), appear to accumulate concomitantly with a decrease in sense-strand mRNA when transformed cells are stimulated to differentiate *in vitro*.

It will be interesting to evaluate the contribution of c-myc antisense transcription in a similar context. Transcription of the sense strand of the c-myc gene is initiated from two promoters, P1 and P2. In normal cells, transcripts generated from P2 predominate over those arising from P1. Transcripts originating from the complementary strand of the c-myc gene have also been reported (Dean *et al.*, 1983; Bentley *et al.*, 1986; Nepveu *et al.*, 1986; Kindy *et al.*, 1987). The relative use of the P1 and P2 promoters is strongly influenced by interleukin 3 (IL3) in the IL-3-dependent pre-B-cell line Ba/F3. Removal of IL3 leads to a rapid and reversible drop in the levels of c-myc RNA, and a concomitant increase in antisense transcription (Chang *et al.*, 1991), perhaps indicating a switch in commitment step decision or in the control of proliferation. In addition, antisense transcription of the thymidine kinase gene is upregulated in serum starved mouse fibroblasts, while sense transcription is stimulated under conditions favouring growth (Sutterluety *et al.*, 1998).

A recent study has demonstrated the requirement of induced antisense transcription during neuronal differentiation of NT2 cells upon treatment with retinoic acid (Kawasaki and Taira, 2003). Stimulated expression of a small (23 nucleotides in length) RNA, called miR-23, predisposes NT2 cells to differentiate by targeting the translational repression of Hes1, a negative regulator of differentiation normally expressed in neural stem cells.

The above studies highlight the presence of sense and antisense transcription for genes normally implicated in the control of proliferation and differentiation. It will be interesting to investigate further their relationship in normal tissue growth and under neoplastic conditions. A general question in such cases is the role of the antisense transcript: does it lead to decreased proliferation or help maintain a committed step in differentiation ?

3.2.6.6 The conflicting arrows of time: bi-directional transcription in health and disease – Deregulated imprinting is implicated in complex diseases such as the Prader-Willi, Angelman, and Beckwith-Wiedemann syndromes (Rougeulle *et al.*, 1998; Yang *et al.*, 1998; Lee *et al.*, 1999; Brannan and Bartolomei, 1999; Nicholls and Knepper, 2001; Dennis, 2003; Gabellini *et al.*, 2003). From a standpoint of cell proliferation, it is tempting to speculate that deregulated sense-antisense mRNA ratios or loss of imprinting (LOI) may underlie growth factor overexpression and participate in the etiology of cancer. For example, *gFg* transcription is readily detectable in T47D breast cancer cells, which are known to express very low levels of bFGF mRNA. In contrast, the antisense transcript is undetectable in U87-MG glioma cells, which overexpress the bFGF sense mRNA (Murphy and Knee, 1994).

In Wilms' tumour, IGF2 overexpression does not appear to directly result from a loss of imprinting (Ogawa *et al.*, 1993; Wang *et al*, 1996), although there is apparent uncoupling of reciprocal imprinting of IGF2 and antisense transcription in this disease (Vu *et al.*, 2003). The apparent discrepancy may reflect sample variability since it is difficult to evaluate whether there is a temporal requirement for the contribution of LOI in the various stages of tumour progression. Further, the expression of the Wilms' tumour suppressor locus *WT1* is enhanced by the expression of an antisense RNA (Moorwood *et al.*, 1998), providing evidence that antisense transcription does not invariably lead to downregulated protein levels.

Another possible example involves the candidate tumour suppressor and transducer of TGF-β/BMP inhibitory signals, SMAD5, which maps to chromosome 5q31.1, a region deleted in human neoplasms of the myeloid lineage (Riggins *et al.*, 1996; Riggins *et al.*, 1997). Interestingly, an antisense transcript to SMAD5, known as DAMS, is expressed during development and in tumours of the pancreas, but not detected in adult tissues (Zavadil *et al.*, 1999). This raises the possibility that DAMS is normally involved in modulating SMAD levels. Similarly, an antisense transcript to hypoxia-inducible factor 1alpha (HIF-1α) detected in normal and cancer cell lines may modulate sense transcript levels (Rossignol *et al.*, 2002).

In an inherited form of α-thalassemia in humans, which is characterized by the loss of the HBA1 and HBQ1 α-globin genes as part of an 18 kb deletion, an aberrant transcript, in antisense orientation to the intact HBA2 globin gene, is produced (Tufarelli *et al.*, 2003). Expression of this antisense RNA leads to the DNA methylation and silencing of the remaining HBA2 gene, and is the causative agent underlying the genetic defect resulting from this chromosomal rearrangement. The activity of a non-coding RNA transcript in inducing *de novo* methylation at CpG islands in *cis*, leading to promoter silencing, is reminiscent of paternal expression of the *Air* transcript (see above, Sleutels *et al.*, 2002). Further, this provides the first reported case of antisense transcription causally implicated in human disease.

Such varied examples as those presented in this discussion highlight the importance of carefully evaluating the functional contributions of each case where antisense transcription is encountered, since not all necessarily involve RNA interference. However, caution should be applied in interpreting the mechanism of antisense transcription-mediated gene silencing, since it was demonstrated that RNAi itself is capable of inducing *de novo* methylation at CpG islands (Wassenegger *et al.*, 1994).

3.2.6.7 Antisense transcription epilogue: epigenetics and cancer – It is now apparent that concerted bidirectonal transcription is integrated with the epigenome. It is thus likely that deciphering the modes and consequences of regulated imprinting and of antisense transcription will contribute significantly to our understanding of disease states such as unregulated cell growth. In fact, recent studies suggest that we are witnessing the tip of the iceberg, since many genes with deregulated expression in cancer are under the influence of the imprinting marks, namely those of DNA methylation and histone acetylation (Cameron *et al.*, 1999; Jones and Baylin, 2002; Shi *et al.*, 2003; Zardo *et al.*, 2002). Formal proof of this interplay was recently reported through the elegant demonstration that tumour cells can be epigenetically

reprogrammed into normal cells using a somatic nuclear transfer technique (Li *et al.*, 2003) and by modulating DNA methylation (Gaudet *et al.*, 2003).

Also, a link now exists between cell proliferation and biological timing, underscoring the importance of understanding the positive and negative signalling influences acting on concerted sense and antisense transcription (Matsuo *et al.*, 2003). Thus, if the transformed phenotype is indeed sometimes symptomatic of deregulated transcription of sense/antisense gene pairs, then the requirement for the orchestrated interaction of the pairs of opposites may be revealed through the combined studies of biological clocks and epigenetics. Indeed, animals with disrupted core clock have a high incidence of sporadic tumours of various origins (reviewed in Fu and Lee, 2003). Furthermore, diurnal rhythmicity in histone acetylation underlies circadian variations in the transcription of clock components (Etchegaray *et al.*, 2003) and, to make matters more interesting, clock function, at least now demonstrated in *Neurosporra crassa*, is regulated by antisense transcription (Kramer *et al.*, 2003). This finding is intriguing since it was recently shown, using fibroblast cells isolated from zebrafish, that entry into S-phase of the cell cycle can be directly regulated by light (Dekens *et al.*, 2003). This is the first report supporting a cell autonomous contribution of the circadian clock, rather than daily fluctuations in hormone levels, in generating rhythms of cell cycle progression.

3.2.7 General discussion of the role played by non-coding RNA genes – *AS progranulin-1/-2* is a *bona fide* non-coding RNA gene. In order to provide a context to evaluate better its relationship to other non-coding RNAs, and its role in the genome of zebrafish, a brief overview of known non-coding RNAs and their biology is presented below.

When evoking transcripts that encode no protein, the classic ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and the five spliceosomal small nuclear uridine-rich RNA (snRNA) components of the spliceosome (U1, U2, U4, U5 and U6) come to mind. Genes encoding these diverse classes of RNAs are usually found as clusters and are represented in multiple copies ranging from < 50 to several hundred in eukaryotic genomes. Perhaps not surprising, nearly 98% of the human transcriptional output is non-coding RNA.

3.2.7.1 The non-coding RNA genes – Another category of non-coding RNAs is expanding in size, and includes RNAs known to perform various essential roles for the organism. Several interesting examples exist. For instance, most eukaryotes utilize telomerase RNA as the template for adding new telomeres (Singer and Gottschling, 1994). Signal recognition particle RNA is involved in translocating proteins across the endoplasmic reticulum (Walter and Blobel, 1982; Bovia and Strub, 1996). The yeast *meiRNA* regulates the onset of meiosis (Watanabe and Yamamoto, 1994). The small nucleolar RNAs (snoRNAs) constitute an intriguing class of non-coding RNAs involved in the post-transcriptional processing and modification of ribosomal RNA (Smith and Steitz, 1997; Tollervey and Kiss, 1997; Kiss, 2001). The surprising number of non-coding RNA genes has warranted the creation of a database that catalogues these sequences: http://www.man.poznan.pl/5SData/ncRNA/index.html) (Erdman *et al.*, 1999). In fact, the discovery of non-coding RNAs has spurred yet another exciting field of research, dubbed RNomics (Filipowics, 2000; Huttenhofer *et al.*, 2001).

The majority of vertebrate small nucleolar RNAs (snoRNAs) derive from intronic sequences of pre-mRNA primarily belonging to the 5' terminal

oligopyrimidine (5'-TOP) gene family (Tollervey and Kiss, 1997; Weinstein and Steitz, 1999). Individual snoRNA molecules thus do not contain their own promoters, and are released from their resident transcripts through nucleolytic processing. Although originally thought to encode protein (Schneider *et al.*, 1988; Coccia *et al.*, 1992), the human growth arrest-specific gene (*gas5*) hosts ten snoRNAs and illustrates an extreme case whereby a snoRNA host gene may be a non-coding RNA itself (Tycowski *et al.*, 1996; Smith and Steitz, 1998). Further, besides having a low probability of encoding protein, the exonic sequences of the *gas5* gene are not recognizably similar between mouse and man (Tycowski *et al.*, 1996). Thus, this brings an interesting perspective to the notion that lack of sequence conservation is not necessarily associated with non-functionality.

3.2.7.2 Non-coding RNAs and disease – Due to intensive searches of the genetic determinants, numerous non-coding RNAs have been discovered in the imprinted region of Prader-Willi syndrome (Meguro *et al.*, 2001). Several of these genes are in antisense orientation to protein-coding genes, and may operate through mechanism already discussed. For instance, this region contains clusters of unusual C/D box snoRNAs that apparently do not modify rRNA and whose expression is as expected for a Prader-Willi candidate gene: it is brain-specific and is turned on as an imprinted gene only from the paternal chromosome (Cavaille *et al.*, 2000; De los Santos *et al.*, 2000). This observation is counterintuitive to the usually unregulated expression of intronic snoRNAs. Moreover, one of these snoRNAs (*HBII-52*) has sequence complementarity to 5-hydroxytryptamine 2C (5-HT_{2C}) receptor mRNA, and has been proposed to direct the methylation of a site known to be subjected to mRNA editing. It thus remains to be proven if a snoRNA is involved in a complex set of interactions leading to the editing of an mRNA transcript.

The contribution of such non-coding RNAs in disease is being actively investigated. Some patients with spinocerebellar ataxia where found to carry mutations in their *SCA8* (*spinocerebellar ataxia 8*) cis-antisense RNA (Nemes *et al.*, 2000). Most notably, the short-limbed dwarfism *cartilage-hair hypoplasia* (CHH) is associated with

mutations in the 267-nucleotide non-coding RNA derived from the RMRP gene, an essential component of the ribonucleoprotein endoribonuclease MRP (mitochondrial RNA processing) (Clayton, 2001; Ridanpaa *et al.*, 2001). Furthermore, mutations were found in the telomerase RNA gene, positioning it as a candidate gene in the etiology of autosomal-dominant dyskeratosis congenital disorder (Vulliamy *et al.*, 2001). Supporting evidence for the involvement of this RNA is its interaction with dyskeratin, a protein already known to be implicated in X-linked dyskeratosis.

3.2.8 The transcriptome nexus: an evolving definition – Protein-coding genes can give rise to products as small as seven amino acids in length (Gonzales-Pastor, *et al.*, 1994). At a first glance, when considering the estimated protein-coding gene repertoire of various organisms – 6 000 in yeast (Goffeau *et al.*, 1996), 12 000 each in *Drosophila* and *C.elegans* (Rubin *et al.*, 2000), 15 000 in Ciona intestinalis (Dehal *et al.*, 2002), 35 000 each in mouse and man (Lander *et al.*, 2001; Venter *et al.*, 2001), and perhaps more in zebrafish – there is a paradox between the apparent complexity of life as one progresses from unicellular to higher vertebrates and the number of players performing the various biochemical functions required for day to day life. This is best exemplified by the 99% conservation of protein-coding genes between *homo sapiens* and *mus musculus* (Mouse Genome Sequencing Consortium, 2002). Admittedly, both morphological and physiological differences are likely achieved through the context-dependent and reiterative use of selected gene components as a consequence of domain accretion events permitting higher order interactions (Duboule and Wilkins, 1998). In this regard, it is conceivable that the construction of functionally distinct proteins by combining pre-existing disparate exonic modules (e.g. exon shuffling, alternative cis- and trans-splicing) and gene duplications might have fundamental repercussions on an organism's physiology and on our conceived notions of the transcriptome (Graveley, 2001).

It may be that a key element contributing to the complexity of life is the number of unique mRNA molecules, whether protein-coding or not, originating within the respective transcriptomes (Black, 2000). Interestingly, from a purely genetic

perspective, less than 0.5 % of the estimated ~ 3 million sequence differences per haploid genome between human individuals occurs within protein-coding regions, and most of these are silent (third base) changes (Venter *el al.*, 2001).

Certainly, attempts made towards determining the size of the exon universe are validated still: life *is* predicated on the presence of proteins that act as fundamental biological effectors of cellular function. However, there is ample support for the hypothesis that phenotypic differences (e.g. X-chromosome inactivation), as well as likely conserved physiological processes (e.g. heterochronic gene regulation), are clearly contributed by the action of non-coding RNA genes. Thus, from the viewpoint of the transcriptome, non-coding RNA, including introns, can be regarded as "architects of eukaryotic complexity" (Mattick *et al.*, 2001; Federova and Federov, 2003).

3.2.8.1 Transcription dark side – It is difficult to predict the number of authentic non-coding transcripts that remain to be discovered. Surprisingly, of the 60770 full-length complementary DNA sequences annotated in the mouse transcriptome, 11665 are non-coding (The FANTOM Consortium *et al.*, 2002). Most new non-coding RNA genes will presumably be selectively expressed during unconventional situations, such as in periods of starvation. For instance, recent analyses of bacterial genome intergenic regions has identified several segments giving rise to non-coding RNA transcripts, some that are selectively expressed in growth conditions other than "normal" (Argaman *et al.*, 2001; Wassarman *et al.*, 2001). While analyzing the expression profile of chick pineal tissue entrained to a 12-hour light, 12-hour dark photoperiod, as well as in constant darkness, 902 transcripts were noticed with no significant BLAST match (Bailey *et al.*, 2003). Implying that all of these RNAs represent heretofore-unrecognized transcripts would be misleading, since the chick transcriptome is not yet fully annotated. However, it can be anticipated that a small fraction of these transcripts will be derived from the growing number of non-coding RNA genes.

Finally the living genome can be viewed as two primordial manifestations. One is a malleable canvas that has been expanded in size as a consequence of the seismic activities, throughout the entire evolutionary process, of transposable elements giving rise to the enigmatic C-value or lack of correspondence between genome size and biological complexity (Comeron, 2001). The other is concerted transcriptional output resulting in the con-joined activities of both protein- and non-coding transcripts (Eddy, 1999; Koonin *et al.*, 2000; Eddy, 2001). The zebrafish progranulin gene family, along with the associated complementary non-coding RNA gene and trans-spliced transcript provide a general illustration of these evolutionary forces that shape the destiny and complexity of life.

3.3 Expression of zebrafish progranulins

3.3.1 Analysis of progranulin gene expression in adult zebrafish tissues by RT-PCR – All primers used in the RT-PCR analyses, as well as the expected size of cDNA amplicons, are indicated in **TABLE 5**. Because progranulin-1 and progranulin-2 possess a high degree of sequence identity throughout their entire sequence, including the UTRs, and since they also give rise to a chimeric transcript, it became essential to determine whether these can be discriminated in the PCR. After numerous attempts, conditions favouring the selective amplification of either of these transcripts were achieved (**Figure 30**). It should be noted, however, that slight changes in the conditions used, such as the preparation of PCR master mix solutions or presumed temperature fluctuations encountered in the thermal cycler, have often resulted in the appearance of non-specific amplicons (cross-amplification) or in non-amplification of target cDNA as judged by the inclusion of positive and negative controls (i.e. cloned cDNAs as template) in each experiment. For this reason, the amplification of progranulin-1, progranulin-2 and the hybrid are presented specifically whenever the RT-PCRs were successful at least twice. In some instances, the amplification of progranulin-1-like sequences using a combination of primers, that generate a product dubbed as "generic", whose sequences do not discriminate between progranulin-1 and progranulin-2, and that possess no sequence complementarity to the AS progranulin-1/-2 transcript is also included (**TABLE 5**).

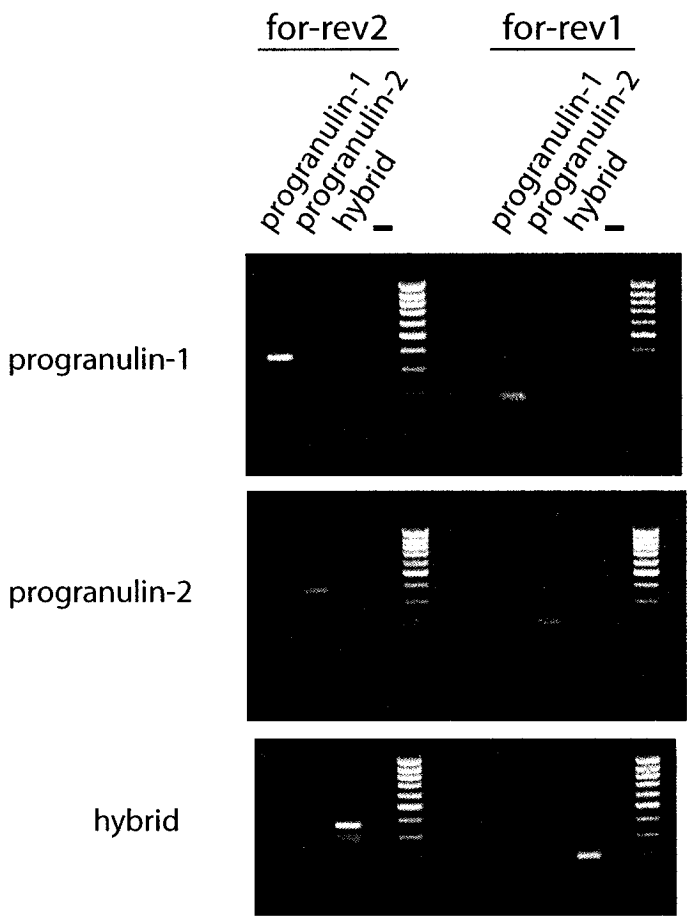
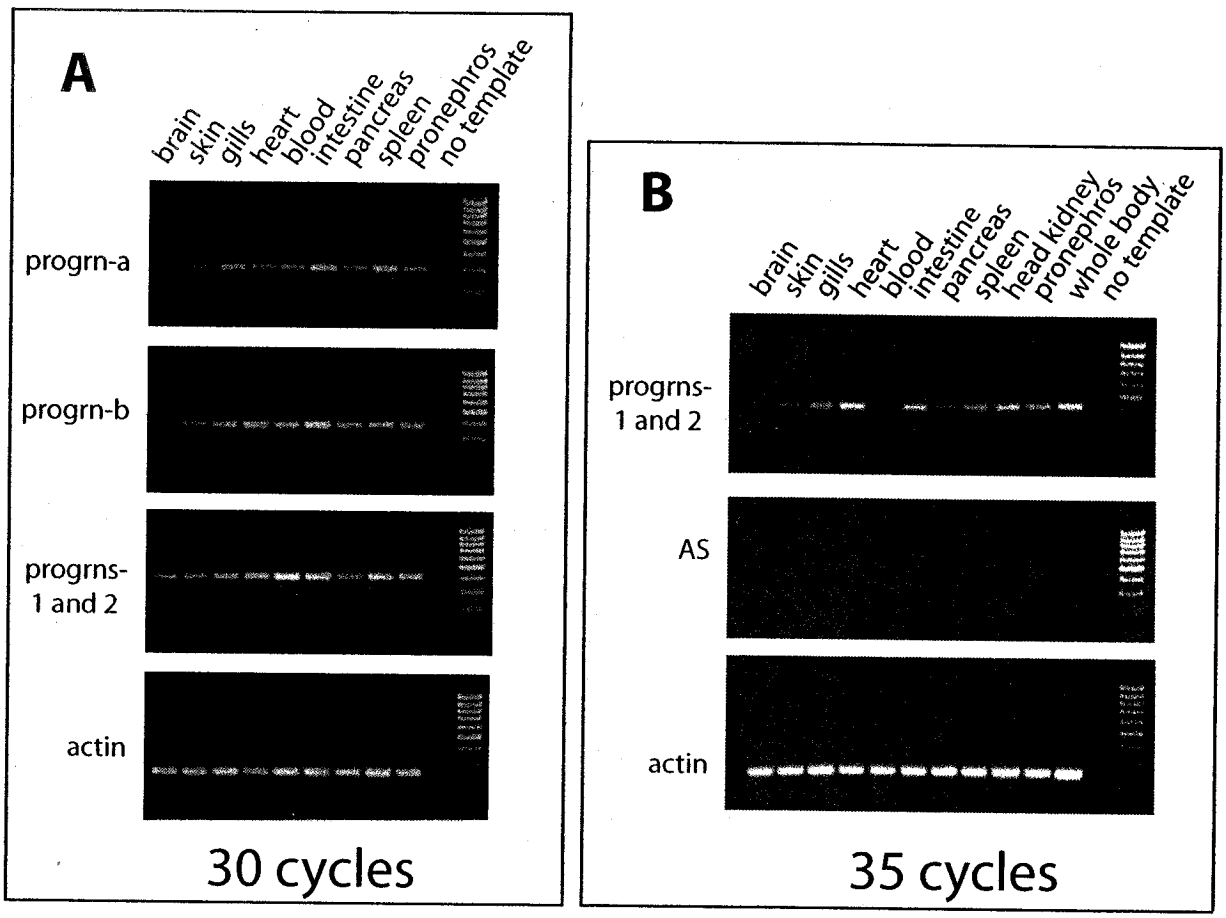


Figure 30. **Discrimination of Progranulin-1 and Progranulin-2 transcripts in the PCR.** Individual primer combinations of primers listed in Table R4 (panel B) were used for the amplification of gene-specific transcripts and for the detection of the hybrid transcript (forward-reverse1: 190 bp; forward reverse2: 372 bp). Approximately 50ng of template DNA cloned into pbluescript was used per lane. Each reaction was carried according to the protocol described in the Methods section 2.17, and aliquots were run next to a 100 bp DNA ladder.

In mammals, progranulin gene expression is known to be widespread. Although this thesis project primarily concerns embryonic development using the zebrafish as model, progranulin gene expression was studied in the adult for several reasons. First, due to the fact that granulins form an extended gene family in zebrafish, it was of interest to determine whether progranulin-a and -b expression differed significantly from one another both qualitatively and quantitatively in selected adult tissues, and also relative to their smaller counterparts (progranulins-1 and -2) (**Figure 31, panel A**). At first glance, the results suggest that progranulins are ubiquitously expressed in various adult organs microdissected from the adult zebrafish. Second, it was important to monitor the expression of the AS transcript relative to progranulins-1 and –2: while the combined expression of the latter genes is detected in every organ, the AS progranulin-1/-2 transcript is detected primarily in the blood and intestine (**Figure 31, panel B**). Third, since it was previously shown that carp granulin-1 and granulin-2 peptides are differently distributed in the spleen and head kidneys of fish (Belcourt *et al.*, 1993), it was important to investigate whether the expression of the orthologous structures in zebrafish were similarly uncoupled at the level of mRNA (**Figure 31, panel C**). Interestingly, as with the carp spleen, the equivalent zebrafish organ shows exclusively progranulin-1 gene expression; alternatively viewed, the reported absence of granulin-2 peptide in carp spleen extracts (Belcourt *et al.*, 1993; and **Figure 7**), is reflected by a similar lack of detectable progranulin-2 mRNA expression in adult zebrafish spleen. In addition, progranulin-1 appears to be the predominant form in the heart, while the eyes express higher levels of progranulin-2 (**Figure 31, panel C**). Interestingly, the hybrid transcript is detected weakly in the intestine, a tissue where antisense expression was also detected (**Figure 31, panel B**). The coincidental detection of both the AS progranulin-1/-2 transcript and hybrid progranulin mRNA in the same tissue may suggest a requirement for AS expression in the formation of a *trans*-spliced transcript. Co-localization of these mRNAs at the cellular level is required to provide more solid evidence for this possible scenario. Discrepancies between progranulin-a and progranulin-b expression are also apparent (**Figure 31, compare panel A vs panel C**). However, cDNAs for these panels were prepared from total RNA isolated from adult zebrafish obtained from different sources,

and different amounts of starting material were used for these different experiments (as reflected by the inclusion of actin as control). Moreover, due to the presence of progranulin-a expression in migrating granulocytes (see whole-mount mRNA *in situ* hybridizations results, sections 3.3.8 to 3.3.11) it is possible that the two sets of animals used were under different environmental conditions, resulting in heightened progranulin-a expression in **Figure 31, panel C**.

In view of their small size, there might be mistakes in the actual microdissection of individual organs from zebrafish. For instance, presumed male and female gonads were shown to be positive for *progranulin-a*, *progranulin-b*, as well as the generic amplicon for *progranulin-1* and *progranulin-2* and their *AS* transcript. It would be interesting to visualize cellular expression of zebrafish progranulin genes expression by *in situs* hybridization of adult organs.


A
brain
skin
gills
heart
blood
intestine
pancreas
spleen
pronephros
no template
progrn-a
progrn-b
progrns-
1 and 2
actin
30 cycles
B
brain
skin
gills
heart
blood
intestine
pancreas
spleen
head kidney
pronephros
whole body
no template
progrns-
1 and 2
AS
actin
35 cycles

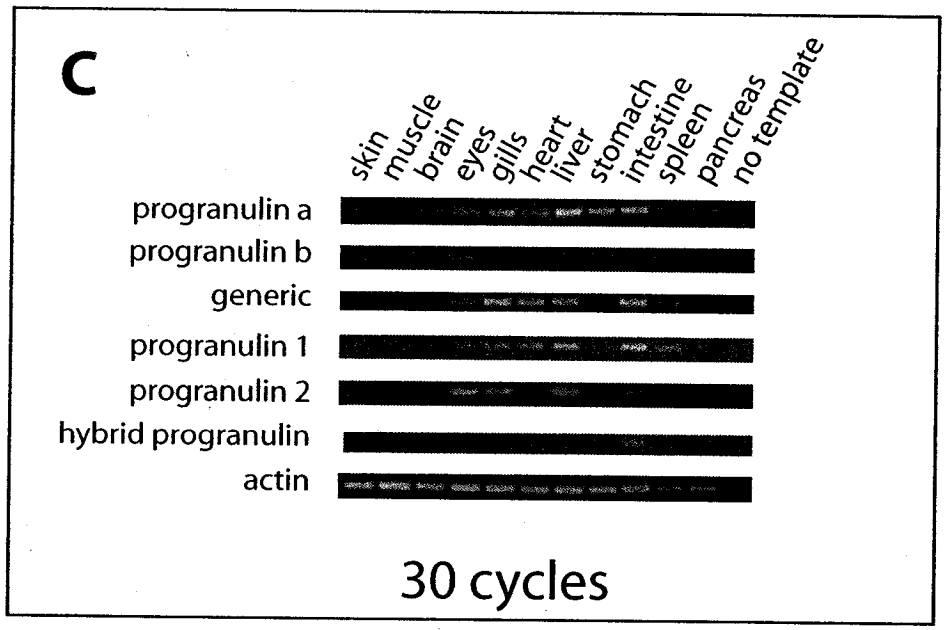

C
skin
muscle
brain
eyes
gills
heart
liver
stomach
intestine
spleen
pancreas
no template
progranulin a
progranulin b
generic
progranulin 1
progranulin 2
hybrid progranulin
actin
30 cycles

Figure 31

Figure 31. **RT-PCR analysis of zebrafish progranulins expression in adult tissues.** *Panel A*: Individual progranulins are ubiquitously expressed in a variety of selected tissues. Note that there is no discrimination between the progranulin-1 and progranulin-2 genes. *Panel B*: The antisense transcript is weakly expressed in the intestine and blood. *Panel C*: Progranulin-1 and progranulin-2 are differently expressed in various adult organs, while the hybrid transcript is weakly detected in the intestine.

3.3.2 Analysis of progranulin gene expression during zebrafish development by RT-PCR – During development, *progranulin-a* and *progranulin-b* are both readily detectable prior to the onset of zygotic expression which is initiated at the high stage (3 hours post-fertilization – hpf) in the zebrafish (**Figure 32, panel A**), and are expressed throughout all subsequent developmental stages. *Progranulin-b* is more abundant than *progranulin-a* in the initial stages of embryogenesis until gastrulation, as judged by a slightly higher level of *progranulin-a* expression at the shield stage which coincides with the formation of the zebrafish organizer. This trend continues in the more advanced gastrula (tailbud stage), prior to neurulation and early segmentation stages where both transcripts appear to be expressed at similar levels (3-7 somites). From the late segmentation stage or pharyngula period (24 hpf) and onwards, the two genes are expressed at a similar level.

Interestingly, *progranulin-1* and/or *progranulin-2* are not readily detectable in the early stages of zebrafish development, reflecting a fundamental difference between the usages of the various progranulin genes, at least during development (**Figure 32, panel A**). Inclusion of primers for the combined detection of these smaller progranulin genes and their antisense transcript reveals that *progranulin-1* and/or *progranulin-2* is expressed by 30hpf, while *AS* RNA levels are too low to be detected by ethidium bromide stain (**Figure 32, panel B**). Southern transfer of this gel and analysis using labelled primer that detects both the sense and antisense amplicons as probe reveals that antisense transcripts are present at low levels from approximately 72 hpf and become more elevated in the 5 day old fish (120 hpf) (**Figure 32, panel B**). Also, this strategy detects very low levels of progranulin-1-like transcripts during earlier stages of development. Interestingly, *progranulin-1*, but not *progranulin-2* or the *hybrid* transcript is expressed in the 1-cell stage embryo (**Figure 32, panel C**). Attempts to specifically amplify *progranulin-1*, *-2* and the *hybrid* during zebrafish development are ongoing.

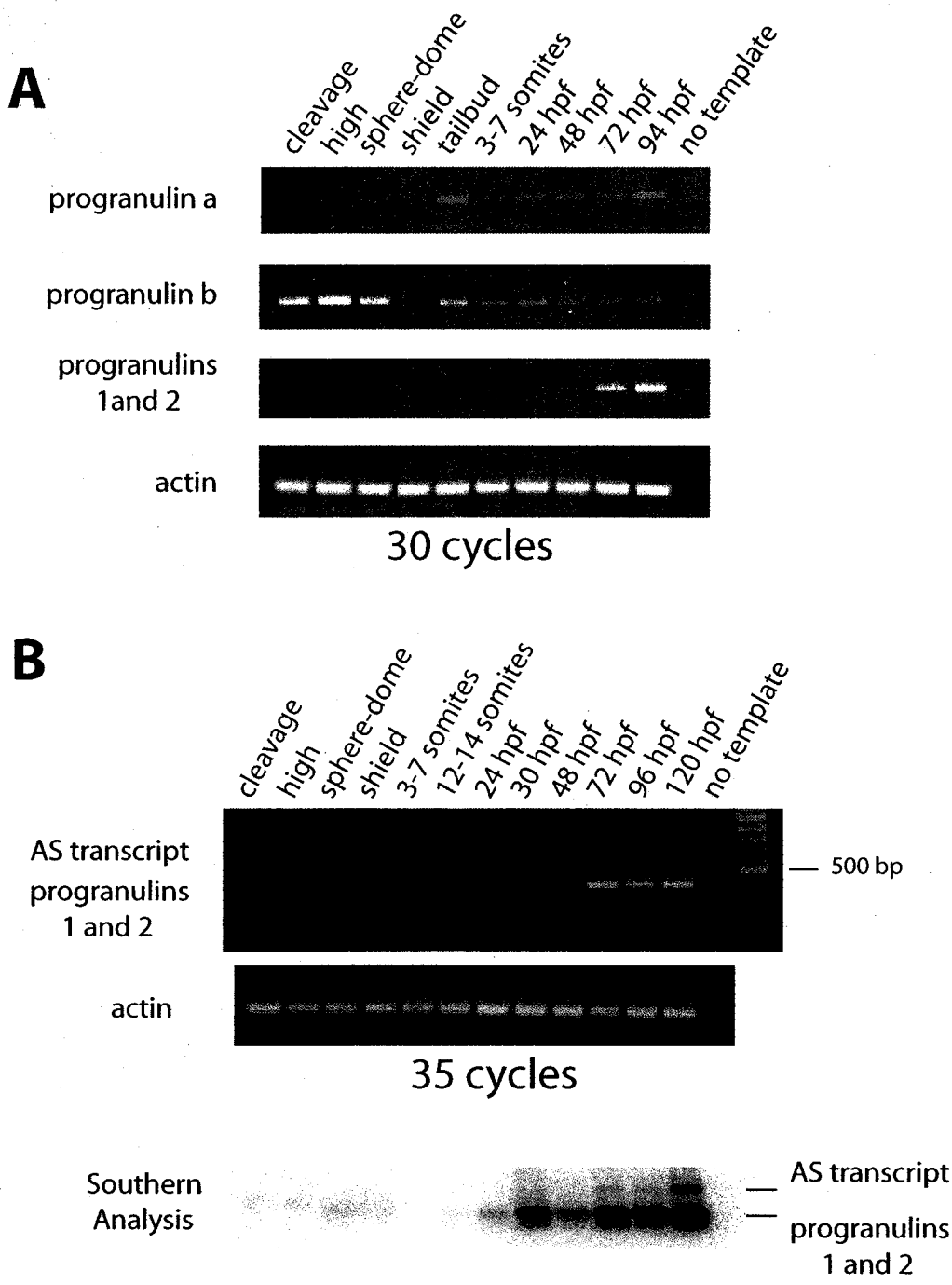

A
cleavage
high
sphere-dome
shield
tailbud
3-7 somites
24 hpf
48 hpf
72 hpf
94 hpf
no template
progranulin a
progranulin b
progranulins 1and 2
actin
30 cycles
B
cleavage
high
sphere-dome
shield
3-7 somites
12-14 somites
24 hpf
30 hpf
48 hpf
72 hpf
96 hpf
120 hpf
no template
AS transcript progranulins 1 and 2
500 bp
actin
35 cycles
Southern Analysis
AS transcript
progranulins 1 and 2
C
grn1
grn2
hybrid
1
1
1

Figure 32

Figure 32. **RT-PCR analysis of zebrafish progranulins during zebrafish development**. *Panel A*: Progranulin-a and progranulin-b transcripts are detected throughout all the stages of development, but maternal expression of progranulin-b is more abundant. Progranulin-1 and/or progranulin-2 expression occurs at later stages of development. These results provide evidence for the devolution of function of the four progranulins paralogues through distinct temporal gene regulation. *Panel B*: Combined progranulins-1 and –2 expressions relative to their antisense transcript. Ethidium bromide stain reveals the presence of the sense transcripts only (top). Detection of the antisense transcript is revealed by Southern blot analysis using a ^{32}P-labelled oligo that recognizes both sense and antisense amplicons as probe (bottom). Note weak expression of progranulin-1/-2 at earlier stages of development. *Panel C*: Progranulin-1, but not progranulin-2, is maternally expressed at the 1 cell stage. Numbers of cycles used for the PCR are indicated.

3.3.3 Northern blot analysis of zebrafish progranulin transcripts – To assess the size of the respective progranulin transcripts, we are now performing Northern blot analyses on total RNA derived from staged zebrafish larvae (data not shown). Since we have detected the presence of an antisense transcript to progranulin-1 and progranulin-2, and also have evidence for a transcript sharing complementarity to the progranulin-a transcript (data not shown), we are using single-stranded RNA probes to ensure specific detection of the targeted transcript. This method is favoured over conventional random-priming methods for labelling of cDNA probes for the following reason: the possible detection of bands whose size did not correspond to that predicted from cDNA cloning could potentially be mis-interpreted as alternative-splice variants or a transcript derived from the complementary DNA strand, in addition to spurious non-specific hybridization.

3.3.4 Analysis of zebrafish progranulin gene expression by whole-mount *in situ* mRNA hybridization – The ontogenic expression of individual progranulin genes has been investigated by *in situ* hybridization, and is discussed next in small segments. For the presentation of the results in this section, anatomical structures have been carefully annotated and attempts were made to highlight uncertainties whenever these arise. Clearly, the resolution of this issue is critical for the interpretation of biological experiments presented in the last section of this thesis. For this reason, we are preparing cross-sections of our processed embryos to better visualize the various structures embedded within the animals. Furthermore, all of the expression studies involving whole-mount *in situ* hybridization here are currently being independently assessed by our colleague Dr. Bernard Thisse (IGBMC, Illkirch, France), in whose laboratory is being conducted the expression analysis of the zebrafish transcriptome (Thisse *et al.*, 2001; http://www.zfin.org).

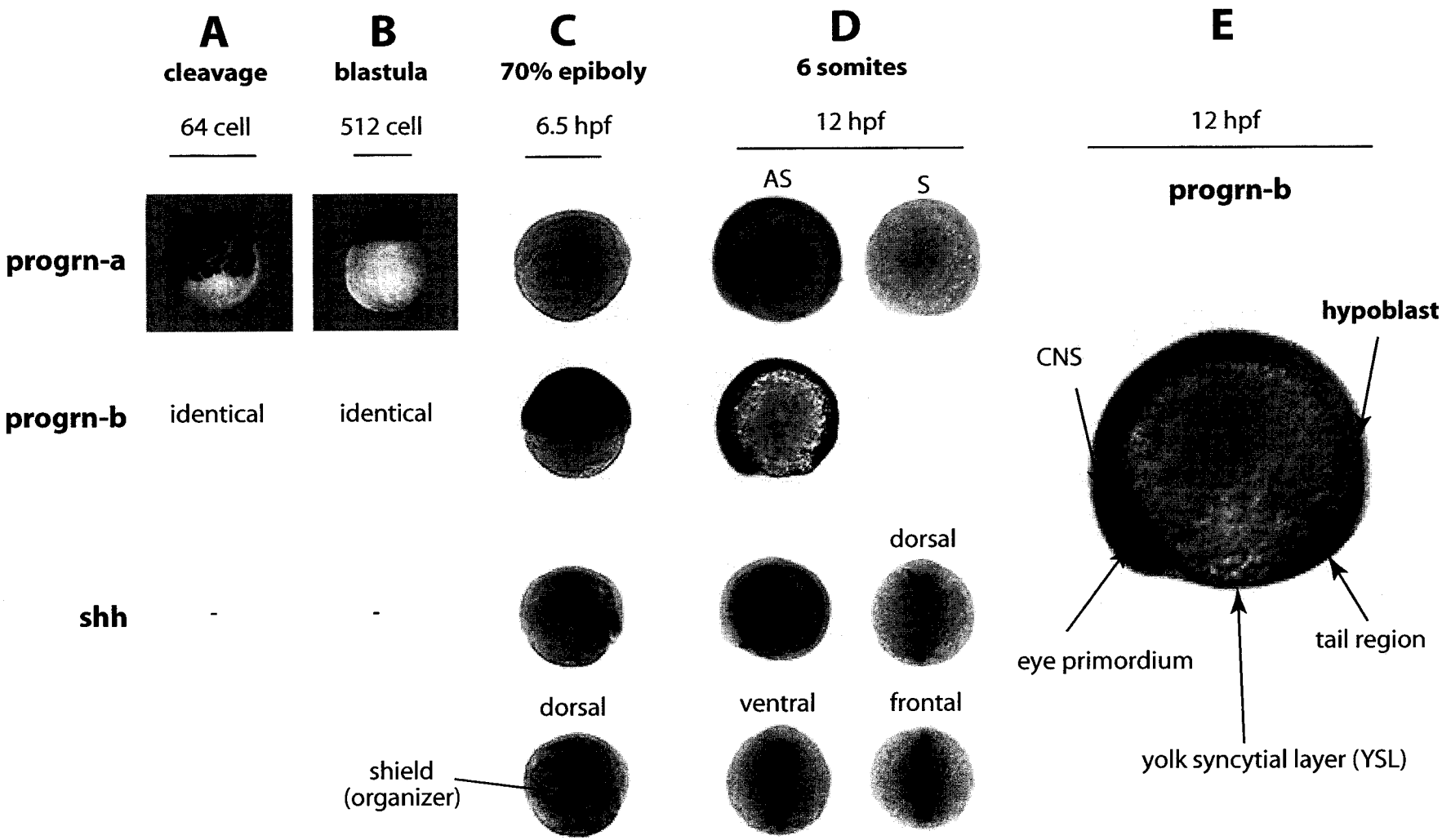
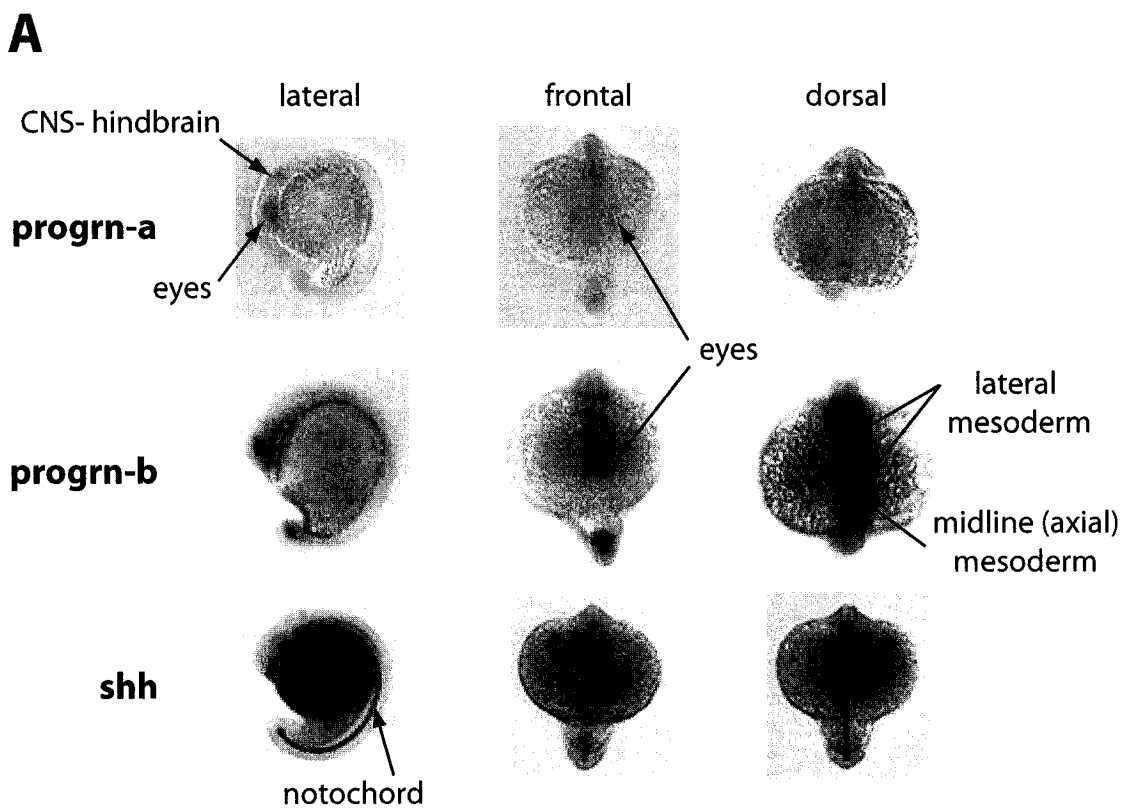

A
cleavage
B
blastula
C
70% epiboly
D
6 somites
E
64 cell
512 cell
6.5 hpf
12 hpf
12 hpf
AS
S
progrn-b
progrn-a
progrn-b
identical
identical
hypoblast
CNS
dorsal
shh
-
-
tail region
eye primordium
dorsal
ventral
frontal
shield
(organizer)
yolk syncytial layer (YSL)

Figure 33

Figure 33. **Progranulin-a and progranulin-b expression in the early stages of zebrafish embryogenesis and in the early segmentation period**. Overall, these expression patterns are similar for these genes; with the exception that progranulin-a is generally weaker than progranulin-b. *Panel A*: Ubiquitous expression in all blastomeres of the cleavage stage zygote. *Panel B*: In the blastula period, prior to zygotic onset of expression, maternal expression of progranulins can still be detected, albeit at a lower intensity. *Panel C*: During epiboly, progranulin-b expression is more intense than progranulin-a, and appears to be uniformly distributed over the blastomeres. *Panel D*: In the early segmentation period (6 somites), progranulin staining is detected fairly ubiquitously, but regionalizations are now apparent, such as in domains of the central nervous system (CNS). Progranulin-b staining is intense in the epithelium of the eye primordium, in the developing CNS, and in the tail region. *Panel E*: Magnification of progranulin-b embryo in panel D. Note the staining in the YSL and that trunk expression is localized to the hypoblast layer. Controls using Shh are shown. Panels C-E, views are lateral, anterior (rostral) to the left, dorsal at the top, except where indicated otherwise.

3.3.5 Progranulin-a and progranulin-b: cleavage stage to early segmentation period – As expected, *progranulin-a* and *-b* are both detected in every blastomeres of cleavage-stage embryo (**Figure 33, panel A**) and prior to the onset of zygotic expression (mid-blastula) (**Figure 33, panels B**). Expression of these genes is still apparently uniformly distributed to all blastomeres (the enveloping layer, the deep cells, and cells of the yolk syncytial layer (YSL)) during epiboly stages, when individual blastomeres are migrating from the animal to the vegetal pole resulting from the pulling movements of the YSL cells (**Figure 33, panel C**). This is in contrast to the localized expression of sonic hedgehog to the hypoblast layer of the embryonic shield at 70% epiboly. During early segmentation (6 somites stage or 12 hpf), *progranulin-a* and *-b* are still ubiquitously expressed, but differences can be seen (**Figure 33, panel D**). While both appear to be generally more highly expressed rostrally (anterior) than caudally (posterior), *progranulin-a* is more diffuse than *progranulin-b*. In fact, rostrally, *progranulin-b* is more intensely detected in the epithelial layers of the eye primordium and in the developing central nervous system (**Figure 33, panel E**) than is *progranulin-a*. In the trunk region, where individual somites are gradually accumulating, *progranulin-b* expression can be detected in the midline extending to the tail region, in the presumed hypochord and notochord (the hypoblast layer - see section 3.4.4.2.5 for discussion); note that expression is not localized to the surface. In the tail region, *progranulin-b* expression is also elevated, perhaps suggesting a role in the tail organizer or its proliferation (axis elongation, but not formation). *Progranulin-b* is also detected in the yolk syncytial layer (**Figure 33, panel E**). Although *progranulin-a* expression appears to reflect that of *progranulin-b*, its detection within the hypoblast derivatives and yolk syncytial layer was not clearly noticed. This may reflect fundamental differences between the duplicated genes in the zebrafish, or simply a lower overall abundance of *progranulin-a* in these structures. Unless stated otherwise throughout these studies, the respective sense riboprobes to *progranulin-a* and *progranulin-b* do give rise to detectable signal (data not shown). As expected from the RT-PCR data (**Figure 32**), *progranulin-1* and *progranulin-2* expression is not detected at these stages (data not shown).


A
lateral
frontal
dorsal
CNS- hindbrain
progrn-a
eyes
eyes
progrn-b
lateral
mesoderm
midline (axial)
mesoderm
shh
notochord

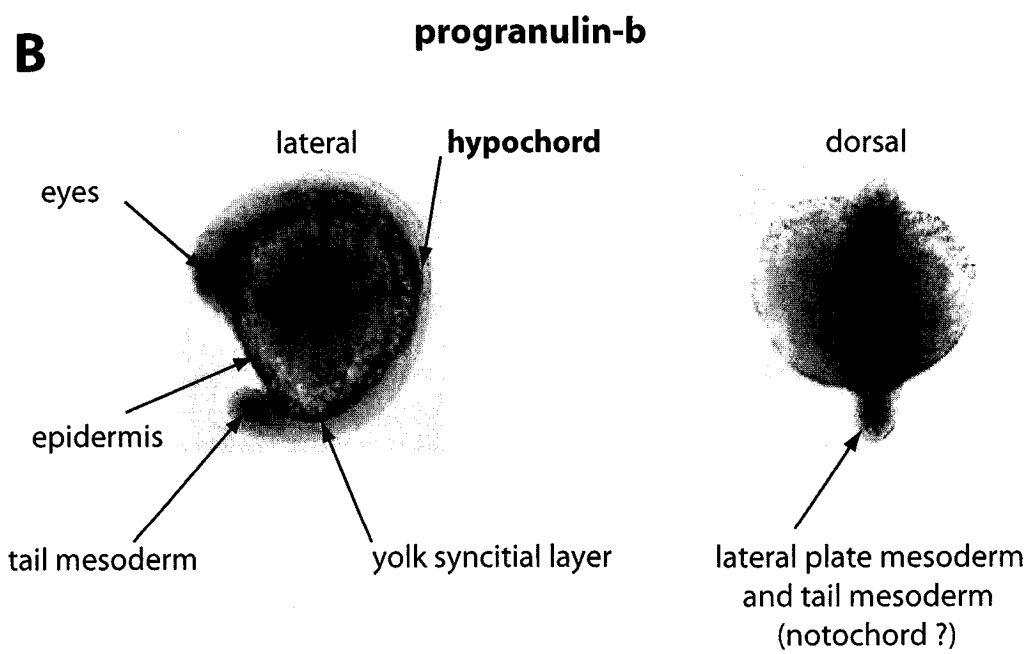

B
progranulin-b
lateral
hypochord
dorsal
eyes
epidermis
tail mesoderm
yolk syncitial layer
lateral plate mesoderm
and tail mesoderm
(notochord ?)

Figure 34

Figure 34. **Progranulin-a and progranulin-b expression in the mid- to late-segmentation stages.** *Panel A*: Progranulin-a and progranulin-b become more restricted as segmentation proceeds. Weak expression of progranulin-a is observed in the CNS, in particular the hindbrain region, and in the developing eyes. Faint expression is also detected in the prospective trunk and tail mesoderm. Progranulin-b expression is strong in the eyes, in the ventral region of the CNS and in the prospective endoderm and/or lateral plate mesoderm. Shh expression is shown as control. *Panel B*: Magnified view of progranulin-b embryo shown in panel A. Note the trunk expression is more ventral and appears to be localized to the hypochord, not the hyperchord. Expression in the epidermis, the YSL, and tail mesoderm, are indicated.

3.3.6 Progranulin-a and progranulin-b: mid- to late-segmentation period (18-20 hpf) – In later stages of somitogenesis, *progranulin-a* expression is again more diffuse and less intense than *progranulin-b* (**Figure 34, panel A**). While *progranulin-a* was found to be more concentrated in the rostral regions of the embryo, such as in the developing eyes and in the CNS (see lateral and frontal views), its expression was very weakly detected in the trunk region (dorsal view). In contrast, *progranulin-b* was intense in the eyes, the tail region and in the presumed yolk syncytial layer (lateral and frontal views). The apparent different expression pattern in the CNS and in the eyes between *progranulin-a* and *progranulin-b* (frontal view) was interpreted as being the result of age differences (the embryo for *progranulin-a* is ~20 hpf, while that for *progranulin-b* is 18hpf). Various timepoints were investigated for each probe with each giving similar results. *Progranulin-b* was also detected in the prospective lateral plate mesoderm (LPM) and in the midline (dorsal view) throughout the entire axis of the embryo (i.e. rostral to caudal). Note that this expression pattern is also restricted ventrally, as revealed by the lateral view (i.e. next to the yolk surface). *Shh* was used as a positive control (**Figure 34, panel A**). A magnified view of the embryo for *progranulin-b* is shown in **Figure 34, panel B**. Closer inspection of the tail region reveals that progranulin-b is expressed in two lateral stripes and one weaker midline stripe (ventral view) at least in this region, in a pattern predicted to be the LPM.

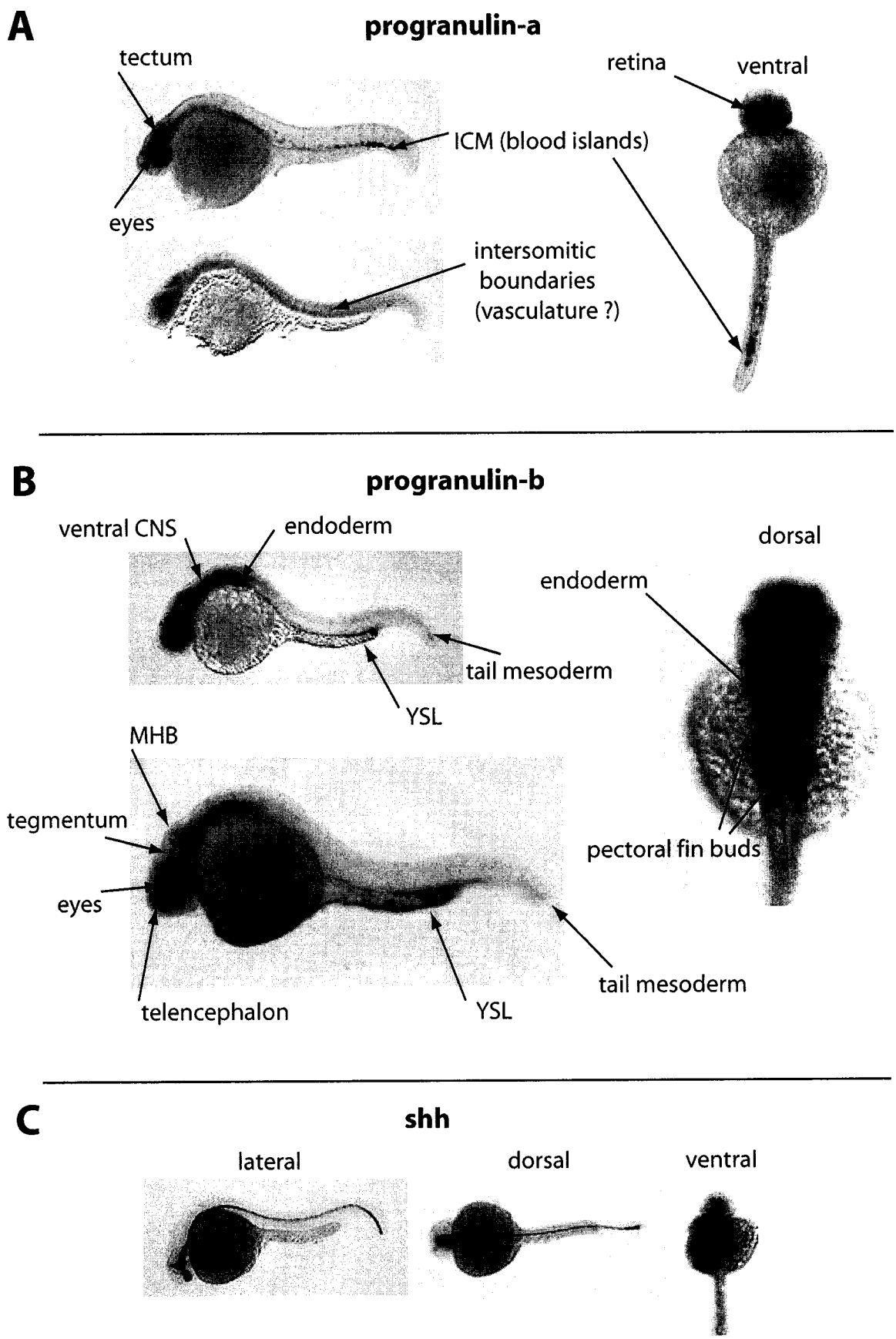

A
progranulin-a
tectum
eyes
ICM (blood islands)
intersomitic
boundaries
(vasculature ?)
retina
ventral
B
progranulin-b
ventral CNS
endoderm
tail mesoderm
YSL
dorsal
endoderm
pectoral fin buds
MHB
tegmentum
eyes
telencephalon
YSL
tail mesoderm
C
shh
lateral
dorsal
ventral

Figure 35

Figure 35. **Progranulin-a and progranulin-b expression in the early pharyngula period (24-26 hpf).** *Panel A*: (lateral view) Progranulin-a expression is detected in the ventral tail region in a population of cells defining hematopoietic stem cells of the intermediate cell mass (ICM). Occasionally, progranulin-a is also detected in single bands demarcating the somite borders. Weak expression in the visceral endoderm on the ventral region of the embryo next to the yolk surface can also be noticed. In the head region, strong progranulin-a expression can be detected in the tectum and the eyes. A ventral view indicates that progranulin-a is expressed in the retina. *Panel B*: (lateral view) Progranulin-b expression is stronger in the prospective visceral endoderm and brain, but not as intense in the eyes (including retina) and in the tectum. In the brain, progranulin-b is more intense in the telencephalon, the tegmentum, and in the midbrain-hindbrain boundary (MHB) region. Expression in the YSL is detected, especially in the yolk extension region. Expression in the trunk mesoderm increases caudally, and is found more intensely in the tail region. (dorsal view) Expressions in the ectodermal surface of the presumptive pectoral fin buds, as well as in the prospective visceral endoderm, are indicated. *Panel C*: Control Shh expression.

3.3.7 Progranulin-a and progranulin-b: early pharyngula stage (24-26 hpf) – Later in the pharyngula period, *progranulin-a* and *progranulin-b* were seen to display clearly diverging expression patterns (**Figure 35, panels A and B, respectively**). Both genes were expressed in the brain, but to varying degrees. *Progranulin-a* was predominantly concentrated in the eyes (ventral view) including a region presumed to be the retina, and in the tectum (**Figure 35, panel A**). *Progranulin-a* was also detected in the prospective visceral endoderm and, occasionally, at the somite borders. Interestingly, *progranulin-a* and not *progranulin-b*, was expressed in cells of the intermediate cell mass (ICM) which define a population of hematopoietic stem cells during the primitive phase of hematopoiesis in the zebrafish (see 3.4.4.2.5.2 for discussion). *Progranulin-b* appeared to be more abundantly expressed in various regions of the zebrafish brain at this stage (**Figure 35, panel B**). Staining on the yolk surface, especially in the yolk extension region, was a feature of *progranulin-b* expression. These do not represent migrating leukocytes, but rather cells of the yolk syncytial layer. Also, note that *progranulin-b* was not expressed in the ICM. A dorsal view for *progranulin-b* revealed staining in the prospective visceral and pharyngeal (more anteriorly located) endoderm, as well the epithelial surface, but not mesoderm, of the growing pectoral fin buds (**Figure 35, panel B, dorsal view**). *Progranulin-a* was observed to be similarly expressed in these regions, but slightly less intense. The use of *Shh* as positive control is represented (**Figure 35, panel C**).

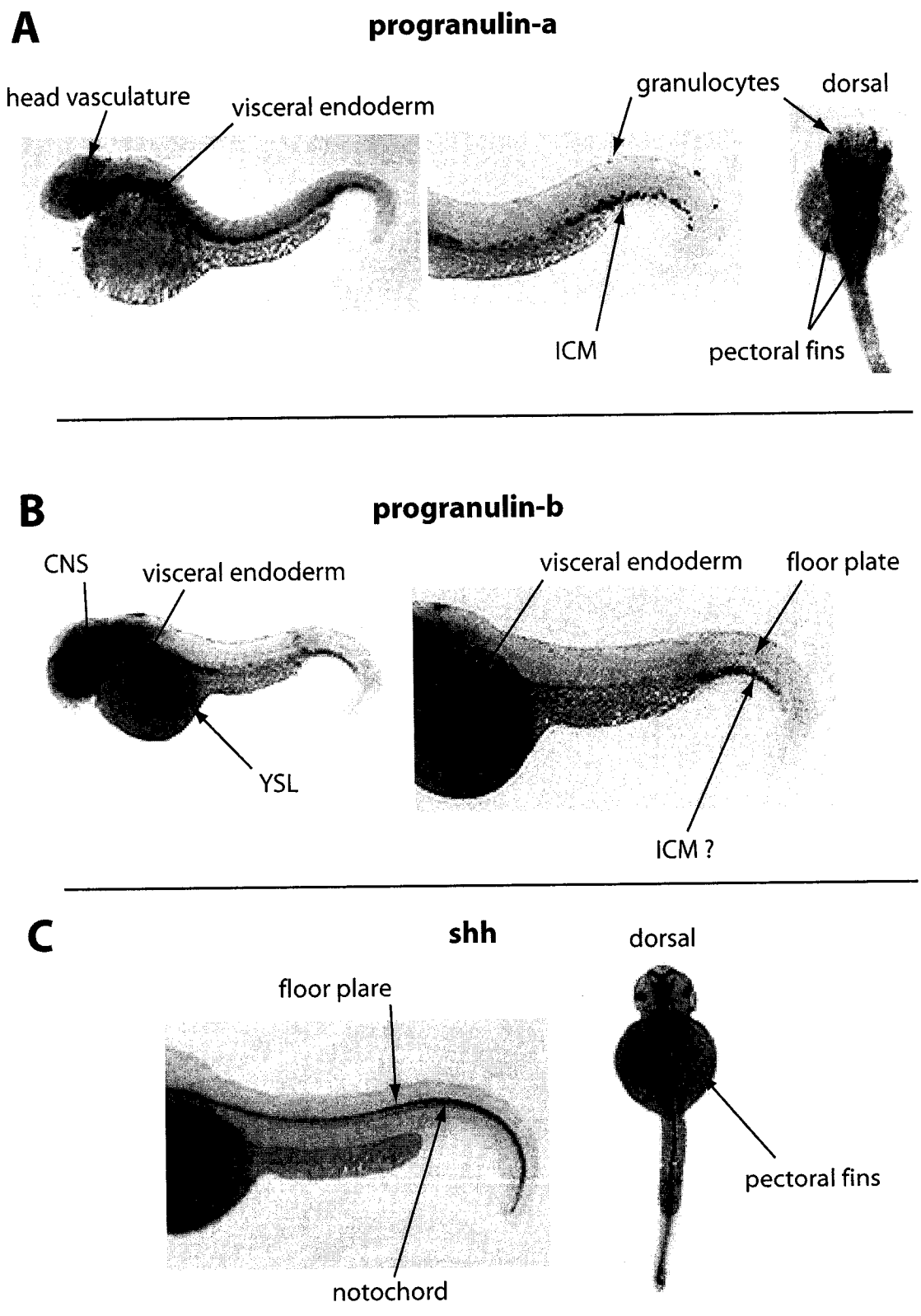

A
progranulin-a
head vasculature
visceral endoderm
granulocytes
dorsal
ICM
pectoral fins
B
progranulin-b
CNS
visceral endoderm
YSL
visceral endoderm
floor plate
ICM ?
C
shh
floor plare
notochord
dorsal
pectoral fins

Figure 36

Figure 36. **Progranulin-a and progranulin-b expression in the mid-pharyngula period (36 hpf).** *Panel A*: Progranulin-a expression is detected in the head vasculature and in dispersed leukocytes. Detection in the ventral region of the tail (ICM) persists, with sporadic expression extending caudally in the ventral region of the trunk. Weak expression in the pectoral fins can be noticed (dorsal view). *Panel B*: Progranulin-b is strongly expressed in the brain, the prospective visceral endoderm, and in the YSL. Expression in the ventral region is also noticed, but in a cell population distinct from that observed for progranulin-a (see text for details). Progranulin-b expression was not detected in leukocytes. Weak expression in the presumed floor plate is indicated. *Panel C*: Control Shh expression.

3.3.8 Progranulin-a and progranulin-b: mid-pharyngula stage (36 hpf) – Expression patterns already observed in the one day old embryo were maintained at 36 hpf. This included the specific ICM and yolk syncytial layer expression for *progranulin-a* and *progranulin-b*, respectively, and their combined expression in the visceral endoderm (**Figure 36, panels A and B**). In addition, *progranulin-a* expression was no longer detected in the tectum and in the eyes, but rather now detected in the head vasculature and in migrating leukocytes throughout the body of the animal (**Figure 36, panel A**). In contrast, *progranulin-b* expression was not detected in leukocytes, persisted at high levels in the brain, and was detectable in the forming head vasculature, and in the posterior region of the ICM, (**Figure 36, panel B**). We have not observed progranulin-b expression in the trunk vasculature.

Inspection of the respective tail regions for *progranulin-a* and *progranulin-b* highlighted that *progranulin-b* expression was found in the presumed endoderm in the ventral region of the trunk, as opposed to the punctate stem cell-like expression of *progranulin-a* in caudal region of the ICM (**Figure 36, panels A and B**). Control expression using *Shh* is represented (**Figure 36, panel C**). The switch from a predominant notochord expression to the floor plate is indicated. Note also *Shh* expression in the pectoral fin mesoderm (dorsal view).

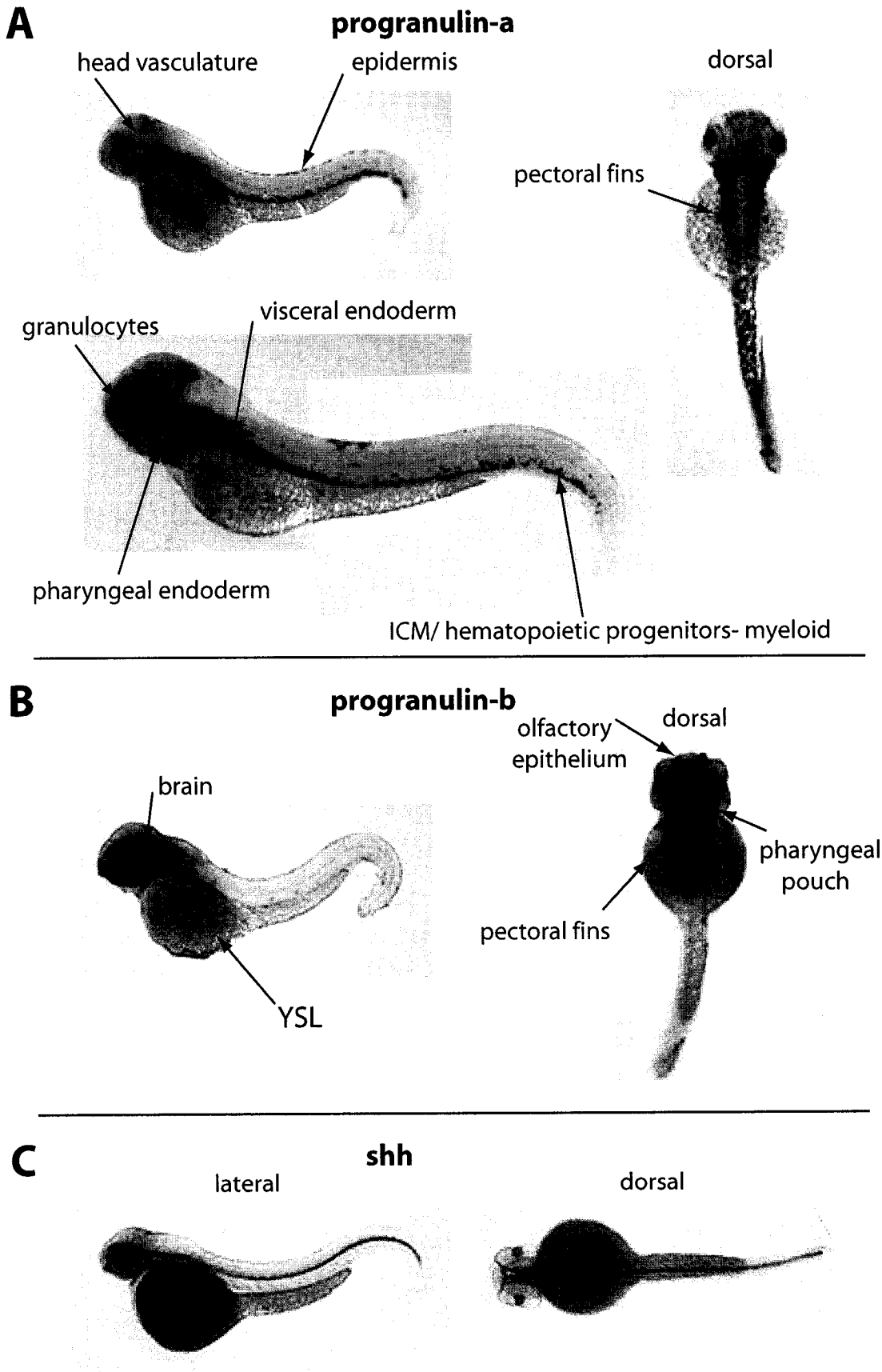

A
progranulin-a
head vasculature
epidermis
dorsal
pectoral fins
granulocytes
visceral endoderm
pharyngeal endoderm
ICM/ hematopoietic progenitors- myeloid
B
progranulin-b
olfactory epithelium
dorsal
brain
pharyngeal pouch
pectoral fins
YSL
C
shh
lateral
dorsal

Figure 37

Figure 37. **Progranulin-a and progranulin-b expression in the hatching period (48 hpf).** *Panel A*: Progranulin-a is detected in blood vessels of the head, the pharyngeal pouch and visceral endoderm, the skin epidermis, as well as in leukocytes (granulocytes), ventral tail ICM progenitors, and pectoral fins. *Panel B*: Progranulin-b expression in various locations in the brain, and on the YSL, is maintained. (dorsal view) Progranulin-b expression is also detected in the pectoral fins and presumptive pharyngeal and visceral endoderm, as well as the olfactory epithelium. *Panel C*: Control Shh expression.

3.3.9 Progranulin-a and progranulin-b: hatching period (48 hpf) – *Progranulin-a* and *progranulin-b* expression was seen to persist in the vasculature of the head at 2 days post-fertilization (**Figure 37, panels A and B**). In addition to ICM and visceral and pharyngeal endoderm expression, *progranulin-a* was now detected in the skin epidermis (**Figure 37, panel A**). *Progranulin-b* expression was seen to be similar to that observed for the 36hpf embryo (**Figure 37, panel B**). Note the pectoral fin expression for both *progranulin-a* and *progranulin-b* (dorsal views). Control *Shh* expression is represented (**Figure 37, panel C**). *Progranulin-1* and *progranulin-2* expression was detected in the proctodeum region of the intestine (data not shown).

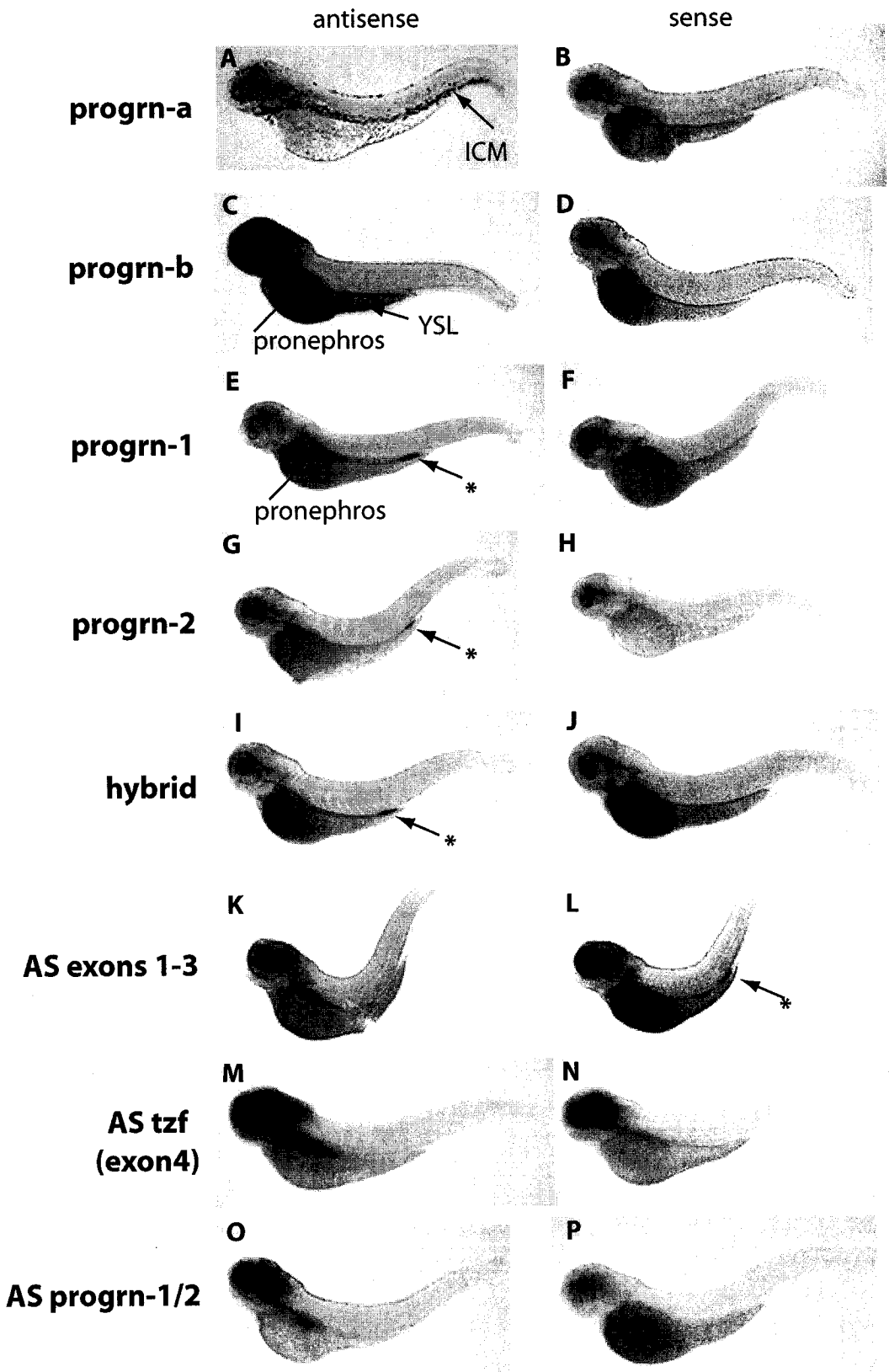
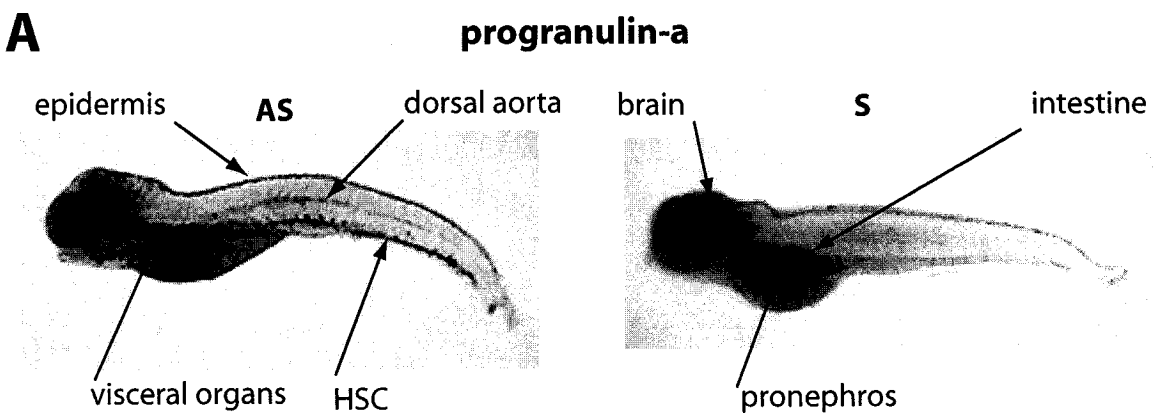


Figure 38. **Zebrafish progranulins, and associated AS gene, expression in the hatching period (72 hpf).** For each transcript, antisense (panels A, C, E, G, I, K, M, O) and sense (panels B, D, F, H, J, L, N, P) riboprobes are included. All views are lateral, rostral to the left, dorsal at the top. The (*) denotes the proctodeum (panels E, G, I, L).

3.3.10 Progranulin gene expression in the hatching period (72 hpf) – At the 72hpf stage, *progranulin-a* was widely expressed in leukocytes, in presumed myeloid progenitors of the ICM located in the ventral region of the tail, and in the epidermis (**Figure 38, panel A**). *Progranulin-a* (and *progranulin-b*) expression in the pectoral fins was no longer detected. *Progranulin-b* was still strongly expressed in the brain and in the yolk syncytial layer and was more abundant in the differentiating visceral organs than *progranulin-a*, including the pronephros (**Figure 38, panel C**). *Progranulin-1* was strongly expressed in the developing pronephros (**Figure 38, panel E**), while *progranulin-2* was not (**Figure 38, panel G**). In contrast, *progranulin-2* was often detected in a few migrating leukocytes (data not shown). Proctodeum expression for *progranulin-1*, *progranulin-2* and the *hybrid* transcript was detected with varying levels, in a highly reproducible manner: *progranulin-2* (**Figure 38, panel G**) was invariably less abundant than *progranulin-1* (**Figure 38, panel E**) and *hybrid progranulin* (**Figure 38, panel I**). Sense riboprobes for these transcripts did not detect a signal for the antisense transcript (**Figure 38, panels F, H, J**), as expected (**see Figure 32, panel B**). Interestingly, the sense riboprobe corresponding to the first three exons of the AS transcript (AS exons 1-3; see methods) also detected expression in the proctodeum (**Figure 38, panel L**). A segment corresponding to part of the fourth exon of the AS transcript (AS tzf; see methods, section 2.20), which corresponds to the tzf transposase segment, was used for the synthesis of antisense and sense riboprobes. These riboprobes resulted in the detection of strong ubiquitous expression (**Figure 38, panels M and N**, respectively) when the reaction was allowed to progress for an equal duration required for the normal detection of the progranulin transcripts using the progranulin riboprobes (data not shown). Similar results were observed when riboprobes corresponding to the entire sequence of the *AS progranulin-1/-2* transcript were used (**Figure 38, panels O and P**), masking the detection of the selective expression of *progranulin-1* and *progranulin-2* transcripts with the sense probe (data not shown).


A
progranulin-a
epidermis
AS
dorsal aorta
visceral organs
HSC
brain
S
intestine
pronephros

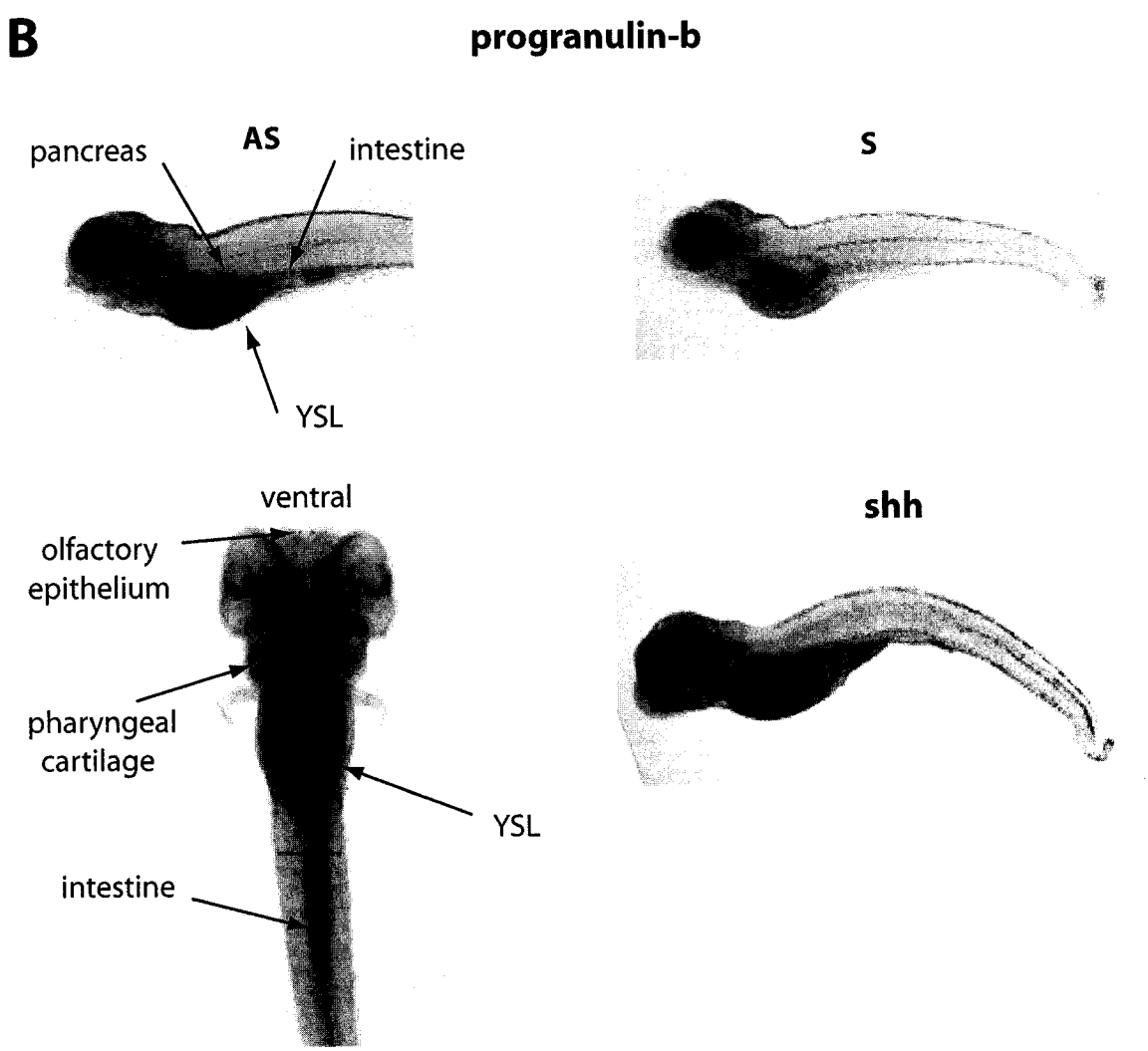

B
progranulin-b
pancreas
AS
intestine
YSL
S
ventral
olfactory epithelium
pharyngeal cartilage
YSL
intestine
shh

Figure 39

Figure 39. **Progranulin-a and progranulin-b expression at 5 dpf**. *Panel A*: Progranulin-a is expressed in several visceral organs, in the skin epidermis, in hematopoietic stem cells (HSC) located in the ventral region of the tail, and in leukocytes (left). Expression is also detected in the dorsal aorta, but not in the pericardinal vein. A sense riboprobe detects signal in the brain, in the intestinal region and in the pronephric ducts (kidneys) (right). *Panel B*: Progranulin-b expression is detected in the pancreas, pronephros (not shown), intestine, and in the YSL (left). A ventral view (bottom) shows the distinction between the YSL and intestine, as well as expression in the olfactory epithelium. Unlike progranulin-a, a sense riboprobes to progranulin-b does not detect expression (right). A control Shh expression is included.

3.3.11 Progranulin gene expression at 5 dpf

3.3.11.1 Progranulin-a and progranulin-b – In the 5 day-old larva, progranulin-a was detected in several visceral organs, including the epithelial lining of the intestine, and pronephric ducts (kidneys) (**Figure 39, panel A**) as revealed by cross-sectioning (data not shown). Staining in the dorsal aorta but not the posterior cardinal vein was also noticeable through sectioning of the animal. *Progranulin-a* was still detected in the epidermis of the skin, in presumed leukocytes, as well as in hematopoietic progenitor cells that are located in the ventral region of the tail and that extend rostrally along the ventral surface of the entire animal (see discussion) (**Figure 39, panel A**). Interestingly, intense staining with the use of the sense riboprobe for *progranulin-a* was noticed in the brain, the intestine and pronephros (**Figure 39, panel A**).

Progranulin-b expression was located in the intestine, the yolk syncytial layer (YSL) and in the pancreas, as well as in the presumptive thymus as bilateral patches located caudal to the eyes, albeit more weakly (**Figure 39, panel B**). A ventral view for *progranulin-b* highlights the difference between the YSL staining and the intestine. Staining (expression) was also seen in the olfactory epithelium. In contrast to *progranulin-a*, the sense riboprobe for *progranulin-b* did not detect any localized expression (**Figure 39, panel B**). A typical result for *Shh* expression is also included as positive control (**Figure 39, panel B**).

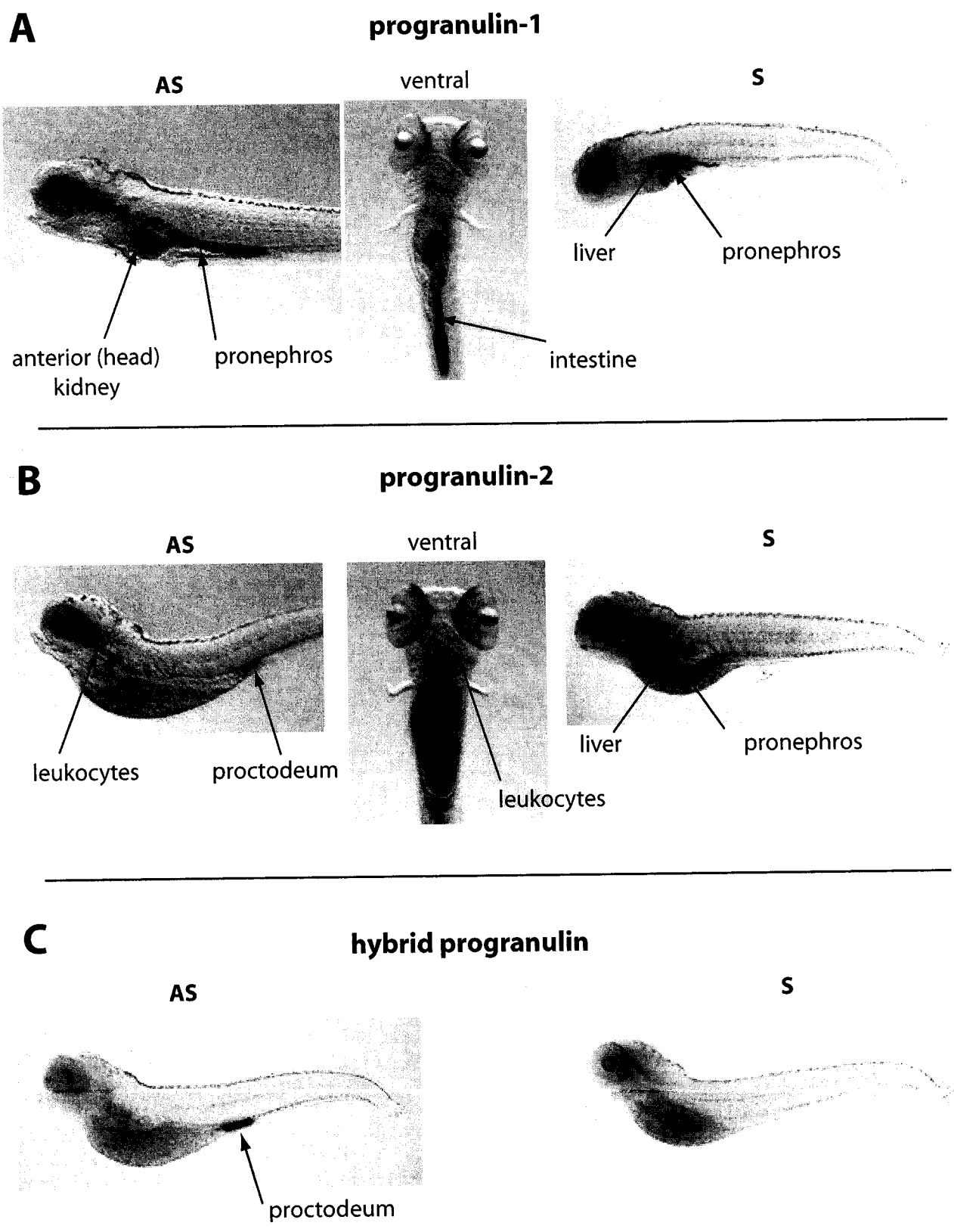

A
progranulin-1
AS
ventral
S
liver
pronephros
anterior (head) kidney
pronephros
intestine
B
progranulin-2
AS
ventral
S
liver
pronephros
leukocytes
proctodeum
leukocytes
C
hybrid progranulin
AS
S
proctodeum

Figure 40

Figure 40. **Progranulin-1, progranulin-2, and hybrid progranulin expression at 5 dpf.** *Panel A*: Progranulin-1 is expressed in the intestine and in the pronephric ducts, including the anterior kidney region (left). Sense riboprobe detects the antisense transcript in the liver and pronephros, but does not extend to the anterior kidney region (right). *Panel B*: Similarly, a sense riboprobe to progranulin-2 detects expression of the antisense transcript in the pronephros and liver (right). Progranulin-2 mRNA is detected in leukocytes, and is restricted to the proctodeum region of the intestine (right). *Panel C*: Hybrid progranulin is expressed strongly in the proctodeum, in contrast to progranulin-2, but not in leukocytes (left). A sense riboprobe to the hybrid transcript does not detect antisense progranulin expression (right).

3.3.11.2 Progranulin-1, progranulin-2 and hybrid progranulin – *Progranulin-1* was detected primarily in two organs: the pronephros (lateral view) and the intestine (ventral view) (**Figure 40, panel A**). Although not easily distinguishable through the pictures provided, the bilateral staining in the pronephros can be noticed in some embryos. Note also that *progranulin-1* expression extends from the proximal region of the pronephric ducts (also called the anterior or head kidney) to its caudal end. Coincident with the observation made by RT-PCR analyses (**Figure 32, panel B**), antisense transcript expression was detected in the 5 dpf zebrafish. Specifically, sense riboprobe for *progranulin-1* detected expression in two areas: the presumptive liver and the middle region of the pronephros (**Figure 40, panel A**). There was no indication that the AS progranulin-1/-2 transcript is expressed in the head kidney region, or in the caudal part of the pronephros (**Figure 40, panels A and B, sense riboprobes; Figure 41, panel D**).

In contrast to the data expressed with *progranulin-1*, *progranulin-2* was weakly detected in the proctodeum and was observed in migrating leukocytes than can be detected throughout the animal (**Figure 40, panel B**). The *progranulin-2* sense riboprobe gave an identical pattern to that noticed for the sense *progranulin-1* riboprobe.

The *hybrid progranulin* RNA interestingly was restricted to the proctodeum, where it is more abundantly expressed than *progranulin-2* (**Figure 40, panel C**). Furthermore, a sense riboprobe did not give rise to a detectable hybridization signal, lending credence to the specificity of the *hybrid* transcript detection, as well as to the *AS progranulin-1/-2* transcript hybridization signal using the sense riboprobes for *progranulin-1* and *progranulin-2*, respectively.

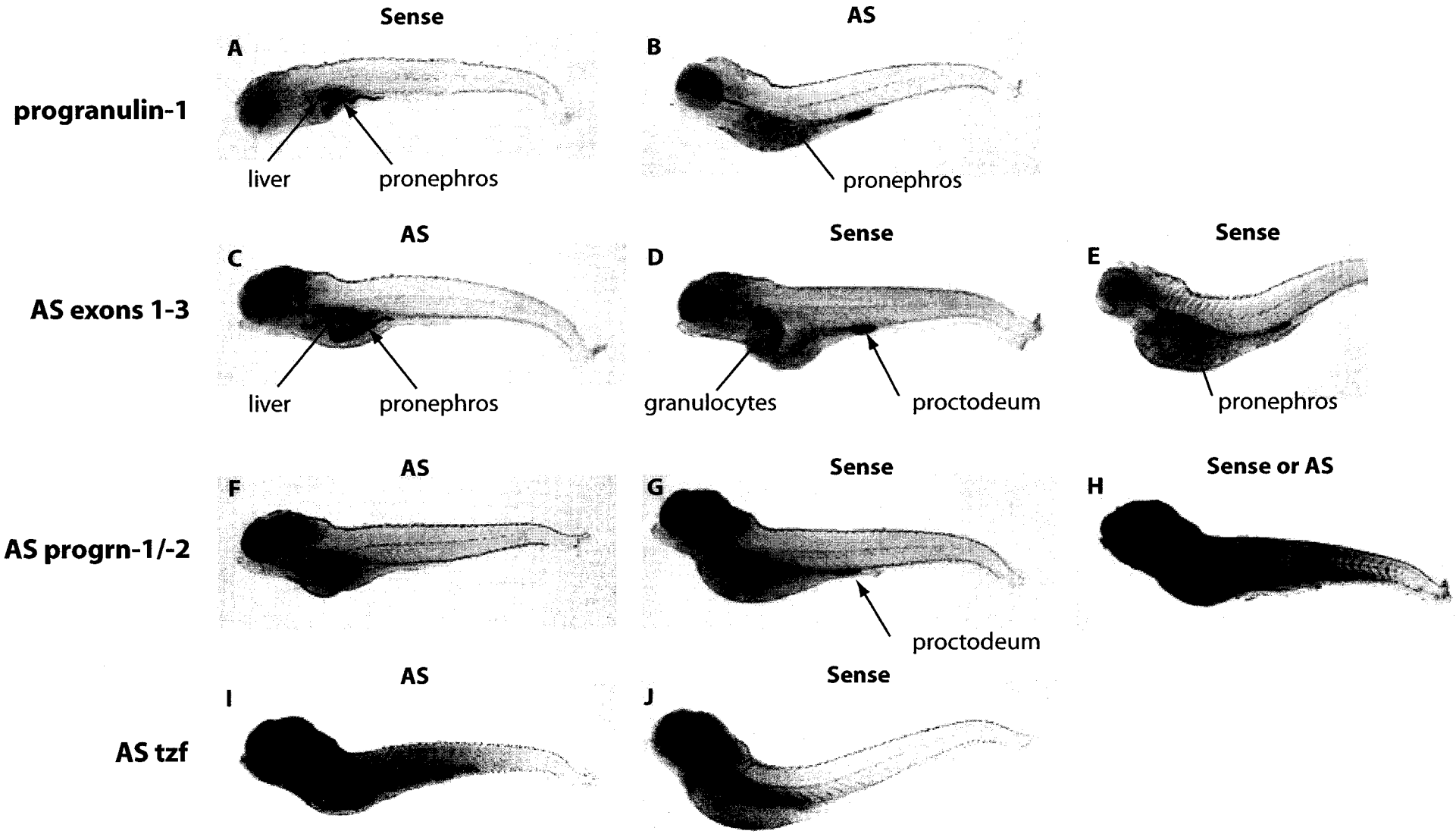

progranulin-1
AS exons 1-3
AS progrn-1/-2
AS tzf
Sense
A
liver
pronephros
AS
B
pronephros
AS
C
liver
pronephros
Sense
D
granulocytes
proctodeum
Sense
E
pronephros
AS
F
Sense
G
proctodeum
Sense or AS
H
AS
I
Sense
J

Figure 41

Figure 41. **Antisense transcript (AS progranulin-1/-2) expression at 5dpf**. *Panels A*: Sense riboprobe for progranulin-1 detects expression in the liver and pronephros. *Panel B*: Antisense riboprobe for progranulin-1 detects expression in the pronephros. *Panel C*: Antisense riboprobe for the AS progranulin-1/-2 transcript (minus the tzf portion) detects expression in the liver and pronephros (compare to panel A). *Panels D and E*: Sense riboprobe for the AS progranulin-1/-2 transcript (minus the tzf portion) detects expression in the proctodeum and leukocytes (panel D) and the pronephros (panel E). *Panel F*: Antisense riboprobe corresponding to the full sequence of the AS prgranulin-1/-2 transcript detects non-specific hybridization due to the tzf sequence. *Panel G*: A sense riboprobe corresponding to the full AS progranulin-1/-2 transcript similarly detects non-specific hybridization, but also in the proctodeum. *Panel H*: Prolongation of the staining reaction for the riboprobes used in panels F and G leads to the detection of ubiquitous expression. *Panels I and J*: Riboprobes corresponding to the tzf portion of the AS progranulin-1/-2 transcript do not detect signal in the proctodeum, but ubiquitous staining if the reaction is prolonged.

3.3.11.3 AS progranulin-1/-2 – In **Figure 41**, progranulin gene expression from the viewpoint of the antisense transcript is presented. Using an antisense riboprobe corresponding to the full-length portion of *AS progranulin-1/-2* devoid of the tzf portion (AS exons 1-3) (**Figure 41, panel C**), pronephric duct and liver expression was detected. This was directly comparable to the data obtained with the sense riboprobe of progranulin-1 (**Figure 41, panel A**). The sense riboprobe for the AS exons 1-3 construct detected expression in leukocytes and in the proctodeum (**Figure 41, panel D**) as well in the pronephros (**Figure 41, panel E**). This recapitulated the overall combined expression of *progranulin-1* and *progranulin-2*, as expected (**Figure 41, panel B**; see also **Figure 40**).

The use of riboprobes corresponding to the full sequence of the *AS progranulin-1/-2* transcript (**Figure 41, panels F and G**) brought about overall whole-body expression if the reaction was prolonged (**Figure 41, panel H**). Nonetheless, proctodeum expression was noticed with the sense probe prior to the detection of overall background (**Figure 41, panel G**). Use of the region derived from exon 4 of the antisense transcript (AS tzf) suggested that the background observed for the full antisense transcript sequence as probe resulted from ubiquitous tzf transposon sequence expression in zebrafish tissues (**Figure 41, panels I and J**).

3.4 Discussion of zebrafish progranulin genes expression

3.4.1 Introduction – In the next section, observations made concerning the tissue distribution of individual zebrafish progranulins gene transcripts are discussed. These studies were undertaken for several reasons. First, since the zebrafish possess an extended progranulin gene family, it was important to address whether all members are widely expressed, as is the case for their mammalian counterpart. Second, it was critical to determine the granulin genes are differentially regulated relative to one-another, possibly reflecting a partitioning of function from the ancestral form onto the duplicated paralogues, and thereby facilitating their retention in the fish genome (see section 3.2.3). Third, knowledge of the expression patterns for these genes was essential in order to provide a context for studying the role(s) that these factors play during development.

The expression of the *AS progranulin-1/2* gene, as well as the localization of the chimeric progranulin mRNA, were analysed in order to examine their spatio-temporal relationship with the progranulins gene. The implications of these results are also discussed in this section.

For clarity, and continuity, this discussion is subdivided into three categories: 1) RT-PCR analyses in zebrafish adult tissues; 2) RT-PCR analyses during zebrafish development; 3:) whole-mount *in situ* hybridization analyses during zebrafish development.

3.4.2 RT-PCR analysis of progranulins expression in the adult zebrafish

3.4.2.1 Progranulin-a and progranulin-b vs mammalian progranulin – The human progranulin gene is widely expressed in various tissues and several cell lines of epithelial, mesenchymal, and hematopoietic origin (Bhandari *et al.*, 1992; Plowman *et al.*, 1992; Daniel *et al.*, 2000). Although interspecies differences do occur, the orthologous gene identified in rat (Bhandari *et al.*, 1993), mouse and guinea pig (referred to as acrogranin) (Baba *et al.*, 1993a), is similarly ubiquitously expressed in many adult tissues and in germ cells (Baba *et al.*, 1993a).

It should be emphasized that the ubiquitous nature of progranulin gene expression does not invariably imply constitutive gene regulation, but rather denotes a widespread distribution resulting from a multitudinous requirement of progranulin gene function in several contexts. A case in point has been provided by Bateman and co-workers, who were the first to demonstrate a restricted expression pattern for progranulin among epithelia using mRNA *in situ* hybridization: in the rat kidney, epithelial expression is detected in the region of the proximal and distal convoluted tubules, but not in the medulla (Bhandari *et al.*, 1993). In addition, the epithelium of adult epididymis and urothelium, tissues that possess a low mitotic index when compared to epithelium of skin and gut, retain high levels of progranulin gene expression (Daniel *et al.*, 2000). This is in contrast to the positive correlation that exists between the rate of epithelial cell proliferation and the levels of progranulin expression (He and Bateman, 1999). It can therefore be surmised that discrepancies between the cellular localization of progranulin transcripts and the proliferative index of the examined tissue are indicative of a pleiotropic requirement for progranulin gene expression in normal cellular homeostasis. This notion finds support in the detection of progranulin mRNA in immune cells located outside of their normal sites of proliferation, including the lung, spleen and intestine (Daniel *et al.*, 2000).

The expression pattern of zebrafish progranulin-a and progranulin-b in all assayed adult organs (**Figure 31, panel A**) is in agreement with observations made for

the equivalent gene in mammals. Despite this, it is advisable to interpret these data with caution since the use of RT-PCR often cannot distinguish the cellular context of gene expression. Confirmation that tissue expression for zebrafish progranulins does not reflect leukocyte entrapment in various organs requires the localization of mRNAs using *in situ* hybridization approaches. Despite this caveat, we can infer that the duplicated zebrafish co-orthologues are similarly regulated in various adult tissues, but sometimes vary in their relative abundance: progranulin-a expression is more abundant overall than that of progranulin-b (**Figure 31, panel C**).

3.4.2.2 Progranulin-1 and progranulin-2 – The RT-PCR analyses of adult tissues reveal that progranulin-1 and progranulin-2 genes are regulated differently from that of the zebrafish co-orthologues to mammalian progranulin (**Figure 31, panel C**). This may have facilitated their retention within the teleost genome. Specifically, progranulin-1 and progranulin-2 are expressed in a pattern consistent with that reported for their orthologous carp peptides (Belcourt *et al.*, 1993), namely granulin-1 is present in spleen, but granulin-2 is not, lending credence to our observations. Similarly, progranulin-1, but not progranulin-2, is exclusively detected in the heart, at least at the mRNA level. Conversely, progranulin-2 is more abundant in the eyes than progranulin-1. It is possible that progranulin-2 expression in the eyes is derived from peripheral macrophages which are often detected in this region during whole-mount *in situ* hybridization of zebrafish larvae (data not shown).

3.4.2.3 AS progranulin-1/-2 and hybrid (chimeric) progranulin – *AS progranulin-1/-2* and *hybrid progranulin* are both detected in adult zebrafish intestine (**Figure 31, panels B and C**). This suggests a possible role for the presence of the AS transcript in facilitating the formation of the chimeric progranulin mRNA by juxtaposing the primary transcripts for progranulin-1 and progranulin-2. In fact, the latter mRNAs are both detected in the intestine. It is unknown if the presence of the hybrid structure confers an advantage to the fish intestine by subserving a specific role such as controlling proliferation or mediating the innate immune response.

In addition to the intestine, *AS progranulin-1/-2* was found to be weakly expressed in the brain and in the blood (**Figure 31, panels B and C**). Southern blot analysis has revealed more widespread expression for this AS transcript (data not shown), suggesting that it serves the purpose of attenuating *progranulin-1* and *progranulin-2* mRNA translation. More evidence for this activity requires the co-localization of these transcripts in the same cellular context. It is also possible that the AS transcript performs an independent function, as a non-coding RNA gene, whose expression and role can be uncoupled from that of the progranulins (section 3.2.7).

3.4.2.4 Another antisense transcript ? – Northern analysis of RNA isolated from human (Bhandari *et al.*, 1992) and guinea pig testis (Baba *et al.*, 1993a) has indicated the presence of additional progranulin transcripts of smaller size than the full-length mRNA. This does not suggest the occurrence of antisense transcription, since a complementary RNA probe was used for the guinea pig study (Baba *et al.*, 1993a). Therefore, the progranulin gene is suspected to undergo alternative splicing. Consistent with this is the observation that the same tissue detects immunoreactive acrogranin of various forms (Baba *et al.*, 1993a). Other evidence for alternative splicing of the progranulin gene is the expression of a lower molecular weight form referred to as transforming growth factor-e (TGF-e), which is 22 kDa and possesses an amino-terminal sequence identical to that of granulin-A peptide (Parnell *et al.*, 1992). There is no doubt that zebrafish progranulin-1 and progranulin-2 share an antisense transcript. Whether this is a feature unique to the smaller progranulin genes or extends to progranulin-a is not known at present. The existence of three ESTs whose sequences are complementary to the 3'UTR region of the zebrafish progranulin-a transcript would suggest that this may be so (see end of section 3.2.5.9). This issue will be resolved by future Northern blot analysis using cRNA probes and by unidirectional 5' and 3' RACE approaches. This will be important in order to resolve the nature of the signal detected in the intestine and liver of the 5 day-old fish using a sense riboprobe to progranulin-a (**Figure 39**).

3.4.2.5 Summary of adult tissue RT-PCRs – In summary, the zebrafish progranulin genes show a ubiquitous expression pattern within various adult organs; a feature shared with mammalian progranulin. Progranulin-a and progranulin-b are similarly regulated, but progranulin-1 and progranulin-2 show different expression patterns, revealing a peculiar use of the progranulins in fish. The differential expression progranulins-1 and –2 is suggestive of an additional level of functional devolution. This is on top of the recognized phenomenon of duplicated paralogues (-a and –b) that display a co-orthologous relationship to a common mammalian gene (in this case progranulin).

3.4.3 RT-PCR analysis of progranulins expression during zebrafish development

3.4.3.1 Progranulin-a and progranulin-b – Differences in expression of the various zebrafish progranulins were also observed during development (**Figure 32**). Progranulin-a and progranulin-b are maternally expressed, albeit at differing levels (**Figure 32, panel A**). This may reflect initial low levels for progranulin-a mRNA coming from the egg, or its more rapid degradation than progranulin-b subsequent to fertilization. Higher levels for progranulin-b persist even after the onset of zygotic expression (high stage) and at least prior to the commencement of the morphogenetic movements associated with epiboly (sphere-dome stage). This suggests that the early function of the ancestral progranulin gene has been partitioned mainly onto this duplicated paralogue (i.e. progranulin-b). Progranulin-b levels then drop, and progranulin-a levels become slightly more elevated than progranulin-b during gastrulation (shield stage). Both transcripts are equally expressed by the end of epiboly (tailbud) and this trend continues in subsequent stages of development (**Figure 32, panel A**). As was noted previously, tissue levels of progranulin-a mRNA appear to be more elevated than those for progranulin-b in the adult zebrafish. It is tempting to speculate that the reverse situation observed early in embryogenesis underlies a partitioning of function between these duplicates.

Although these expression results are qualitative, they do share features found for their murine orthologues, which is similarly expressed in the early mouse embryo (Diaz-Cueto *et al.*, 2000). In fact, embryonic acrogranin mRNA levels fall rapidly after egg fertilization, reaching negligible levels as early as the 2-cell stage, but rise again to detectable levels by the eight-cell stage. This preceeds the morula stage and subsequent blastocyst stage when the epithelium is first formed. This pattern of expression typically reflects the replacement of maternal mRNAs with embryonically (zygotically) expressed transcripts (Schultz, 1993). The rapid decline and subsequent low expression levels of progranulin-a and progranulin-b after fertilization appear to parallel the observations made in mouse embryos (data not shown). As was mentioned previously, however, it can be observed that progranulin-a levels, unlike progranulin-b, do not appear to rise significantly at the onset of zygotic expression (high stage), suggesting that regulation of expression has been partitioned between the two paralogues relative to the mammalian gene. The shared expression between progranulin-a and progranulin-b in early zebrafish embryogenesis is reminiscent of the pattern often detected between other duplicated zebrafish paralogues (see discussion section 3.2.3.7: e.g. zebrafish *squint* and *cyclops*). The early expression of these genes also suggests a possible role in the control of embryonic growth, as was observed for the mouse counterpart (discussed later, section 3.6.6).

3.4.3.2 Progranulin-1, progranulin-2 and AS progranulin-1/-2 – In contrast with these results, progranulin-1 and/or progranulin-2 was detected at a stage of development when organogenesis is already well advanced (**Figure 32, panel A**). This indicates that these additional genes have been co-opted for additional functions and will not necessarily be redundant with respect to progranulin-a or –b expression. It may be that these additional genes are required for the proliferative phase of organogenesis. Nevertheless, the presence of these additional genes and their divergent expression patterns presumably reflects a key difference between species. Of course, a role for progranulin-1 or progranulin-2 in the early stages of embryonic development cannot be ruled out, since these were detected at very low levels after Southern transfer of the electrophoresed gel (**Figure 32, panel B**), and progranulin-1 is maternally

expressed but rapidly degraded during the early cleavage period (**Figure 32, panel C and data not shown**).

Possible regulation of gene expression by the antisense transcript may also reflect a fundamental difference between the different members of the zebrafish progranulin gene family (**Figure 32, panel B**). Antisense progranulin-1/-2 is detected by 5 days post-fertilization (dpf), and may be necessary for regulating sense gene expression or translation. We find no evidence suggesting a role for AS progranulin in facilitating hybrid progranulin formation (see later).

3.4.4 *In situ* analysis of zebrafish progranulin mRNAs expression in embryo whole-mounts

3.4.4.1 Introduction – The expression of the zebrafish progranulin genes during development using whole-mount *in situ* hybridisations analysed and discussed next. Overall, the expression patterns of zebrafish progranulins faithfully replicate those observed for the mouse counterpart in a similar context (Daniel *et al.*, 2003). For instance, expression was widely detected in the epithelial lining of several ductal structures including those of the pronephric ducts (kidneys) and digestive organs. Other conserved sites of expression include the developing central nervous system, sprouting angiogenesis, the skin epidermis, and apparently the fin bud ectoderm.

However, the observed expression patterns for progranulins were not always congruent between mouse and zebrafish. A notable discrepancy is the finding of expression of mouse progranulin in the blood islands that define primitive hematopoiesis, and the apparent absence of transcripts for the zebrafish co-orthologues within this animal's equivalent structure, the intraembryonic intermediate cell mass (ICM).

A more direct comparison with the murine expression awaits sectioning of zebrafish whole-mounts processed for progranulin expression, which will allow for a more complete histological analysis of gene expression. Issues, such as the apparent absence of expression in the zebrafish epicardium unlike the mouse counterpart (Daniel *et al.*, 2003), will be resolved in this way. Whenever possible, efforts have been made to compare the noted similarities and differences. Nevertheless, our results suggest the likelihood of uncovering important conserved functions for progranulin(s) during vertebrate development using the anamniote zebrafish as model.

In the following section, *progranulin-a* and *progranulin-b* are discussed first and follow a logical progression based on standardized developmental transitions. As development proceeds, the expression of these genes is discussed in terms of developing organ systems. This section ends with a brief discussion concerning the expression of *progranulin-1* and *progranulin-2* and of their corresponding antisense gene.

3.4.4.2 Progranulin-a and progranulin-b

3.4.4.2.1 Cleavage – The detection of *progranulin-a* and *progranulin-b* in all blastomeres, as revealed by *in situ* hybridization in whole-mounts, is congruent with the maternal deposition of these transcripts (**Figure 33**). Differences in the relative intensity of the signals during the staining reaction were noted: progranulin-b transcript detection is achieved sooner and is more intense (not shown). Although the ubiquitous expression for these genes persists subsequent to zygotic transcription and remains so until the end of epiboly (**Figure 33, panels C,** and data not shown), it remains possible that subtle regionalizations do occur by the blastula period. This possibility is discussed in detail next in the context of the successive embryonic transitions observed subsequent to zygotic onset of transcription in the zebrafish.

3.4.4.2.2 Mid-blastula transition to late blastula – At 3hpf, defining the mid-blastula transition, the global synchrony of mitosis observed at earlier stages of development is gradually lost and the embryo is reorganized into three domains (Kane *et al.*, 1992). Although this transition is governed by changes in the nucleo-cytoplasmic ratio of individual blastomeres (Edgar *et al.*, 1986; Kane and Kimmel, 1993), this event is important to consider in light of the fact that it presages early commitment steps during development (Foe, 1989).

In the blastoderm, these domains or regions are morphologically recognized as an outer enveloping monolayer of cells (**EVL**), that surrounds the more loosely organized deep cell layer (**DEL**), and a layer of cells located at the yolk-blastoderm margin that collapse into the yolk cell, the yolk syncytial layer (**YSL**) (Kimmel *et al.*, 1995; Kane, 1999). The YSL adopts the most rapid cell cycle, while the EVL is the slowest. Deep cells adopt an intermediate cycle between the two. It is currently not known if there is a regional difference in the expression of progranulins between these three domains, and thus a possible relationship with the differing mitotic indices that emerge by the mid-blastula transition.

Nonetheless, progranulin-b expression appears to be more abundant in the EVL than in the deep cells and the YSL during late blastula - a trend that continues during the epiboly stage (progranulin-a is less intense than progranulin-b, but still detectable) (**Figure 33, panel C**).

3.4.4.2.3 Epiboly and the gastrula – The first morphogenetic movements of the blastoderm are characterized by the epibolic spread of the blastoderm towards the vegetal pole. During this period, extending from 4.5 hpf to 10 hpf (dome to tailbud stage), the three "mitotic" domains interestingly acquire distinct roles. Cells of the EVL, for instance, will form an epithelium that will cover the blastoderm. Expression of the progranulins in these cells is reminiscent of progranulin gene expression in the mammalian placenta. There is stronger progranulin expression in the apical surface of the blastocyst epithelium, the trophectoderm, relative to the inner cell mass population

(Diaz-Cueto *et al.*, 2000). Of course this does not suggest functional equivalence for these tissues between these animal forms – the fish does not form a placenta.

Deep cells contribute to the formation of the primordial germ layers. By approximately 6 hpf, or halfway through epiboly, the deep cells located at the rim of the blastoderm will coalesce to form a structure termed the germ ring. The germ ring is constituted of two layers - the inner *hypoblast* and the outer *epiblast* – and its appearance indicates the commencement of gastrulation. The ectoderm derives from the epiblast, while the mesoderm and endoderm derive from the hypoblast (Warga and Kimmel, 1990; Kimmel *et al.*, 1995).

Axis formation is also first morphologically noticeable at this timepoint. Coincident with the formation of the germ ring is its accentuated thickening on one side of the otherwise symmetrical embryo, a structure called the shield. The zebrafish shield is the equivalent of Spemmann and Mangold's dorsal blastopore lip amphibian organizer (Ruiz I Altaba, 1998; DeRobertis *et al.*, 2000), and of the mouse node (Beddington, 1994; Lemaire and Kodjabachian, 1996; Lu *et al.*, 2001), and is the first morphologically detectable sign of dorsalisation. Thus structured, the fish body masterplan is laid out, and, like every model of vertebrate development, illustrates eloquently the necessary intricate balance between pairs of opposites: First, the cells at the margin of the blastoderm (hypoblast) will form mesodermal and endodermal structures, and the cells located further away towards the surface (epiblast) give rise to ectodermal derivatives (Ho and Kimmel, 1993). This, in turn, is intermingled with establishment of polarity: cells located on the dorsal side of the animal (shield) will give rise to axial structures and will tend to contribute more to anterior (rostral) structures, while cells located on the ventral side of the blastoderm make the paraxial and lateral structures and tend to contribute more to posterior (caudal) structures (Kessler and Gerhart, 1994; Harland and Gerhart, 1997; Kanki and Ho, 1997; Wilson and Hemmati-Brivanlou, 1997; Agathon *et al.*, 2003).

The gastrula period, like that of epiboly, ends at approximately 10 hpf (tailbud stage), except for the ventral side of the tailbud where it is thought to continue (Kanki and Ho, 1997). Interestingly, terminal divisions of differentiating cells start occurring in the gastrula. In the vicinity of the axial midline, birth of several structures including the notochord, floorplate, prospective polster cells (hatching gland), some cells of the somite muscle, and several large motorneurons of the brain and spinal cord (Kimmel et al., 1994; Kane, 1999), can be seen.

An example of the restricted expression of a gene within the hypoblast layer of the shield region was provided by Shh (**Figure 33, panel C**). This contrasts sharply with the diffuse expression for progranulin-a and progranulin-b during this period (**Figure 33, panel C**). The ubiquitous expression pattern argues against an inductive activity for the granulins in general during gastrulation, unless a gradient is demonstrated for its mRNA expression or its translated product. However, this does not exclude the need for progranulin action during gastrulation. Rather, this pattern is easier to reconcile with a permissive function, such as in the control of cell proliferation and survival, and possibly cell motility.

Since cell movements and cell mitosis are typically antagonistic behaviours (Trinkets, 1980), it would be interesting to investigate more closely progranulin expression in the YSL cells, since these cells are mitotically arrested as they pull the rest of the blastoderm towards the vegetal pole during epiboly (Kimmel and Law, 1985; Kane, 1999).

3.4.4.2.4 The early segmentation period – From the tailbud stage (10hpf) to the remainder of the first day of development when the animal has between 8,000 and 10,000 cells, the zebrafish embryonic body plan undergoes subdivisions. One of the most fascinating developmental events during this period is the action of a segmentation clock affecting the paraxial mesoderm in its sequential and rhythmic organization into somites, which form at a rate of about two per hour (Jiang *et al.*, 2000; Pourquie, 2001). In addition, primitive hematopoiesis and vasculogenesis, both

originating from a common structure called the intermediate cell mass (ICM), will be established. This is coupled to the appearance of a beating heart by the end of the day. Other organ systems for which the development will be already laid out include the pronephros (kidneys) and part of the sensory organs of the central nervous system (CNS) and the appearance of the touch sensitive response. Notable exceptions to this general plan include endodermally derived structures of the gut tube, and neural-crest-derived structures of the jaw, which appear later in development.

Progranulin-a and progranulin-b were both detected in several locations in the early segmentation stage embryo – again, *progranulin-a* expression was weaker but generally parallels patterns observed for *progranulin-b* (**Figure 33, panel D**). Interestingly, both appear to be expressed at higher levels at the rostral and caudal ends, relative to the trunk region of the embryo. This pattern is interesting since these are known sites of high cellular proliferation in the growing animal during the segmentation stage: the brain neuroectoderm thickens in rostral regions, and caudally, the notochord and tail organizer (tail mesoderm) extend to elongate the axis of the embryo (Kimmel *et al.*, 1995; Kanki and Ho, 1997). Inductive signals emanating from the tail mesoderm also influence the rhythmic appearance of somites from the paraxial mesoderm (Pourquie, 2001).

Early in segmentation, brain subdivisions are not visible (Kimmel *et al.*, 1995). Strong progranulin expression at this stage may either suggest a role in general brain growth or subsequent patterning of the brain into individual neuromeres (telencephalon, diencephalons, mesencephalon and hindbrain rhombomeres), noticeable by approximately 18hpf (18-somite stage). Further, in zebrafish, neuroectoderm cells are selected to become neuroblasts early in development: the acquisition of neural identity begins in the early segmentation-stage embryo (1-5 somites) (Chitnis and Dawid, 1999). However, progranulin gene expression is not specifically localized within the three bilateral longitudinal domains, commonly referred to as the proneural domains, where neural fate specification occurs (Blader *et al.*, 1997; Higashijima *et al.*, 2000; Appel *et al.*, 2002).

A closer inspection of progranulin-b expression in the 6-somite stage embryo reveals its localization to the hypoblast layer (mesoderm and endoderm) in the trunk region (**Figure 33, panel E**). Both the eye primordium (5-somite stage) and ear primordium (10-somite stage) appear at the segmentation stage (Kimmel *et al.*, 1995). Strong expression in the epithelial layer surrounding the eye primordium is noticeable for progranulin-b, suggesting its involvement in proper eye formation. Expression of progranulin-b in the YSL suggests a role for this gene in the control of metabolism or endoderm formation (discussed later).

3.4.4.2.5 The late segmentation period – Towards the end of segmentation stage, progranulin-a expression is again less intense than that of progranulin-b (**Figure 34, panel A**). Continued expression in the CNS and developing eyes is consistent with the ongoing but attenuated cellular proliferation in these areas.

Progranulin-b was detected on the ventral region of the trunk embryo in the presumptive midline region and paraxial (lateral plate) mesoderm. This is different from the strong axial mesoderm (notochord) expression seen for Shh (**Figure 34, panel A**). Signals originating from the midline are known to pattern the embryo (Kodjabachian *et al.*, 1999), but the significance of progranulin gene expression in the midline region at this stage is unknown. As was pointed out (section 3.3.4), there are expression patterns for the progranulins whose anatomical structures are currently uncertain, and measures are being taken to resolve potential discrepancies. It is possible that progranulin expression is specifically located within the hypochord, one cell type that lies ventral to the notochord in anamniote embryos (Latimer *et al.*, 2002), but of endodermal origin (Lofberg and Collazo, 1997). In zebrafish, the hypochord is detectable at the 9-somite stage as a single row of cells in the dorsomedial endoderm immediately ventral to the notochord (Eriksson and Lofberg, 2000). Rostrally, it can be recognized from the second or third somite, and extends caudally along the trunk to the same extent as the somites. Once formed, the hypochord stays in close association with the notochord, and this axial complex gradually moves dorsally, separating the

hypochord from the endoderm. Interestingly, this is consistent with the pattern of expression for progranulin-b at the late somite stage (**Figure 34, panel B, lateral view**). It was also shown that within the gap between the hypochord and the endoderm, angioblast cells aggregate and start to form the dorsal aorta, which becomes intimately associated with the hypochord (Eriksson and Lofberg, 2000). Interestingly, localized expression of VEGF in the *Xenopus* hypochord has been shown to mediate angioblast migration during formation of the dorsal aorta (Cleaver and Krieg, 1998). Thus, progranulin-a and progranulin-b expression in the hypochord could possibly influence the proper development of this endodermal structure, or similarly to VEGF, participate in the migration or assembly of angioblasts into the dorsal aorta. Providing more support for a role for progranulin in hypochord formation is the continued correlation between the expression patterns for these duplicated genes and the localization of this structure as development proceeds. Later, the size of the hypochord decreases, and, in the pharyngula stage, is still recognizable in the posterior trunk, but has apparently vanished in anterior regions.

A role for *progranulin-b* in the paraxial mesoderm can only be speculated upon. For example, the paraxial mesoderm is known to emit signals that will posteriorise the neural plate not only during gastrulation, but also later during segmentation stages (Woo and Frazer, 1997; Moens and Fritz, 1999). As pointed out earlier, somites are derived from the paraxial mesoderm. Other laterally and ventrally derived mesodermal tissues that could be potentially influenced by progranulin gene expression include pronephric kidneys, the blood and the vasculature (Kimmel *et al.*, 1990). The expression of zebrafish progranulins in these contexts is briefly discussed next.

3.4.4.2.5.1 Pronephric kidneys – Pronephric kidneys develop bilaterally in the vicinity of the third somite pair. Each pronephric duct primordium is usually without a lumen and grows posteriorly during the early part of the segmentation period, and ventrally around the posterior end of the yolk sac extension (Drummond *et al.*, 1998; Serluca and Fishman, 2001). Progranulins are not clearly

detected in the developing pronephric tubules at this stage, preventing a proposal of a direct involvement for these genes in kidney development, at least during segmentation (**Figure 33, panel E and Figure 34**).

3.4.4.2.5.2 Hematopoiesis and vasculogenesis (ICM) – The establishment of hematopoietic stem cells, as well as of endothelial cell progenitors (angioblasts), occurs at the onset of somitogenesis as the lateral plate begins to detach from the somites and differentiates into the splanchnic and somatic mesodermal layers in order to form the first site of hematopoiesis, the intermediate cell mass (ICM) (Al-Adhami and Kunz, 1977; Willet *et al.*, 1999). The intraembryonic ICM, located between the notochord and endoderm of trunk in zebrafish, is analogous to the mammalian primitive blood islands surrounding the yolk sac (Orkin, 1995), and gives rise primarily to embryonic erythropoiesis (Adhami and Kunz, 1977; Weinstein *et al.*, 1996; Willet *et al.*, 1999). However, it is not clear if progranulin gene expression is localized to the converging mesodermal cells during the segmentation stage (**Figure 34**) presaging the formation of the ICM like other better characterized genes in this context such as stem cell leukaemia (SCL) factor (Liao *et al.*, 1998; Gering *et al.*, 1998), and GATA-1 (Amatruda and Zon, 1999). Thus, whether or not progranulins are expressed early enough in a pattern to suggest their involvement in the formation of the ICM *per se,* or in primitive hematopoiesis (embryonic erythropoiesis), remains questionable. A second site of primitive hematopoiesis, distinct from the ICM, arises from the lateral mesoderm in the anterior part of the embryo (Parker *et al.*, 1999; Willett *et al.*, 1999; Hsu *et al.*, 2001). In fact, Herbomel and co-workers demonstrated that macrophages appear early in zebrafish development and originate separately from other hematopoietic cells (Herbomel *et al.*, 1999). In addition to early macrophage precursors, other myeloid cell types are derived from this anterior lateral mesoderm site (Bennett *et al.*, 2000). Again, progranulin gene expression in dispersed macrophages is not detected during the segmentation period.

Nevertheless, during segmentation, zebrafish progranulin gene expression was localized to the ventral surface of the embryo in a pattern that closely resembles the diffuse distribution of the VEGF receptor family member, flk-1 (data not shown) (Liao *et al.*, 1997; Fouquet *et al.*, 1997). This includes expression in the lateral mesoderm (**Figure 33**), although expression could not be unambiguously detected in the bilateral angiogenic cell clusters in the trunk, or the endocardial and cephalic angioblasts in the more rostral regions of the embryo (not shown). This confusion is due to the close proximity between the hypoblastic mesoderm and endoderm during segmentation, and the diffuse ventral progranulin expression covering the entire rostral-caudal length of the embryo (juxtaposed to the yolk surface) (**Figure 34**).

The endoderm becomes morphologically distinctive at about the onset of segmentation as a few disorganized-looking cells that are in contact with the YSL at or near the midline and deep to the main part of the hypoblast (Kimmel *et al.*, 1995; Warga and Nusslein-Volhard, 1999). Progressively, more endodermal cells appear, and form a more compact and orderly looking epithelial-like arrangement. Importantly, the endoderm develops on only the dorsal side (i.e. trunk region) of the embryo, beneath the axial and paraxial mesoderm. Endodermal expression of progranulins was observed during later stages of development, prior to the organization of prospective visceral endoderm into a gut tube (**Figures 35 and 36**) (Warga and Nusslein-Volhard, 1999; Ober *et al.*, 2003). It will be important to carefully analyse the endodermal expression of zebrafish progranulins since their mammalian counterpart is absent from the endoderm of the early mouse embryo (Daniel *et al.*, 2003).

3.4.4.2.5.3 Caudal tail – As was noticed during the early segmentation period, *progranulin-b*, and to a lesser extent *progranulin-a*, were expressed in the caudal region of the tail. This pattern becomes more refined in the tail bud at later stages of segmentation, in a manner suggesting a role for these genes in axis elongation in the posterior region of the embryo (**Figure 34, panel B**). This pattern was maintained at 24 hpf (**Figure 35, panels A and B**), but was no longer noticeable at 36

hpf (**Figure 36, panels A and B**), coinciding with the gradual disappearance of the tail bud by 42 hpf (Kimmel *et al.*, 1995).

Whether the caudal expression of progranulins can be attributed to the notochord (mesoderm) or the hypochord (endoderm) at the midline is not certain. The lateral stripes, however, suggest lateral plate mesoderm expression (**Figure 34, panel B, dorsal view**). It is also possible that this caudal expression was localized to Kupffer's vesicle, a transient structure that is found only in teleost embryos, and that makes its appearance in the tailbud at the 16-somite stage (17hpf) (Kimmel *et al.*, 1995). Fate mapping studies indicate that the epithelial cells lining this vesicle will later form tail mesoderm derivatives, including notochord and muscle (Melby *et al.*, 1993).

3.4.4.2.5.4 Yolk syncytial layer – Progranulin gene expression in the yolk syncytial layer itself denotes a presence within the extraembryonic endoderm. The specific and selective localization of progranulin-b expression on the yolk cell surface (YSL) (**Figures 34 to 39**) may also suggests a role for this gene in lipid metabolism (Poupard *et al.*, 2000), and also highlights a possible partition of function between progranulin-a and progranulin-b.

3.4.4.2.6 The pharyngula and hatching periods (24 to 72 hpf) – During the pharyngula period, clear qualitative distinctions between progranulin-a and progranulin-b expression were seen.

With regional brain morphogenesis completed and sculpted into five lobes, progranulin-a and progranulin-b expressions were more easily detected in distinct areas within the anterior CNS of the 1 day old zebrafish, presumably reflecting the epithelial lining of the brain (**Figure 35, panels A and B**). Interestingly, progranulin-b expression appeared to be more intense in regions of the midbrain-hindbrain boundary, the tegmentum (ventral midbrain), and the telencephalon, while intense progranulin-a expression was located within the tectum (dorsal midbrain). The significance of these

regional differences is at present not understood, but might involve a participation in proliferation or selective apoptosis within different areas of the brain. For instance, at 24 hpf, the tectum is composed of a single zone of mitotic cells called the ventricular zone (Stuermer, 1988). In addition, the optic tectum contains the highest number of apoptotic cells of any brain region (Cole and Ross, 2001).

In the eyes, progranulin-a and progranulin-b were similarly expressed in the epithelial lining surrounding the eye, and in the lens, which has detached from the epidermis by this stage (Schmitt and Dowling, 1996; Malicki, 1999). Due to the strong progranulin-a expression in both the retina and tectum, it is tempting to speculate a role for this gene in the establishment of retinotectal projections (**Figure 35, panel A**). Retinal neurons, specifically the ganglion cells, project axonal outgrowths that navigate through the middle of the diencephalon into the dorsal part of the midbrain - the optic tectum. However, it has been observed that axonal projections of the ganglion cells leave the eye between 34 and 36 hpf (Stuermer, 1988). This suggests that progranulin-a expression may be involved in the proliferation or survival of cells located in the eyes and tectum of the 1 day-old embryo, rather than in axon pathfinding *per se*, since its expression in this locale is not clearly detected at 36 hpf (**Figure 36, panel A**). In the retina, for instance, the expression of progranulin-a at 24 hpf might help prevent cell death which peaks only at 36 hpf: the timing of retinal ganglion axonal outgrowth (Cole and Ross, 2001). In this sense progranulin-a expression could ensure a regulated onset of apoptosis, which is presumed to be necessary to create space for the axons to navigate outside of the retina. In the lens and in the future cornea (the overlying epithelium), in contrast, the number of apoptotic cells peaks at 24 hpf (Cole and Ross, 2001). This peak correlates with the timing of lens/ectoderm separation, and progranulin-a may in this context, play a role in separation of the lens from the skin ectoderm (Ishizaki *et al.*, 1993) and to anucleate the lens (Wride, 2000). Thus, it remains to be explored whether progranulins participate in or counteract morphogenetic apoptosis (Glücksmann, 1951). However, providing support for a role for progranulin-a in tissue growth is the concomitant high expression of proliferating

cell nuclear antigen (PCNA), a marker for mitotic cells, in the tectum and eyes at this stage.

3.4.4.2.6.1 Pectoral fins – Progranulin-a and progranulin-b are expressed early in the pectoral fin buds (**Figure 35**). Similarly, progranulin gene expression has been noticed in the ectodermal surface of the limb bud of the mouse (Daniel *et al.*, 2003). Whether the zebrafish co-orthologues are similarly located to the apical ectodermal region of the developing fins and excluded from the underlying mesenchyme is currently being investigated. This can be directly compared to Shh expression (mesoderm) in the same developing structure (**Figure 36**).

Expression on the ectodermal portion of the developing fin/limb bud would suggest a possible involvement of progranulin gene function in apical ectodermal ridge (AER) formation or outgrowth. Experiments aimed at monitoring progranulin gene expression in response to retinoic acid treatment in this context might be informative, since progranulin gene expression levels were shown to be elevated in response to all-trans retinoic acid (ATRA) (Chen *et al.*, 2001; Kohroki *et al.*, 2001).

Observations made in the developing fins are currently being compared with analyses of progranulin gene expression in the wound healing/regeneration model of partial caudal fin amputation. For instance, a specific recruitment of progranulin-1-expressing cells at the apical ectodermal surface in the initial days following amputation has been detected (data not shown); a pattern consistent with the involvement of progranulins in the innate immune response and wound repair (Carter and Halper, 1996; Zhu *et al.*, 2002; He *et al.*, 2003). It should be noted that wound healing itself may not require the presence of active inflammatory cells at the site of injury (Martin *et al.*, 2003), and that a clear role for progranulin-1 in this context cannot be evoked with confidence on the basis of expression analysis.

3.4.4.2.6.2 Vasculogenesis and Angiogenesis – The major arteries and veins have begun to form by 24 hpf from endothelial progenitors which are thought to arise from the ICM (Kimmel *et al.*, 1995; Parker *et al.*, 1999). In addition, intersomitic vessels, as well as vessels of the eyes and brain, begin to form through angiogenesis during the pharyngula period. Weak progranulin-a expression was detected between somites in the trunk region at 24 and 36 hpf (**Figures 35, panel A; Figure 36, panel A**) in a pattern that resembles flk-1 (Fouquet *et al.*, 1997). These are presumably angioblasts that are forming the intersomitic vessels. In the head region, progranulin-a and progranulin-b expression was also detected in a pattern predicted to reflect the vascular endothelium (**Figure 36**). It is important to note that progranulin gene expression in the context of vasculogenesis and angiopoiesis is not an exact replica of that observed for flk-1, suggesting either a misinterpretation of progranulin gene expression in this context, or that progranulins are either transiently expressed or present in only a subset of the embryonic vasculature. In fact, the VEGF receptor expression is consistently more pronounced and persistent in all vascular cells as early as the 5- to 7-somite stage until at least 36 hpf (Liao *et al.*, 1997). Also, upon inspection of the reported expression for zebrafish angiopoietin and its receptors (Tie genes), genes known to be required for later stages of endothelial cell differentiation in the mouse (Sato *et al.*, 1995), it can be inferred that endothelial marker gene expression can sometimes be weak and scattered (diffuse), especially on the ventral surface of the trunk (Pham *et al.*, 2001). It is therefore important to consider that progranulin-a and progranulin–b gene expression is apparently not present in the early stages of vasculature formation, implying that these duplicated paralogues may be specifically required in vascular integrity and endothelial cell survival, rather than vascular patterning *per se*. Clearly, more work is needed to clarify this important issue.

3.4.4.2.6.3 Hematopoietic organs – As was discussed earlier, primitive hematopoiesis is located in two areas in the zebrafish: the ICM, which gives rise primarily to embryonic erythropoiesis, and a distinct, more anteriorly located area arising from the lateral mesoderm that gives rise to early myeloid cells including macrophages.

The posterior region of the ICM contains cells that are delayed during differentiation and enter the circulation at a later stage than cells derived from the anterior and trunk domain of the ICM (Thompson *et al.*, 1998). These posterior cells have been proposed to contribute to a second wave of hematopoiesis in the developing embryo, since the ICM eventually wanes by approximately 30 hpf. Primitive hematopoiesis is replaced by a first wave of definitive hematopoiesis located in the ventral wall of the dorsal aorta, as judged by conserved c-myb expression at 48 hpf (Thompson et al., 1998), as well as a region in the tail posterior to the ICM and ventral to the axial vein (posterior cardinal vein) (Liao *et al.*, 1998; Thompson *et al.*, 1998). These sites are the zebrafish equivalents of the para-aortic splanchopleura and the related aorta-gonad-mesonephros (AGM) region of mammals (Medvinsky and Dzierzak, 1996). Starting at 96 hpf, and for the rest of the zebrafish adult life, blood is derived from the kidneys, even if well-developed renal tubules and glomeruli are observed at 72 hpf (Willet *et al.*, 1999). The kidneys are capable of giving rise to additional blood cells, including those of the lymphoid lineage (Medvinsky and Dzierzak, 1996). Although the adult zebrafish kidneys subserve the equivalent function of mammalian bone marrow in hematopoiesis, the zebrafish fetal liver does not appear to participate in blood cell formation (c.f. mammals).

In the pharyngula period and later stages even after the ICM has subsided, but not during segmentation, progranulin-a was expressed in the ventral tail region in a pattern that most closely resembles that for the granulocytic markers myeloperoxidase (MPO) (Bennett *et al.*, 2001), and Pu.1 (Bennett *et al.*, 2001; Lieschke *et al.*, 2002). Pu.1, also known as SPI-1, is a member of the E26 transformation-specific (Ets) family of transcription factors that plays a critical role in mammalian myelopoiesis. Direct evidence for this comes from the observed deficiencies in macrophages, granulocytes, and lymphocytes manifest in Pu-1-deficient mice immediately after birth (McKercher *et al.*, 1996). This is in contrast with the macrophage marker L-plastin, which detects an early wave of cells on the yolk surface of the zebrafish embryo (Herbomel *et al.*, 1999).

In regards to erythropoiesis, progranulin-a expression in the caudal-ventral tail region most closely resembles that of c-myb, a gene that is not required for primitive erythropoiesis in zebrafish, even though it is expressed in these cells (Thompson *et al.*, 1998). This contrasts with the pronounced and continuous expression of other erythropoietic markers also located within the ICM, such as alpha-globin (Brownlie *et al.*, 2003), GATA-1 and GATA-2 (Abdelilah and Driever, 1997; Thompson *et al.*, 1998).

The resemblance between the expression of progranulin-a and of myeloid marker genes, specifically in the ventral tail region, may indicate a regulatory role in the commitment of these progenitors and their continued proliferation (**Figure 35 to 39**). In addition, progranulin-a expression in presumed granulocytes all over the body of the animal is suggestive of a role in the innate immune response of the host (**Figures 36 to 39**).

Taken together, these interpretations suggest a role for progranulin-a in the control of myelopoiesis; a role in erythropoiesis is less certain. A possible role for progranulins in lymphopoiesis has not yet been addressed, since their expression within the zebrafish thymus (starting at 72 hpf) has not been conclusively determined. This possibility is currently being addressed by analysing myeloid marker gene expression – ikaros, rag1 and rag2 (Trede *et al.*, 1998; Trede *et al.*, 2001; Willet *et al.*, 2001) in progranulin-a knockdown experiments.

3.4.4.2.6.4 Pharyngeal and visceral endoderm – Expression of progranulin-a and progranulin-b were both detected in the anterior endoderm, including the pharyngeal endoderm and endoderm that presumably lies in between the posterior pharynx and the foregut region, presaging their possible contribution to jaw and gut tube formation, respectively (**Figures 35 to 37**). Although these patterns need to be confirmed, the diffuse and scattered distribution observed for zebrafish progranulins in the endoderm is suggestive of the presence of the presence of undifferentiated endoderm. A similar broad distribution is seen for some, but not all

molecular markers used to reveal endodermal patterning and organ primordia within the zebrafish digestive system (Wallace and Pack, 2003). For instance, a widespread expression pattern is observed for the pharyngeal endoderm marker foxA2 (Strahle *et al.*, 1993) in the anterior pharynx region after day 1 of development (similar to that observed for the progranulins), that becomes better defined into two broad lateral bands by 34 hpf (Wallace and Pack, 2003). In contrast, several other gut tube markers indicate that, by 24hpf, the endoderm has begun its convergence at the midline in order to form a solid endodermal rod that will progressively give rise to the alimentary canal and contribute to its attached visceral organs (Ober *et al.*, 2003). This is suggestive of a misallocation of our expression analysis for progranulin-a and progranulin-b to the visceral endoderm. In fact, some organs such as the pancreas, do display a dispersed appearance early in their ontogeny, but are usually resolved into well-defined structures by the first day of development (Biemar *et al.*, 2001). However, it was recently discovered that rostral to the gut, the endoderm is not organized into a radial configuration and rather has no apparent histological organization at 34 hpf, and may contribute to the pharynx and oesophagus (Wallace and Pack, 2003). At 58 hpf, the pharynx, oesophagus and developing intestine are contiguous. This is interesting since we no longer detected progranulin expression in a broad diffuse pattern in this area within the 72 hpf-stage animal (progranulin expression between the 48 and 72 hpf timepoints was not investigated). This pattern would suggest a role for progranulins in the formation of the anterior digestive tract, which was recently shown to develop separately from the gut tube in zebrafish, a unique aspect of zebrafish digestive tract morphogenesis (Wallace and Pack, 2003). This finding also provides a possible explanation for *in situ* results indicating the diffuse expression of progranulins-a and –b in the anterior visceral endoderm. This is not to say that zebrafish gut morphogenesis is otherwise significantly different form that observed in mammals. In fact, conserved features include an orderly sequence of gut tube formation: the rostral gut forms first and is located at the level of the fin buds rather than the stomodeum (future mouth), followed by the hindgut (posterior gut or anorectum), and last the midgut.

The common ductular system of the liver and pancreas is contiguous with the digestive tract by 58 hpf in the zebrafish (Wallace and Pack, 2003). Subsequently, during larval development, there is further growth and differentiation of all digestive organs primordia such that the digestive system is functional by 5 days postfertilization (dpf).

Efforts are currently being made to characterize better progranulin expression in these regions. An alternative possibility for the observed expression for progranulins in the area presumed to be the anterior pharyngeal endoderm is the yolk cell surface, whose nuclei express genes generally associated with digestive organs. Nevertheless, because of the observed combined morphological defects in the digestive organs, the branchial arches and the jaw of progranulin-a knockdown animals observed at 5 days post-fertilization (see later – section 3.5), the observed pattern of expression is provisionally ascribed to the endoderm. Concurrently, the issue of whether progranulin-a or -b is specifically expressed within the anterior arches (mandibular and hyoid), which form the jaws and operculum, and/or the posterior arches (branchial) that contribute to gill formation is being examined.

3.4.4.2.7 Larval stage (5dpf) – After 5 days of development, progranulin-a and progranulin-b were observed to be both expressed in the digestive organs, including the liver, intestine and pancreas (**Figure 39**). Expression in the swim bladder was less certain. Sectioning of these animals has revealed that the expression is localized to the epithelial lining of these organs (data not shown). Digestive epithelium is derived from gut (visceral) endoderm, whereas the adjacent mesenchyme, which is derived from lateral plate mesoderm, is constituted of muscle and connective tissue surrounding the epithelial parenchyma (Wallace and Pack, 2003).

Expression of these progranulin genes was detected in the lumen of the pronephric ducts, indicating a possible role in the regulation of the growth of this organ or participation in controlling definitive hematopoiesis (see above).

The significance of sense riboprobe signal detection in the intestine and pronephros is not known at present, but may indicate the presence of antisense transcripts to zebrafish progranulin-a (**Figure 39, panel A**). The brain expression appeared diffuse and may represent background. Attempts to confirm the authenticity of the three deposited EST sequences (accession CD585963, CD586001, CD588938) derived from adult zebrafish kidneys and corresponding to the reverse complement sequence of the 3'end of zebrafish *progranulin-a*, which would confirm antisense transcription for this gene, are currently underway. If confirmed, it will be interesting to determine if this feature is conserved for mammalian progranulin, since it may have repercussions in the selective upregulation of this gene in neoplastic states (see section 3.2.6.6).

3.4.4.3 Progranulin-1 and progranulin-2 – Evidence is provided that *progranulin-1* and *progranulin-2*, which have not been detected in the tetrapod lineage, are uniquely regulated during zebrafish development and are specifically localized in the posterior region of the alimentary canal of zebrafish (proctodeum) (**Figures 38 and 40**). The possibility that this expression might be localized to the fused domain of the pronephric ducts, whose cavity is also open to the external environment, is being considered. Indeed, during the course of their ontogeny, the growing pronephric tubes reach the ventral midline site near the future anus where the bilateral pair of ducts will join together (Kimmel *et al.*, 1995; Drummond *et al.*, 1998). Subsequently, the lumen appears, and eventually the ducts open to the outside. The consequence of *progranulin-1* expression and of the transpliced chimeric transcript in this area is unknown (*progranulin-2* mRNA is also expressed, but to a lesser extent). These genes may perform an antimicrobial function, participate in definitive hematopoiesis (kidney), or control epithelial cell proliferation in these organs (**Figure 38**). In contrast, progranulin-2 is more often detected in dispersed cells presumed to be macrophages based on the reported immunolocalization of granulin-1 peptide to these cells in carp (Belcourt *et al.*, 1995).

The analysis of the expression of these genes, as well as for progranulin-a and –b, has not been sufficient to determine whether these are localized specifically within all the structures of the zebrafish kidney (pronephric ducts, pronephric tubules, and glomerulus), or in the zebrafish equivalent of the adrenal cortex, called the "interrenal". Indeed, the interrenal glands were recently localized in zebrafish using molecular probes measuring P450scc (CYP11A1) gene expression, and shown to undergo a developmental process parallel to that of the pronephros, but under the influence of different signalling events (Hsu *et al.*, 2003). In the adult zebrafish, the mammalian adrenocortical equivalent of zebrafish is embedded in both lobes of the head kidney.

Determining the exact localization of the zebrafish progranulin genes in the zebrafish kidney will offer more insight about its developmental role within these structures. This is particularly important since adrenal deficiencies are often associated with problems arising during their development. The adrenal cortex is known for its engagement in the synthesis of corticosteroids, which in turn are secreted to control homeostasis of glucose and electrolytes (Jones and Bellamy, 1964; Dallman *et al.*, 1989). Knockout mice deficient in steroidogenic genes like STAR and P450scc have been generated to model the molecular mechanisms of adrenal insufficiencies (Caron *et al.*, 1997; Hu *et al.*, 2002; Ishii *et al.*, 2002). Despite this, little is known about the early events leading to the formation of functional adrenal glands (Hsu et al., 2003).

Interestingly, *progranulin-1* and *progranulin-2* are differentially regulated, as exemplified by their distinct expression pattern in the 5 day-old fish (**Figure 40**): while *progranulin-1* is expressed in the epithelial lining of the gut and kidneys (not shown), *progranulin-2* is confined to the presumed proctodeum and in peripheral leukocytic cells.

3.4.4.4 Chimeric progranulin and AS progranulin-1/-2 – Although the use of a corresponding sense riboprobe for the chimeric transcript did not result in *AS progranulin-1/-2* transcript detection, suggesting specificity of our detection (**Figure 40, panel C**), attempts are being made to confirm the presence the hybrid progranulin transcript by RT-PCR. *AS progranulin-1/-2* expression was detected in the liver and pronephros (**Figures 40 and 41**), in a location that is not consistent with its role in the formation of a trans-spliced granulin transcript. Its restriction to the middle portion of the pronephros, and not to anterior and posterior extremities of this organ may reflect antisense-mediated translational repression of the *progranulin-1* transcript or evidence for some other unknown function for this non-coding RNA gene. The transcriptional relationship of the sense and antisense genes is currently being examined using a double *in situ* hybridization approach according to a published protocol (Jowett, 2001).

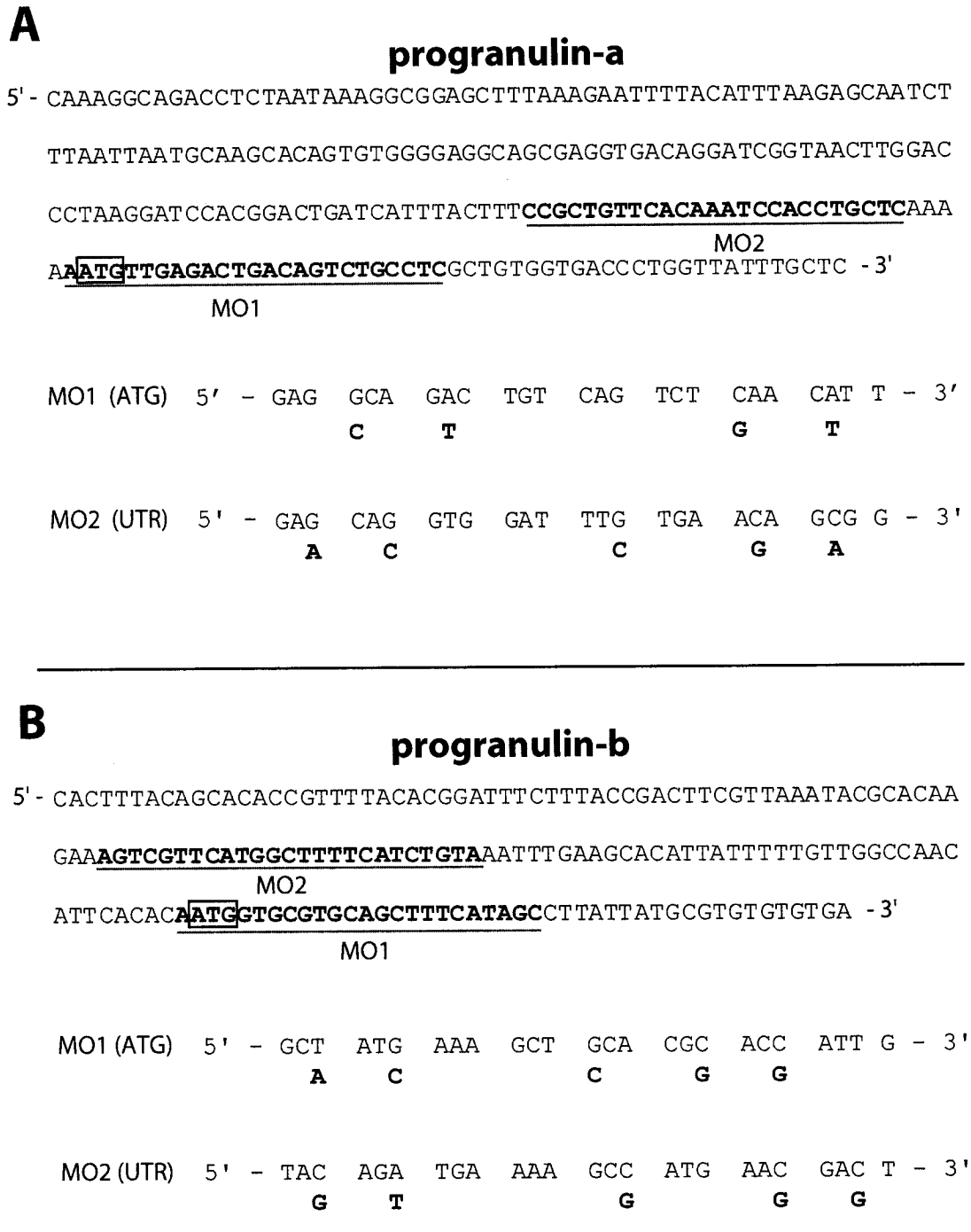


Figure 42. **Sequence of morpholinos used for individual zebrafish progranulin gene knockdown.** 5'UTR sequences used for the design of individual morpholinos were verified by 5'RACE (progranulin-a) and deposited EST sequences (progranulin-a and progranulin-b). *Panel A*: Location and sequence of morpholinos used for progranulin-a knockdown. *Panel B*: Location and sequence of morpholinos used for progranulin-b knockdown.

3.5 Progranulin-a gene knock-down: results – For gene knock-down studies, morpholinos targeting the 5-UTR region of *progranulin-a* and *progranulin-b* transcripts, respectively, as well as corresponding mismatch controls, were designed (**Figure 42**). Injections into the yolk of 1- or 2-cell stage embryos were controlled for potential non-specific developmental defects resulting from toxicity using a standardized MO whose sequence displays complementarity to no known zebrafish transcript. In all stages examined, injection of this "scramble" MO leads to no observable defects at doses as high as 10 ng per embryo (**Figure 43, panel A, left**). Injections of 15 ng of scramble MO per animal leads to mild brain necrosis and slight developmental delay (data not shown). Embryos injected with a MO targeting the *chordin* transcript develop with the characteristic ventralized phenotype (**Figure 43, panel A, right**) with the severity ranging from mild to severe in a dose-dependent manner as was previously reported (Nasevicius and Ekker, 2000). Having established the permissible dose range of morpholino per embryo before the appearance of toxicity effects, the efficacy of progranulin-A knockdown was assessed by whole-mount immunohistochemistry. Using a polyclonal antibody raised against a synthetic peptide whose sequence corresponds to part of the progranulin-A polypeptide but that shares no conservation with progranulin-B (see methods section 2.21), the presence of immunoreactive progranulin-A in MO2-injected vs scramble MO-injected animals at blastula stage was monitored (**Figure 43, panel B**). The use of progranulin-a MO1 gave similar results, but injection of the control mismatch MOs occasionally lead to the disappearance of immunoreactive progranulin-A in a manner less dramatic than that observed for MO1 or MO2 (data not shown – see section 3.6.1 for discussion).

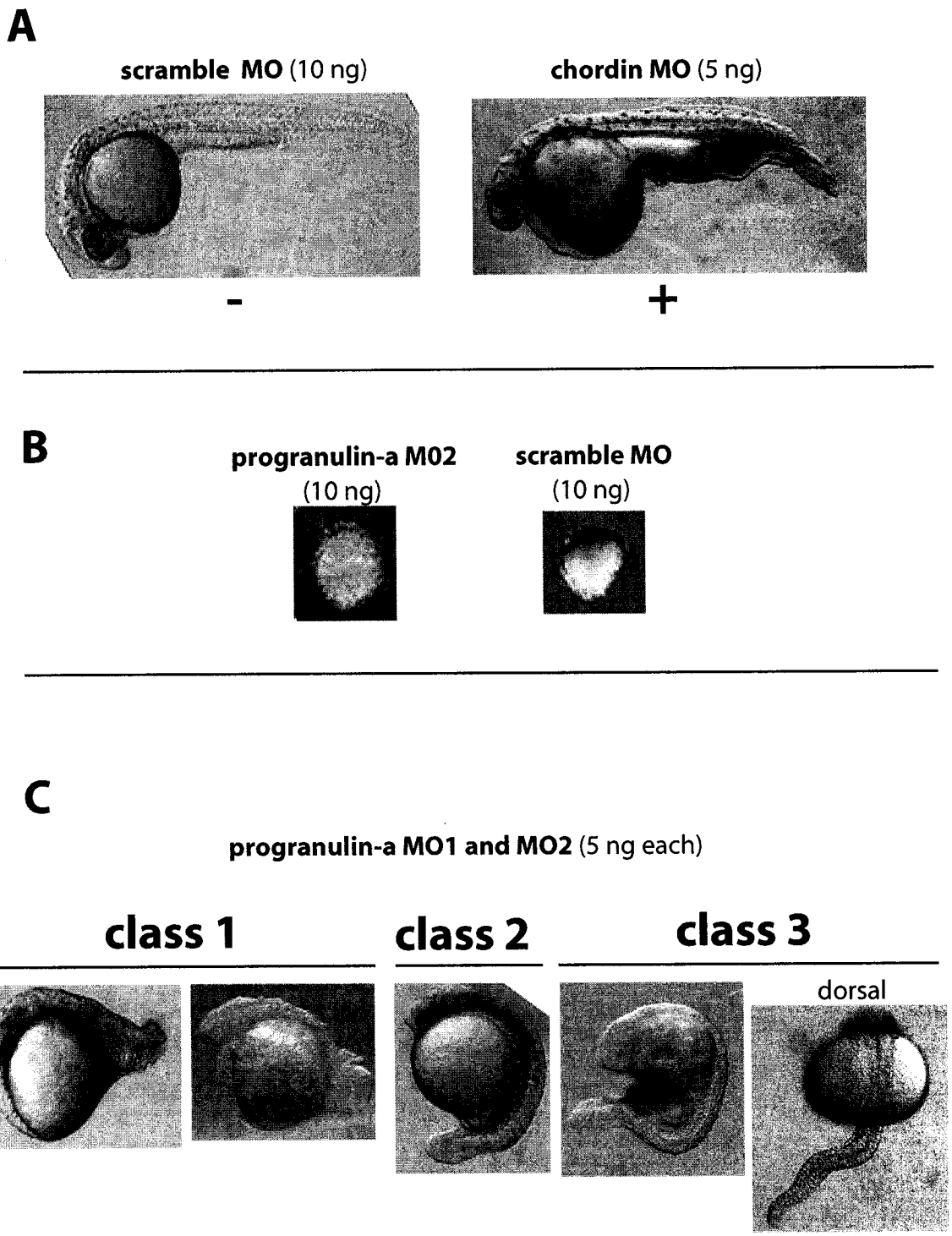
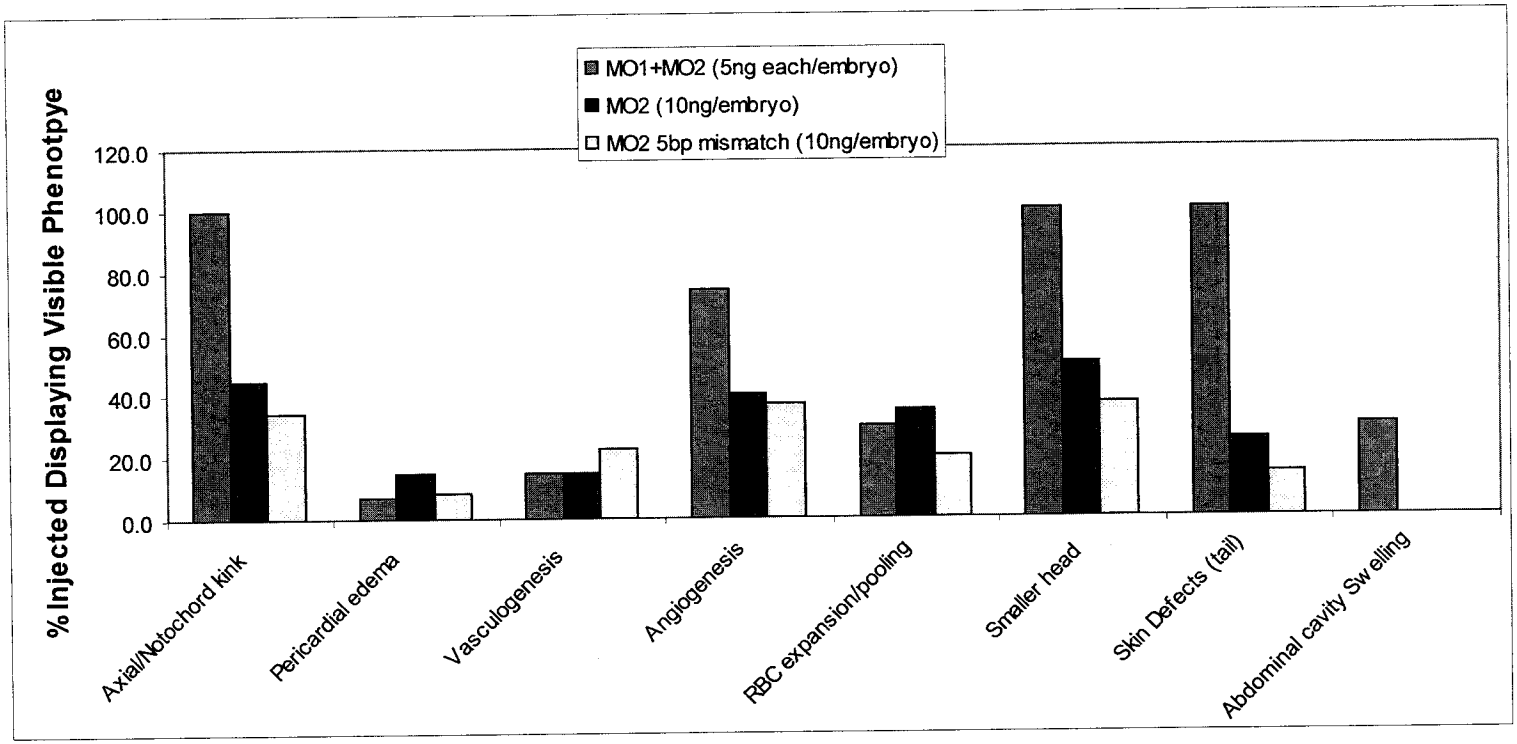

A
scramble MO (10 ng)
chordin MO (5 ng)
-
+
B
progranulin-a M02
(10 ng)
scramble MO
(10 ng)
C
progranulin-a MO1 and MO2 (5 ng each)
class 1
class 2
class 3
dorsal

Figure 43

Figure 43. **Progranulin-a gene knockdown phenotypes.** *Panel A*: Negative (scramble sequence) and positive (chordin) morpholino (MO) controls used in the validation of progranulin-a gene knockdown. Doses up to 10 ng of scramble MO do not result in observable morphological defect (left). Embryos injected with chordin MO display a typical ventralized phenotype with reduced dorso-anterior structures (head and eyes) and expanded ventral regions (e.g. blood islands). *Panel B*: Progranulin-a knockdown efficacy evaluated by whole-mount immunohistochemistry. Note that progranulin-a MO2, but not the scramble control, inhibits progranulin-a mRNA translation. *Panel C*: Range of observable phenotypes observed for progranulin-a knockdown by 24 hpf. Embryos dying prior to 24 hpf are not included in these categories, and define the presumed extreme consequence of progranulin-a knockdown. Surviving embryos fall into three categories, based on decreasing severity: smaller heads and severe axis truncation, lack of somites, and cellular adhesion defects (class 1); truncated axis but presence of somites (class 2); axis of normal length but with kinks, reduced size, a granular appearance, and a lack of proper CNS development (class 3). Phenotype severity and penetrance varies with dose of morpholino injected.

Although the data is not yet strengthened by rigorous statistical analysis, injections of the doses presented below were performed several times (n >4 each) on hundreds of embryos. However, reproducible and consistent results were obtained. In all cases, the scramble MO was used as negative control. Overall, although MO1 is less efficient than MO2, both morpholinos to *progranulin-a* give rise to identical phenotypes with varying penetrance when injected individually at 10 ng per embryo, providing evidence for the specificity of the results obtained. Typically, mild defects observable at approximately 24 hpf included a truncated axis with slight kinks in the trunk and tail region including the notochord, a smaller head and eyes, and occasional skin defects on the surface of the embryo. These defects persisted and sometimes worsened as development proceeded such that, by 48 hpf, the absence of head and trunk vasculature together with complementary blood pooling were noticed (data not shown). A graphic representation of typical phenotypic observations made at 48 hpf is found in **Figure 44**.

The combined injection of MO1 and MO2 at 5 ng each per embryo resulted in a dramatic increase in the severity and penetrance of the phenotypes after one day of development. Although a significant number of embryos died by the early- to mid-segmentation stage (11-15 hpf) using these conditions, those surviving the first day of development often displayed severe truncated axes, small heads with dark cellular masses especially in the prospective eye regions and a granular appearance of the skin surface suggesting cell adhesion defects. Other phenotypes included a lack of somitic boundaries, developmental delay, and prominent oedema of the pericardial sac. A range of phenotypic consequences typically observed in a single clutch of injected embryos is presented in (**Figure 43, panel C**). To facilitate phenotype analysis of surviving embryos by 24 hpf, these have been categorized into three classes (class 1-3) based on severity; class 1 being the most severe (**Figure 43, panel C, bottom**). Typically, animals that received 10 ng of either MO1 or MO2 alone, instead of the combined dose of 5 ng of each MO, displayed morphological defects falling in the class 3 category (not shown).



	MO1 and MO2 (5ng each/embryo)	**MO2** (10ng/embryo)	**5-mm MO2** (10ng/embryo)
Axial/Notochord kink	100.0	45.0	34.3
Pericardial oedema	7.4	15.0	8.6
Disrupted Vasculogenesis	14.8	15.0	22.9
Disrupted Angiogenesis	74.1	40.0	37.1
RBC expansion/pooling	29.6	35.0	20.0
Smaller head	100.0	50.0	37.1
Skin Defects (tail)	100.0	25.0	14.3
Abdominal cavity Swelling	29.6	0.0	0.0
	n = 28	n = 14	n = 35

Figure 44

Figure 44. **Graphic representation of the progranulin-a knockdown phenotypes observed at 48 hpf**. Phenotypes are based on morphological criteria, and vary in severity between injected animals. The combined use of MO1 and MO2 leads to increased phenotype penetrance relative to the use of each morpholino singly. Results obtained with the use of a control 5-nucleotide mismatch based on progranulin-A MO2 sequence are shown. Note the overall reduced number of affected animals in the majority of criteria evaluated. The use of the scramble control leads to no observable phenotype (not shown). Prominent defects are axis bifurcations and kinks in the notochord, a lack of normal blood circulation in the intersomitic vessels (angiogenic defects), reduced anterior structures (head and eyes), skin defects in the distal region of the tail, and swelling of the abdominal cavity. Vasculogenic properties denote the main vessels of the vasculature, namely the dorsal artery and the ventral pericardinal vein. Among the double MO1 and MO2-injected animals, embryos that died prior to 24 hpf were not considered in the analysis (n= 5). Number of injected animals: 5 ng each MO1 and MO2, n = 28; 10 ng MO2, n = 14; 10 ng 5-MM MO2, n = 35.

Interestingly, embryos that had little or no obvious morphological abnormalities during the first 3 days of development displayed several phenotypes by 5 dpf (**Figure 45**). These included dramatic pericardial oedema and a severely perturbed vasculature, reduced anterior structures including small and immobile eyes, overall craniofacial dysmorphogenesis, smaller pectoral fins that are often kinked at an angle, and most notably, swelling of the body cavity with a lack of well-defined visceral organs. Again, these defects were generally weaker and less penetrant for animals receiving either MO1 or MO2 alone (**Figure 45, panel A**), when compared to animals injected with an equivalent dose of both MO1 and MO2 (**Figure 45, panel B**).

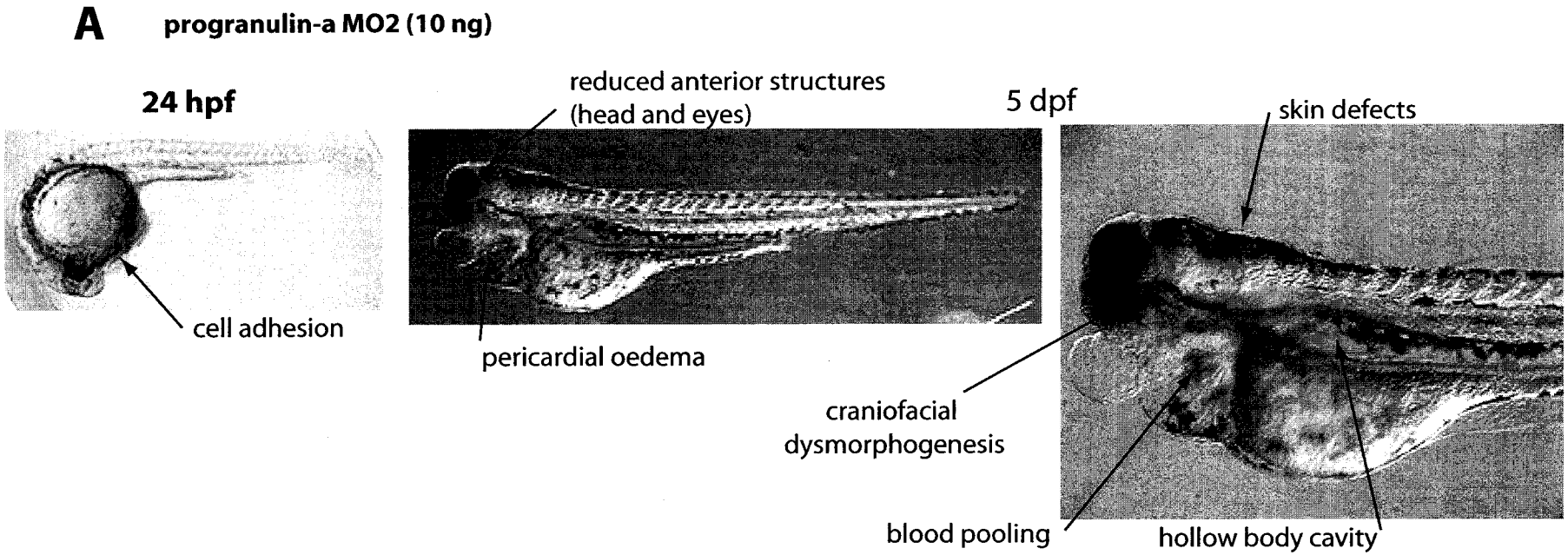

A
progranulin-a MO2 (10 ng)
24 hpf
reduced anterior structures
(head and eyes)
5 dpf
skin defects
cell adhesion
pericardial oedema
craniofacial
dysmorphogenesis
blood pooling
hollow body cavity
B
progranulin-a MO1 and MO2 (5 ng each)
distended visceral cavity (ill-defined organs)

Figure 45

Figure 45. **Progranulin-a knockdown phenotypes at 5 dpf.** *Panel A*: Injection of 10 ng of single morpholino (MO2) give rise to no discernable or mild phenotypes at 24 hpf (see text and Figure 43). By 5 dpf, several morphological abnormalities can be observed: Pericardial oedema, pooling of blood, reduced anterior structures, craniofacial defects, reduced body size, and loose cells on skin surface, occasional pectoral fins dysmorphogenesis. Note that no such phenotypes are observed using 10 ng of scramble MO (not shown). *Panel B*: Co-injection of progranulin-a MO1 and MO2 at a final dose of 5 ng each (total 10ng) leads to increased severity and penetrance of phenotypes described in panel A. In addition, these animals display a distended visceral cavity with ill-defined or absent internal organs.

3.6 Progranulin-a gene knockdown: discussion

3.6.1 Introductory remarks: validating the phenotypes – A programme has been instituted with the goal of defining the functional characterization of individual progranulin genes during zebrafish embryonic development and organogenesis, the interpretation of which is predicated on the proper localization of progranulin gene expression during development. Presently, the focus is on the clear orthologue of mammalian progranulin, namely zebrafish progranulin-a. Since a technique exploiting homologous recombination for the generation of germline transmission has not been perfected in the zebrafish, a widely used technique that employs morpholine-based oligonucleotides for the targeted translational inhibition of selected mRNA populations has been adopted (Nasevicius and Ekker, 2000). In zebrafish, morpholinos appear to be effective in gene knock-down for at least the first two days of development (Ekker and Larson, 2001), allowing for the analysis of the effects of expression of a given gene upon virtually all developmental processes. An interesting feature of this technique is that it allows for a direct appreciation of the various dose requirements of a factor in time and at any given location during the growth and modelling of an embryo, tissue or organ. Thus, phenotypic consequences ranging from the wild-type condition to those reflecting null alleles for a given gene, while progressively revealing all the possible hypomorphic conditions can be expected using this technique. This is a boon to the researcher if phenotypic severity can be evaluated by increments, or a daunting task if the target is truly pleiotropic in nature. In this case, several rigorous controls are required to ascertain whether observed phenotypes are a direct consequence of targeted gene knockdown or in the realm of non-specific targeting or toxicity. For this reason, the use of this technique has often been limited to the direct validation of candidate genes defined by mutations, a process called phenocopying, or in instances where the null condition of the orthologous murine allele is already known. Nevertheless, progranulin gene expression during murine development supports a role for this gene in angiogenesis, and in CNS and epidermis development (Daniel *et al.*, 2003), and this information will assist in the interpretation of phenotypes resulting from progranulin gene knockdown.

Key controls consist of the use of morpholinos whose sequences do not recognize any other known zebrafish mRNAs, and that should not lead to any observable effect (Headsman, 2002). Some investigators have reported widespread cell death (Braat *et al.*, 2001), defects in epiboly (Imai and Talbot, 2001), and neural degeneration (Nasevicius and Ekker, 2000) with the use of morpholinos at the high end of concentrations ranging from ~0.1-0.8 mM; effects that would not be expected from genetic mutants in which the genes are completely inactivated. For comparison, the concentration employed for the progranulin-a knockdown is at 0.3 mM. It is therefore important to consider that the effective dose for progranulin-a knockdown may be close to the toxic dose. In addition, the use of a morpholino whose sequence differs by only 4 or 5 nucleotides from the targeting sequence is considered to be a rigorous control; however, a sufficient amount of base-pairing can be effectively achieved leading to translational inhibition, albeit at reduced efficiency. The extent of this depends on the annealing temperature of any given morpholino and its mismatch counterpart, and is dose-dependent. As an example, in the case where high picogram - low nanogram amounts of MO are effective in generating an observable phenotype, the use of a control morpholino with as little as one mismatch may be sufficient (Marc Ekker, personal communication). Another level of assurance of data quality is the comparable achievement of phenotypic consequences using two morpholinos of unrelated sequence but specifically targeting the same transcript.

Importantly, at all doses assayed, it was noticed that progranulin-a knockdown using MO2 was more efficacious than MO1 (the MO structures are depicted in **Figure 42, panel A**), both in number of affected animals and in their consequent phenotypic severity. It should be noted however, that the observed consequences are the same between the two morpholinos, and these appear to form a continuum when considered from the viewpoint of severity. This is not surprising, since a similar phenomenon is often encountered between independent morpholinos targeting a common transcript, with reported efficacies varying by up to three fold, as in the case with Shh (Ekker and Larson, 2001).

In the case of *progranulin-a* knock-down, the occasional weak penetrance of the class 3 (smaller with an overall granular appearance and kinked axis) or class 4 (pleiotropic phenotypes observed by 5 dpf) categories – was observed with as little as 5-10 ng per embryo of MO1 or MO2 (**Figure 43, panel C; Figure 45**), the majority of which were noticeable at 5 dpf (data not shown). In other words, the use of a lower dose of morpholino resulted in phenotypes difficult to witness at 24 hpf, but easily observable at 5dpf, and include weaker versions of pericardial oedema and hollowed visceral cavity, as well as the occasional appearance of skin blebbing or blistering. The use of a 5-nucleotide mismatch (5-MM) does not lead to such observations. It could thus be argued that these represent the extent of the observable phenotypic consequences for progranulin-a gene knockdown (i.e. classes 3 and 4). However, the observed 10-15% percentage of embryos with these phenotypes is unsatisfactory (n = 4, where n is the number of injection sittings with at least 20 embryos per set), since this technique offers the potential for at least 80% phenotype penetrance (Ekker and Larson, 2001). In fact, the corresponding figure for the chordin MO which is used for the estimation of injection dose and phenotypic penetrance (also a measure of the precision of the experimenter) exceeds 90% with a direct correspondence between calculated dose of morpholino injection and expected severity of phenotype

Embryos injected with a dose of 10 ng of scramble MO were observed to be developmentally unaffected at any given stage (**Figure 43, panel A**). This observation, and those obtained with several other genes whose transcripts were targeted using a similar dose of MO, indicated to us that the embryos could tolerate a higher dose of progranulin-a morpholino. At 10 ng, MO1 and MO2 independently led to an increasing number of animals whose development was severely perturbed (**classes 1-3, Figure 43**), while those injected with the corresponding 5-MM control showed a small number of weakly affected embryos that are morphologically indistinguishable from those injected with 5 ng of MO1 or MO2. Importantly, among the animals that survived the higher (10 ng) dose of a single morpholino, the phenotypes resembled those observed using 5 ng, but with increased severity and penetrance. Thus, these phenotypes appear to form a continuum, lending more credence to these observations.

Morpholinos have been shown to synergise when co-injected (Ekker and Larson, 2001), thus lowering the required dose for generating a higher efficacy of targeted knockdown. This is manifested phenotypically by a maximal number of affected animals displaying a reproducible defect. Using MO1 and MO2 at a combined dose of 10 ng (5 ng each), it was possible to achieve 100% penetrance of the class 1 and 2 phenotypes, with the resulting survivors at 5 dpf displaying the predicted characteristic morphological defects (**Figure 44, panel B**). The minimal combined dose necessary to reproduce these results is currently being assessed. Plans for monitoring the efficiency of translation inhibition using whole-mount immunohistochemistry are in place.

3.6.2 Monitoring knockdown efficacy – When using MO2 singly at 10 ng, the absence of progranulin-A immunoreactivity was noted, while the scramble MO led, as expected, to no change in progranulin protein level (**Figure 43, panel B**). Unfortunately, injections using the corresponding 5-MM control MO led to ambiguous results in this regard and may indicated a varying degree of effective knockdown (data not shown).

Changes in progranulin-A translation subsequent to morpholino injection will also be monitored by Western Blot analysis. A construct encoding the zebrafish progranulin-A ORF followed by the V5 epitope as tag has been prepared and transfected into an insect cell line according to a published protocol (Curtis *et al.*, 2002). Western blot analysis of the recovered material produced from these cells, using either an anti-V5 antibody or the anti-progranulin-A antibody, results in the detection of a protein of approximately 100 kDa consistent with the calculated mass for progranulin-A (91.3 kDa) (data not shown), providing an indirect means of assessing the specificity of the antisera.

3.6.3 Do the phenotypes make sense ? – Overall, the knockdown phenotypes that have been found are consistent with the expression patterns for *progranulin-a.* In light of the known biological activities of mammalian progranulin, namely its involvement in the control of proliferation and its role in cell-adhesion/motility through the activation of focal adhesion kinase (He *et al.*, 2002), we conclude that the observed defects are likely valid.

It will be important to determine to what extent the range of phenotypes observed for progranulin-a knockdown resembles observable defects uncovered through the large-scale mutagenesis screens (as an example, see special issue of *Development* 123, 1996). It will be essential to establish the correct phenotypes that would most closely reflect a null allele for this gene, which would imply that all milder phenotypes are valid. The data accumulated thus far would suggest that the most severe phenotype appears to be early lethality which occurs prior to 24 hpf but after the onset of segmentation (approximately mid-somitogenesis). Embryos dying as a consequence of the injection procedure are rarely encountered, and these usually collapse prior to gastrulation.

The broad range of the phenotypes observed early (class 1-3) (**Figure 43, panel C**), and the pleiotropic defects later uncovered during organogenesis (**Figures 44 and 45**) allow for predictions to be made regarding the requirement for progranulin gene activity during development. Experiments are sufficiently advanced to allow for hypotheses to be made, and these are discussed below: first, an early requirement for progranulin activity might influence proper axis formation and definition of the endoderm from the mesendoderm; second, there is a requirement for progranulin action in the control of proliferation/cell adhesion and possibly cell survival.

3.6.4 Does progranulin activity influence endoderm formation ? – Based on expression data, it is at present not possible to predict a role for progranulins in the formation of zebrafish endoderm during gastrulation with certainty. The specific localization of progranulin expression to the YSL, the latter structure suspected of emitting signals that pattern the mesendoderm (Ho, 1997; Chen, 2000), or exclusively to the marginal zone in the late blastula or early gastrula of zebrafish, has not been conclusively demonstrated. As was mentioned earlier, the ubiquitous expression pattern of progranulin-a is more in keeping with a permissive role during development. Nevertheless, ways in which progranulin activity may affect endoderm formation have remained unexplored. A molecular pathway leading to the formation of the endoderm in zebrafish and other vertebrate models is emerging. Briefly, important players are activin-related growth factors (*squint* and *cyclops* – Nodal) and their receptors (including TARAM-A, a type I transforming growth factor β receptor) and a co-receptor (EGF-CFC or *one-eyed pinhead* (Oep), a member of the Cripto family of epidermal growth factor related proteins). The localized signalling activities of these factors results in the activation of *casanova* (Kikuchi *et al.*, 2001) which in turn promotes the salt and pepper expression pattern of Sox17, a Sry-related transcriptional regulator involved in endodermal differentiation (Hudson *et al.*, 1997; Alexander and Stainier, 1999; Kanai-Azuma *et al.*, 2002; Tam *et al.*, 2003).

Interestingly, in zebrafish at least, Sox17 activation is partially Nodal-independent (Shivdasani, 2002). In light of this and of observed morphological consequences of progranulin-a gene knockdown, it will be interesting to determine whether the maternally expressed progranulins influence endoderm and axis formation. In this respect, preliminary observations indicate a possible change in the stability of beta-catenin in *progranulin-a* knockdown animals; Western blot analysis revealed a lower molecular weight form for this protein compared to control-injected animals (not shown). Although not yet confirmed this activity is in keeping with the known requirement for maternal beta-catenin activity in the axis formation of zebrafish as was demonstrated in the *ichabod* mutant background (Kelly *et al.*, 2000), as well as in mesendoderm formation of the *Xenopus* embryo (Schohl and Fagotto, 2003). Also,

responding to a signal that remains to be identified, beta-catenin accumulates in the nuclei of the mesendodermal progenitors between the 16- and 32-cell stages of sea urchin embryos (reviewed in Stainier, 2002). It thus remains to be demonstrated if progranulin activity is required for the nuclear localization of beta-catenin in vegetally-located blastomeres in the newly fertilized zebrafish embryo, thus leading to the formation of a Nieuwkoop-like centre and subsequent axis formation (Kimelman and Griffin, 2000). Significantly, the activation of beta-catenin occurs in a graded manner from the ventral posterior side to the anterior dorsal of the endoderm, which peaks at late gastrulation (Schohl and Fagotto, 2002). In this way, early progranulin activity would indirectly contribute to the establishment of cell populations capable of responding appropriately to endoderm inducers.

An alternative is the requirement for progranulin(s) in the early formation of the endoderm (see above), or later in its development. In the latter case, the observed endoderm defects would be indirect and result from a requirement for progranulin activity in other locations, such as early progranulin signalling from the notochord/hypochord acting on the underlying endoderm (Kim *et al.*, 1997; Cleaver and Krieg, 2001). It will be necessary to evaluate all phenotypic possibilities using a combination of histochemical techniques and an appropriate battery of marker genes (see methods, sections 2.25 and 2.26), and through the use of complementary *progranulin-a* overexpression studies. Similarly, experiments to rescue the progranulin-a knockdown using the human orthologue have been planned. This may have the added advantage of determining the functional consequence of the sequence divergence of the progranulin gene that took place through evolutionary time.

Uncovering developmental roles for growth factors is not a straightforward process. Again, endoderm formation provides an excellent example. For instance, when the mouse Nodal ligand is disrupted, an early phenotype observed is severely perturbed gastrulation whereby not only the definitive endoderm but other germ-layer tissues are also deficient (Brennan *et al.*, 2001). However, studies in *Xenopus* have shown a requirement for graded Nodal activity in endoderm differentiation of the

mesendoderm (Faure *et al.*, 2000). Subsequently, a more specific impact for mouse Nodal on endoderm formation was revealed in hypomorphic mutants displaying different degrees of loss of Nodal activity. For instance, a greater level of residual Nodal activity leads to the formation of some endoderm, but a greater reduction in activity leads to a loss of endoderm at the expense of favouring mesoderm differentiation (Low *et al.*, 2001). A graded activity of the zebrafish duplicated Nodals, *squint* and *cyclops*, is similarly required in this process – *squint* and *cyclops* compound homozygotes lack endoderm entirely (Feldman *et al.*, 1998).

The progranulin-a knockdown leads to apparent defects that cover a wide range of endodermally-derived structures, including the jaw and visceral organs (**Figure 45**). Interestingly, it has been hypothesized that gut formation occurs non-cell-autonomously, through signals transmitted to the endoderm from the gut mesenchyme, adjacent midline structures, or the YSL (Wallace and Pack, 2003). Significantly, a possible role for axial mesoderm in gut morphogenesis is supported by the observation that *spadetail* mutants, in which axial mesoderm is expanded laterally, can have gut duplications (Verkade H. and Stainier D.Y., unpublished observations, in Wallace and Pack, 2003). In this way, progranulin activity could be provided from mesoderm derivatives and act indirectly on nearby endoderm, thus affecting the formation of the visceral organs.

3.6.5 Communication breakdown: an indirect effect on endoderm-derived structures through impaired cellular adhesion – In addition to the previous discussion, it is known that epithelial polarization is coincident with digestive organ morphogenesis (Leung *et al.*, 1999). The formation of cell-cell junctions is an essential step in separating apical and basolateral membrane domains, and in giving an epithelium its vectorial or polarized property. Cadherins are the major proteins of the adherens junctions, which, via calcium-dependent, homophilic interactions, maintain the integrity of epithelial sheets and separate apical and basolateral membrane domains. Impaired stabilization of beta-catenin at the adherens junctions could be an indirect consequence of a loss of proper cell adhesion in a context of reduced

progranulin activity. This would lead to the pleiotropic phenotypes observed late in development such as the loss of vascular integrity and disorganized epithelium of skin, as well improper formation of the digestive organs and craniofacial dysmorphogenesis (**Figure 45**). It will be interesting to determine whether this phenotype is due to an arrest of morphogenesis and cytodifferentiation of the gut tube in a primordial state, as was recently reported for the zebrafish *nil per os* (npo) mutation (Mayer and Fishman, 2003). This scenario might involve progranulin gene expression in cellular transitions (epithelial-mesenchymal or endothelial-epithelial, such as in endoderm-intestine transitions), via the regulation of cell adhesion (Thiery, 2003). Similarly, Wnt signalling appears to be required for proper differentiation of the gut epithelium (Korinek *et al.*, 1998; Lickert *et al.*, 2000). Similarly, perturbation of cellular adhesion may underlie the defective vasculature observed in progranulin-A-depleted zebrafish embryos. For instance, a signalling pathway elicited by progranulin, namely the activation of FAK, is involved in the proper assembly, and presumably required for maintaining the integrity, of the vasculature (Ilic *et al.*, 2003). It will be important to consider the progranulin-a knockdown phenotypes in relation to phenotypes observed for defects in other signalling activities for this gene in the zebrafish once these are uncovered.

3.6.6 Riding the wave: proliferation and cellular integrity – Exogenous addition of purified recombinant acrogranin to mouse blastocyst explants results in an early onset of the appearance, and stimulated growth, of the trophectodermal layer (Diaz-Cueto *et al.*, 2000). In contrast, the use of a function-blocking antibody to mouse progranulin (acrogranin) delays progression of the mouse embryo to the blastocyst stage (Diaz-Cueto *et al.*, 2000). These studies, along with extensive data accumulated from *in vitro* and *in vivo* experiments (Shoyab *et al.*, 1990; Zhang and Serrero, 1998; He and Bateman, 1999; Lu and Serrero, 2001; Brown Jones *et al.*, 2003; reviewed in Bateman and Bennett, 1998; He and Bateman, 2003; Ong and Bateman, 2003) argue strongly for an early requirement for progranulin activity in the control of cell proliferation during development. In support of this hypothesis and of the observed progranulin-a knock-down phenotypes, studies using *Drosophila* have demonstrated

that mitosis and morphogenesis are not necessarily uncoupled and unrelated events (Groβhans and Wieschaus, 2000).

Of particular interest, during the cell cycle 16 of zebrafish, which occurs during the early segmentation stage, many cells are leaving the cell cycle as part of a major wave of differentiation (Kimmel *et al.*, 1994). Mutants that arrest early in development, such as *zombie*, *speed bump*, and *ogre* first show their phenotypes during this stage (Kane *et al.*, 1996). It has been suggested that many of these mutations occur in genes that are necessary for the cell cycle, since many of these mutant animals produce abnormal nuclei (Kane *et al.*, 1996). If so, it is likely that mitotic arrest is occurring at the end of cycles 15 or 16. Interestingly and consistent with this hypothesis, cells that have their terminal divisions before cycle 16, such as notochord and hatching gland cells, are unaffected in these early arrest mutants. Interestingly, the observed severe morphological defects resulting from progranulin-a knockdown resemble these mutants. To reiterate, these phenotypes consist of early embryo lethality subsequent to the onset of segmentation and presumed anoikis judged by the loose cells detaching from the surface of the embryos and the overall granular appearance (class 1) (**Figure 43, panel C**). These defects are suggestive of mitotic and cell adhesion defects – anoikis is defined as apoptosis resulting from improper interaction between the cell and its extracellular environment (reviewed in Frish and Screaton, 2001). Significantly, work conducted in Dr. Bateman's laboratory showed that progranulin stimulation of SW-13 cells leads to a reduction of cell death resulting from anoikis, and a concomitant increase in the phosphorylation of the cell migration factor focal adhesion kinase (FAK) (He *et al.*, 2002). Efforts are being made to characterize the extent of proliferation (using marker gene expression such as PCNA and immunohistochemistry with a anti-phosphohistone H3 antibody) and of apoptotic death (see methods, section 2.26), as well as the state of cell adhesion (by monitoring the expression of several cadherins), in the progranulin-a knockdown animals.

3.7 Contributions to knowledge

The following experimental observations represent original contributions to science:

1. Characterization of the zebrafish progranulin gene family:

   - Cloning of the cDNA and gene encoding zebrafish progranulin-1
   - Cloning of the cDNA and gene encoding zebrafish progranulin-2
   - Cloning of a cDNA that encodes the zebrafish progranulin-a precursor
   - Cloning of a cDNA that encodes the zebrafish progranulin-b precursor

2. Resolution of the chromosomal location of the zebrafish progranulins, and demonstration that zebrafish progranulin-a is a clear orthologue of mammalian progranulin through conserved synteny.

3. Evidence through comparative genomics that an ancestral progranulin gene was duplicated at the base of the vertebrate radiation and that one copy was subsequently lost in the lineage leading to tetrapods but retained in the lineage leading to teleosts after the split between lobe-finned (sarcopterygians) and ray-finned (actinopterygians) fish.

4. Cloning of a single non-coding RNA gene with antisense complementarity to both progranulin-1 and progranulin-2.

5. Evidence for *trans*-splicing between the primary transcripts for progranulin-1 and progranulin-2 through the characterization of a cDNA encoding a chimeric transcript between the amino-terminal sequence of progranulin-1 and the carboxyl-terminus between progranulin-2.

6. Zebrafish progranulin-a and progranulin-b are widely expressed in adult tissues, but differently regulated. Progranulin-1 and progranulin-2 are expressed in hematopoietic organs in a pattern that is consistent with observations made for the distribution of orthologous peptides in carp. Also, the antisense transcript is primarily expressed in the intestine and blood of the adult fish, and the hybrid progranulin is detected only in the intestine.

7. Zebrafish progranulin-a and progranulin-b are expressed maternally in the embryo like mouse progranulin, but progranulin-1 is only detectable transiently early during cleavage stage, suggesting rapid turnover of this mRNA. Also, progranulin-b is more abundant than progranulin-a in early development indicating a possible devolution of function between these paralogues.

8. Described the ontogeny of expression of the zebrafish progranulin genes in development using whole-mount *in situ* hybridization, and provided evidence for extensive overlap in their expression patterns. Notable divergence in expression was also observed between the duplicated paralogues, notably myeloid expression for progranulin-a and yolk syncytial layer expression for progranulin-b.

9. These studies are the first to provide evidence for pleiotropic requirements of progranulin gene activity (progranulin-a) in vertebrate development, as assessed by preliminary gene knockdown in the zebrafish.

3.8 References

Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. **Evidence of *en bloc* duplication in vertebrate genomes.** *Nature Genet* **31**:100-105 (2002)

Abdelilah S, Driever W. **Pattern formation in janus-mutant zebrafish embryos.** *Dev Biol* **184**:70-84 (1997)

Adelman JP, Bond CT, Douglass J, Herbert E. **Two Mammalian Genes Transcribed from Opposite Strands of the Same DNA Locus.** *Science* **235**:1514-1517 (1987)

Agabian N. **Trans splicing of nuclear pre-mRNAs.** *Cell* **61**:1157-1160 (1990)

Agathon A, Thisse C, Thisse, B. **The molecular nature of the zebrafish tail organizer.** *Nature* **424**:448-452 (2003)

Ahlberg P, Milner A. **The origin and early diversification of the tetrapods.** *Nature* **368**:507-514 (1994)

Akopian AN, Okuse K, Souslova V, England S, Ogata N, Wood JN. **Trans-splicing of a voltage-gated sodium channel is regulated by nerve growth factor.** *FEBS Lett* **445**:177-182 (1999)

Al-Adhami MA, Kunz YW. **Ontogenesis of hematopoietic sites in *Brachydanio rerio* (Hamilton-Buchanana) (*Teleostei*).** *Develop Growth Differ* **19**:171-175 (1977)

Alexander J, Stainier DYR. **A molecular pathway leading to endoderm formation in zebrafish.** *Curr Biol* **9**:1147-1157 (1999)

Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, Volff J-N, Schartl M. **Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish.** *Genetics* 161: 259-267 (2002)

Amatruda JF, Zon LI. **Dissecting hematopoiesis and disease using the zebrafish.** *Dev Biol* **216**:1-15 (1999)

Amaya E, Offield MF, Grainger RM. **Frog genetics: *Xenopus tropicalis* jumps into the future.** *Trends Genet* **14**:253-255 (1998)

Ambros V. **Control of developmental timing in *Caenorhabditis elegans*.** *Curr Opin Genet Dev* **10**:48-433 (2000)

Amemiya CT, Zon LI. **Generation of a zebrafish P1 artificial chromosome library.** *Genomics* **58**:211-213 (1999)

Ames RS, SarauHM, Chambers JK, Willette RN, *et al.* **Human urotensin-II is a potent vasoconstrictor and agonist for the orphan receptor GPR14.** *Nature* **401**:282-286 (1999)

Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L, Westerfield M, *et al.* **Zebrafish Hox clusters and vertebrate genome evolution.** *Science* **282**:711-714 (1998)

Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S. **Organization of the *Fugu rubripes* Hox clusters: Evidence for continuing evolution of vertebrate Hox clusters.** *Nat Genet* **16**:79-83 (1997)

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S. **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*.** *Science* **297**:1301-1310 (2002)

Appel B, Givan LA, Eisen JS. **Delta-Notch signalling and lateral inhibition in zebrafish spinal cord development: Correction.** *BMC Dev Biol* **2**:3 (2002)

Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. **Novel small RNA-encoding genes in the intergenic regions of Escherichia coli.** *Curr Biol* **11**:941-950 (2001)

Atchley WR, Fitch WM, Bronner-Fraser M. **Molecular evolution of the MyoD family of transcription factors.** *Proc Natl Acad Sci USA* **91**:11522-11526 (1994)

Arney KL. ***H19* and *igf2* – enhancing the confusion** ? *Trends Genet* **19**:17-23 (2003)

Ashkenas J, Muschler J, Bissell MJ. **The extracellular matrix in epithelial biology: shared molecules and common themes in distant phyla.** *Dev Biol* 180:433-444 (1996)

Baba, T, Hoff III HB, Nemoto H, Lee H, Orth J, Arai Y, Gerton GL. **Acrogranin, an acrosomal cysteine-rich glycoprotein, is the precursor of the growth-modulating peptides, granulins, and epithelins, and is expressed in somatic as well as male germ cells.** *Molec Reprod Dev* **34**:233-243 (1993a)

Baba T, Nemoto H, Watanabe K, Arai Y and Gerton GL. **Exon/intron organization of the gene encoding the mouse epithelin/granulin precursor (acrogranin).** *FEBS Letters* **322**: 89-94 (1993b)

Bailey JA, Gu Z, Royden CA, Reinart K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. **Recent segmental duplications in the human genome.** *Science* **297**:1003-1007 (2002)

Bailey WJ, Kim J, Wagner GP, Ruddle FH. **Phylogenetic reconstruction of vertebrate Hox cluster duplications.** *Mol Biol Evol* **14**:843-853 (1997)

Bailey JA, Carrel L, Chakravarti A, Eichler EE. **Molecular evidence for a relationship between LINE-1 elements and X-chromosome inactivation: the Lyon repeat hypothesis.** *Proc Natl Acad Sci USA* **97**:6634-6639 (2000)

Bailey MJ, Beremand PD, Hammer R, Bell-Pederson D, Thomas TL, Cassone VM. **Transcriptional profiling of the chick pineal gland, a photoreceptor circadian oscillator and pacemaker.** *Mol Endocrinol* **17**:2084-2095 (2003)

Bakiri L, Lallemand D, Bossy-Wetzel E, Yaniv M. **Cell cycle-dependent variations in c-Jun and JunB phosphorylation: a role in the control of cyclin D1 expression.** *EMBO J* **19**:2056-2068 (2000)

Ball ST, Williamson CM, Hayes C, Hacker T, Peters Jo. **The spatial and temporal expression pattern of *Nesp* and its antisense *Nespas*, in mid-gestion mous embryos.** *Mech Dev* **100**:79-81 (2001)

Barbazuk WB, Korf I, Kadavi C, Heyen J, Tate S, Wun E, Bedell JA, McPherson JD, Johnson SL. **The syntenic relationship of the zebrafish and human genomes.** *Genome Res* **10**:1351-1358 (2000)

Barlow DP, Stöger R, Herrmann BG, Saito K, Schweifer N. **The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus.** *Nature* **349**:84-87 (1991)

Barlow DP. **Competition: a common motif for the imprinting mechanism ?** *EMBO J* **16**:6899-6905 (1997)

Baserga R. **The IGF-1 receptor in normal and abnormal growth.** In *Hormones and Growth Factors in Development and Neoplasia.* Edited by Robert B. Dickson and David S. Salomon. New York: Wiley-Liss (1998)

Bass, B. L. **Double-stranded RNA as a template for gene silencing.** *Cell* **101**:235-238 (2000)

Bateman AJ. **Interpretation of linkage in somatic cell hybrids.** *Nature* **233**:480-1 (1971)

Bateman A, Belcourt D, Bennett H, Lazure C, Solomon S. **Granulins, a novel class of peptide from leukocytes.** *Biochem Biophys Res Comm* **173**:1161-1168 (1990)

Bateman A, Bennett HPJ. **Granulins: the structure and function of an emerging family of growth factors.** *J Endocrinol* **158**:145-151 (1998)

Bates S, Phillips AC, Clarke PA, Stott F, Peters G, Ludwig RL, Vousden KH. **$p14^{ARF}$ links the tumour suppressors RB and p53.** *Nature* **395**:124-125 (1998)

Beddington RS. **Induction of a second neural axis by the mouse node.** *Development* **120**:613-620 (1994)

Bedford M, Arman E, Orr-Urtreger A, Lonai, P. **Analysis of the Hoxd-3 gene: structure and localization of its sense and natural antisense transcripts.** *DNA Cell Biol* **14**:295-304 (1995)

Belcourt, DR, Lazure, C, Bennett, HPJ. **Isolation and primary structure of the three major forms of granulin-like peptides from hematopoietic tissues of a teleost fish (*cyprinus carpio*).** *J Biol Chem* **268**:9230-9237 (1993)

Belcourt DR, Okawara Y, Fryer JN, Bennett HPJ. **Immunocytochemical localization of granulin-1 to mononuclear phagocytic cells of the teleost fish Cyprinus carpio and Carassius auratus.** J Leuk Biol 57:94-100 (1995)

Bell AC and Felsenfeld G. **Methylation of a CTCF-dependent boundary controls imprinted expression of the igf2 gene.** *Nature* **405**:486-489 (2000)

Bell GI, Fong NM, Stempien MM, Wormsted MA, Caput D, Ku LL, Urdea MS, Rall LB, Sanchez-Pescador R. **Human epidermal growth factor precursor: cDNA sequence, expression in vitro and gene organization.** *Nucleic Acid Research* **14**:8427-46 (1986)

Ben-Baruch N, Alroy I, Yarden Y. **Developmental and physiologic roles of ErbB receptors and their ligands in mammals.** In *Hormones and Growth Factors in Development and Neoplasia.* Edited by Robert B. Dickson and David S. Salomon. New York: Wiley-Liss (1998)

Bennett CM, Kanki JP, Rhodes J, Liu TX, Paw BH, Kieran MW, Langenau DM, Delahaye-Brown A, Zon LI, Fleming MD, Look AT. **Myelopoiesis in the zebrafish, Danio rerio.** *Blood* **98**: 643-651 (2001)

Bennett HPJ, Browne CA, Solomon S. **Purification of the two major forms of rat pituitary corticotropin using only reversed-phase liquid chromatography.** *Biochemistry* **20**:4530-4538 (1981)

Benton WD, Davis RW. **Screening λgt recombinant clones by hybridization to single plaques in situ.** *Science* **196**:180 (1977)

Berget SM. **Exon recognition in vertebrate splicing.** *J Biol Chem* **270**:2411-2414 (1995)

Bester AJ, Kennedy DS, Heywood SM. **Two Classes of Translational Control RNA: Their Role in the Regulation of Protein Synthesis.** *Proc Natl Acad Sci USA* **72**:1523-1527 (1975)

Bhandari V, Bateman A. **Structure and chromosomal location of the human granulin gene.** *Biochem Biophys Res Comm* **188**:57-63 (1992)

Bhandari V, Palfree RGE, Bateman A. **Isolation and sequence of the granulin precursor cDNA from human bone marrow reveals tandem cysteine-rich granulin domains.** *Proc Natl Acad Sci USA* **89**:1715-1719 (1992)

Bhandari V, Giaid A, Bateman A. **The complementary deoxyribonucleic acid sequence, tissue distribution, and cellular localization of the rat granulin precursor.** *Endocrinology* **133**:2682-2689 (1993)

Bi W, Deng JM, Zhang Z, Behringer RR and de Crombrugghe B. **Sox9 is required for cartilage formation.** *Nature Genetics* **22**:85-89 (1999)

Bi W, Huang W, Whitworth DJ, Deng JM, Zhang Z, Behringer RR and de Crombrugghe B. **Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization.** *Proc Natl Acad Sci USA* **98**:6698-6703 (2001)

Biemar F, Argenton F, Schnidtke R, Epperlein S, Peers B, Driever W. **Pancreas development in zebrafish: early dispersed appearance of endocrine hormone expressing cells and their convergence to form the definitive islet.** *Dev Biol* **230**:189-203 (2001)

Birnboim HC, Doly J. **A rapid alkaline extraction procedure for screening recombinant plasmid DNA.** *Nucleic Acids Res* **7**:1513-1517 (1979)

Bisbee CA, baker MA, Wilson, AC, Haji-Azimi I, Fischberg M. **Albumin phylogeny for clawed frogs (*Xenopus*).** *Science* **195**:785-787 (1977)

Black DL. **Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology.** *Cell* **103**:367-370 (2000)

Blader P, Fischer N, Gradwohl G, Guillemont F, Strahle U. **The activity of neurogenin1 is controlled by local cues in the zebrafish embryo.** *Development* **124**:4557-4569 (1997)

Blumenthal T. **Trans-splicing and polycistronic transcription in Caenorhabditis elegans.** *Trends Genet* **11**:132-136 (1995)

Boeke JD. **LINEs and *Alus* – the polyA connection.** *Nature Genet* x:6-7 (1997)

Boeke JD, Pickeral OK. **Retroshuffling the genomic deck.** *Nature* **398**:108-111 (1999)

Bonen L. **Trans-splicing of pre-mRNA in plants, animals, and protists.** *FASEB J* **7**:40-46 (1993)

Boukis LA, Bruzik JP. **Functional selection of splicing enhancers that stimulate trans-splicing in vitro.** *RNA* 7:793-805 (2001)

Boumil MR and Lee JT. **Forty years of decoding the silence in X-chromosome inactivation.** *Hum Mol Genet* **10**:2225-2232 (2001)

Bovia F, Strub K. **The signal recognition particle and related small cytoplasmic ribonucleoprotein particles.** *J Cell Sci* **109**:2601-2608 (1996)

Braat AK, van De Water S, Korving J, Zivkovic D. **A zebrafish vasa morphant abolishes Vasa protein but does not affect the establishement of the germlaine.** *Genesis* **30**:183-185 (2001)

Brakenhoff RH, Schoenmakers JG, Lubsen NH. **Chimeric cDNA clones: a novel PCR artifact.** *Nucleic Acids Res* **19**:1949 (1991)

Brannan CI and Bartolomei MS. **Mechanisms of genomic imprinting.** *Curr Opin Genet Dev* **9**:164-170 (1999)

Branciforte D, Martin SL. **Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition.** *Mol Cell Biol* **14**:2584-2592 (1994)

Braun T, Buschhausen-Denker G, Bober E, Tannich E, Arnold HH. **A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts.** *EMBO J* **8**:701-709 (1989)

Breathnach R, Chambon P. **Organization and expression of eucaryotic split genes coding for proteins** *Annu Rev Biochem* **50**:349-383 (1981)

Breen MA, Aschroft SJ. **A truncated isoform of Ca2+/calmodulin-dependent protein kinase II expressed in human islets of Langerhans may result from trans-splicing.** *FEBS Lett* **409**:375-379 (1997)

Brennan J, Lu CC, Norris DP, Rodriguez TA, Beddington RSP, Robertson EJ. **Nodal signalling in the epiblast patterns the early mouse embryo.** *Nature* **411**:965-969 (2001)

Brennan J, Norris DP, Robertson EJ. **Nodal activity in the node governs left-right asymmetry**. *Genes Dev* **16**:2339-2344 (2002)

Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. **Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila***. *Cell* **113**:25-36 (2003)

Britten RJ. **Mobile elements inserted in the distant past have taken on important functions**. *Gene* **205**:177-182 (1997)

Bronner-Fraser M. **Molecular analysis of neural crest formation**. *J Phyisiol Paris* **96**:3-8 (2002)

Brown Jones M, Michener CM, Blanchette JO, Kuznetsov VA, Raffeld M, Serrero G, Emmert-Buck MR, Petricoin EF, Krizman DB, Liota LA, Kohn EC. **The granulin-epihtelin precursor/PC-cell-derived growth factor is a growth factor for epithelial ovarian cancer**. *Clin Canc Res* **9**:44-51 (2003)

Brownlie A, Donovan A, Pratt SJ, Paw BH, Oates AC, Brugnara C, Witkowska HE, Sassa S, Zon LI. **Positional cloning of the of the zebrafish *sauternes* gene: A model for congenital sideroblastic anaemia.** *Nat Genet* **20**:244-250 (1998)

Brownlie A, Hersey C, Oates AC, Paw BH, Falick AM, Witkowska HE, Flint J, Higgs, D, Jessen J, Bahary N, Zhu H, Lin S, Zon LI. **Characterization of embryonic globin genes of the zebrafish**. *Dev Biol* **255**:48-61 (2003)

Brunet F, Godin F, David JR, Capy P. **The *mariner* transposable element in the Drosophilidae family**. *Heredity* **73**:377-385 (1994).

Bruzik JP, Doren K, Hirsh D, Steitz JA. **Trans splicing involves a novel form of small nuclear ribonucleoprotein particles**. *Nature* **335**:559-562 (1988)

Bruzik JP. **Splicing glue: a role for SR proteins in *trans* splicing**? *Microbial Pathogen* **21**:149-155 (1996)

Bruzik JP, Van Doren K, Hirsch D, Steitz JA. **Trans splicing involves a novel form of small nuclear ribonucleoprotein particles**. *Nature* **335**:559-562 (1988)

Bruzik JP, Maniatis T. **Spliced leader RNAs from lower eukaryotes are trans-spliced in mammalian cells**. *Nature* **360**:692-695 (1992)

Bruzik JP, Maniatis T. **Enhancer-dependent interaction 5' and 3' splice sites in trans**. *Proc Natl Acad Sci USA* **92**:7056-7059 (1995)

Buchner K, Roth P, Schotta G, Krauss V, Saumweber H, Reuter G, Dorn R. **Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila.** *Genetics* **155**:141-157 (2000)

Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubel A, Spitz L, Lewis P, Goldfarb D, Emerman M, Stevenson M. **A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells** *Nature* **365**: 666-669 (1993)

Burd CG, Dreyfuss G. **Conserved structures and diversity of functions of RNA-binding proteins.** *Science* **265**:615-621 (1994)

Burwinkel B, Kilimann MW. **Unequal homologous recombination between LINE1 elements as a mutational mechanism in human genetic disease.** *J Mol Biol* **277**:513-517 (1998)

Cameron EE, Bachman KE, Myohanen S, Herman JG, Baylin SB. **Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer.** *Nat Genet* **21**:103-107 (1999)

Cameron FJ, Hageman RM, Cooke-Yarborough C, Kwok C, Goodwin LL, Sillence DO, Sinclair AH. **A novel germ line mutation in Sox9 causes familial campomelic dysplasia and sex reversal.** *Hum Mol Genet* **5**:1625-1630 (1996)

Campbell DA, Thornton DA, Boothroyd JC. **Apparent discontinuous transcription of Trypanosoma brucei variant surface antigens.** *Nature* **311**:350-355 (1984)

Cañestro C, Bassham S, Postlethwait JH. **Seeing chordate evolution through the *Ciona* genome sequence.** *Genome Biol* **4**:208 (2003)

Caron KM, Soo SC, Wetsel WC, Stocco DM, Clark BJ, Parker KL. **Targeted disruption of the mouse gene encoding steroidogenic acute regulatory protein provides insights into congenital lipoid adrenal hyperplasia.** *Proc Natl Acad Sci USA* **94**:11540-11545 (1997)

Carpenter G. **EGF: new tricks for and old growth factor.** *Curr Opin Cell Biol* **5**:261-264 (1993)

Carpenter EM, Goddard JM, Chisaka O, Manley NR, Capecchi MR. **Loss of Hox-A1 (Hox1.6) function results in the reorganization of the murine hindbrain.** Development **118**:1063-1075 (1993)

Carter BJ, Halper J. **Transforming growth factor type e is a novel mediator of wound repair.** *Wound Rep Reg* **4**:259-268 (1996)

Carthew RW. **Gene silencing by double-stranded RNA.** *Curr Opin Cell Biol* **13**:244-248 (2001)

Caudevilla C, Serra D, Miliar A, Codony C, Asins G, Bach M, Hegardt FG. **Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver.** *Proc Natl Acad Sci USA* **95**:12185-12190 (1998)

Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A. **Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization.** *Proc Natl Acad Sci USA* **97**:14311-14316 (2000)

Cedar H. **DNA methylation and gene activity.** *Cell* **53**:3-4 (1988)

Chan SJ, Cao QP, Steiner DF. **Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus.** *Proc Natl Acad Sci USA* **87**:9319-9323 (1990)

Chang J, Taylor J. **In vivo RNA-directed transcription, with template switching, by a mammalian RNA polymerase.** *EMBO J* **21**:157-164 (2002)

Chao H, Mansfield G, Bartel RC, Hiriyanna S, Mitchell LG, Garcia-Blanco MA, Walsh CE. **Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing.** *Nature Med* **9**:1015-1019 (2003)

Chaput M, Claes V, Portetelle D, Cludts I, Cravador A, Burny A, Gras H, Tartar A. **The neurotrophic factor neuroleukin is 90% homologous with phosphohexose isomerase.** *Nature* **332:** 454-455 (1988)

Chau YM, Pando S, Taylor HS. **Hoxa11 silencing and endogenous Hoxa11 antisense ribonucleic acid in the uterine endometrium.** *J Clin End Met* **87**:2674-2680 (2002)

Chen K-S, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR. **Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome** *Nature Genet* **17**:154-164 (1997)

Chen S, Kimmelman D. **The role of the yolk syncytial layer in germ layer patterning in zebrafish.** *Development* **127**:4681-4689 (2000)

Chen Y and Schier AF. **The zebrafish Nodal signal Squint functions as a morphogen.** *Nature* **411**:607-10 (2001)

Chetverina HV, Demidenko AA, Ugarov VI, Chetverin AB. **Spontaneous rearrangements in RNA sequences.** *FEBS Lett* **450**:89-94 (1999)

Cheung J, Wilson MD, Zhang J, Khaja R, MacDonald JR, Heng HHQ, Koop BF, Scherer SW. **Recent segmental and gene duplications in the mouse genome.** *Genome Biol* **4**:R47 (2003)

Chiara MD, Reed RA. **A two-step mechanism for 5' and 3' splice-site pairing.** *Nature* **375**:510-513 (1995)

Chisaka O and Capecchi MR. **Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*.** *Nature* **350**:473-479 (1991)

Chitnis AB, Dawid IB. **Neurogenesis in zebrafish embryos.** *Methods Cell Biol* **59**:367-386 (1999)

Chizzonite RA and Zak R. **Regulation of myosin isoenzyme composition in fetal and neonatal rat ventricle by endogenous thyroid hormones.** *J Biol Chem* **259**:12628-12632 (1984)

Chao H, Mansfield SG, Bartel RC, Hiriyanna S, Mitchell LG, Garcia-Blanco MA, Walsh CE. **Phenotype correction of hemophilia A mice by spliceosome-mediated RNA *trans*-splicing.** *Nat Med* **9**:1015-1019 (2003)

Chmoczynski P, Sacchi N. **Single-step method of RNA isolation by acid guanidium-thiocynanate-phenol-chloroform extraction.** *Anal Biochem* **162**:156-159 (1987)

Christian SL, Fantes JA, Mewborn SK, Huang B, Ledbetter DH. **Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11-q13).** *Hum Mol Genet* **8**:1025-1037 (1999)

Clayton DA. **A big development for a small RNA.** *Nature* **410**:29-31 (2001)

Cleaver O, Krieg PA. **VEGF mediates angioblast migration during development of the dorsal aorta in *Xenopus*.** *Development* **125**:3905-3914 (1998).

Cleaver O, Seufert DW, Krieg PA. **Endoderm patterning by the notochord: development of the hypochord in *Xenopus*.** *Development* **127**:869-879 (2000)

Cleaver O, Krieg PA. **Notochord patterning of the endoderm.** *Dev Biol* **234**:1-12 (2001)

Clemens JC, Worby CA, Simonson-Leff N, Muda m, Maehama T, Hemmings BA, Dixon JE. **Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways.** *Proc Natl Acad Ssci USA* **97**:6499-6503 (2000)

Coccia E, Cicala C, Charlesworth A, Ciccarelli C, Rossi G, Philipson L, Sorrentino V. **Regulation and expression of a growth arrest-specific gen (*gas5*) during growth, differentiation, and development.** *Mol Cell Biol* **12**:3514-3521 (1992)

Cohen SN, Chang AC, Hsu L. **Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA.** *Proc Natl Acad Sci USA* **69**:2110-2118 (1972)

Cole LK, Ross LS. **Apoptosis in the developing zebrafish embryo.** *Dev Biol* **240**:123-142 (2001)

Comeron JM. **What controls the length of noncoding DNA?** *Curr Opin Genet Dev* **11**:652-659 (2001)

Comings DE. **Evidence for ancient tetraploidy and conservation of linkage groups in mammalian chromosomes.** *Nature* **238**:455-457 (1972)

Communi D, Suarez-Huerta N, Dussossoy D, Savi P, Boeynaems JM. **Cotranscription and intergenic splicing of human P2Y11 and SSF1 genes.** *J Biol Chem* **276**:16561-16566 (2001)

Condie BG and Capecchi MR. **Mice homozygous for a targeted disruption of *Hoxd-3* (*Hox-4.1*) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and the axis.** *Development* **119**: 579-595 (1993)

Condie BG and Capecchi MR. **Mice with targeted disruptions in the paralogous genes hoax-3 and hoxd-3 reveal synergistic interactions.** *Nature* **370**:304-307 (1994).

Conlon FL, Lyons KM, Takaesu N, Barth KS, Kispert A, Herrmann B, Robertson EJ. **A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse.** *Development* **120**:1919-1928 (1994)

Conlon JM. **Singular contributions of fish neuroendocrinology to mammalian regulatory peptide research.** *Regul Peptides* **93**:2-12 (2000)

Corness JD, Demchyshyn LL, Seeman P, Van Tol Hh, Srikant CB, Kent G, Patel YC, Niznik HB. **A human somatostatin receptor (SSTR3), located on chromosome 22, displays preferential affinity for somatostatin-14 like peptides.** *FEBS Lett* **321**:279-284 (1993)

Crnogorac-Jurcevic T, Brown JR, Lehrach H, Schalkwyk LC. ***Tetraodon fluviatilis*, a new puffer fish model for genome studies.** *Genomics* **41**:177-184 (1997).

Culouscou J-M, Carlton GW, Shoyab M. **Biochemical analysis of the epithelin receptor.** *J Biol Chem* **268**:10458-10462 (1993)

Curtis H, Sandoval C, Oblin C, Difalco MR, Congote LF. **Insect cell production of a secreted form of human α1-proteinase inhibitor as a bifunctional protein which inhibits neutrophil elastase and has growth factor-like activities.** *J Biotech* **93**:35-44 (2002)

Danilevskaya ON, Traverse KL, Hogan NC, DeBaryshe PG, Pardue ML. **The two Drosophila telomeric transposable elements have very different patterns of transcription.** *Mol Cell Biol* **19**:873-881 (1999)

Dallman MF, Darlington DN, Suemaru S, Cascio CS, Levin N. **Corticosteroids in homeostasis.** *Acta Physiol Scand* Suppl **583**:27-34 (1989)

Daniel R, Daniels E, He Z, Bateman A. **Progranulin (acrogranin/PC cell-derived growth factor/granulin-epithelin precursor) is expressed in the placenta, epidermis, microvasculature, and brain during murine development.** *Dev Dyn* **227**:593-599 (2003)

Dearry A, Gingrich JA, Falardeau P, Fremeau RT Jr, Bates MD, Caron MG. **Molecular cloning and expression of the gene for a human D1 dopamine receptor.** *Nature* **347**:72-76 (1990)

DeBlock M, Debrouwer D. **RNA-RNA *in situ* hybridisation using DIG-labeled probes: the effect of high molecular weight polyvinyl alcohol on the alkaline phosphatase indoxyl-nitoblue tetrazolium reaction.** *Anal Biochem* **215**:86-89 (1993)

DeChiara TM, Efstratiadis A, Robertson EJ. **A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting.** *Nature* **345**:78-80 (1990)

DeChiara TM, Robertson EJ, Efstratiadis A. **Parental imprinting of the mouse insulin-like growth factor II gene.** *Cell* **64**:849-859 (1991)

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS. **The draft**

genome of *ciona intestinalis*: **insights into chordate and vertebrate origins.** *Science* 298:2157-2167 (2002)

Deininger PL, Batzer MA. **Minireview: Alu repeats and human disease.** *Mol Gen Metab* **67**:183-193 (1999)

Dekens MP, Santoriello C, Vallone D, Grassi G, Whitmore D, Foulkes NS. **Light regulates the cell cycle in zebrafish.** *Curr Biol* **13**: 2051-2057 (2003)

De los Santos T, Schweizer J, Rees CA, Francke U. **Small evolutionary conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader-Willi deletion region, which is highly expressed in brain.** *Am J Hum Genet* **67**:1067-1082 (2000)

Dennis C. **Epigenetics and disease: Altered states.** *Nature* **421**:686-688 (2003)

D'Ercole AJ, Ye P, Calikoglu AS, Gutierrez-Ospina G. **The role of insulin-like growth factors in the central nervous system.** *Mol Neurobiol* **13**:227-255 (1996)

Derynek R, Roberts AB, Winkler ME, chen EY, Goeddel D. **Human transforming growth factor-α: precursor structure and expression in *E coli*.** *Cell* **38**:287-297 (1984)

DeRobertis EM, Larrain J, Oelgeschlager M, Wessely O. **The establishement of spemann's organizer and patterning of the vertebrate embryo.** *Nat Rev Genet* 1:171-181 (2000)

Detrich HW III, Kieran MW, Chan FY, Barone LM, Yee K, Rundtadler JA, Pratt S, Ransom D, Zon LI. **Intraembryonic hematopoietic cell migration during vertebrate development.** *Proc Natl Acad Sci USA* **92**:10713-10717 (1995)

Dewannieux M, Esnault C, Heidman T. **LINE-mediated retrotransposition of marked Alu sequences.** *Nature Genet* **35**:41-48 (2003).

Dickson B, Salomon DS, editors. **Hormones and Growth Factors in Development and Neoplasia.** New York: Wiley-Liss (1998)

Diebold SS, Montoya M, Unger H, Alexopoulou L, Roy P, Haswell LE, Al-Shamkhani A, Flavell R, Borrow P, Reis e Sousa C. **Viral infection switches non-plasmacytoid dendritic cells into high interferon producers.** *Nature***424**:324-328(2003)

Dingwall C, Laskey RA. **Nuclear targeting sequences--a consensus?** *Trends Biochem Sci* **16**:478-481 (1991)

Divorsky V et al., **A novel mechanism of β-thalassemia. The insertion of L1 retrotransposable element into β-globin IVSII.** *Blood* **88**:148a (1996)

Doak TG, Doerder FP, Jahn CL, Herrick. **A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "DD35E" motif.** *Proc Natl Acad Sci USA* **91**:942-946 (1994)

Donovan A, Brownlie A, Zhou Y, Shepard J, Pratt SJ, Moynihan J, Paw BH, Drejer A, Barut B, Zapata A, Law TC, Brugnara C, Lux SE, Pinkus GS, Pinkus JL, Kingsley PD, Palis J, Fleming MD, Andrews NC, Zon LI. **Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter.** *Nature* **403**:776-781 (2000)

Doolittle RF. **The genealogy of some recently evolved vertebrate proteins.** *Trends Biochem Sci* **10**:233-237 (1985)

Dorn R, Krauss V. **The modifier of mdg4 locus in Drosophila: functional complexity is resolved by *trans* splicing.** *Genetica* **117**:165-177 (2003)

Dorn R, Reuter G, Loendorf A. **Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila.** *Proc Natl Acad Sci USA* **98**:9724-9729 (2001)

Doolittle WF, Sapienza C. **Selfish genes, the phenotype paradigm and genome evolution.** *Nature* **284**:601-603 (1980)

Dougan ST, Warga RM, Kane DA, Schier AF, Talbot WS. **The role of the zebrafish *nodal*-related genes *squint* and *cyclops* in patterning of mesendoderm.** *Development* **130**: 1837-1851 (2003)

Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, Zwartkruis F, Abdelilah S, Rangini Z, Belak J, Boggs C. **A genetic screen for mutations affecting embryogenesis in zebrafish.** *Development* **123**:37-46 (1996)

Drummond IA, Majumdar A, Hentshel H, Elger M, Solnica-Kretzel L, Schier AF, Neuhauss SC, Stemple DL, Zwartkruis F, Rangini Z, Driever W, Fishman MC. **Early development of the zebrafish pronephros and analysis of mutations affecting pronephric function.** *Development* **125**:4655-4667 (1998).

Duboule D, Wilkins AS. **The evolution of "bricolage".** *Trends Genet* **14**:54-59 (1998)

Durand D. **Vertebrate evolution: doubling and shuffling with a full deck.** *Trends Genet* **19**:2-5 (2003)

Eddy SR. **Non-coding RNA genes and the moderm RNA world.** *Nat Rev Genet* **2**:919-929 (2001)

Eddy SR. **Non-coding RNA genes.** *Curr Opin Genet Dev* **9**:695-699 (1999)

Edgar BA, Kiehle CP, Schubiger G. **Cell cycle control by the nucleo-cytplasmic ratio in early *Drosophila* development.** *Cell* **44**:365-372 (1986)

Edmonston DG, Olson EN. **A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program.** *Genes Dev* **3**:628-640 (1989)

Eferl R, Sibilia M, Hilberg F, Fuchsbichler A, Kufferath I, Guertl B, Zenz R, Wagner EF, Zatloukal K. **Functions of c-Jun in liver and heart development.** *J Cell Biol* **145**:1049-1061 (1999).

Efstratiadis A. **Genetics of mouse growth.** *Int J Dev Biol* **42**:955-976 (1998)

Eickbush TH. **Repair by retrotransposition.** *Nature Genet* 31:126-127 (2002)

Eickbush TH. **Telomerase and retrotransposons: which came first** ? *Science* **277**:911-912 (1997)

Eichler EE. **Recent duplication, domain accretion and the dynamic mutation of the human genome.** *Trends Genet* **17**:661-669 (2001)

Eizirik E, Murphy WJ, O'Brien SJ. **Molecular dating and biogeography of the early placental mammal radiation.** *J. Hered.* **92**: 212-219 (2001)

Ehrlich J, Sankoff D, Nadeau JH. **Synteny conservation and chromosome rearrangements during mammalian evolution.** *Genetics* **147**: 289-296 (1997)

Ejima Y, Yang L. ***Trans* mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling.** *Hum Mol Genet* **12**:1321-1328 (2003)

Ekker SC, Larson JD. **Morphant technology in model developmental systems.** *genesis* **30**:89-93 (2001)

Elbashir SM, Lendeckel W, Tuschi T. **RNA interference is mediated by 21- and 22-nucleotide RNAs.** *Genes Dev* **15**:188-200 (2001)

Erdmann V, Szymanski M, Hochberg A, deGroot N, Barciszewski J. **Collection of mRNA-like non-coding RNAs.** *Nucleic Acids Res* **27**:192-195 (1999)

Eriksson J, Lofberg J. **Development of the hypochord and dorsal aorta in the zebrafish embryo (*Danio rerio*).** *J Morphol* **244**:167-176 (2000)

Esnault C, Maestre J, Heidmann T. **Human LINE retrotransposons generate processed pseudogenes.** *Nature Genet* **24**:363-367 (2000)

Etchegaray J-P, Lee C, Wade PA, Reppert SM. **Rhythmic histone acetylation underlies transcription in the mammalian circadian clock.** *Nature* 421:177 - 182 (2003)

Eul J, Graessmann M, Graessmann A. **Experimental evidence for RNA trans-splicing in mammalian cells.** *EMBO J* **14**:3226-3235 (1995)

Fagotto F, Gumbiner BM. **Cell contact-dependent signaling.** *Dev Biol* **180**:445-454 (1996)

Fahey ME, Moore TF, Higgins DG. **Overlaping antisense transcription in the human genome.** *Comp Funct Genom* **3**:244-25 (2002)

Faik P, Walker JIH, Redmill AAM, Morgan MJ. **Mouse glucose-6-phosphate isomerase and neuroleukin have identical 3' sequences** *Nature* **332**: 455 (1988)

Faure S, Lee MA Keller T, Dijke P, Whitman M. **Endogenous patterns of TGFbeta superfamily signalling during early Xenopus development. 127**:2917-2931 (2000)

Federova L, Federov A. **Introns in gene evolution.** *Genetica* **118**:123-131 (2003)

Feinberg AP, Vogelstein, B. **A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity.** *Anal Biochem* **132**:6-13 (1983)

Feinberg AP, Vogelstein, B. **"A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum.** *Addendum Anal Biochem* **137**:266-267 (1984)

Fehrmann F, Welker R, Krausslich H-G. **Intracisternal A-type particles express their proteinase in a separate reading frame by translational frame-shifting, similar to D-type retroviruses.** *Virology* **235**:352-259 (1997).

Feldman B, Gates MA, Egan ES, Dougan ST, Rannebeck G, Sirotkin HI, Schier AF and Talbot WS. **Zebrafish organizer development and germ-layer formation require nodal-related signals.** *Nature* **395**: 185-189 (1998)

Feldman B, Dougan ST, Schier AF, Talbot WS. **Nodal–related signals establish mesendodermal fate and trunk neural identity in zebrafish.** *Curr Biol* **10**:531-534 (2000)

Ferier DEK and Holland PWH. **Ancient origin of the Hox gene cluster.** *Nature Rev Genet* **2**:33-38 (2001)

Ferris SD and Whitt GS. **Loss of duplicate gene expression after polyploidisation.** *Nature* **265**:258-260 (1977)

Filipowics W. **Imprinted expression of small nucleolar RNAs in brain: time for RNomics.** *Proc Natl Acad Sci USA* **97**:14035-14037 (2000)

Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G. **Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA.** *Nature* **408**, 86-89 (2000)

Finta C and Zaphiropoulos. **Intergenic mRNA molecules resulting from *trans*-splicing.** *J Biol Chem* **8**:58825890 (2002)

Finta C, Zaphiropoulos PG. **The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons.** *Gene* **260**:13-23 (2000)

Finta C, Zaphiropoulos PG. **The human CYP2C locus: a prototype for intergenic and exon repetition splicing events.** *Genomics* **63**:433-438 (2000)

Fire A, Xu SQ, Montgomery MK, Kostas SA, Driver SE, Mello CC. **Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*.** *Nature* **391**:806-811 (1998)

Fishman MC. **Genomics. Zebrafish – The canonical vertebrate.** *Science* **294**:1290-1291 (2001)

Foe VE. **Mitotic domains reveal early commitment of cells in Drosophila embryos.** *Development* **107**:1-22 (1989)

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* **151**: 1531-1545 (1999).

Forsdyke DK. **A stem-loop "kissing" model for the initiation of recombination and the origin of introns** *Mol Biol Evol* **12**:949-958 (1995)

Foster JW, Dominguez-Steglich MA, Guioli S, Kowk G, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, *et al.* **Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene.** *Nature* **372**: 525-530 (1994)

Fouquet B, Weinstein BM, Serluca FC, Fishman MC. **Vessel patterning in the embryo of the zebrafish: guidance by notochord.** *Dev Biol* **183**:37-48 (1997)

Frantz SA, Thiara AS, Lodwick D, Ng LL, Eperon IC, Samani NJ. **Exon repetition in mRNA.** *Proc Natl Acad Sci USA* **96**:5400-5405 (1999)

Frisch SM, Screaton RA. **Anoikis mechanisms.** *Curr Opin Cell Biol* **13**:555-562 (2001)

Fryer J, Lederis K, Rivier J. **Urotensin-I, a CRF-like neuropeptide, stimulates ACTH from the teleost pituitary.** *Endocrinology* **113**:2308-2310 (1983)

Fu X-D. **The superfamily of arginine/serine-rich splicing factors.** *RNA* 1:663-680 (1995)

Fu L and Lee CC. **The circadian clock: pacemaker and tumour suppressor.** *Nat Rev Cancer* 3:350-361 (2003)

Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. **The RAG proteins and V(D)J recombination: complexes, ends and transposition.** *Annu Rev Immunol* **18**:495-527 (2000)

Gabellini D, Tupler R, Green MR. **Transcriptional derepression as a cause of genetic diseases.** *Curr Opin Genet Dev* **13**:239-245 (2003)

Garcia-Blanco MA. **Messenger RNA reprogramming by spliceosome-mediated RNA trans-splicing.** *J Clin Invest* **112**:474-480 (2003)

Garcia-Blanco MA, Puttaraju M, Mansfield SG, Mitchell LG. **Spliceosome-mediated RNA trans-splicing in gene therapy and genomics.** *Gene Ther Reg* **1**:141-163 (2000)

Gates MA, Kim L, Egan ES, Cardozo T, Sirotkin HI, Dougan ST, Lashkari D, Abagyan R, Schier AF, and Talbot WS. **A genetic linkage mp for zebrafish: Comparative analysis and localization of genes and expressed sequences.** *Genome Res* **9**:334-347 (1999).

Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R. **Induction of Tumors in Mice by Genomic Hypomethylation.** *Science* **300**:489-492 (2003)

Gaufo GO, Flodby P, Capecchi MR. **Hoxb1 controls effectors of sonic hedgehog and Mash1 signaling pathways.** *Development* **127**:5343-5354 (2000)

Geetha-Habib M, Noiva R, Kaplan HA, Lennarz WJ. **Glycosylation site binding protein, a component of oligosaccharyl transferase, is highly similar to three other 57 kd luminal proteins of the ER.** *Cell* **54**:1053-1060 (1988)

Genereux DP, Logsdon JM. **Much ado about bacteria-to-vertebrate lateral gene transfer.** *Trends Genet* **19**:191-195 (2003)

Gerhart E, Wagner H, Flärdh. **Antisense RNAs everywhere ?** *Trends Genet* **18**:223-226 (2002)

Gering M, Rodaway AR, Gottgens B, Patient RK, Green AR. **The SCL gene specifies haemangioblast development from early mesoderm.** *EMBO J* **17**:4029-4045 (1998)

Girard L, Freeling M. **Regulation changes as a consequence of transposon insertion.** *Dev Genet* **25**:291-296 (1999)

Glücksmann A. **Cell deaths in normal vertebrate ontogeny.** *Biol Rev* **26**:59-86 (1951)

Glusman G, Sosinsky A, Ben-asher E, Avidan N, Sonkin D, Bahar A, Rosenthal A, Clifton S, Roe B, Ferraz C, Demaille J, Lancet D. **Sequence, structure, and evolution of a complete human olfactory receptor gene cluster.** *Genomics* **63**:227-245 (2000)

Goddard JM, Rossel M, Manley NR, Capecchi MR. **Mice with targeted disruption of hoxb1 fail to form the motor nucleus of the V11th nerve.** *Development* **122**:3217-3228 (1996)

Goeke S, Greene EA, Grant PK, Gates MA, Crowner D, Aigaki T, Giniger E. **Alternative splicing of lola generates 19 transcription factors controlling axon guidance in Drosophila.** *Nat Neurosci* **6**:917-924 (2003)

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP.**Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence.** *Mol. Phylogenet. Evol.* **9**:585-598 (1998).

Gonzalez-Pastor JE, San Millan JL, Moreno F. **The smallest known gene.** *Nature* 413:432-435 (2001)

Graessmann M, Eul J, Berg B, Zimmermann C. Graessmann A. **SV40 T1-mRNA trans-splicing and translation requires that the in vitro synthesized cRNA is capped before microinjection.** *FEBS Lett* **394**:233-236 (1996)

Graham A. **The evolution of the vertebrates – genes and development.** *Curr Opin Genet Dev*10:624-628 (2000)

Grapin-Botton A, Melton DA. **Endoderm development: From patterning to organogenesis.** *Trends Genet* **16**:124-130 (2000).

Graveley BR. **Alternative splicing: increasing diversity in the proteomic world.** *Trends Genet* **17**:100-107 (2001)

Gray A, Dull TJ, Ullrich A. **Nucleotide sequence of epidermal growth factor precursor cDNA predicts a 128,000-molecular weight protein precursor**. *Nature* **303**:722-725 (1983)

Greally JM. **Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome.** *Proc Natl Acad Sci USA* **99**:327-332 (2002)

Greenwald I, Rubin GM. **Making a difference: the role of cell-cell interactions in establishing separate identities for equivalent cells.** *Cell* **68**:271-281 (1992)

Greer JM, Puetz J, Thomas KR and Capecchi MR. **Mainenance of functional equivalence during paralogous Hox gene evolution.** *Nature* **403**:661-665 (2000)

Groβhans J, Wieschaus E. **A genetic link between morphogenesis and cell division during formation of the ventral furrow in Drosophila.** *Cell* **101**:523-531 (2000)

Gu X. **Early metazoan divergence was about 830 million years ago.** *J Mol Evol* **47**:369-371 (1998)

Gu X, Wang Y, Gu J. **Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution.** *Nature Genet* **31**:205-209 (2002)

Gu Z, Steinmetz, LM, Gu X, Scharfe C, Davis, RW, Li W-H. **Role of duplicate genes in genetic robustness aginst null mutations.** *Nature* **421**:63-66 (2003)
Guo S, Kemphues KJ. **par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed.** *Cell* **81**:611-620 (1995)

Gupta R, Jung E, Brunak S. **Prediction of N-glycosylation sites in human proteins.** In preparation, 2002. (http://www.cbs.dtu.dk/services/NetNGlyc/)

Haddad F, Madatsugu M, Bodell PW, Qin A, McCue SA, Baldwin KM. **Role of thyroid hormone and insulin in control of cardiac isomyosin expression.** *J Mol Cell Cardiol* **29**:559-69 (1997)
Haddad F, Bodell PW, Qin AX, Giger JM, Baldwin KM. **Role of antisense RNA in coordinating cardiac myosin heavy chain gene switching.** *J Biol Chem* **278**: 37132-37138 (2003)

Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, van Eeden FJM, Jiang Y-J, Heisenberg C-P, Kelsh RN, Furutani-Seiki M, Vogelsang E, Beuchle D, Schach U, Fabian C, Nusslein-Volhard C. **The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*.** *Development* **123**:1-36 (1996)

Haig D and Graham C. **Genomic imprinting and the strange case of the insulin-like growth factor II receptor**. *Cell* **64**:1045-1046 (1991)

Haig D and Westoby M. **Parent-specific gene expression and the triploid endosperm**. *Am Nat* **134**:147-155 (1989)

Hall CJ, Flores MV, Davidson AJ, Crosier KE, Crosier PS. **Radar is required for the establishment of vascular integrity in the zebrafish**. *Dev Biol* **251**:105-117 (2002)

Hamilton AJ and Baulcombe DC: **A species of small antisense RNA in postranscriptional gene silencing in plants**. *Science* **286**:950-952 (1999)

Hammond SM Bernstein E, Beach D, Hannon GJ. **An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells**. *Nature* **404**:293-296 (2000)

Hammond SM, Caudy AA, Hannon GJ. **Post-transcriptional gene silencing by double-stranded RNA**. *Nature Rev Genet* **2**:110-119 (2001)

Haren L, Ton-Hoang B, Chandler M. **Integrating DNA: transposases and retroviral integrases**. *Annu Rev Microbiol* **53**:245-281 (1999)

Harland R, Gerhart J. **Formation and function of spemann's organizer**. *Annu Rev Cell Dev Biol* **13**:611-667 (1997)

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilgham SM. **CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.** *Nature* **405**:486-489 (2000)

Hartl DL. **The most unkindest cut of all**. *Nat Genet* **12**:227-229 (1996)

Hastings ML, Milcarek C, Martincic K, Peterson ML, Munroe SH. **Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels**. *Nucleic Acids Res* **25**: 4296-4300 (1997)

Hastings ML, Ingle HA, Lazar MA, Munroe SH. **Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occuring antisense RNA**. *J Biol Chem* **275**:11507-11513 (2000)

Hatta K, Kimmel CB, Ho RK, Walker C. **The cyclops mutation blocks specification of the floor plate of the zebrafish central nervous system**. *Nature* **350**: 339-341 (1991)

Hattori *et al.* **The DNA sequence of human chromosome 21**. *Nature* **405**:311-319 (2000)

He Z, Bateman A. **Progranulin gene expression regulates epithelial cell growth and promotes tumor growth in vivo.** *Cancer Res* **59**:3222-3229 (1999)

He Z, Ismail A, Kriazhev L, Sadvakassova G, Bateman A. **Progranulin (PC-cell-derived growth factor/acrogranin) regulates invasion and cell survival.** *Cancer Res* **62**:5590-5596 (2002)

He Z, Bateman A. **Progranulin (granulin-epithelin precursor, PC-cell-derived growth factor, acrogranin) mediates tissue repair and tumorigenesis.** *J Mol Med* **81**:600-612 (2003)

He Z, Ong CH, Halper J, Bateman A. **Progranulin is a mediator of the wound response.** *Nat Med* **9**:225-229 (2003)

Heasman J. **Morpholino oligos: making sense of antisense ?** *Dev Biol* **243**:209-214 (2002)

Heid CA, Stevens J, Livak KJ, Williams PM. **Real time quantitative PCR.** *Genome Res* **6**:986-994 (1996)

Herbomel P, Thisse B, Thisse C. **Ontogeny and behaviour of early macrophages in the zebrafish embryo.** *Development* **126**:3735-3745 (1999)

Hey, J. **Selfish genes, pleiotropy and the origin of recombination.** *Genetics* **149**:2089-2097 (1998)

Heywood SM. **tcRNA as a naturally occurring antisense RNA in eukaryotes** *Nucleic Acids Res* **14**: 6771-6772 (1986). Published erratum appears in Nucleic Acids Res 15:384 (1987)

Higashijima S, Hotta Y, Okamoto H. **Visualization of cranial motor neurons inn live transgenic zebrafish expressing green fluorescent protein under the control of the islet-1 promoter/enhancer.** J Neurosci 20:206-218 (2000)

Hill RE and Hastie ND. **Accelerated evolution in the reactive centre regions of serine protease inhibitors.** *Nature* **326**:96-99 (1987)

Hiorn K, Melek M, Gellert M. **DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations.** *Cell* **94**:463-470 (1998)

Hirose Y, Manley JL. **RNA polymerase II and the integration of nuclear events.** *Genes Dev* **14**:1415-1429 (2000)

Ho CY, Houart C, Wilson SW, Stainier DY. **A role for the extraembryonic yolk syncytial layer in patterning the zebrafish embryo suggested by properties of the hex gene.** *Curr Biol* **9**:1131-1134 (1999)

Ho RK, Kimmel CB. **Commitment of cell fate in the early zebrafish embryo.** *Science* **261**:109-111 (1993)

Hodgkinson CA, Moore KJ, Nakayama A, Steingrimsson E, Copeland NG, Jenkins NA, Arnheiter H. **Mutations at the mouse microphtalmia locus are associated with defects in a gene encoding a novel basic helix-loop-helix zipper protein.** *Cell* **74**:395-404 (1993)

Holland PWH, Garcia-Fernandez J, Williams NA, Sidow A. **Gene duplications and origins of vertebrate development.** *Development Suppl.* :125-133 (1994)

Holland PHW, Garcia-Fernandez JG. ***Hox* genes and chordate evolution.** *Dev Biol* **173**:382-395 (1996)

Holland PWH. **Vertebrate evolution: something fishy about Hox genes.** *Curr Biol* **7**:R570-R572 (1997)

Holmes SE, Dombrowski BA, Krebs CM, Boehm CD, Kazazian HH. **A new retrotransposable element from the LRE locus on chromosome 1q produces a chimaeric insertion.** *Nature Genet* **17**:143-148 (1994)

Horowitz DS, Krainer R. *Trends Genet* **10**:100-106 (1993)

Houck MA, Clark JB, Peterson KR, Kidwell MG. **Possible horizontal transfer of *Drosophila* genes by mite *Proctolaellaps regalis*.** *Science* **253**:1125-1129 (1991)

Hrabal R, Chen Z, James S, Bennett HPJ, and Ni F. **The hairpin stack fold, a novel protein architecture for a new family of protein growth factors.** *Nat Struct Biol* **3**:747-752 (1996)

Hsieh-Li HM, Witte DP, Weinstein M, Branford W, Li H, Small K, Potter SS. **Hoxa 11 structure, extensive antisense transcription, and function in male and female fertility.** *Development* **121**:1373-1385 (1995)

Hsu K, Kanki JP, Look AT. **Zebrafish myelopoiesis and blood cell development.** *Curr Opin Hematol* **81**:245-251 (2001)

Hu JF, Oruganti H, Vu, TH, Hoffman AR. **Tissue-specific imprinting of the mouse insulin-like growth factor II receptor gene correlates with differential allele-specific DNA methylation.** *Mol Endocrinol* **12**:220-32 (1998)

Hu JF, Oruganti H, Vu TH, Hoffman AR. **The Role of Histone Acetylation in the Allelic Expression of the Imprinted Human Insulin-like Growth Factor II Gene.** *Biochem Biophys Res Comm* **251**:403-408 (1998)

Hu JF, Balaguru KA, Ivaturi RD, Oruganti H, Li T, Nguyen BT, Vu TH, Hoffman AR. **Lack of reciprocal genomic imprinting of sense and antisense RNA of mouse insulin-like growth factor II receptor in the central nervous system.** *Biochem Biophys Res Comm* **257**:604-608 (1999)

Hu JF, Pham J, Dey I, Li T, Vu TH, Hoffman AR. **Allele-specific histone acetylation accompanies genomic imprinting of the insulin-like growth factor II receptor gene.** *Endocrinology* **141**:4428-4435 (2000)

Hu M-C, Hsu N-C, El Hadj NB, Pai C-I, Chu H-P, Wang C-KL, Chung B-C. **Steroid deficiency syndromes in mice with targeted disruption of Cyp11a1.** *Mol Endocrinol* **16**:1943-1950 (2002)

Huang J, Van der Ploeg JH. **Requirement of a polypyrimidine tract for trans-splicing in trypanosomes: discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site.** *EMBO J* **10**:3877-3885 (1991)

Hudson C, Clements D, Friday RV, Stott D, Woodland HR. **Xsox17alpha and –beta mediate endoderm formation in Xenopus.** *Cell* **91**:397-405 (1997)

Hughes MK and Hughes AL. **Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*.** *Mol Biol Evol* **10**:1360-1369 (1993)

Hughes AL. **The evolution of functionally novel proteins after gene duplication.** *Proc R Soc Lon B* **256**:119-124 (1994)

Hughes AL. **Phylogenetic tests of the hypothesis of block duplication of homologous genes on human chromosomes 6, 9, and 1.** *Mol Biol Evol* **15**:854-870 (1998)

Hughes AL. **Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history.** *J Mol Evol* **48**:565-576 (1999)

Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, Epstein JA, Barbazuk, WB, Li F N, Paw B, Postlethwait JH, Hudson,TJ, Zon, LI, McPherson JD, Chevrette M, Dawid IB, Johnson SL, Ekker M. **Radiation hybrid mapping of the zebrafish genome.** *Proc. Natl. Acad. Sci. USA* **96**:9745-9750 (1999)

Hummel HS, Gillespie RD, Swindle J. **Mutational analysis of 3' splice site selection during *trans*-splicing.** *J Biol Chem* **45**:35522-35531 (2000)

Hunter CP. **Gene silencing: Shrinking the black box of RNAi.** *Curr Biol* **10**:R137-R140 (2000)

Hurst GDD, Werren, JH. **The role of selfish genetic elements in eukaryotic evolution.** Nat Rev Genet 2:597-606 (2001)

Huttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie JP, Brosius J. **RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse.** *EMBO J* **20**:2943-2953 (2001)

Hutvagner G, et al. **A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.** *Science* **293**:834-838 (2001)

IlicD, Kovacic B, McDonagh S, Jin F, Baumbusch C, Gardner DG, Damsky CH. **Focal adhesion kinase is required for blood vessel morphogenesis.** *Circ Res* **92**:300-307 (2003)

Imai Y, Talbot WS. **Morpholino phenocopies of the bmp2b/swirl and bmp7/snailhouse mutations.** *Genesis* **30**:160-163 (2001)

Inouye M. **Antisense RNA: its functions and applications in gene regulation - a review.** *Gene* **72:**25-34 (1988)

Irwin MS, Kaelin WG. **p53 family update: p73 and p63 develop their own identities.** *Cell Growth Differ* **12**:337-349 (2001)

Ishii T, Hasegawa T, Pai CI, Yvgi-Ohana N, Timberg R, Zhao L, Majdic G, Chung BC, Orly J, Parker KL. **The roles of circulating high-density lipoproteins and trophic hormones in the phenotype of knockout mice lacking the steroidogenic acute regulatory protein.** *Mol Endocrinol* **16**:2297-2309 (2002)

Ishizaki Y, Voyvodic JT, Burne JF, Raff MC. **Control of lens epithelial cell survival.** *J Cell Biol* **121**:899-08 (1993)

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. **Molecular reconstruction of Sleeping Beauty, a *Tc1*-like transposon from fish, and its transposition in human cells.** *Cell* **91**:501-510 (1997)

Ivics Z, Izsvak Z, Minter A, Hackett PB. **Identification of functional domains and evolution of Tc1-like transposable elements.** *Proc Natl Acad Sci USA* **93**:5008-5013 (1996)

Iwabe N, Kuma K, Miyata T. **Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates.** *Mol Biol Evol* **13**:483-493 (1996)

Izsvak Z, Ivics, Z, Garcia-Estefania D, Fahrenkrug SC, Hackett PB. **DANA elements: a family of composite , tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish.** *Proc Natl Acad Sci USA* **93**:1077-1081 (1996)

Izsvak Z, Ivics Z, Hackett PB. **Repetitive elements and their genetic applications in zebrafish.** *Biochem Cell Biol* **75**:507-523 (1997)

Izsvak Z, Ivics Z, Shimoda N, Mohn D, Okamoto H, Hackett PB. **Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification** *J Mol Evol* **48**:13-21 (1999)

Jacobs LL and Pilbeam D. **Of mice and men: fossil-based divergence dates and molecular 'clocks'.** *J Hum Evol* **9**:551-555 (1980).

Jenne DE, Tinschert S, Reimann H, Lasinger W, Thiel G, Hameister H, Kehrer-Sawatzki H. **Molecular Characterization and Gene Content of Breakpoint Boundaries in Patients with Neurofibromatosis Type 1 with 17q11.2 Microdeletions** *Am J Hum Genet* **69**:516 (2001)

Jenuwein T and Allis CD. **Translating the Histone Code.** *Science* **293**:1074-1080 (2001).

Jiang YJ, Aerne BL, Smithers L, Haddon C, Ish-Horowicz D, Lewis J. **Notch signalling and the synchronization of the somite segmentation clock.** *Nature* **408**:475-479 (2000)

Jochum W, Passegué E, Wagner EF. **AP-1 in mouse development and tumorigenesis.** *Oncogene* **20**:2401-2412 (2001)

Johnson PJ, Kooter JM, Borst P. **Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene.** *Cell* **51**:273-281 (1987)

Jones CM, Kuehn MR, Hogan BL, Smith JC, Wright CV. **Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation.** *Development* 121:3651-3662 (1995).

Jones IC, Bellamy D. **Hormonal mechanisms in the homeostatic regulation of the vertebrate body with special reference to the adrenal cortex.** *Symp Soc Exp Biol* **18**:195-236 (1964)

Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. **Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription.** *Nat Genet* **19**:187-191 (1998)

Jones PA and Baylin SB. **The fundamental role of epigenetic events in cancer.** *Nat Rev Genet* **3**:415-428 (2002)

Jowett T. **Double in situ hybridization techniques in zebrafish.** *Methods* **23**:345-358 (2001)

Kamijo T, Zindy F, Roussel MF, Quelle DE, Downing JR, Ashmun RA, Grosveld G, Sherr CJ. **Tumor suppression at the mouse INKa locus mediated by the alternative reading frame product p19ARF.** *Cell* **91**:649-659 (1997)

Kanai-Azuma M. Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PPL, Hayashi Y. **Depletion of definitive gut endoderm in Sox17-null mutant mice.** *Development* **129**:2367-2379 (2002)

Kane DA, Warga RM, Kimmel CB. **Mitotic domains in the early embryo of the zebrafish.** *Nature* **360**:735-737 (1992)

Kane DA, Kimmel CB. **The zebrafish midblastula transition.** *Development* **119**:447-456 (1993)

Kane DA, Maischein HM, Brand M, van Eeden FJ, Furutani-Seiki M, Granato M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kelsh RN, Mullins MC, Odenthal J, Warga RM, Nusslein-Volhard C. **The zebrafish early arrest mutants.** *Development* **123**:57-66 (1996)

Kane DA. **Cell cycles and development in the embryonic zebrafish.** *Methods Cell Biol* **59**:11-26 (1999)

Kanki JP, Ho RK. **The development of the posterior body in zebrafish.** *Development* **124**:881-893 (1997)

Kaplow LS. **Simplified myeloperoxidase stain using benzidine dihydrochloride.** *Blood* **26**: 215-219 (1965).

Kasahara M, Nakaya J, Satta Y, Takahata N. **Chromosomal duplication and the emergence of the adaptive immune system.** *Trends Genet* **3**:90-92 (1997)

Kashkush K, Feldman M, Levy AA. **Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat.** *Nature Genet* **33**:102-106 (2003)

Kasschau KD, Carrington JC. **A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing.** *Cell* **95**:461-470 (1998)

Katz MJ. **Comparative anatomy of the tunicate tadpole, *Ciona intestinalis*.** *Biological Bulletin* **164**:1-27 (1983)

Kawakami K, Koga A, Hori H, Shima A. **Excision of the *Tol2* transposable element of the medaka fish, *Oryzias latipes*, in zebrafish, *Danio rerio*.** *Gene* **225**:17-22 (1998)

Kawasaki H and Taira K. **Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells.** *Nature* **423**:838-842 (2003)

Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. **Haemophilia A resulting from the de novo insertion of L1 sequiences represents a novel mechanism for mutation in man.** *Nature* **332**:164-166 (1988)

Kazazian HH. **An estimated frequency of endogenous insertional mutations in humans.** *Nature Genet* **22**:130 (1999)

Kelly C, Chin AJ, Leatherman JL, Kozlowski DJ, Weinberg ES. **Maternally controlled (beta)-catenin-mediated signaling is required for organizer formation in the zebrafish.** *Development* **127**:3899-3911 (2000)

Kennerdell JR and Carthew RW. **Heritable gene silencing in Drosophila using double-stranded RNA.** *Nat Biotechnol* **18**:896-898 (2000)

Kessler DS, Melton DA. **Vertebrate embryonic induction: mesodermal and neural patterning.** *Science* **266**:596-604 (1994)

Ketting RF, et al., **Target choice determinants of the Tc1 transposon of *Caenorhabditis elegans*.** *Nucleic Acids Res* **25**:4041-4047 (1997)

Ketting R, Haverkamp T, van Luenen H, Plasterk R. *Cell* **99**:133-141 (1999)

Ketting RF and Plasterk RHA. **A genetic link between co-suppression and RNA interference in *C. elegans*.** *Nature* **404**:296-298(2000)

Khochbin A and Lawrence JJ. **An antisense RNA involved in p53 mRNA maturation in murine erythroleukemia cells induced to differentiate.** *EMBO J* **8**:4107-4114 (1989)

Khochbin S, Brocard MP, Grunwald D, Lawrence JJ. **Antisense RNA and p53 regulation in induced murine cell differentiation.** *Ann N Y Acad Sci* **660**:77-87 (1992)

Kiess W, Haskell JF, Lee L, Greenstein LA, Miller BE, Aarson AL, Rechler MM, Nissley SP. **An antibody that blocks insulin-like growth factor (IGF) binding to the type II IGF receptor is neither an agonist nor an inhibitor of IGF- stimulated biologic responses in L6 myoblasts.** *J Biol Chem* **262**:12745-12751 (1987)

Kiess W, Blickenstaff GD, Sklar MM, Thomas MM, Nissley SP, Sahagian GG. **Biochemical evidence that the type II insulin-like growth factor receptor is identical to the cation-independent mannose 6-phosphate receptor.** *J Biol Chem* **263**:9339-9344 (1988)

Kiger AA, Baum B, Jones S, Jones MR, Coulson A, Echeverri C, Perrimon N. **A functional genomic analysis of cell morphology using RNA interference.** *J Biol* **2**:27 (2003)

Kikuchi Y, Trinh LA, Reiter JF, Alexander J, Thisse C, Waldron S, Yelon D, Thisse B, Stainier DY. **Casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish.** *Genes Dev* **15**:1493-1505 (2001)

Kim J, Bergmann A. Stubbs L. **Exon sharing of a novel human zinc-finger gene, ZIM2, and paternally expressed gene 3 (PEG3).** *Genomics* **64**:114-118 (2000)

Kim SK, Hebrok M, Melton DA. **Notochord to endoderm signalling is required for pancreas development.** *Development* **124**:4243-4252 (1997)

Kimelman D, Griffin KJ. **Vertebrate mesendoderm induction and patterning.** *Curr Opin Genet Dev* **10**:350-356 (2000).

Kimmel CB, Law RD. **Cell lineage of zebrafish blastomeres:II. Formation of the yolk syncytial layer.** *Dev Biol* **108**:86-93 (1985)

Kimmel CB, Warga RM, Schilling TF. **Origin and organization of the zebrafish fate map.** *Development* **108**:581-594 (1990)

Kimmel CB, Warga RM, Kane DA. **Cell cycles and clonal strings during the formation of the zebrafish nervous system.** *Development* **120**: 265-276 (1994)

Kimmel CB, Ballard WW, Kimmel SR. Ullmann B, Schilling TF. **Stages of embryonic development of the zebrafish.** *Dev. Dyn.* **203**, 253-310 (1995)

Kiss T. **Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs.** *EMBO J* **20**:3617-3622 (2001)

Kiss-Laszio Z, Henry Y, Bachellerie JP, Caizergues-Ferrier M, Kiss T. **Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs.** *Cell* **85**:1077-1088 (1996)

Kist R, Schrewe H, Balling R, Scherer G. **Conditional inactivation of Sox9: a mouse model for campomelic dysplasia.** *Genesis* **32**:121-123 (2002).

Kiyosawa H, Chance PF. **Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot.** *Hum Mol Genet* **5**:745-753 (1996)

Kiyosawa H, Yamanaka I, Osato N, Kondo S, RIKEN GER Group and GSL Members, Hayashizaki Y. **Antisense transcripts with FANTOM2 clone set and their implications for gene regulation.** *Genome Res* **13**:1324-1334 (2003)

Knee RS, Pitcher SE, Murphy PR. **Basic Fibroblast Growth Factor Sense (FGF) and Antisense (*GFG*) RNA Transcripts Are Expressed in Unfertilized Human Oocytes and in Differentiated Adult Tissues.** *Biochem Biophys Res Comm* **205**:577-583 (1994)

Knee RS, Li AW, Murphy PR. **Characterization and tissue-specific expression of the rat basic fibroblast growth factor antisense mRNA and protein.** *Proc Natl Acad Sci USA* **94**:4943-4947 (1997)

Kodjabachian L, Dawid IB, Toyama R. **Gastrulation in zebrafish: what mutants teach us.** *Dev Biol* **213**:231-245 (1999)

Koga A, Suzuki M, Inagaki H, Bessho Y, Hori H. **Transposable element in fish.** *Nature* **383**:30 (1996)

Konarska MM, Padgett RA, Sharp PA. **Trans splicing of mRNA precursors in vitro.** *Cell* **42**:165-171 (1985)

Kookin EV, Aravind L, Kondrashov AS. **The impact of comparative genomics on our understanding of evolution.** *Cell* 101:573-576 (2000)

Kooter JM, De Lange T, Borst P. **Discontinuous synthesis of mRNA in trypanosomes.** *EMBO J* **3**:2387-2392 (1984)

Korinek V, Barker N, Moerer P, vanDonselaar E, Huls G, Peters PJ, Clevers H. **Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4.** *Nat Genet* **19**:379-383 (1998)

Kornfeld S and Mellman I. **The Biogenesis of Lysosomes.** *Annu Rev Cell Biol* **5**:483-525 (1989)

Kozak M. **An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.** *Nucleic Acids Res* **15**:8125-8148 (1987)

Kozak M. **Initiation of translation in prokaryotes and eukaryotes.** *Gene* **234**:187-208 (1999)

Kramer C, Loros JJ, Dunlap JC, Crosthwaite SK. **Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*.** *Nature* **421**:948-952 (2003)

Krausse M, Hirsh D. **A trans-spliced leader sequence on actin mRNA in *C. elegans*.** *Cell* **49**:753-761 (1987)

Krenz M, Sanbe A, Bouyer-Dalloz F, Gulick J, Klevitsky R, Hewett TE, Osinska HE, Lorenz JN, Brosseau C, Federico A, Alpert NR, Warshaw DM, Perryman MB, Helmke SM, Robbins J. **Analysis of Myosin Heavy Chain Functionality in the Heart** *J Biol Chem* **278**: 17466-17474 (2003)

Kumar M and Carmichael GG. **Antisense RNA: Function and fate of duplex RNA in cells of hugher eukaryotes.** *Microbiol Mol Biol Rev* **62**:1415-1434 (1998)

Kumar S and Hedges SB. **A molecular timescale for vertebrate evolution.** *Nature* **392**:917-920 (1998).

Kumazawa Y., *et al.* **Mitochondrial molecular clocks and the origin of euteleostean biodiversity: familial radiation of perciforms may have predated the cretaceous/tertiary boundary.** In *The Biology of Biodiversity.* Edited by Kato M., *et al.* Springer-Verlag. p.35-52 (1999).

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterson RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC. **The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men.** *Nature Genet* **29**:279-286 (2001)

Kurosawa G, Yamada K, Ishiguro H, Hori H. **Hox gene complexity in Medaka fish may be similar to that in pufferfish rather than zebrafish.** *Biochem Biophys Res Commun* **260**:66-70 (1999)

Kwok C, Weller PA, Guioli S, Foster JW, Mansour S, Zuffardi O, Punnett HH, Dominguez-Steglich MA, Brook JD, Young ID, *et al.* **Mutations in SOX9, the gene responsible for Campomelic dysplasia and autosomal sex reversal.** *Am J Hum Genet* **57**:998-1001 (1995)

Labrador M, Mongelard F, Plata-Rengifo P, Baxter EM, Corces VG, Gerasimova TI. **Protein encoding by both DNA strands.** *Nature* **409**:1000 (2001).

Laemmli UK. **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* **227**:680-685 (1970)

Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T. **New microRNAs from mouse and human.** *RNA* **9**:175-179 (2003)

Lam WL, Lee T-S, Gilbert W. **Active transposition in zebrafish.** *Proc Natl Acad Sci USA* **93**:10870-10875 (1996).

Lamm GM, Lamond AI. **Non-snRNP protein splicing factors.** *Biochem Biophys Acta* **1173**:247-265 (1993)

Lamond AI. **A glimpse into the spliceosome.** *Curr Biol* **3**:62-64 (1993)

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowki J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ;International Human Genome Sequencing Consortium. **Initial sequencing and analysis of the human genome.** *Nature* **409**, 860-921 (2001)

Langenau DM, Traver D, Ferrando AA, Kutok JL, Aster JC, Kanki JP, Lin S, Prochownik E, Trede NS, Zon LI Look AT. **Myc-induced T cell leukemia in transgenic zebrafish.** *Science* **299**:887-890 (2003)

Larhammer D, Risinger C. **Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpo*.** *Mol Phylogenet Evol* **3**:59-68 (1994)

Lavigueur A, La Branche H, Kornblihtt AR, Chabot B. **A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding.** *Genes Dev* **7**:2405-2417 (1993)

Laski FA, Rio DC, Rubin GM. **Tissue specificity of *Drosophia P* element transposition is regulated at the level of mRNA splicing.** *Cell* **44**:7-19 (1986)

Latimer AJ, Dong X, Markov Y, Appel B. **Delta-Notch signaling induces hypochord development in zebrafish.** *Development* **129**:2555-2563 (2002)

Lazar MA, Hodin RA, Cardona G, Chin WW. **Gene expression from the c-erbA alpha/Rev-ErbA alpha genomic locus. Potential regulation of alternative splicing by opposite strand transcription** *J Biol Chem* **265**:12859-12863 (1990)

Leaver MJ. **A family of *Tc1*-like transposons from the genomes of fishes and frogs: evidence for horizontal transmission.** *Gene* **271**:203-214 (2001)

LeBowitz JH, Smith HQ, Rusche L, Beverley SM. **Coupling of poly(A) site selection and trans-splicing in *Leishmania*.** *Genes Dev* **7**:996-1007 (1993)

Lee J, Davidow L, Warshawsky D. ***Tsix*, a gene antisense to *Xist* at the X-inactivation centre.** *Nat Genet* **21**:400-404 (1999)

Lee JT. **Disruption of imprinted X inactivation by parent-of-origin effects at Tsix.** *Cell* **103**:17-27 (2000)

Lee MG, Van der Ploeg LH. **Transcription of protein-coding genes in trypanosomes by RNA ploymerase I.** *Annu Rev Microbiol* **51**:463-489 (1997)

Lee MP, DeBaun MR, Mitsuya K, Galonek HL, Brandenburg S, Oshimura M, Feinberg AP. **Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Widemann syndrome and is independent of insulin-like growth factor imprinting.** *Proc Natl Acad Sci USA* **96**:5203-5208 (1999)

Lee H, Kimelman D. **A dominant-negative form of p63 is required for epidermal proliferation in zebrafish.** *Dev Cell* **2**:607-616 (2002)

Lehner B, William G, Campbell RD, Sandrson CM. **Antisense transcripts in the human genome.** *Trends Genet* **18**:63-65 (2002)

Lemaire P, Kodjabachian L. **The vertebrate organizer: structure and molecules.** *Trends Genet* **12**:525-531 (1996)

Leslie GA, Clem LW. **Phylogen of immunoglobulin structure and function. 3. Immunoglobulins of the chicken.** *J Exp Med* **130**:1337-52 (1969)

Leung B, Hermann GJ, Priess JR. **Organogenesis of the *Caenorhabditis elegans* intestine.** *Dev Biol* **216**:114-134 (1999)

Levi S, Corsi B, Bosisio M, Invernizzi R, Volz A, Sanford D, Arosio P. **A human mitochondrial ferritin encoded by an intronless gene.** *J Biol Chem* **276**:24437-24440 (2001)

Levken AC, Thorpe CJ, Waxman JS, Moon RT. **Zebrafish wnt8 encodes two wnt8 proteins on a bicistronic transcript and is required for mesoderm and neurectoderm patterning.** *Dev Cell* **1**:103-114 (2001)

Levrero M, De Laurenzi V, Costanzo A, Gong J, Wang JY, Melino G. **The p53/p63/p73 family of transcrption factors: overlapping and distinct functions.** *J Cell Sci* **113**:1661-1670 (2000)

Li W-H. **Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes.** *Genetics* **95**:237-258 (1980).

Li E, Beard C, Jaenisch R. **Role for DNA methylation in genomic imprinting.** *Nature* **366**:362-365 (1993)

Li AW, Too CKL, Murphy PR. **The Basic Fibroblast Growth Factor (FGF-2) Antisense RNA (GFG) Is Translated into a MutT-Related Protein*in Vivo*.** *Biochem Biophys Res Comm* **223**: 19-23 (1996)

Li BL, Li XL, Duan ZJ, Lee O, Lin S, Ma ZM, Chang CC, Yang XY, Park JP Mohandas TK, Noll W, Chan L, Chang TY. **Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes.** *J Biol Chem* **274**:11060-11071 (1999)

Li AW and Murphy PR. **Expression of FGF-2 antisense RNA transcripts in the central nervous system: regulation of FGF-2 mRNA translation.** *Mol Cell Endocrinol* **162**:69-78 (2000)

Li Y-X, Farrell MJ, Liu R, Mohanty N, Kirby ML. **Double-stranded RNA injection produces null phenotypes in zebrafish.** *Dev Biol* **217**, 394-405 (2000).

Li WH, Gu Z, Wang H, Nekrutenko A. **Evolutionary analyses of the human genome.** *Nature* **409**:847-849 (2001)

Li L, Connelly MC, Wetmore C, Curran T, Morgan JI. **Mouse Embryos Cloned from Brain Tumors.** *Cancer Res* **63**:2733-2736 (2003)

Liao W, Bisgrove BW, Sawyer H, Hug B, Bell B, Peters K, Grunwald DJ, Stainier DY. **The zebrafish gene cloche acts upstream of a flk-1 homologue to regulate endothelial cell differentiation.** *Development* **124**:381-389 (1997).

Liao EC, Paw BH, Pratt SJ, Postlethwait JH, Zon LI. **SCL/Tal-1 transcription factor acts downstream of cloche to specifiy hematopoietic and vascular progenitors in zebrafish.** *Genes Dev* **12**:621-626 (1998)

Liau LM, Lallone RL, Seitz RS, Buznikov A, Gregg JP, Kornblum HI, Nelson SF, Bronstein JM. **Identification of human glioma-associated growth factor gene, *granulin*, using differential immuno-absorption.** *Cancer Res* **60**:1353-1360 (2000)

Lickert H, Domon C, Huls G, Wehrle C, Duluc I, Clevers H, Meyer BI, Freund JN, Kemler R. **Wnt/(beta)-catenin signalling regulates the expression of the homeobox gene Cdx1 in embryonic intestine.** *Development* **127**:3805-3813 (2000)

Lieschke GJ, Oates AC, Crowhurst MO, Ward AC, Layton JE. **Morphologic and functional characterization of granulocytes and macrophages in embryonic and adult zebrafish.** *Blood* **98**, 3087-3096 (2001)

Lieschke GJ, Oates AC, Paw BH, Thompson MA, Hall NE, Ward AC, Ho RK, Zon LI, Layton JE. **Zebrafish SPI-1 (PU.1) marks a site of myeloid development independent of primitive erythropoiesis: implications for axial patterning.** *Dev Biol* **246**:274-295 (2002)

Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP. **Vertebrate microRNA genes.** *Science* **299**:1540 (2003)

Lister JA, Robertson CP, Lepage T, Johnson SL, Raible DW. **nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate.** *Development* **126**:3757-3767 (1999)

Lister JA, Close J, Raible DW. **Duplicate *mitf* genes in zebrafish: complementary expression and conservation of melanogenic potential.** *Dev Biol* **237**:333-344 (2001)

Liu X, Jiang Q, Mansfield SG, Puttaraju M, Zhang Y, Zhou W, Cohn JA, Garcia-Blanco MA, Mitchell LG, Engelhardt JF. **Partial correction of endogenous ΔF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA *trans*-splicing.** *Nat Biotech* **20**:47-52 (2002)

Lofberg J, Collazo A. **Hypochord, an enigmatic embryonic structure: study of the axolotl embryo.** *J Morphol* **232**:57-66 (1997)

Lohe AR, Moriyama EN, Lidholm D-A, Hartl DL. **Horizontal transmission, vertical inactivation, and stochastic loss of *mariner*-like transposable elements.** *Mol Biol Evol* **12**:62-72 (1995)

Lohe AR, Hartl DL. **Autoregulation of *mariner* transposase actvity by overproduction and dominant-negative complementation.** *Mol Biol Evol* **13**:549-555 (1996)

Lohe AR, Sullivan DT, Hartl DL. **Genetic evidence for subunit interactions in the transposase of the transposable element mariner.** *Genetics* 144: 1087-1095 (1996)

Lompré AM, Nadal-Ginard B, Mahdavi V. *J Biol Chem* **259**: 6437-6446 (1984)

Lowe LA, Yamada S, Kuehn MR. **Genetic dissection of nodal function in patterning the mouse embryo.** *Development* **28**:1831-1843 (2001)

Lu CC, Brennan J, Robertson EJ. **From fertilization to gastrulation: axis formation in the mouse embryo.** *Curr Opin Genet Dev* **11**:384-392 (2001)

Lu M, Wagner GF, Renfro JL. **Stanniocalcin stimulates phosphate reabsorption by flounder renal proximal tubule in primary culture.** *Am J Physiol* **267**:R1356-R1362 (1994)

Lu R, Serrero G. **Stimulation of PC cell-derived growth factor (epithelin/granulin precursor) expression by estradiol in human breast cancer cells.** *Biochem Biophys Res Comm* **256**:204-207 (1999)

Lu R, Serrero G. **Inhibition of PC cell-derived growth factor (PCDGF, epithelin/granulin precursor) expression by antisense PCDGF cDNA transfection inhibits tumorigenicity of the human breast carcinoma cell line MDA-MB-468.** *Proc Natl Acad Sci USA* **97**:3993-3998 (2000)

Lu R, Serrero G. **Mediation odf estrogen mitogenic effect in human breast cancer MCF-7 cells by PC-cell-derived growth factor (PCDF/granulin precursor).** *Proc Natl Acad Sci USA* **98**:142-147 (2001)

Lund PK, Goodman RH, Habener JF. **Pancreatic pre-proglucagons are encoded by two separate mRNAs.** *J Biol Chem* **256**:6515-6518 (1981).

Lund PK, Goodman RH, Dee PC, Habener JF. **Pancreatic pre-proglucagon cDNA contains two glucagons-related coding sequences arranged in tandem.** *Proc Natl Acad Sci USA* **79**:345-349 (1982).

Luo G, Ivics Z, Izsvak Z, Bradley A. **Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells.** *Proc Natl Acad Sci USA* **95**:10769-10773 (1998)

Lynch M and Conery JC. **The evolutionary fate and consequences of duplicate genes.** *Science* **290**:1151-1154 (2000)

Lynch M and Force A. **The origin of interspecific genomic incompatibility via gene duplication.** *Am Nat* **156**:590-605 (2000)

Lynch M, O'Hely M, Walsh B, Force A. **The probability of preservation of a newly arisen gene duplicate.** *Genetics* **159**:1789-1804 (2001)

Maestre J, Tcheinio T, Dhelin O, Heidmann T. **mRNA retrotransposition in human cells: processed pseudogene formation.** *EMBO J* **14**:6333-6338 (1995)

Magrangeas F, Pitiot G, Dubois S, Bragado-Nilsson E, Cherel M, Jobert S, Lebeau B, Boisteau O, Lethe B, Mallet J, Jacques Y, Minvielle S. **Cotranscription and intergenic splicing of human galactose-1-phosphate uridylyltransferase and interleukin-11 receptor alpha-chain genes generate a fusion mRNA in normal cells. Implication for the production of multidomain proteins during evolution.** *J Biol Chem* **273**:16005-16010 (1998)

Malicki J. **Development of the retina.** *Methods Cell Biol* **59**:273-2999 (1999)

Mandel M, Higa A. **Calcium-dependent bacteriophage DNA infection.** *J Mol Biol* **53**:159-163 (1970)

Manley NR and Capecchi MR. **Hox group 3 paralogous genes act synergistically in the formation of somatic and neural crest-derived structures.** *Dev Biol* **192**: 274-288 (1997)

Manley NR and Capecchi MR. **Hox group 3 paralogs regulate the development and migration of the thymus, thyroid and parathyroid glands.** *Dev Biol* **195**:1-15 (1998)

Mansfield SG, Kole J, Puttaraju M, Yang CC, Garcia-Blanco MA, Cohn JA, Mitchell LG. **Repair of CFTR mRNA by spliceosome-mediated RNA *trans*-splicing.** *Gene Ther* **7**:1885-1895 (2000)

Manzon LA. **The role of prolactin in fish osmoregulation: a review.** *Gen Comp Endocrinol* **125**:291-310 (2002)

Markert JM, Fuller CM, Gillespie GY, Bubien JK, McLean LA, Hong RL, Lee K, Gullans SR, Mapstone TB, Benos DJ. **Differential gene expression profiling in human brain tumors.** *Physiol Genomics* **5**:21-33 (2001)

Maroney PA, Denker JA, Darzynkiewicz E, Laneve R, Nilsen TW. **Most mRNAs in the nematode *Ascaris lumbricoides* are trans-spliced: A role for spliced leader addition in translational efficiency.** *RNA* **1**:714-723 (1995)

Martin P, D'Souza D, Martin J, Grose R, Cooper L, Maki R, McKercher SR. **Wound healing in the PU.1 null mouse--tissue repair is not dependent on inflammatory cells.** *Curr Biol* 13:1122-1128 (2003)

Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H. **Control mechanism of the circadian clock for timing of cell division in vivo.** *Science* **302**:255-259 (2003)

Mattick JS. **Non-coding RNAs: the architects of eukaryotic complexity.** *EMBO Rep* **2**:986-991 (2001).

Matzke M, Matzke AJM, Kooter JM. **RNA: guiding gene silencing.** *Science* **293**:1080-1083 (2001)

Maxwell AI, Morrison GM, Dorin JR. **Rapid sequence divergence in mammalian beta-defensins by adaptive evolution.** *Mol Immunol* **40**:413-21 (2003)

Mayer AN, Fishman MC. ***nil per os* encodes a conserved RNA recognition motif protein required for morphogenesis and cytodifferentiation of digestive organs in zebrafish.** *Development* **130**:3917-3928 (2003)

McGowan RA and Martin CC. **DNA methylation and genome imprinting in zebrafish, Danio rerio: some evolutionary ramifications.** *Biochem Cell Biol* **75**:499-506 (1997)

McKercher SR, Torbett BE Anderson KL, Henkel GW, Vestal DJ, Baribault H, Klemsz M, Feeney AJ, Wu GE, Pauge CJ, Maki RA. **Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities.** *EMBO J* **15**:5647-5658 (1996)

McLysaght A, Hokamp K, Wolfe KH. **Extensive genomic duplication during early chordate evolution.** *Nat Genet* **31**:200-204 (2002)

Medvinsky A, Dzierzak E. **Definitive hematopoiesis is autonomously initiated by the AGM region.** *Cell* **86**:897-906 (1996)

Mefford HC, Trask BJ. **The complex structure and dynamic evolution of human subtelomeres.** *Nature Rev Genet* **3**:91-102 (2002)

Meguro M, Mitsuya K, Nomura N, Kohda M, Kashiwagi A, Nishigaki R, Yoshioka H, Nakao M, Oishi M, Oshimura M. **Large-scale evaluation of imprinting status in the Prader-Willi syndrome region: an imprtinted direct repeat cluster resemblingsmall nucleolar RNA genes.** *Hum Mol Genet* **10**:383-394 (2001)

Meischi C, de Boer M, Ahlin A, Roos D. **Intronic insertion of a LINE-1 fragment as the cause of chronic granulomatous disease.** *Eur J Hum Genet* **8**:697-703 (2000)

Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. ***roX1* RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system.** *Cell* **88**:45-457 (1998)

Meyer A, Malagra-Trillo E. **Vertebrate genomics: more fishy tales about Hox genes.** *Curr Biol* **9**:R210-R213 (1999)

Meyer A, Schartl M. **Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions.** *Curr Opin Cell Biol* **11**:699-704 (1999)

Migeon BR, Lee CH, Chowdhury AK, Carpenter H. **Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation.** *Am J Hum Genet* **71**:286-293 (2002)

Mighell AJ, Smith NR, Robinson PA, Markham AF. **Vertebrate pseudogenes.** *FEBS Lett* **468**:109-114 (2000)

Miki Y et al. **Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer.** *Cancer Res* **52**:643-645 (1992)

Miklos GLG, Rubin GM. **The role of the genome project in determining gene function: insight from model organisms.** *Cell* **86**:521-529 (1996)

Milhausen M, Nelson RG, Sather S, Selkirk M, Agabian N. **Identification of a small RNA containing the trypanosome spliced leader: A donor of shared 5' sequences of trypanosomatid mRNAs?** *Cell* **38**:721-729 (1984)

Millar R, Conklin D, Lofton-Day C, Hutchinson E, Troskie B, Illing N, Sealfon SC, Hapgood J. **A novel human GnRH receptor homolog gene: abundant and wide tissue distribution of the antisense transcript.** *J Endocrinol* **162**:117-126 (1999)

Miller CL, Burmeister M, Thompson RC. **Antisense expression of the human pro-melanin-concentrating hormone genes.** *Brain Res* **803**:86-94 (1998)

Mills AA, Zheng B, Wang XJ, Vogel H, Roop DR, Bradley A. **p63, is a p53 homologue required for limb and epidermal morphogenesis.** *Nature* **398**:708-713 (1999)

MiyazawaK, Wang Y, Minoshima S, Shimizu N, Kitamura N. **Structural organization and chromosomal localization of the human hepatocyte growth factor activator gene. Phylogenetic and functional relationship with blood coagulation factor XII, urokinase, and tissue-type plasminogen activator.** *Eur J Biochem* **258**:355-361 (1998)

Moens CB, Fritz A. **Techniques in neural development.** *Methods Cell Biol* **59**:253-272 (1999)

Mohideen M-APK, Beckwith LG, Tsao-Wu GS, Moore JL, Wong ACC, Chinoy MR, Cheng KC. **Histology-based screen for zebrafish mutants with abnormal cell differentiation.** *Dev Dyn* **228**:414-423 (2003)

Mongelard F, Labrador M, Baxter EM, Gerasimova TI, Corces VG. ***Trans*-splicing as a novel mechanism to explain interallelic complementation in Drosophila.** *Genetics* **160**:1481-1487 (2002)

Montero JA, Kilian B, Chan J, Bayliss PE, Heisenberg CP. **Phosphoinositide 3-kinase is required for process outgrowth and cell polarization of gastrulating mesendodermal cells.** *Curr Biol* **13**:1279-1289 (2003)

Montgomery MK and Fire A. **Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression.** *Trends Genet* **14**:255-258 (1998)

Moore T and Haig D. **Genomic imprinting in mammalian development: a parental tug of war.** *Trends Genet* **7**:45-49 (1991)

Moore T, Constancia M, Zubair M, Bailleul B, Feil R, Sasaki H, Reik W. **Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*.** *Proc Natl Acad Sci USA* **94**:12509-12514 (1997)

Moorman S.J. **Anatomical nomenclature for zebrafish.** *Zebrafish Science Monitor* 7-8 (2001)

Moorwood K, Charles AK, Salpekar A, Wallace JI, Brown KW, Malik K. **Antisense *WT1* transcription parallels sense mRNA and protein expression in fetal kidney and can elevate protein levels *in vitro*.** *J Pathol* **185**:352-359 (1998)

Moran JV, DeBerardinis RJ, Kazazian HH. **Exon shuffling by L1 retrotransposition.** *Science* **283**:1530-1534 (1999)

Morgan DO, Edman JC, Standring DN, Fried VA, Smith MC, Roth RA, Rutter WJ. **Insulin-like growth factor II receptor as a multifunctional binding protein.** *Nature* **329**:301-307 (1987)

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. **DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.** *Nature Genet* **31**:159-165(2002)

Moschetti R, et al. **Intra- and interspecies variation among *Bari*-1 elements of the *melanogaster* species group.** *Genetics* **150**:239-250 (1998)

Moss EG. **Noncoding RNAs: lightning strikes twice.** *Curr Biol* **10**:R436-R439 (2000)

Mount SM. **A catalogue of splice junction sequences.** *Nucleic Acids Res* **10**:459-472 (1982)

Mouse Genome Sequencing Consortium. **Initial sequencing and comparative analysis of the mouse genome.** *Nature* **420**:520-562 (2002)

Mullins MC, Hammerschmidt M, Hafter P, Nusslein-Volhard C. *Curr Biol* **4**:189-202 (1994)

Munroe SH and Lazar MA. **Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA.** *J Biol Chem* **266**:22083-22086 (1991)

Murashov AK and Wolgemuth DJ. **Sense and antisense transcripts of the developmentally regulated murine hsp70.2 gene are expressed in distinct and only partially overlapping areas in the adult brain.** *Mol Brain Res* **37**:85-95 (1996)

Maruyama K, Hartl DL. **Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*.** *J Mol Evol* **33**:514-524 (1991)

Murphy PR and Knee RS. **Identification and characterization of an antisense RNA transcript (gfg) from the human basic fibroblast growth factor gene.** *Mol Endocrinol* **8**: 852-859 (1994)

Nadeau JH and Sankoff D. **Comparable rates of of gene loss and functional divergence after genome duplications early in vertebrate evolution.** *Genetics* **147**: 1259-1266 (1997).

Nakakura N, Hietter H, Van Dorsselaer A, Luu B. **Isolation and structural determination of three peptides from the insect *Locusta migratoria*. Identification of a deoxyhexose-linked peptide.** *Eur J Biochem* **204**:147-153 (1991)

Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR. **Telomerase catalytic subunit homologs from fission yeast and human.** *Science* **277**:955-959 (1997)

Nakamura TM, Cech TR. **Reversing time: origin of telomerase.** *Cell* **92**:587-590 (1998)

Nakao K, Minobe, W, Roden R, Bristow MR, Leinwand LA. **Myosin Heavy Chain Gene Expression in Human Heart Failure.** *J Clin Invest* **100**:2362-2370 (1997)

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. **Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.** *Nature* **393**: 386-389 (1998)

Narita N et al., **Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a caee of Duchenne muscular dystrophy.** *J Clin Invest* **91**:1862-1867 (1993)

Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, Shimada A, Wada H, Kusakabe T, Suzuki N, Kinoshita M, Kanamori A, Terado T, Kimura H, Nonaka M, Shima A. **A detailed linkage map of Medaka, *Oryzias latipes*: Comparative genomics and genome evolution.** *Genetics* **154**:1773-1784 (2000)

Nasevicius A, Ekker SC. **Efficient targeted gene 'knockdown' in zebrafish.** *Nature Genetics* **26**:216-220 (2000)

Nei M, Gu X, Stinikova T. **Evolution by the birth-and-death process in multigene families of the vertebrate immune system.** *Proc Natl Acad Sci USA* **94**:7799-7806 (1997)

Nemes JP, Benzow KA, Moseley ML, Ranum LP, Koob MD. **The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel acting-binding protein (KLHL1).** *Hum Mol Genet* **9**:1543-1551 (2000)

Nepveu A and Marcu KB. **Intragenic pausing and anti-sense transcription within the murine c-myc locus.** *EMBO J* **5**:2859-2865 (1986)

Neve EP, Ingelman-Sundberg M. **Molecular basis for the transport of cytochrome P450 2E1 to the plasma membrane.** *J Biol Chem* **275**:17130-17135 (2000)

Nicholls RD and Knepper JL. **Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes.** *Annu Rev Genomics Hum Genet* **2**:153-75 (2001)

Nicholson RH, Nicholson AW. **Molecular characterization of a mouse cDNA encoding Dicer, a ribonuclease III ortholog involved in RNA interference.** *Mamm Genome* **13**:67-73 (2002)

Nielsen H, Engelbrecht J, Brunak S, von Heijne G. **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Engineering* **10**:1-6 (1997)

Nigumann P, Redik K, Matlik K, Speek M. **Many human genes are transcribed from the antisense promoter of L1 retrotransposon.** *Genomics* **79**:628-634 (2002)

Nilsen TW. **Trans-splicing of nematode premessenger RNA.** *Annu Rev Microbiol* **47**:413-440 (1993)

Nilsen TW. **RNA-RNA interactions in the spliceosome: unravelling the ties that bind.** *Cell* **78**:1-4 (1994)

Noguchi M, Miyamoto S, Silverman TA, Safer B. **Characterization of an antisense Inr element in the eIF-2 alpha gene.** *J Biol Chem* **269**:29161-29167 (1994)

Nothacker HP, Wang Z, McNeill AM, Saito Y, Merten S, O'Dowd B, Duckles SP, Civelli O. **Identification of the natural ligand of an orphan G-protein-coupled receptor involved in the regulation of vasoconstriction.** *Nat Cell Biol* **1**:383-385 (1999)

Oates AC, Bruce AEE, Ho RK. **Too much interference: injection of double-stranded RNA has nonspecific effects in the zebrafish embryo.** *Dev Biol* **224**, 20-28 (2000).

Ober EA, Field HA, Stainier DY. **From endoderm to liver and pancreas development in zebrafish.** *Mech Dev* **120**:5-18 2003-12-17 (2003)

Oettinger MA, Schatz DG, Gorka C, Baltimore D. **RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.** *Science* **248**:1517-1523 (1990)

Ogawa O, Becroft DM, Morison IM, Eccles MR, Skeen JE, Mauger DC, Reeve AE. **Constitutional relaxation of insulin-like growth factor II gene imprinting associated with Wilm's tumour and gigantism.** *Nat Genet* **5**:408-412 (1993)

Ohno S. *Evolution by Gene Duplication.* New York: Springer Verlag (1970)

Ohta T. **Role of gene duplication in evolution.** *Genome* **31**:304-310 (1989).

Ong CHP, Bateman A. **Progranulin (Granulin-epithelin precursor, PC-cell derived growth factor, Acrogranin) in proliferation and tumorigenesis.** *Histol Histopathol* **18**:1275-1288 (2003)

Orkin S. **Hematopoiesis: how does it happen?** *Curr Opin Cell Biol* **7**:870-877 (1995)

Palopoli MF, Patel NH. **Neo-Darwinian developmental evolution: can we bridge the gap between pattern and process?** *Curr Opin Genet Dev* **6**: 502-508 (1996).

Panning B and Jaenisch R. **RNA and the epigenetic regulation of X chromosome inactivation.** *Cell* **93**:305-308 (1998)

Panopoulou G, Hennig S, Groth D, Krause A, Poustka AJ, Herwig R, Vingron M, Lehrach H. **New evidence for genome-wide duplications at the origin of vertebrates using an Amphioxus gene set and completed animal genomes.** *Genome Res* **13**:1056-1066 (2003)

Papp B, Pál C, Hurst LD. **Evolution of cis-regulatory elements in duplicated genes of yeast.** *Trends Genet* **19**:417-422 (2003)

Pardue M-L, Danilevskaya K, Lowenhaupt K, Slot F, Traverse KL. **Drosophila telomeres: new views on chromosome evolution.** *Trends Genet* **12**:48-52 (1996)

Parker LH, Zon LI, Stainier DY. **Vascular and blood gene expression.** *Methods Cell Biol* **59**:313-336 (1999)

Parnell PG, Wunderlich J, Carter B, Halper J. **Transforming growth factor e: amino acid analysis and partial amino acid sequence.** *Growth Factors* **7**:65-72 (1992)

Parsons MJ, Campos I, Hirst EMA, Stemple DL. **Removal of dystroglycan causes severe muscular dystrophy in zebrafish embryos.** *Development* **129**:3505-3512 (2002)

Passegué E and Wagner EF. **JunB suppresses cell proliferation by transcriptional activation of p16^{INK4a} expression.** *EMBO J* **19**:2969-2979 (2000)

Passegué E, Jochum W, Behrens A, Ricci R, Wagner EF. **JunB can substitute for Jun in mouse development and cell proliferation.** *Nat Genet* **30**:158-166 (2002).

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Haywards DC, Ball EE, Degnan B, Muller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G. **Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA.** *Nature* **408**:86-89 (2000)

Pedone PV, Pikaart MJ, Cerrato F, Vernucci M, Ungaro P, Bruni CB, Riccio A. **Role of histone acetylation and DNA methylation in the maintenance of the imprinted expression of the H19 and Igf2 genes.** *FEBS Lett* **458**:45-50 (1999)

Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. **The genetic architecture of divergence between threespine stickleback species.** *Nature* **414**:901-905 (2001).

Pera EM, Wessely O, Li S-Y, De Robertis EM. **Neural and head induction by insulin-like growth factor signals.** *Dev Cell* **1**:655-665 (2001)

Peri S, Pandey A. **A reassessment of the translation initiation codon in vertebrates.** *Trends Genet* **17**:685-687 (2001)

Perry KL, Watkins KP, Agabian N. **Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader.** *Proc Natl Acad Sci USA* **84**:8190-8194 (1987)

Pfarr CM, Mechta F, Spyrou G, Lallemand D, Carillo S, Yaniv M. **Mouse JunD negatively regulates fibroblast growth and antagonizes transformation by ras.** *Cell* **76**:747-70 (1994)

Pham VN, Roman BL, Weinstein BM. **Isolation and expression analysis of three zebrafish angiopoietin genes.** *Dev Dyn* **221**:470-474 (2001)

Piatigorsky J. **Gene sharing by δ-crystallin and argininosuccinate lyase.** *Proc Natl Acad Sci USA* **85**:3479-3483 (1988)

Piatigorsky J and Wistow G. **The recruitment of crystallins: New functions precede gene duplication.** *Science* **252**, 1078-1079 (1991)

Pirrotta V. ***Trans*-splicing in *Drosophila*.** *BioEssays* **24**:988-991 (2002)

Plasterk RHA, Izsvak Z, Ivics Z. **Resident aliens: the *tc1/mariner* superfamily of transposable elements.** *Trends Genet* **15**:326-332 (1999)

Plowman GD, Green JM, Neubauer MG, Buckley SD, McDonald VL, Todaro GJ, Shoyab M. **The epithelin precursor encodes two proteins with opposing activities on epithelial cell growth.** *J Biol Chem* **267**:13073-13078 (1992)

Ponting CP. **Plagiarized bacterial genes in the human book of life.** *Trends Genet* 17:235-237 (2001)

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan Y-L, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS. **Zebrafish comparative genomics and the origins of vertebrate chromosomes.** *Genome Res* **10**:1890-1902 (2000)

Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Egan ES, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar TS, Yelick P, Beier D, Joly JS, Larhammar D, Rosa F, et al. **Vertebrate genome evolution and the zebrafish gene map.** *Nat Genet* **18**:345-349 (1998)

Potter SS and Branford WW. **Evolutionary conservation and tissue-specific processing of Hoxa11 antisense transcripts.** *Mammal Gen* **9**:799-806 (1998)

Poupard G, Andre M, Durliat M, Ballagny C, Boeuf G, Babin PJ. **Apolipoprotein E gene expression correlates with endogenous lipid nutrition and yolk syncytial layer lipoprotein synthesis during fish development.** *Cell Tissue Res* **300**:251-261 (2000)

Pourquie O. **Vertebrate somitogenesis.** *Annu rev Cell Dev Biol* **17**:311-350 (2001)

Prak ETL, Kazazian Jr HH. **Mobile elements and the human genome.** *Nature Rev Genet* **1**:134-144 (2000)

Pubesque MJ, Coulier F, Birnbaurm D, Pontarotti P. **Ancient large-scale genome duplications: phylogenetic linkage analyses shed light on chordate genome evolution.** *Mol Biol Evol* **15**:1145-1159 (1998)

Puttaraju M, DiPasquale J, Baker CC, Mitchell LG, Garcia-Blanco MA. **Messenger RNA repair and restoration of protein function by spliceosome-mediated RNA *trans*-splicing.** *Mol Ther* **4**:105-114 (2001)

Puttaraju M, Jamison SF, Mansfield SG, Garcia-Blanco MA, Mitchell JG. **Spliceosome-mediated RNA *trans*-splicing as a tool for gene therapy.** *Nat Biotech* **17**:246-252 (1999)

Query CC, Bentley RC, Keene JD. **A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70 K U1snRNP protein.** *Cell* **57**:89-101 (1989)

Rainier S, Johnson LA, Dobry CJ, Ping AJ, Grundy PE, Feinberg AP. **Relaxation of imprinted genes in human cancer.** *Nature* **362**:747-749 (1993)

Rajkovic A, Davis RE, Simonsen JN, Rottman FM. **A spliced leader is present on a subset of mRNAs from the human parasite Schistosoma mansoni.** *Proc Natl Acad Sci USA* **87**:-8883 (1990)

Ratcliff F, Harison BD, Baulcombe DC. **A similarity between viral defense and gene silencing in plants.** *Science* **276**:1558-1560 (1997)

Regev A, Lamb MJ, Jablonka E. **The role of DNA methylation in invertebrates: developmental regulation or genome defense** ? *Mol Biol Evol* **15**:880-891 (1998)

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. **The 21-nucleotide let-*7* RNA regulates developmental timing in *C. elegans***. *Nature* **403**:901-906 (2000)

Reiser PJ, Portman MA, Ning X-H, Moravec CS. **Human cardiac myosin heavy chain isoforms in fetal and failing adult atria and ventricles**. *Am J Physiol Heart Circ Physiol* **280**:H1814-H1820 (2001)

Reiter LT, Marukami T, Koeuth T, Pentao L, Muzny DM, **Gibbs RA, Lupski JR. A recombination hotspot responsible for two inherited peripheral neuropathies is located near a *mariner* transposon-like element**. *Nat Genet* **12**:288-297 (1996)

Requena JM, Quijada L, Soto M, Alonso C. **Conserved nucleotides surrounding the trans-splicing acceptor site and thje translation initiation codon in Leishmania genes**. *Exp Parasitol* **103**:78-81 (2003)

Riggins GJ, Thiagalingam S, Rozenblum E, Weinstein CL, Kern SE, Hamilton SR, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. **Mad-related genes in the human**. *Nat Genet* **13**:347-349 (1996)

Riggins GJ, Kinzler KW, Vogelstein B, Thiagalingam S. **Frequency of Smad gene mutations in human cancers**. *Cancer Res* **57**:2578-80 (1997)

Riggs AD and Pfeifer GP. **X-inactivation and cell memory**. *Trends Genet* **8**:169-174 (1992)

Rihs HP, Jans DA, Fan H, Peters R. **The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking the nuclear localization sequence of the SV40 T-antigen.** *EMBO J* **10**:633-639 (1991)

Ridanpaa M, van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, vanVenrooij W, Pruijn G, Salmela R, Rockas S, Makitie O, Kaitila I, de la Chapelle A. **Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia.** *Cell* **104**:195-203 (2001)

Robertson HM. **The *mariner* transposable element is widespread in insects**. *Nature* **362**:241-245 (1993)

Robertson HM, Lampe DJ. **Distribution of transposable elements in arthropods.** *Annu Rev Entomol* **40**:333-357 (1995)

Rodin SN and Riggs AD. **Epigenetic silencing may aid evolution by gene duplication.** *J Mol Evol* **56**:718-729 (2003)

Rodriguez-Niedenfuhr M, Papoutsi M, Christ B, Nicolaides KH, von Kaisenberg CS, Tomarev SI, Wilting J. **Prox1 is a marker of ectodermal placodes, endodermal compartments, lymphatic endothelium and lymphangioblasts.** *Anat Embryol (Berl)* **204**:399-406 (2001)

Rogers CS, Vanoye CG, Sullenger BA, George Jr AL. **Functional repair of a mutant chloride channel using trans-splicing ribozyme.** *J Clin Invest* **110**:1783-1789 (2002)

Romani A, Guerra E, Trerotola M, Alberti S. **Detection and analysis of spliced chimeric mRNAs in sequence databanks.** *Nucleic Acids Res* **31**:e17 (2003)

Rossignol F, Vaché C, Clottes E. **Natural anisense transcripts of hypoxia-inducible factor 1alpha are detected in different normal and tumour human tissues.** *Gene* **299**:135-140 (2002)

Rougeulle C, Cardoso C, Fontés M, Colleaux, L Lalande M. **An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript.** *Nature Genet* **19**:15-16 (1998)

Rougeulle C and Heard E. **Antisense RNA in imprinting: spreading silence through *Air*.** *Trends Genet* **18**:434-437 (2002)

Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LS, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJ, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XH, Lewis S. **Comparative genomics of the eukaryotes.** *Science* **287**:2204-2215 (2000)

Rudnicki MA, Braun T, Hinuma S, Jaenisch R. **Inactivation of MyoD in mice leads up to up-regulation of myogenic HLH gene Myf-5 and results in apparently normal muscle development.** *Cell* **71**:383-390 (1992)

Rudnicki MA, Schnegelsberg PNJ, Stead RH, Braun T, Arnold H-H, Jaenisch R. **MyoD or Myf-5 is required for the formation of skeletal muscle.** *Cell* **75**:1351-1359 (1993)

Ruiz I Altaba A. **Deconstructiong the organizer.** *Nature* **391**:748-749 (1998)

Ryu S-L, Murooka Y and Kaneko Y. **Reciprocal translocation at duplicated RPL2 loci might cause speciation of *Saccharomyces bayanus* and *Saccharomyces cerevisiae*.** *Curr Genet* **33**:345-351 (1998)

Sado T, Wang Z, Sasaki H, Li E. **Regulation of imprinted X-chromosome inactivation in mice by Tsix.** *Development* **128**:1275-1286 (2001)

Sambrook J, Fritsch EF, Maniatis T. **Molecular cloning - A laboratory manual.** Second edition. Cold Spring Harbor Laboratory Press (1989)

Sampath K, Rubinstein, AL, Cheng AMS, Liang JO, Fekany K, Solnica-Krezel L, Korzh V, Halpern ME and Wright CVE. **Induction of the zebrafish ventral brain and floor plate requires cyclops/nodal signaling.** *Nature* **395**:185-189 (1998)

Santini F, Tyler JC. **A new phylogenetic hypothesis for the order Tetraodontiformes (Teleostei, Pisces), with placement of the most fossil basal lineages.** *Am Zool* **39**:10A (1999)

Sato TN, Tozawa Y, Deutsch U, Wolburg-Buchholz K, Fujiwara Y, Gendron-Maguire M, Gridley T, Wolburg H, Risau W, Qin Y. **Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation.** *Nature* **376**:70-74 (1995)

Schmid CW. **Alu: a parasite's parasite.** *Nature Genet* **35**:15-16 (2003)

Schmitt E, Dowling J. **Early eye morphogenesis in the zebrafish, Brachydanio rerio.** *J Comp Neurol* **344**:532-542 (1994)

Schohl A, Fagotto F. **A role for maternal beta-catenin in early mesoderm induction in Xenopus.** *EMBO J* **22**:3303-3313 (2003)

Schorpp-Kistner M, Wang Z-Q, Angel P, Wagner EF. **JunB is essential for mammalian placentation.** *EMBO J* **18**: 934-948 (1999)

Schreiber M, Kolbus A, Liu F, Szabowski A, Mohle-Steinlein U, Tian J, Karin M, Angel P, Wagner EF. **Control of cell cycle progression by c-Jun is p53 dependent.** *Genes Dev* **13**: 607-619 (1999)

Schulte-Merker S, Ho RK, Hermann BG, Nusslein-Volhard C. **The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo.** *Development* **116**:1021-1032 (1992)

Schulte-Merker S, van Eeden FJM, Halpern ME, Kimmel CB, Nusslein-Volhard C. ***no tail* (*ntl*) is the zebrafish homologue of the mouse *T* (Brachyury) gene.** *Development* **120**:1009-1015 (1994)

Schultz RM. **Regulation of zygotic gene activation in the mouse.** *BioEssays* **15**:531-538 (1993)

Scott J, Urdea M, Quiroga M, Sanchez-Pescador R, Fong N, Selby M, rutter JW, Bell GI. **Structure of amouse submaxillary messenger RNA encoding epidermal rowth factor and sevenrelated proteins.** *Science* **221**:236-240 (1983)

Serluca FC, Fishman MC. **Pre-pattern in the pronephric kidney field of zebrafish.** *Development* **128**:2233-2241 (2001)

Serrero G. **Autocrine growth factor revisited: PC-cell-derived growth factor (progranulin), a critical player in breast cancer tumorigenesis.** *Biochem Biophys Res Commun* **308**:409-413 (2003)

Sharp PA. **RNA interference – 2001.** *Genes Dev* **15**:485-490 (2001)

Sharp PA, Burge CB. **Classification of introns: U2-type or U12-type.** *Cell* **91**:875-879 (1997)

Sharp PA, Konarska MM. **Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs.** *Cell* **46**:211 (1986)

Shi M, Yan X, Ryan DH, Harris RB. **Identification of urocortin mRNA antisense transcripts in rat tissue.** *Brain Res Bull* **53**:317-324 (2000)

Shi H, Wei SH, Leu Y-W, Rahmatpanah F, Liu JC, Yan PS, Nephew KP, H-M T Huang. **Triple Analysis of the Cancer Epigenome: An Integrated Microarray System for Assessing Gene Expression, DNA Methylation, and Histone Acetylation.** *Cancer Res* **63**:2164-2171 (2003)

Shim S, Lee S-K, Han J-K. **A novel family of retrotransposons in *Xenopus* with a developmentally regulated expression.** ***genesis*** **26**:198-207 (2000)

Shimeld SM and Holland PW. **Vertebrate innovations.** *Proc Natl Acad Sci USA* **97**:4449-4452 (2000)

Shivdasani RA. **Molecular regulation of vertebrate early endoderm development.** *Dev Biol* **249**:191-203 (2002)

Sidow A. **Gen(om)e duplications in the evolution of early vertebrates.** *Curr Opin Genet Dev* **6**: 715-722 (1996)

Simmen MW, Leitgeb S, Charlton J, Jones SJ, Harris BR, Clark VH, Bird A. **Nonmethylated transposable elements and methylated genes in a chordate genome.** *Science* **283**:1164-1167 (1999)

Simons RW. **Naturally occurring antisense RNA control – a brief review.** *Gene* **72**:35-44 (1988)

Singer MS, Gottschling DE. **TLC1: template RNA component of Saccharomyces cerevisiae telomerase.** *Science* **266**:404-409 (1994)

Skrabanek L, Wolfe KH. **Eukaryote genome duplication – where's the evidence?** *Curr Opin Genet Dev* **8**: 694-700 (1998)

Slack F and Ruvkun G. **Temporal pattern formation by heterochronic genes.** *Annu Rev Genet* **31**:611-634 (1997)

Sleutels F, Barlow DP, Lyle R. **The uniqueness of the imprinting mechanism.** *Curr Opin Genet Dev* **10**:229-233 (2000)

Sleutels F, Zwart R, Barlow DP. **The non-coding Air RNA is required for silencing autosomal imprinted genes.** *Nature* **415**:810-813 (2002)

Smit AF, Riggs AD. **Tiggers and DNA transposon fossils in the human genome.** *Proc Natl Acad Sci USA* **93**:1443-1448 (1996)

Smit AFA. **Interspersed repeats and other mementos of transposable elements in mammalian genomes.** *Curr Opin Genet Dev* **9**:657-663 (1999)

Smith CM, Steitz JA. **Classification of *gas5* as a multi-small-nucleolar RNA (snoRNA) host gene and a member of the 5'terminal oligopyrimidine gene family reveals common features of snoRNA host genes.** *Mol Cell Biol* **18**:6897-6909 (1998)

Smith CM, Steitz JA. **Sno storm in the nucleolus: new role for myriad small RNPs.** *Cell* **89**:669-672 (1997)

Smith SD, Kelley PM, Kenyon JB and Hoover D. **Tietz syndrome (hypopigmentation/deafness) caused by mutation of *MITF*.** *J Med Genet* **37**:446-448 (2000)

Snell EA, Scemama JL, Stellwag EJ. **Genomic organization of the Hoxa4-Hoxa10 region from *Morone saxatilis*: implications for Hox gene evolution among vertebrates.** *J Exp Zool (or Mol Dev Evol)* **285**:41-49 (1999)

Solnick D. **Trans splicing of mRNA precursors.** *Cell* **42**:157-164 (1985)

Solnick D. **Does trans splicing in vitro require base pairing between RNAs?** *Cell* **44**:211 (1986)

Spina M, Merlo-Pich E, Chan RK, Basso AM, Rivier J, Vale W, Koob GF. **Appetite-suppressing effects of urocortin, a CRF-related neuropeptide.** *Science* **273**:1561-1564 (1996)

Spring J. **Vertebrate evolution by interspecific hybridisation - are we polyploid ?** *FEBS Lett.* **400**:2-8 (1997)

Stainier DYR. **A glimpse into the molecular entrails of endoderm formation.** *Genes Dev* **16**:893-907 (2002)

Staknis D, Reed R. **SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex.** *Mol Cell Biol* **14**:7670-7682 (1994)

Stankiewicz JR, Lupski JR. **Genome architecture, rearrangements and genomic disorders.** *Trends Genet* **18**:74-82 (2002)

Stellwag EJ. **Hox gene duplication in fish.** *Semin Cell Dev Biol* **10**:531-540 (1999)

Sterrer S, Koenigstorfer A, Hoffmann W. **Biosynthesis and expression of ependymin homologous sequences in zebrafish brain.** *Neuroscience* **37**: 277-284 (1990)

Stevens L. **Twin peaks: Spitz and Argos star in patterning of the Drosophila egg.** *Cell* **95**:291-294 (1998)

Strahle U, Blader P, Henrique D, Ingham PW. **Axial, a zebrafish gene expressed along the developing body axis, shows altered expression in cyclops mutant embryos.** *Genes Dev* **7**:1436-1446 (1993)

Studer M, Lumsden A, Ariza-McNaughton L, Bradley A, Krumlauf R. **Altered segmental identity and abnormal migration of motor neurons in mice lacking Hoxb1.** *Nature* **384**:630-634 (1996)

Stuermer C. **Retinotopic organization of the developing retinotectal projection in the zebrafish embryo.** *J Neurosci* **8**:43-4530 (1988)

Sudhof TC, Russell DW, Goldstein JL, Brown MS, Sanchez-Pescador R, Bell GI. **Cassette of eight exons shared by genes for LDL receptor and EGF precursor.** *Science* **17**:893-5 (1985)

Sullenger BA, Gilboa E. **Emerging clinical applications of RNA.** *Nature* **418**:252-258 (2002)

Sullivan PM, Petrusz P, Szpirer C, Joseph DR. **Alternative processing of androgen-binding protein RNA transcripts in fetal rat liver. Identification of a transcript formed by trans splicing.** *J Biol Chem* **266**:143-154 (1991)

Summerton J, Weller D. **Morpholino antisense oligomers: design, preparation and properties.** *Antisense Nuc Acid Drug Dev* **7**:187-195 (1997)

Sutterluety H, Bartl S, Doetzlhofer A, Khier H, Wintersberger E, Seiser C. **Growth-regulated antisense transcription of the mouse thymidine kinase gene.** *Nucleic Acids Res* **26**:4989-4995 (1998)

Schneider C, King R, Philipson L. **Genes specifically expressed at growth arrest of mammalian cells.** *Cell* **54**:787-793 (1988)

Schohl A, Fagotto F. **Beta-catenin, MAPK, and Smad signalling during early Xenopus development.** *Development* **129**:37-52 (2002)

Schwann U et al., **Positional cloning of the gene for X-linked retinitis pigmentosa 2.** *Nature Genet* **19**:327-332 (1998)

Tabara H, Grishok A, Mello CC. **RNAi in *C. elegans*: soaking in the genome sequence.** *Science* **282**:430-431 (1998)

Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC. **The rde-1 gene, RNA interference, and transposon silencing in C. elegans.** *Cell* **99**:123-132 (1999)

Tacke R, Manley JL. **The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities.** *EMBO J* **14**:3540-3551 (1995)

Takahara T, Kanazu SI, Yanagisawa S, Akanuma H. **Heterogeneous Sp1 mRNAs in human HepG2 cells include a product of homotypic trans-splicing.** *J Biol Chem* **275**:38067-38072 (2000)

Talmadge K, Vamvakopoulos NC, Fiddes JC. **Evolution of the genes for the β subunits of human chorionic gonadotropin luteinizing hormone.** *Nature* **307**:37-40 (1984)

Tam PPL, Kanai-Azuma M, Kanai Y. **Early endoderm development in vertebrates: lineage differentiation and morphogenetic function.** *Curr Opin Genet Dev* **13**:393-400 (2003)

Tang X, Nakata Y, Li H-O, Zhang M, Gao H, Fujita A, Sakatsume O, Ohta T, Yokoyama K. **The optimization of preparations of competent cells for transformation of *E.coli*.** *Nucleic Acids Res* **22**:2857-2858 (1994)

Tassabehji M, Newton VE and Read AP. **Waardenburg syndrome type2 caused by mutations in the human microphtalmia (MITF) gene.** *Nature Genet* **8**:251-255 (1994)

Tarn WY, Steitz JA. **Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns.** *Science* **273**: 1824-1832 (1996)

Tavernarakis N, Wang SL, Dorovkov M, Ryanzanov A, Driscoll M. **Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes.** *Nat Genet* **24**:180-183 (2000)

Tazi J and Bird A. **Alternative chromatin structure at CpG islands.** *Cell* **60**:909-920 (1990)

Tchenio T, Casella J-F, Heidmann T. **Members of the SRY family regulate the human LINE retrotransposons.** *Nucleic Acids Res* **28**:411-415 (2000)

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I and II Team. **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.** *Nature* **420**:563-573 (2002)

Thiery JP. **Cell adhesion in development: a complex signaling network.** *Curr Opin Genet Dev* **13**:365-371 (2003)

Thisse C, Thisse B, Schilling TF, Postlethwait JH. **Structure of the zebrafish snail1 gene and its expression in wild-type, *spadetail* and *no tail* mutant embryos.** *Development* **119**:1203-1215 (1993)

Thisse B, Pflumio S, Furthauer M, Loppin B, Hever V, Degrave A, Woehl R, Lux A, Steffan T, Charbonnier XQ, Thisse C. **Expression of the zebrafish genome during embryogenesis.** ZFIN Direct DATA Submission (2001)

Thomas G, Zelenka PS, Cuthbertson RA, Norman BL, Piatigorsky J. **Differential expression of the two delta-crystallin/argininosuccinate lyase genes in lens, heart, and brain of chicken embryos.** *New Biol* **2**:903-914 (1990)

Thompson MA, Ransom DG, Pratt SJ, MacLennan H, Kieran MW, Detrich III HW, Vail B, Huber TL, Paw B, Brownlie AJ, Oates AC, Friztz A, Gates MA, Amores A, Bahary N, Talbot WS, Her H, Beir DR, Postlethwait JH, Zon LI. **The cloche and spadetail genes differentially affect hematopoiesis and vasculogenesis.** *Dev Biol* **125**:248-269 (1998)

Thummel, C. S. **Molecular mechanisms of developmental timing in *C. elegans* and *Drosophila*.** *Dev Cell* **1**, 453-465 (2001)

Tollervey D, Kiss T. **Function and synthesis of small nucleolar RNAs.** *Curr Opin Cell Biol* **9**:337-342 (1997)

Topol L, Jiang X, Choi H, Garrett-Beal PJ, Carolan PJ, Yang Y. **Wnt-5a inhibits the canonical Wnt pathway by promoting GSK-3 independent β-catenin degradation.** *J Cell Biol* **162**:899-908 (2003)

Tosic M, Roach A, de Rivaz JC, Dolivo M, Matthieu JM. **Post-transcriptional events are responsible for low expression of myelin basic protein in myelin deficient mice: role of natural antisense RNA.** *EMBO J* **9**:401-406 (1990)

Trede NS, Zon LI. **Development of T-cells during fish embryogenesis.** *Dev Comp Immunol* **22**:253-263 (1998)

Trede NS, Zapata A, Zon LI. **Fishing for lymphoid genes.** *Trends Immunol* **22**:302-307 (2001)

Trinkaus JP. **Formation of protrusions of the cell surface during tissue cell movement.** *Prog Clin Biol Res* **41**:887-906 (1980)

Tufarelli C, Sloane Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR. **Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease.** *Nature Genet* **34**:157-165 (2003)

Tycowski K, Yao ZH, Graham P, Steitz J. **Modification of U6 spliceosomal RNA is guided by other small RNAs.** *Mol Cell* **2**:629-638 (1998)

Uyeno T and Smith G. **Tetraploid origin of the karyotype of catostomid fishes.** *Science* **175**:644-646 (1972)

Vale W, Spiess J, Rivier C, Rivier J. **Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and β-endorphin.** *Science* **213**:1394-1397 (1981)

Vandenberghe AE, Meedel TH, Hastings KEM. **mRNA 5'-leader *trans*-splicing in the chordates.** *Genes Dev* **15**:294-303 (2001)

Van Doren K, Hirsh D. ***Trans*-spliced leader RNA exists as small nuclear ribonucleoprotein particles in *Caenorhabditis elegans*.** *Nature* **335**:556-559 (1988)

van Luenen HG, Colloms SD, Plasterk RH. **The mechanism of transposition of Tc3 in *C. elegans*.** *Cell* **79**:293-301 (1994)

Varlet I, Collignon J, Norris DP, Robertson EJ. **Nodal signaling and axis formation in the mouse.** *Cold Spring Harb Symp Quant Biol* **62**:105-13 (1997)

Vaughan J, Donaldson C, Bittencourt J, Perrin MH, Lewis K, Sutton S, Chan R, Turnbull AV, Lovejoy D, Rivier C, et al. **Urocortin, a mammalian neuropeptide related to fish urotensin I and to corticotropin-releasing factor.** *Nature* **378**:287-292 (1995).

Vasant G, Reynolds WF. **The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element.** *Proc Natl Acad Sci USA* **92**:8229-8233 (1995)

Vastenhouw NL, Fischer SE, Robert VJ, Thijssen KL, Fraser AG, Kamath RS, Ahringer J, Plasterk RH. **A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*.** *Curr Biol* **13**:1311-1316 (2003)

Vellard M, Soret J, Viegas-Pequinot E, Galibert F, Nguyen VC, Dutrillaux B, Perbal B. **C-myb proto-oncogene: evidence for intermolecular recombination of coding sequences.** *Oncogene* **6**:505-514 (1991)

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly

M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. **The Sequence of the Human Genome.** *Science* **291**:1304-1351 (2001)

Vogel G. **Genomics: Sanger will sequence zebrafish genome.** *Science* **290**:1671 (2000)

Vogelstein B, Gillespie D. **Preparative and analytical purification of DNA from agarose.** *Proc Natl Acad Sci USA* **76**:615-619 (1979)

Volk R, Koster M, Poting A, Hartmann L, Knochel W. **An antisense transcript from the Xenopus laevis bFGF gene coding for an evolutionarily conserved 24 kd protein.** *EMBO J* **8**:2983-2988 (1989)

von Heijne G. **A new method for predicting signal sequence cleavage sites.** *Nucleic Acid Res* **14**:4683-4690 (1986)

Vulliamy T, Marrone A, Goldman F, Dearlove A, Bessler M, Mason PJ, Dokal I. **The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenital.** *Nature* **413**:432-435 (2001).

Wada H, Saiga H, Satoh N, Holland PWH. **Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian *Pax-2/5/8*, *Hox* and *Otx* genes.** *Development* **125**:1113-1122 (1998)

Wagner GF, Milliken C, Friesen HG, Copp DH. **Studies on the regulation and characterization of plasma stanniocalcin in rainbow trout.** *Mol Cell Endocrinol* **79**:129-138 (1991)

Wallace KN, Pack M. **Unique and conserved aspects of gut development in zebrafish.** *Dev Biol* **255**:12-29 (2003)

Wallis M. **The expanding growth hormone/prolactin family.** *J Mol Endocrinol* **9**:185-188 (1992)

Wallis M. **Episodic evolution of protein hormones in mammals.** *J Mol Evol* **53**:10-18 (2001)

Walsh CP, Chaillet JR, Bestor TH. **Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.** *Nature Genet* **20**:116-117 (1998)

Walter P, Blobel G. **Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum.** *Nature* **299**:691-698 (1982)

Wang H, Long Q, Marty SD, Sassa S, Lin S. **A zebrafish model for hepatoerythropoietic porphyria.** *Nat Genet* **20**:239-243 (1998)

Wang WH, Duan JX, Vu TH, Hoffman AR. **Increased expression of the insulin-like growth factor-II gene in Wilms' tumor is not dependent on loss of genomic imprinting or loss of heterozygosity.** *J Biol Chem* **271**:27863-27870 (1996)

Wang Y and Gu X. **Evolutionary patterns of gene families generated in the early stage of vertebrates.** *J Mol Evol* **51**:88-96 (2000)

Warga RM, Kimmel CB. **Cell movements during epiboly and gastrulation in zebrafish.** *Development* **108**:569-580 (1990)

Warga RM, Nusslein-Volhard C. **Origin and development of the zebrafish endoderm.** *Development* **126**:827-838 (1999).

Wargelius A, Ellingsen S, Fjose A. **Double-stranded RNA induces specific developmental defects in zebrafish embryos.** *Biochem Biophys Res Commun* **263**:156-161 (1999).

Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. **Identification of novel small RNAs using comparative genomics and microarrays.** *Genes Dev* **15**:1637-1651 (2001)

Wassenegger M, Heimes S, Riedel L, Sanger HL. **RNA-directed *de novo* methylation of genomic sequences in plants.** *Cell* **76**:567-576 (1994)

Watanabe Y, Yamamoto M. ***S. pombe* mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA.** *Cell* **78**:487-498 (1994)

Weidinger and Moon. **When Wnts antagonize Wnts.** *J Cell Biol* **162**:753-755 (2003)

Weinstein BM, Schier AF, Adbelilah S, Malicki J, Solnica-Krezel L, Stemple DL, Stainier DY, Zwartkruis F, Driever W, Fishman MC. **Hematopoietic mutations in the zebrafish.** *Development* **123**:303-309 (1996)

Weinstein L, Steitz JA. **Guided tours: from precursor snoRNA to functional snoRNP.** *Curr Opin Cell Biol* **11**:378-384 (1999)

Weintraub H. **The MyoD family and myogenesis: Redundancy, networks and thresholds.** *Cell* **75**:1241-1244 (1993)

Weisinger G, Zinder O, DeCristofaro JD, LaGamma EF. **Novel transcriptional mechanisms are involved in regulating preproenkephalin gene expression in vivo.** *Biochem Biophys Res Comm* **246**:524-531 (1998)

Weinstein L, Steitz JA. **Guided tours: from precursor snoRNA to functional snoRNP.** *Curr Opin Cell Biol* **11**:378-384 (1999)

Westfall TA, Hjertos B, Twedt J, Brrimeyer J, Gladon A, Olberding M, Furutani-Seiki, Slusarski DC. **Wnt-5/pipetail functions in vertbrate axis foamtion as a negative regulator of Wnt/β-catenin activity.** *J Cell Biol* **162**:889-898 (2003)

Whitman M. **Nodal signalling in early vertebrate embryos: themes and variations.** *Dev Cell* **1**:605-617 (2001)

Wianny F, Zernicka-Goetz M. **Specific interference with gene function by double-stranded RNA in early mouse development.** *Nat Cell Biol* 2:70-75 (2000)

Willett CE, Cortes A, Zuasti A, Zapata AG. **Early hematopoiesis and developing lymphoid organs in the zebrafish.** *Dev Dyn* **214**:323-336 (1999)

Willett CE, Kawasaki H, Amemiya CT, Lin S, Steiner LA. **Ikaros expression as a marker for lymphoid progenitors during zebrafish development.** *Dev Dyn* **222**:694-698 (2001)

Wilson PA, Hemmati-Brivanlou A. **Vertebrate neural induction: Inducers, inhibitors, and a new synthesis.** *Neuron* **18**:699-710 (1997)

Wise JA. **Guides to the heart of the spliceosome.** *Science* **262**:1978-1979 (1993)

Wittbrodt J, Meyer A, Schartl M. **More genes in fish ?** *BioEssays* **20**:511-515 (1998)

Wolfe KH. **Yesterday's polyploids and the mystery of diploidization.** *Nature Rev Genet* **2**:333-341 (2001)

Woo K, Fraser SE. **Specification of zebrafish nervous system by nonaxial signals.** *Science* **277**:254-257 (1997)

Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan Y-L, Huang H, Postlethwait JH, Talbot WS. **A comparative map of the zebrafish genome.** *Genome Res* **10**:1903-1914 (2000).

Wride MA. **Minireview: Apoptosis as seen through a lens.** *Apoptosis* **5**:203-209 (2000)

Wu JY, Maniatis T. **Specific interactions between proteins implicated in splice site selection and regulated alternative splicing.** *Cell* **75**:1061-1070 (1993)

Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP. **Imprinted expression of the Igf2r gene depends on an intronic CpG island.** *Nature* **389**:745-749 (1997)

Xia X, Serrero G. **Identification of cell surface binding sites for PC-cell-derived growth factor, PCDGF, (Epithelin/Granulin Precursor) on epithelial cells and fibroblasts.** *Biochem Biophys Res Comm* **245**:539-543 (1998)

Xu S-Q Tang D, Chamberlain S, Pronk G, Masiarz FR, Kaur S, Prisco M, Zanocco-Marani T, Baserga R. **The granulin/epithelin precursor abrogates the requirement for the insulin-like growth factor 1 receptor for growth in vitro.** *J Biol Chem* **273**:20078-20083 (1998)

Yamashita S, Miyagi C, Carmany-Rampey A, Shimizu T, Fujii R, Schier AF, Hirano T. **Stat3 Controls Cell Movements during Zebrafish Gastrulation.** *Dev Cell* **2**:363-375 (2002)

Yan Y-L, Miller CT, Nissen R, Singer A, Liu D, Kirn A, Draper B, Willoughby J, Morcos PA, Amsterdam A, Chung B-C, Westerfield M, Haffter P, Hopkins N, Kimmel C and Postlethwait JH. **A zebrafish sox9 gene required for cartilage morphogenesis.** *Development* **129**:5065-5079 (2002)

Yang A, Kaghad M, Wang Y, Gillett E, Fleming MD, Dotsch V, Andrews NC, Caput D, McKeon F. **p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities.** *Mol Cell* **2**: 305-316 (1998)

Yang A, Schweitzer R, Sun D, Kaghad M, Walker N, Bronson RT, Tabin C, Sharpe A, Caput D, Crum C, McKeon F. **p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development.** *Nature* **398**:714-718 (1999)

Yang A, McKeon F. **p63 and p73: p53 mimics, menaces and more.** *Nat Rev Mol Cell Biol* **1**:199-207 (2000)

Yang IC, Yang SK, Park CW, Chung JH. ***De novo* methylation at intronic CpG islands of Igf2r is associated with decreased expression of antisense RNA in aged mice.** *Biochem Biophys Res comm* **274**:722-726 (2000)

Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI. **A mouse model for Prader-Willi syndrome imprinting-centre mutations.** *Nature Genet* **19**:25-31 (1998)

Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R. **Apolipoprotein (a) gene enhancer resides within a LINE element.** *J Biol Chem* **273**:891-897 (1998)

Yarnitzky T, Min L, Volk T. **An interplay between two EGF-receptor ligands, Vein and Spitz, is required for the formation of a subset of muscle precursors in Drosophila.** *Mech Dev* **79**:73-82 (1998)

Young WP, Wheeler PA, Coryell VH, Keim P, Thorgaard GH. **A detailed linkage map of rainbow trout produced using doubled haploids.** *Genetics* **148**: 839-850 (1998)

Zamore PD, Green MR. **Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor.** *Proc Natl Acad Sci USA* **86**:9243-9247 (1989)

Zamore PD, Tuschi T, Sharp PA, Bartel DP. **RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21-23 nucleotides intervals.** *Cell* **101**:25-33 (2000)

Zanocco-Marani T, Bateman A, Romano G, Valentinis B, He ZH, Baserga R. **Biological activities and signalling pathways of the granulin/epithelin precursor.** *Cancer Res* **59**:5331-5340 (1999)

Zaphiropoulos PG. **RNA molecules containing exons originating from different members of the cytochrome P450 2C gene subfamily (CYP2C) in human epidermis and liver.** *Nucleic Acids Res* **27**:2585-2590 (1999)

Zardo G, Tiirikainen MI, Hong C, Misra A, Feuerstein BG, Volik S, Collins CC, Lamborn KR, Bollen A, Pinkel D, Albertson DG, Costello JF. **Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors.** *Nature Genet* **32**:453 - 458 (2002)

Zavadil J, Svoboda P, Liang H, Kottickal LV, Nagarajan L. **An antisense transcript to SMAD5 expressed in fetal and tumor tissues.** *Biochem Biophys Res Comm* **255**:668-672 (1999)

Zhang M, Zamore PD, Carmo-Fonseca M, Lamond AI, Green MR. **Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit.** *Proc Natl Acad Sci USA* **89**:8769-8773 (1992)

Zhang H, Serrero G. **Inhibition of tumorigenicity of the teratoma PC cell line by transfection with antisense cDNA for PC cell-derived growth factor (PCDGF, epithelin/granulin precursor).** *Proc Natl Acad Sci USA* 95:14202-14207 (1998)

Zhang Q, Arbuckle J, Wessler SR. **Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize.** *Proc Natl Acad Sci USA* **97**:1160-1165 (2000)

Zhang J, Zhang Y-P, Rosenberg HF. **Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey.** *Nat Genet* **30**:411-415 (2002)

Zhang C, Xie Y, Martignetti JA, Yeo TT, Massa SM, Longo FM. **A Candidate Chimeric Mammalian mRNA Transcript Is Derived from Distinct Chromosomes and Is Associated with Nonconsensus Splice Junction Motifs.** *DNA Cell Biol* **22**:303-315 (2003)

Zhou J, Gao G, Crabb JW, Serrero G. **Purification of an autocrine growth factor homologous with mouse epithelin precursor from a highly tumorigenic cell line.** *J Biol Chem* **268**:10863-10869 (1993)

Zhou X, Sasaki H, Lowe L, Hogan B, Kuehn MR. **Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation.** *Nature* **361**:543-547 (1993)

Zhu J, Nathan C, Jin W, Sim D, Ascroft GS, Wahl SM, Lacomis L, Erdjument-Bromage H, Temps P, Wright CD, Ding A. **Conversion of proepithelin to epithelins: roles of SLPI and elastase in host defense and wound repair.** *Cell* **111**: 867-878 (2002)

Zhuang Y, Weiner AM. **A compensatory base change in U1 snRNA suppresses a 5' splice site mutation.** *Cell* **46**:827-835 (1986)

Zúñiga A, Borja M, Meijers C, Zeller R. **Expression of Alternatively Spliced bFGF First Coding Exons and Antisense mRNAs during Chicken Embryogenesis.** *Dev Biol* **157**: 110-118 (1993)