Transcriptional Regulation by CUX1 and its Implication in the DNA Damage Response and Wnt/β-Catenin Pathway Activation

Charles Vadnais

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Department of Biochemistry
McGill University
Montréal, Québec, Canada

Abstract

The objective of my research project was to study and characterize the transcriptional role of the CUX1 transcription factor both on a global scale and with regards to specific cellular processes.

I have carried out genome-wide location analysis experiments to identify large numbers of potential direct targets of CUX1 and investigated their regulation by CUX1 using expression profiling experiments following both over-expression of the active p110 CUX1 isoform and following its knockdown using shRNA. This study demonstrated that CUX1 can both activate and repress its targets even when binding at a distance from their promoters and that its consensus binding sequence does play a role in its binding but is not required for it in many cases.

Analysis of the putative targets of CUX1 identified by genome-wide location analysis strongly suggested a role for the factor in the cellular response of cells to DNA damage. I used molecular biology methods to investigate this and demonstrated that CUX1's transcriptional activity is required for the maintenance of adequate levels of several key proteins constituting the machinery necessary for an effective response to DNA damage. Cells lacking CUX1 expression have defective cell cycle arrest and DNA damage repair capacities, reduced survival following DNA damage and show a phenotype of increased genomic instability.

Previous studies of a mouse model of mammary gland tumours showed that a subset of tumours from p110 and p75 CUX1 over-expressing mice display activation of the Wnt/β-Catenin pathway. However, the mechanism by which only some of these tumours displayed this phenotype were not fully understood. I used expression profiling on microdissected material from these tumours to characterize the transcriptional effect of p110 and p75 CUX1 expression in these tumours and to identify collaborating events in Wnt/β-Catenin pathway activation. I identified members of the GLI family of transcription factors as being required for activation of the Wnt/β-Catenin pathway in these tumours. I also showed that these tumours display features of epithelial to mesenchymal transition, which may

have implications for the invasiveness and severity of these tumours. The cooperation between CUX1 and GLI genes was confirmed by meta-analysis of human tumour datasets as well as cell based assays testing the ability of each factor to activate the Wnt/β-Catenin pathway on their own and in combination.

Résumé

L'objectif global de ma thèse était d'étudier et de caractériser l'activité transcriptionelle du facteur de transcription CUX1, autant de façon globale que dans le contexte de processus cellulaires spécifiques.

J'ai effectué des expériences de localisation génomiques à grande échelle afin d'identifier un grand nombre de cibles potentielles directes de transcription de CUX1 et j'ai analysé leur régulation par CUX1 en effectuant des expériences de profilage d'expression génétique à la suite de la surexpression de l'isoform p110 CUX1 ainsi qu'à la suite de la répression de CUX1 par shRNA. Cette étude a démontré que CUX1 peut activer ou réprimer l'expression de ces cibles, même lorsqu'il se lie à une distance considérable du promoteur de ces gènes, et que la séquence consensus de liaison de CUX1 joue un rôle dans sa liaison à l'ADN mais n'est pas nécessaire pour celle-ci dans plusieurs cas.

L'analyse de cibles potentielles de CUX1 identifiées par des expériences de localisation génomique a grande échelle ont fortement suggéré l'implication de CUX1 dans la réponse des cellules au dommage à l'ADN. J'ai utilisé diverses techniques de biologie moléculaire pour étudier ce phénomène et j'ai démontré que l'activité transcriptionelle de CUX1 est nécessaire au maintien de niveaux suffisants de nombreuses protéines constituant la machinerie essentielle à la réponse des cellules au dommage à l'ADN. Les cellules n'exprimant pas ou peu de CUX1 sont déficientes dans leur capacité d'arrêt du cycle cellulaire et de réparation du dommage à l'ADN, sont plus sensibles au dommage et montrent un phénotype d'instabilité génomique accrue.

Des études précédentes de tumeurs des glandes mammaires dans un modèle de souris ont montré qu'une partie des tumeurs provenant de souris surexprimant p110 ou p75 CUX1 montraient une activation du processus de signalement Wnt/β-Catenin. Cependant, le mécanisme par lequel seule une partie des tumeurs montrait ce phénotype n'était pas connu. J'ai donc effectué du profilage d'expression génétique sur ces tumeurs afin de caractériser l'effet transcriptionel de CUX1 dans celles-ci et d'identifier d'autres facteurs qui

collaborent dans l'activation de Wnt/β-Catenin. J'ai ainsi identifié des membres de la famille de facteurs de transcription GLI comme étant requis pour l'activation de Wnt/β-Catenin dans ces tumeurs. J'ai aussi observé des caractéristiques de transition épithélio-mésenchymateuse dans ces tumeurs, ce qui pourrait avoir des implications sur la capacité d'invasion et la sévérité de celles-ci. La coopération entre CUX1 et les gènes GLI a été confirmée par une méta-analyse de données de profilage d'expression de tumeurs humaines ainsi qu'avec des expériences cellulaires testant la capacité de chacun de ces deux facteurs à activer Wnt/β-Catenin, individuellement et en combinaison.

Acknowledgements

I would first like to thank my supervisor Alain Nepveu for his support and guidance, and for the time he invested in my training.

I am grateful to the members of the lab at the time of my arrival, whom helped introduce me to the world of research and accepted me among their team, especially Ryoko Harada, our exceptional technicians Ginette Bérubé and Lam Leduy, as well as Chantal Cadieux and Laurent Sansregret. Your help was truly invaluable.

I also want to thank all the other members of the lab who have helped me in my research over the years, in particular Maria Drossos for her indispensable expertise, Sayeh Davoudi for her help at critical times and Zubaidah Ramdzan for stimulating discussions and help with this manuscript.

I would like to thank the members of my Research Advisory Committee throughout the years: Jose Teodoro, Peter Siegel and Michael Hallett.

Thanks also to Nicholas Bertos, Hong Zhao, Margarita Souleimanova and Naila Chungtai from the BCFGG and to Robert Lesurf, Greg Finak and François Pepin from the bioinformatics group.

I wish to acknowledge financial support from the Goodman Cancer Centre, le Fonds de Recherche en Santé du Québec and the faculty of Medicine of McGill.

Je tiens à remercier mes parents et ma famille pour leur soutien à travers les années, tout particulièrement ma conjointe Josianne pour son support et mes filles Cécilia et Pénélope pour l'inspiration qu'elles m'apportent.

Preface

This thesis is divided in four chapters:

Chapter 1: Introduction

Chapter 2: Methods

Chapter 3: Results

Chapter 4: Discussion

The guidelines concerning thesis preparation issued by the Faculty of Graduate and Postdoctoral Studies at McGill University reads as follows:

Theses must conform to the requirements of Library and Archives Canada. Please reference the Thesis Specifications and the Thesis Formatting sections below to ensure compliance with the requirements of the Library and Archives Canada.

In either format, the thesis must contain methodology, results and scholarly discussion, in accordance with disciplinary norms. It must also contain or conform to the following requirements:

1. Title page:

- The title of the thesis;
- The student's name and Unit* followed by "McGill University, Montreal";
- The month and year the thesis was submitted;
- The following statement: "A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of...."
- The universal copyright notice "©" followed by the student's name and the year the thesis was submitted;

2. A detailed table of contents

3. A brief abstract in both English and French. If the language of the thesis is neither English nor French (only allowed for specific language units) then a third abstract in the language of the thesis is required.

4. Acknowledgements:

- Among other acknowledgements, the student is required to declare the extent to

which assistance (paid or unpaid) has been given by members of staff, fellow students, research assistants, technicians, or others in the collection of materials and data, the design and construction of apparatus, the performance of experiments, the analysis of data, and the preparation of the thesis (including editorial help).

- In addition, it is appropriate to recognize the supervision and advice given by the thesis supervisor(s) and advisors.

5. Preface & Contribution of Authors:

- In the case of collaborative work presented in either a standard format or manuscript-based thesis, there must be an explicit statement of the contributions of all parties, including the student, in the Preface of the thesis.
- The Preface of a Doctoral thesis must also include a statement clearly indicating those elements of the thesis that are considered original scholarship and distinct contributions to knowledge.

6. An introduction that clearly states the rationale and objectives of the research

7. A comprehensive review of the relevant literature, according to disciplinary norms

8. A final conclusion and summary

9. A thorough bibliography or reference list, according to disciplinary norms

Publications arising from this thesis

1) CUX1 Transcription Factor is Required for Optimal ATM/ATR-Mediated Responses to DNA Damage.
Vadnais C, Davoudi S, Afshin M, Harada R, Dudley R, Clermont P-L, Drobetsky E, Nepveu A. Nucleic Acids Res. 2012 May;40(10):4483-95.

2) Long-Range Transcriptional Regulation by the p110 CUX1 Homeodomain Protein on the ENCODE Array.
Vadnais C, Awan A, Harada R, Clermont P-L, Leduy L, Bérubé G, and Nepveu A.
Accepted in BMC Genomics March 26th 2013

3) Transcriptional Regulation of Canonical Wnt/β-Catenin Pathway by CUX1 in Cancer Cells.
Vadnais C, Shooshtarizadeh P, Rajadurai C, Lesurf R, Hulea L, Davoudi S, Cadieux C, Leduy L, Hallet M, Park M, Nepveu A.
Author order not final.
Manuscript in preparation.

Contributions

1)

- Davoudi S: Performed Clonogenic survival assays
- Afshin M: Performed siRNA knockdown and qPCR
- Harada R: Performed the original ChIP experiments
- Dudley R: Performed EMSA
- Clermont P-L: Performed qPCR
- Drobetsky E: Co-senior author, helped design study and experiments

2)

- Awan A: Participated in study design and analysis
- Harada R: Performed parts of the ChAP and ChIPs
- Clermont P-L: Performed qPCR validations
- Leduy L: Generated viruses, performed infections
- Bérubé G: Designed and generated constructs

3)

- Shooshtarizadeh P: Performed TOP/FOP assays and over-expression experiments (Hs578t, MCF7, 293T) and CUX1 siRNA in Breast and Ovarian cell lines
- Hulea L: Performed TOP/FOP Assays
- Davoudi S: Performed cell culture, Western blots and qPCRs on MCF10A cells
- Lesurf R: Designed and wrote the BreSat algorithm
- Cadieux C: Performed IHC on tumour slides
- Leduy L: Generated viruses, performed infections
- Hallet M: Co-senior author
- Park M: Co-senior author

Other Publications

4) Genome-Wide Location Analysis and Expression Studies Reveal a Role for p110 CDP/Cux in the Activation of DNA Replication Genes.
Harada R, **Vadnais C**, Sansregret L, Leduy L, Bérubé G, Robert F and Nepveu A. Nucleic Acids Research. 2008 Jan;36(1):189-202.

5) The p110 isoform of CUX1 Cooperates with E2F Transcription Factors in the Transcriptional Activation of Cell Cycle-Regulated Gene Promoters.
Truscott M, Harada R, **Vadnais C** and Nepveu A. Mol Cell Biol. 2008 May;28(10):3127-38.

6) Transcriptional Activation of the LATS1 Tumor Suppressor in Tumors of CUX1 Transgenic Mice.
Siam R, Harada R, Cadieux C, Battat R, **Vadnais C**, Leduy L, and Nepveu A. Mol Cancer. 2009 Aug 5;8:60.

7) Mouse Mammary Tumour Virus p75 and p110 CUX1 Transgenic Mice Develop Mammary Gland Tumours of Various Histologic Types.
Cadieux C, Kedinger V, Yao L, **Vadnais C**, Drossos M, Paquet M, Nepveu A. Cancer Res. 2009 Sep 15;69(18):7188-97.

8) p110 CUX1 homeodomain protein stimulates cell migration and invasion in part through a regulatory cascade culminating in the repression of E-cadherin and occludin.
Kedinger V, Sansregret L, Harada R, **Vadnais C**, Cadieux C, Fathers K, Park M, Nepveu A. J Biol Chem. 2009 Oct 2;284(40):27701-11.

9) Cux1 Causes Chromosomal Instability By Promoting Bipolar Divisions After Mitotic Failure.
Sansregret L, **Vadnais C**, Livingstone J, Kwiatkowski N, Awan A, Cadieux C, Leduy L, Hallett M, Nepveu A. Proc Natl Acad Sci U S A. 2011 Feb 1;108(5):1949-54

10) Microrna122 Is A Key Regulator Of A-Fetoprotein Expression And Influences The Aggressiveness Of Hepatocellular Carcinoma.
Kojima K, Takata A, **Vadnais C***, Otsuka M, Yoshikawa T, Akanuma M, Kondo Y, Kang YJ, Kishikawa T, Kato N, Xie Z, Zhang WJ, Yoshida H, Omata M, Nepveu A, Koike K. Nat Commun. 2011 Jun 7;2:338.
* Co-first authorship.

Table of Contents

1. Introduction

1.1 The CUX1 Transcription Factor

1.1.1 The CUX family of transcription factors

CUX1 is part of the CDP/Cux/Cut (CCAAT-displacement protein/cut homeobox) family of transcription factors, which is present in all higher eukaryotes(Reviewed in [1]). The gene has previously been referred to as CDP (CCAAT-Displacement Protein), CDP/Cut and *CUTL1* (Cut-like 1). The term "cut" refers to the "cut-wing" phenotype observed in *D. Melanogaster* mutants lacking expression of the gene in cells normally giving rise to the wing margins[2-4].

The full length CUX1 proteins, referred to as p200, contains several conserved domains (Fig. 1): four conserved DNA binding domains, including three Cut Repeats (CR1, CR2 and CR3) and a Cut Homeodomain (HD). The N-terminal region of the protein contains an auto-inhibitory domain (IN) that inhibits DNA binding while the C-terminal region contains two active repression domains (R1 and R2). Finally, a region predicted to form a Coiled Coil domain (CC) is located upstream of CR1[5, 6]; the function of this domain has yet to be determined.

In mice, *Cux1* has been shown to be expressed in almost all tissues of the adult mice, although it appears to be expressed mostly in the epithelial compartment in most organs[7-9].

The related CUX2 protein contains similar domains to CUX1. However, contrarily to CUX1, its expression in mice has been shown to be restricted to the nervous system in the adult mice and during late embryonic development[10]. Also in contrast to CUX1, no processing of CUX2 into shorter isoforms has been detected (see below for CUX1 isoforms).

While the full length isoforms of the 2 proteins have similar *in vitro* DNA binding properties, some experiments have shown that CUX2 fails to activate transcription of a known target of CUX1[11]. While the function of CUX2 is not

clear at the moment, what is known suggests considerable differences with that of CUX1.

1.1.2 Isoforms of CUX1

Shorter isoforms of CUX1 are also named after their apparent molecular weight and contain different combinations of the conserved domains of the full length protein (Fig. 1). p75 is the only isoform that is generated from an alternate transcription start site, which is located within intron 20 of the gene[12]. All other known short isoforms are generated through proteolytic processing of the full length form: The p110 and p90 isoforms are processed by Cathepsin L while the p150 isoform is processed by a yet unidentified cysteine protease[13-15]. The p80 isoform is generated through the action of Cathepsin L at the N-terminus and a caspase at the C-terminus[16].

The DNA binding and regulatory properties of CUX1 isoforms are determined by the DNA binding domains they contain. The full length protein binds to DNA transiently and with fast kinetics through its CR1 and CR2 domains; Its main mode of transcriptional activity, is through the "CAATT-displacement activity", a mechanism of passive repression involving competition for binding site occupancy with transcriptional activators [17-19].

p200 can also mediate repression by the action of its C-terminal active repression domains, even when binding at a distance from transcription start sites, a process proposed to occur through the recruitment of HDAC1 (Histone Deacetylase 1) and the histone lysine methyltransferase, G9a[16, 20-22].

The p150 isoform cannot bind DNA, is not transcriptionally active and appears to function only as a dominant negative isoform in the lactating mammary gland[15].

The p75, p90 and p110 isoforms display slower, stable DNA binding kinetics and were suggested to have a preference for the ATCRAT consensus sequence[12, 14, 17, 23]. These shorter isoforms can function either as transcriptional activators or repressors depending on the promoter context [24,

25]. The p80 isoform lacks the active repression domains and was suggested to functional mainly, if not exclusively, as an activator[16].

1.1.3 Regulation of CUX1 activity

One of the principal modes of regulation of CUX1's activity is the processing of a fraction of p200 into p110. while the DNA binding activity of p220 CUX1 is constant throughout the cell cycle, that of p110 CUX1 is tightly regulated [23]. In mid-G1, Cdc25a dephosphorylates p200 CUX1 in the CR3HD region[26], then 1% to 10% of p200 CUX1 is proteolytically processed by a nuclear isoform of cathepsin L to produce the p110 CUX1 isoform [23, 27]. At some point in G2, p110 CUX1 DNA binding is inhibited by phosphorylation by CyclinA/Cdk1 and further inhibited by additional phosphorylation by CyclinB/Cdk1 during mitosis[28, 29]. Following mitosis, this hyper-phosphorylated form does not localize to the nucleus, requiring active p110 CUX1 to be generated anew from processing of p200. These post-translational modifications circumscribe the transcriptional activity of p110 CUX1 to the period between mid-G1 to sometimes in G2.

1.1.4 Involvement of CUX1 in cancer

CUX1 was first identified as a potential tumour-suppressor in genetic studies, however evidence has since accumulated to more strongly indicate that CUX1 in fact functions as an oncogene(Reviewed in [30]).

The genes was mapped to 7q22[31, 32], a region that was found to be deleted or rearranged in uterine leiomyomas[33], acute myeloid leukemia[34, 35] and myelodysplastic syndrome[36, 37]. Later studies confirmed that CUX1 was present in the smallest commonly deleted region found in some uterine leiomyomas as well as breast and ovarian tumours[38-40], suggesting a tumour suppressor function. However, subsequent studies in uterine leiomyomas showed that while *CUX1* mRNA levels were often reduced in tumours harboring a deletion, the protein levels were more likely to be increased than decreased and

that furthermore, levels of the short p110 isoform of CUX1 were increased in several cases[41].

In human breast cancer, elevated expression of CUX1 mRNA and protein levels was documented in high grade tumours compared to low grade tumours[42]. An increase of CUX1 protein in high grade pancreatic tumours was also reported[43] and a comparison of expression profiles between normal plasma cells and malignant plasma cells from multiple myeloma patients showed a marked increase in *CUX1* mRNA levels[44].

Mouse models of CUX1 overexpression showed an association with cancer-like diseases in multiple organs that varied depending on the tissue type expression of the transgene, including multi-organ organomegaly, glomerulosclerosis and interstitial fibrosis[45, 46] and hepatomegaly[47].

The short isoforms of CUX1 specifically were implicated in cancer progression: a study showed that the p75 CUX1 isoform was activated in many breast tumours and breast tumour derived cell lines[12] and it was also shown to be overexpressed in the polycystic kidneys of a $Pkd1^{-/-}$ mouse model[48]. Furthermore, increased processing of full-length CUX1 by a nuclear isoform of Cathepsin-L into the shorter p110 isoform was document in ras transformed cells[49].

To study the role of short CUX1 isoforms in cancer, mouse models of the overexpression of the p110 and p75 CUX1 isoforms were generated in the laboratory of Dr. Alain Nepveu. One model placed p75 CUX1 under the control of the *cmv* immediate early enhancer and the chicken β-actin promoter. These mice developed a polycystic kidney phenotype, consistent with the $pkd1^{-/-}$ mice results[50].

Another model placed either the p110 or p75 isoforms under the control of the MMTV (Mouse Mammary Tumour Virus) promoter inserted into the hprt locus, in order to study their effects on mammary gland tumour development. Somewhat surprisingly, it was first observed that mice overexpressing the p75 isoform were found to develop a myeloproliferative disease-like myeloid leukemia[51]. In addition, however, mice from both the p75 and p110 model

developed mammary gland tumours of varied histological subtypes at a significantly greater frequency than their wild-type counterparts[52] and a subset of tumours in the model were associated with lung metastasis. The considerable latency preceding tumour formation in both cases suggested that other genetic events are required for tumorigenesis.

1.1.5 CUX1 promotes hallmark properties of cancer

Multiple studies have associated CUX1 activity in the cell with a number of hallmark properties of cancer which may explain parts of the role it plays in tumour initiation and progression.

Cell proliferation

Overexpression of the p110 isoform of CUX1 was shown to promote cell proliferation by accelerating entry into S phase both in asynchronous cell populations and cells exiting quiescence and conversely, cells lacking *CUX1* display an extended G1 and slower proliferation[53].

Cell motility and invasion

siRNA-mediated knockdown approaches first suggested that it was required for the processes of motility and tumour invasion[42] and that its effect was mediated in part by the Src tyrosine kinase, theWNT5A ligand and the glutamate receptor GRIA3[43, 54, 55]. A Subsequent study showed that overexpression of p110 CUX1 enhanced cell migration and invasiveness through both the activation of genes that promote motility and repression of genes that inhibit it[25].

Resistance to apoptotic signals

Overexpression of CUX1 was shown to protect pancreatic cells from TRAIL (tumour necrosis factor related apoptosis-inducing ligand) ligand induced apoptosis while knockdown of CUX1 caused an increase in apoptosis[56].

Furthermore, treatment of cells with IGF1, a growth factor that promotes cell survival, leads to an increase in CUX1 levels[57, 58].

Mitosis and genomic instability

While CUX1 is not active during mitosis, it has nevertheless been shown to play a role in the spindle checkpoint assembly during mitosis. In addition to being required for proper division of normal cells, this activity of CUX1 was shown to be associated with chromosomal instability in cells that undergo cytokinesis failure and CUX1 overexpression was shown to lead to the spontaneous appearance of tetraploid populations in cell culture[59]. Consistent with these results, the majority of mammary gland tumours arising from the MMTV-CUX1 mouse model harbor an aneuploid and generally sub-tetraploid phenotype, suggesting that tetraploidization of cells permitted by CUX1 overexpression and associated chromosomal instability play an important role in CUX1-induced tumorigenesis.

1.2 Transcriptional regulation

1.2.1 Methods to study transcription

The last decade has seen significant advances in the field of transcription. The discovery of nuclear histone acetyltransferases (HATs) in the mid nineties has literally opened a new field of investigation into post-translational modifications that target histones and modulate the chromatin state either locally or over large genomic loci. The multiple types of modifications that take place on specific histone residues and the regulatory cascades that can be triggered in this manner led investigators to propose that a "histone code" regulates gene expression in a manner reminiscent of the genetic code translating nucleic acid coding sequences into protein sequences [60, 61]. In parallel, a number of novel experimental approaches have contributed to move the transcription field from a gene-by-gene approach focused on core promoters to a genome-wide non-biased approach that enables us to study large numbers of transcriptional targets as well as the mechanisms by which these targets are regulated [62]. Recent tools in our arsenal include the increasing availability of genomic microarrays [63], siRNA-mediated gene knockdown [64], more efficient virus-based gene delivery systems[65, 66], and high-throughput sequencing [67]. Importantly, the "rediscovery" of chromatin immunoprecipitation combined with the development of microchip arrays containing large numbers of genomic sequences has opened new horizons. Indeed, chromatin immunoprecipitation was first described in the mid eighties by the group of John T. Lis who used this assay to show that RNA polymerase II molecules were already present at the 5' end of the hsp70 gene in uninduced cells and that heat shock somehow enabled transcription elongation to take place [68]. Curiously, the method was not applied to specific transcription factors before another decade [69]. The information thus gathered has forced us to reconsider our original understanding of basic mechanisms of transcriptional regulation [70]. For example, a common belief was that a specific transcription factor could bind to a few dozen genes whose core promoter contain its consensus binding site as defined *in vitro*, and once recruited to a promoter could almost single-handedly regulate transcription [71]. We now know that c-MYC binds to

approximately 20% of gene promoters and is also capable of regulating genes at a distance [72-74]. Another major conceptual advance concerns the criteria to define a transcriptional target. Experimental evidence typically included the presence of a consensus binding motif within a core promoter, *in vitro* binding assays and luciferase reporter assays. While these assays are still employed, it is clear that they cannot provide definitive evidence that a transcription factor regulates a specific gene. Additional evidence must also include chromatin immunoprecipitation assays to demonstrate "in vivo" DNA binding, and change in expression of the endogenous gene target in response to the knockdown and/or overexpression of the transcription factor.

1.2.2 Studying transcription factors by high throughput methods

As mentioned, the transcriptional activity of a factor of interest, in our case CUX1, can now be studied using genome-scale methods. Expression profiling can be carried out following over-expression and down-regulation of a factor to observe the global effect of that change on transcription and genomic regions bound by the same factor can be identified by genome-wide location analysis.

Expression profiling

Expression profiling consists of measuring the expression levels of tens of thousands of transcript in a given sample, or sets of samples, in order to gain a wide-scope view of the transcriptional activity. One of the methods, which is still the most widely used, is by hybridization of extracted RNA to a microarray. Oligonucleotide probes in the 60bp range of size, each corresponding to a different transcript, are covalently bound to a glass slide. The earliest slides could contain a few tens of thousands of individual spots, but the most recent platforms can contain as many as a few million individual spots, each consisting of several thousands of copies of a single oligonucleotide. Sample RNA is then labelled by incorporation of a fluorescent dye, often Cy5 (which emits a red signal) or Cy3 (which emits a green signal). Labelled RNA is then hybridized to the array, which

will then be scanned by a high-resolution scanner (5μm is common, but a 2μm resolution is required to scan the densest arrays). The intensity of signal at each spot is proportional to the amount of corresponding mRNA in the sample. Arrays with probes corresponding to miRNAs and other non-coding RNAs have been designed in the recent years.

More recently, high-throughput RNA sequencing methods (RNASeq) have been devised which sequence millions of individual RNA fragments from a given sample and thereby provide a global profile of mRNA levels without the inherent limitation of measuring only transcripts for which probes are present on the microarray.

Chromatin Immunoprecipitation and Genome-wide location analysis

Chromatin immunoprecicipitation (ChIP) is a method that allows the identification of genomic locations where a protein of interest is bound. The basic principle is to stabilize the interaction between proteins and DNA using a cross-linking agent, for example formaldehyde, extract the protein-chromatin complexes and isolate the protein of interest using a specific antibody. The proteins are then degraded and the DNA that was formerly associated with the protein of interest is purified.

A variant of this method, called Chromatin Affinity-Purification (ChAP), consist in expressing a tagged variant of the protein of interest and, following cross-linking, isolate it using the tags as opposed to utilising antibodies.

The presence of DNA from genomic locations of interest can then be measured by qPCR, but in order to obtain a genome-wide view of the binding pattern of the immunoprecipitated protein, ChIP-Chip (or ChAP-Chip) can be used. In this case, the DNA is labelled using Cy dyes and hybridized to a microarray containing oligonucleotide probes covering regions of potential interest with the un-enriched labelled chromatin used as a control. For example, some arrays are designed with probes covering promoter regions of genes while others are designed to cover loci of interest across the genome or whole chromosomes.

One design of interest is the ENCODE genomic microarray that was designed as part of the ENCODE project (ENCyclopedia Of DNA Elements) provided a 1% sampling of the human genome that can be interrogated to define the distribution types of transcriptional regulation of specific transcription factors [75].

As with expression profiling, high-throughput sequencing based methods have emerged in the recent years to massively sequence ChIPed DNA (ChIPseq), without the problems associated with probe generation and coverage.

Interestingly, it can be argued that the relative gains in terms of coverage of going from ChIP-Chip to ChIPSeq are greater than those of going from microarray based to RNASeq expression profiling. The ChIP array with the largest number of probes contains 4.6 million of them (www.affymetrix.com). At a density of 1 probe per 35bp, this platform can provide coverage of up to 161Mbp, which could cover 4.8% of the 3.3 billion bases of the human genome. A 6.3% coverage is obtained with the Nimblegen platform containing 2.1 million probes spaced 100bps apart (www.nimblegen.com). Considering that a ChIPSeq experiment can, in principle, cover the entire genome, this is roughly a 20-fold increase in coverage. On the other hand, standard expression array designs, cover in the range of 30 to 50 thousand transcripts, with the possibility of creating custom arrays with hundreds of thousands of unique probes[www.agilent.com]. With recent estimates putting the number human genes around 25,000 and the number of unique human transcripts in the 100 to 120 thousand range[76, 77], the coverage afforded by expression arrays is already considerable.

1.2.3 Transcriptional regulation by CUX1

In light of all this, my first project involved the study of transcriptional regulation by the CUX1 transcription factor.

Many specific transcription factors are able bind to genomic sites that are far away from TSS (Transcription start sites). These studies also revealed that only about up to 10% of putative transcriptional targets showed evidence of

regulation in response to changes in transcription factor concentrations [78-80]. Whether CUX1 binds preferentially to core promoter sequences, like E2F1, or whether it can also bind at a distance from TSS, like c-Myc, had not yet been determined at the time[72, 73]. Also unknown was the proportion of CUX1 targets that is regulated in response to overexpression or silencing of CUX1. To begin to address these questions, we have performed ChAP-chip using ENCODE and promoter microarrays. Putative targets were validated in independent ChIP followed by q-PCR. The regulatory effects of CUX1 on these targets were measured in expression profiling experiments following changes in CUX1 expression and confirmed by RT-qPCR.

1.2.4 Gene annotations and functional classification tools

Since high-throughput methods identify hundreds and possibly thousands of potential genes of interest at a time in a given experiment, it is generally not realistic to manually research the role of each individual gene identified in this manner. To solve this, a number of functional classification databases and analysis tools have emerged that identify biological functions that may be of interest given a set of genes. I present below two such functional analysis tools.

DAVID - The Database for Annotation, Visualization and Integrated Discovery

Created in 2003 by the NIAID (National Institute of Allergy and Infectious Diseases), DAVID is a database of gene classifications and functional annotations. It provides tools allowing users to associate functions to each gene identified by genome-scale studies. The categories of functions that can be associated with genes include biological process, cellular components, molecular pathways, protein domains, etc.

Importantly, DAVID also allows the identification of such functional categories which are significantly over-represented among the list of gene provided. That is, functions that are associated with the genes of interest at a frequency greater than would be expected from a random sampling of genes. The typical workflow of a session is to identify a list of gene from an experiment using

user-specified cut-offs (e.g. expression fold changes, P values, etc.) and to submit this list to DAVID to identify functions.

In general, the use of DAVID and its tools requires a limited computational or statistical expertise [81-83][david.abcc.ncifcrf.gov/].

GSEA – Gene Set Enrichment Analysis

GSEA is a different computational method to calculate the significance of the association of functional categories using results from expression profiling. Contrarily to an approach such as that of DAVID, GSEA does not identify sets of "significant genes" using threshold values. Instead, GSEA evaluates the distribution of genes contained in a "gene set" relative to the overall distribution of genes on the platform used. A gene set can be any list of gene that is associated with a function or pathway of interest, and can be taken from any annotation database such as GO, KEGG or MSigDB.

The software then measures if the genes of interest are significantly concentrated among the most over- or under-expressed genes in a comparison between two groups of samples and calculates the statistical significance of this enrichment.

Since this principle requires a continuous distribution of quantitative data, GSEA is not really amenable to genomic data that provides a Yes/No result, such as identifying genes whose promoters are either bound or not in a ChIP-on-Chip experiment for example.

GSEA is used as a downloadable software and accompanying gene sets that is freely accessible online. It can be modified by users to create custom gene sets and use them with the software. However, GSEA requires a certain computational expertise, with at least a familiarity with spreadsheet manipulation software to transform result datasets and gene sets into the appropriate format and to run the software. [84-86] [www.broadinstitute.org/gsea/]

1.2.5 Public datasets and meta-analyses

As discussed above, numerous high throughput methods of genomic analysis are now available to researchers, and as they become more widespread and accessible, massive amounts of data are being generated by research groups throughout the world.

By their very nature, such methods typically generate more data than the initial experimenters can make use of, allowing them to mine the data in subsequent projects, but also creating datasets that could be useful to other researchers, which, looking at the same set of results in a different context, may make valuable observations in addition to those of the original study. Furthermore, the aggregation of multiple sets of data allows meta-analyses to be carried out, revealing trends that may otherwise have gone unnoticed in individual datasets.

Fortunately, researchers often choose to include the results of their high-throughput experiments either as supplementary material to their publications or on their laboratory website. In fact, many journals, including all Nature journals and Science, now require such data to at least be submitted in a public repository, to allow its critical examination by peers but also its use in future studies.

Unfortunately, this is always not done using a standardized format (see MIAME below) and researchers may provide either raw data or processed data without specifying how the processing is done. In this context, manually acquiring datasets from individual publications, processing the data into the desired format and then combining all these datasets into a usable database is a arduous and time consuming task that requires considerable computational and statistical expertise.

In response to this, multiple organisms have created databases compiling genomic results, making large amounts of data readily available in a standardized format that allows rapid and efficient comparison and meta-analyses of multiple datasets without the need for each person accessing it to perform their own pre-processing of the data.

I present below the most widely used standard for publication of genome scale experiment results as well as a number of databases and repositories that

provide access to large amounts of genomics data, with a focus on databases relevant to research on cancer.

MIAME – A standard format for microarray experiments

MIAME is a standard for the reporting of microarray results proposed in 2001 by the MGED (Microarray Gene Expression Data) Society, which has since become the FGED (Functional GEnomics Data) Society, consisting of researchers from around the world, which advocates for open access to genomic data sets. The goal of the proposal was to correct the problem of microarray data being reported in un-standardized and hard to use formats[87]. Since their original proposal, the group as created specific file formats for data reporting: first MAGE-ML, which had the disadvantage of requiring considerable computational expertise to use, and later MAGE-TAB, a more user-friendly, more accessible spreadsheet-based reporting format, freely accessible to researchers who which to publish their data in a MIAME-compliant manner but do not have sophisticated informatics expertise[88].

More than 50 journals have chosen to require that any article published in their pages apply the MIAME standard to their reports of microarray results, including Nature Publishing Group journals and certain Cell Press journals as well as others[89]. To conform to the latest version of the standard, MIAME 2.0, a study including microarray data requires the inclusion of: 1- The raw signal for each hybridization obtained from the applicable image analysis software. 2- The final normalized data for the set of hybridizations in the study. 3- A detailed annotation (description) of the samples used, including any treatments or parameters used. 4- The experimental design detailing which samples and/or reference samples were hybridized on each arrays. 5- Annotation of the array platform used, such as gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference to commercial array catalog numbers. 6- All relevant experimental protocols and data processing information or references to the methods used if they are well established.(www.mged.org)

GEO – A repository of genomic datasets

The GEO (Gene Expression Omnibus) project was the first online repository created with the objective of providing a system to publicly distribute the results of a broad range of high-throughput experimental methods. It was created in 2000 by the NCBI (National Center for Biotechnology Information) [90]. As of this writing, it includes data from over 30,000 studies, covering all types of high-throughput methods used in any research context, but of which the vast bulk (~26,000) consists of microarray studies of expression profiling[91]. Datasets are submitted by researchers through the repositories online submission system, and access to the GEO database is free and public. Users can search for datasets that match certain criteria of interest (platform, species, experiment type, etc.) and obtain generally both the unprocessed results and the results processed as per the original study, as well as detailed information on the platform used for the experiment. In addition to this, GEO also provides a number of tools to mine the database, such as a tool to measure co-expression between genes or calculate differential expression between groups of samples within a dataset. [92, 93]

ArrayExpress – A MIAME compliant database of genomics results

ArrayExpress is an online database of various genomics experiments operated by the EBI (European Bioinformatics Institute) based at the Wellcome Trust Genome Campus in Hinxton (UK), which is part of EMBL (European Molecular Biology Laboratory). This non-profit academic group is supported mainly by a combination of grants from the European Commissions, The Wellcome Trust, The NIH and the UK Research Councils. ArrayExpress began its operations in 2006 with a focus on microarray-based expression profiling experiments but has since begun archiving datasets of CGH analyses, ChIP-Chip and ChIP-Seq experiments, DNA modifications profiling, miRNA profiling etc. Any researcher publishing the results of genomics analyses can submit their data to the ArrayExpress Archive as long as it is compliant with MIAME standards. Datasets are searchable by experiment types, species, platforms or other keywords. Datasets generally include raw and processed data, detailed

experimental procedure descriptions and platform information. ArrayExpress also features a "gene expression atlas", which is a gene centric search option to quickly retrieve all datasets showing changes in the levels of a given gene of interest. The expression atlas is based on a curated subset of the ArrayExpress Archive and is continuously updated with the most recent available annotations. ArrayExpress and all of its contents and functions are freely available and accessible online at www.ebi.ac.uk/arrayexpress/ [94-96](www.ebi.ac.uk).

Oncomine – A database of expression profiling results

The Oncomine database, currently operated by Compendia Bioscience, provides expression profiling datasets specifically of human tumours or human tumour-derived cell lines. The objective of this endeavour is to integrate microarray data from a vast number of studies into a platform that facilitates access, analysis and comparison of these datasets, in an effort to improve our understanding of cancer genetics, and importantly, to provide the ability to mine these datasets for relevant information without requiring an extensive expertise in computational biology. It was first released in October 2003, and at the time covered 40 gene expression datasets and was operated by a group of researchers at the University of Michigan Medical School. As of this writing, Oncomine now encompasses 953 gene expression datasets covering 19 cancer types and a total of over 60,000 tumour samples, as well as ~10,000 normal tissue samples. In recent years, Oncomine has also started compiling gene copy number data, for example from CGH array experiments, but the number of such datasets is limited in comparison to expression profiling.

Compendia Bioscience was founded in 2006 by Arul Chinnaiyan and Dan Rhodes, the original instigators of the project, with the objective of creating a commercial version of the Oncomine database with a focus on assisting research aimed at improving drug development and clinical practice.

Oncomine differs from the above mentioned databases/repositories in that it does not just provide the original results of studies but rather does its own

processing of all of the datasets it contains and presents them through its own standardized interface.

Every dataset in the database is selected by Oncomine from publicly available data or is requested from the authors of a given study and is then formatted in a standardized manner and expression levels are given in a processed and "ready-to-use" format. Samples are classified by tumour types and subtypes, and calculations of differential expression between either normal and tumour samples or between specific subtypes of tumours are readily available in the database, allowing users to get immediate answers to such questions. In addition, Oncomine compiles all available clinical information for each dataset allowing analyses of links between gene expression and multiple parameters: Tumour stage or grade, patient survival, occurrence of metastasis, responsiveness to treatment, presence of mutations in specific genes, molecular or pathological subtypes, demographics, etc.

Importantly, the database is entirely searchable by genes of interest, cancer types, any clinical parameters or any combination of these. This allows users to quickly identify all datasets relevant to the question being investigated. The data can then be visualized, exported in graphic form or downloaded as raw tables of standardized data for further analysis.

Unfortunately, the free, public version of Oncomine does not include some of these features (such as searching for multiples genes at once), although it provides the same functionalities as the version of Oncomine that existed prior to the creation of the commercial version and can still provide much relevant information about the expression of single genes in different tumour types. [97, 98][www.oncomine.com, www.compendiabio.com]

TCGA – The Cancer Genome Atlas

The Cancer Genome Atlas (TCGA) is a large-scale project aimed at improving our understanding of the molecular basis of cancer, by constructing a databases of genomic information of different types of cancer. Overall, the endeavour consists in obtaining samples from cohorts of patient and subjecting

them to multiple methods of genomic analysis, in order to generate a comprehensive understanding of the genetic events involved in cancer formation and progression.

TCGA differs from databases such as Oncomine in that it does not contain previously published data but creates its own sets of data. This approach allows for a complete integration of the genomics methods since they are carried out in the same sets of samples but obviously involves a considerable amount of work.

TCGA, operated by The Cancer Genome Atlas Research Network, was originally created in 2006 as a pilot project funded by the NCI (National Cancer Institute) and the NHGRI (National Human Genome Research Institute). The 3-year pilot program successfully demonstrated the feasibility of coordinating a network of research teams to combine their results and the benefits of making these results freely available. As a result the program has become permanent and has since collected tumour samples from 25 tumour types. Although the collection and processing of samples and their analysis is an ongoing process, samples are meant to be analyzed by each of the following: 1) Expression profiling by microarray analysis or high-throughput RNA sequencing. 2) miRNA profiling by miRNA sequencing. 3) DNA Methylation patterns by DNA sequencing 4) Copy number variation by CGH microarray analysis. 5) Single Nucleotide polymorphisms and 6) Somatic mutation by high-throughput DNA sequencing. Normal samples are available for all cancer types and matched normal samples are available for most cancer samples.

All data generated by this network is made publicly available (https://tcga-data.nci.nih.gov/tcga/). The data portal provides functionalities to select and export data for the desired types of experiments performed on the samples of interest; users can select groups of samples on which all the desired analyses have been carried out. The results generated by TCGA have been used in over 200 peer-reviewed scientific publications and the TCGA Research Network has also published a number of research articles containing analyses of their own data [99-101] [http://cancergenome.nih.gov/].

ICGC - The International Cancer Genome Consortium

The ICGC (International Cancer Genome Consortium) is a international network of cancer research laboratories launched in 2007, at the behest of 6 organizations: the European Commission, Genome Canada, the National Cancer Institute, the National Human Genome Research Institute, the Ontario Institute for Cancer Research, and the Wellcome Trust. Scientific teams that are part of the ICGC have the responsibility to undertake the genomic characterization of at least 500 samples of a given cancer type or subtype. As of 2012, 47 member teams across North America, Europe, Asia and Australia have obtained genomic data from 23,000 tumour samples, including whole genome sequencing, copy number variations, expression and methylation profiles. The ICGC provides a set of quality control guidelines that must be met in order for sample sets to be accepted and makes the results available through its data portal (http://dcc.icgc.org), but individual research groups are otherwise independent in their research[102, 103] (http://icgc.org/).

1.3 The DNA damage response

The genomic DNA of all cells is under constant assault from a variety of endogenous sources, such as reactive oxygen species generated by the cell's metabolism, as well as exogenous sources such as ultraviolet (UV) irradiation or exposure to damaging chemicals. In order to survive, cells must be able to detect the occurrence of damage, arrest their progression through the cell cycle and repair the damage before proceeding.

The types of DNA lesions to which cells are exposed include single and double strand breaks in the DNA's sugar-phosphate backbone, which can both be induced by Ionizing Radiation (IR), radiomimetic chemicals and to a limited extent by reactive oxygen species. Damage to DNA bases themselves can be caused by reactive oxygen species and exposure to UV radiation, which can induce dimerization of adjacent bases.

1.3.1 Detection of DNA damage and signal transduction

The early DDR is largely mounted in a DNA lesion-specific manner. In the case of DSBs, it is mediated primarily through the protein kinase ATM (Ataxia Telangectasia Mutated), which is a member of the PIKK (phosphoinositide 3-kinase (PI3K)-related protein kinase) family, of which the ATR (ATM and Rad3-Related) protein, and the DNA-PKcs (DNA-dependent Protein Kinase catalytic subunits) are also members.

These lesions are initially detected by the MRN complex, composed of MRE-11, Rad-50 and NBS1, which then recruits inactive ATM[104]. Inactive ATM exists as dimers or possibly multimers which become active by autophosphorylation on Serine 1981[105] and then phosphorylate hundreds of substrate proteins[106]. The Chk2 (Checkpoint Kinase 2) kinase is among the prominent targets of ATM, and acts in transduction of the DDR signal by phosphorylating its own set of downstream targets[107, 108]. The Histone H2A variant H2AX located near sites of damage, are also phosphorylated by ATM[109] into γH2AX, as well as the MDC1 (Mediator of DNA Damage

Checkpoint 1) protein[110]. These two proteins then act as a scaffold to recruit several members of the DDR machinery, including 53BP1 (Tumour Protein p53 Binding Protein), ATM and BRCA1 (Breast Cancer 1), forming the Ionizing Radiation Induced Foci (IRIFs) [111].

53BP1 was shown to be required for efficient phosphorylation of ATM targets following IR, including Chk2, BRCA1 and importantly it is required for the accumulation of p53 after damage. 53BP1 is believed to act as a scaffold between ATM and its substrates[112].

BRCA1 is phosphorylated by ATM[113], localizes to sites of damage and plays a role in activation of downstream checkpoint kinases[114]. It has also been shown to play a role in actual repair of double strand breaks through both homologous recombination[115] and Non-Homologous End Joining[116].

p53 itself is phosphorylated downstream of ATM by both ATM itself and Chk2[107, 117]. Its phosphorylation causes an increase in its nuclear levels by preventing its export from the nucleus and degradation[118], allowing it to play its role in cell cycle checkpoint control (see below).

A partially independent pathway of DDR is that centered around the ATR kinase, which is mainly activated following UV irradiation or other sources that cause base damages.

Single stranded DNA, which can be generated as an intermediate during the repair of damaged DNA bases, or found at stalled replication forks, is bound by RPA (Replication Protein-A) which specifically activates the ATR kinase by facilitating the association of ATR-ATRIP (ATR Interacting Protein), TopBP1, and the 911 complex (Rad9-Rad1-Hus1) [119, 120]. The ATR-ATRIP complex may also bind directly to damaged bases, independently of RPA[121]. Activated ATR, similarly to activated ATM, then phosphorylates a large number of downstream targets[106], notably the Chk1 (Checkpoint Kinase 1) kinase[122, 123], which then propagates the signal to more downstream phosphorylation substrates, in a manner similar to Chk2.

Note that at this point, several targets of the ATM-Chk2 and ATR-Chk1 branches of DDR are common. For instance, H2AX was shown to be phosphorylated into γ-H2AX in a similarly rapid manner as after ATM activation following UV-Irradiation of cells. In this case however, γ-H2AX does not localize to IRIFs as per IR exposure, but rather produces a more widespread signal across the nucleus. The function of γ-H2AX following UV is considerably less clear than after IR, but may serve a similar purposes [124].

BRCA1, also a target of ATR, is also required for the activation of Chk1 kinase during the G2/M checkpoint[125]. (see below)

ATR activation also leads to p53 phosphorylation both directly[126, 127] and through Chk1[128], activating it and leading to long term cell cycle arrest.

1.3.2 Cell Cycle Checkpoints

One of the main consequences of the DDR signalling cascade is to induce cells to stop their progression through the cell cycle, in order to allow repair of damage to occur.

The G1/S checkpoint refers to the arrest of cells prior to their entry in to S phase, it consists of two mechanisms, one which arrest the cells rapidly and a second, slower one, that ensures the duration of the arrest. The initial arrest is caused by the phosphorylation of CDC25A by either Chk1 or Chk2, causing it to be exported from the nucleus and degraded[129, 130].

The maintenance of cell cycle arrest is ensured by the transcriptional activity of p53[131], which is first phosphorylated at Ser15 by either ATM or ATR and then by Chk2 or Chk1 at Ser20[132]. It then causes transcriptional activation of CDKN1A (Cyclin Dependent Kinase Inhibitor 1A; p21), which will then bind the CyclinE-Cdk2 complex ad inhibit its activity, which is required for entry into S phase. CDKN1A also binds the CyclinD-Cdk4 complex, preventing phosphorylation of Rb[133]. Unphosphorylated Rb in turn sequesters the E2F transcription factor away from its transcriptional targets, preventing the activation of genes required for progression into S phase[131, 134].

The intra-S-phase checkpoint is activated by damage that causes a block in ongoing replication[135]. Double strand breaks in this context causes cell cycle arrest through the ATM-Chk2-CDC25A pathway, but also involves phosphorylation of SMC1 by ATM, which, with the involvement of other DDR proteins such as BRCA1, NBS1 and FANCD2, leads to recovery from the damage through recombination repair pathways[114, 136, 137]. On the other hand, Helix-distorting adducts, including UV-induced pyrimidine dimers, as well modified bases caused by oxidative damage strongly block DNA replication which results in the formation of large tracts of single-stranded DNA (ssDNA) due to functional uncoupling of DNA synthetic enzymes at stalled replication forks [119, 138]. This is followed by coating of the ssDNA segments by binding of RPA, which leads to ATR activation and, again, arrest is in part mediated by downstream phosphorylation of CDC25A. [Recovery of staled replication forks again involves BRCA1 and NBS1.]

The G2/M checkpoint prevents the entry of damaged cells into mitosis and blocks them in G2. ATM or ATR activation again leads to CDC25A phosphorylation and degradation, preventing the de-phosphorylation and activation of Cdk1. In parallel, the Wee1 kinase is phosphorylated and activated and in turn phosphorylates Cdk1[139]. These two combined effects lead to increased Cdk1 phosphorylation and keeps the CyclinB1/Cdk1 complex inactive and prevents entry into mitosis[140].

Obviously, the DNA damage response also includes multiple mechanisms for the actual repair of the lesions that induce it. These mechanisms involve a highly complex network of proteins mediating a variety of repair processes each adapted to the type of damage present. The study of these highly complex processes is outside the scope of this thesis.

1.3.3 CUX1 in the DNA damage response

Of major relevance here, one of the categories of genes enriched among transcriptional targets of CUX1 identified in early genome-wide location analysis experiments[24] was that of "Cell Cycle Checkpoint". Further analysis of the list of putative CUX1 targets revealed several key members of the DNA damage response, particularly at the level of detection of damage and signal transduction, including key kinases ATM, ATR, Chk1 and Chk2. This strongly suggested that CUX1's transcriptional activity might play contribute to the ability of cells to respond to mutagenic insult.

Thus, the aim of my second research project was to first confirm the implication of CUX1 in the regulation of genes involved in DDR and then investigate what the implications of this regulation were for the ability of cells to recover and survive following exposure to damaging agents. My initial hypothesis was that cells deficient in CUX1 may lack the necessary proteins to properly respond to DNA damage and as a consequence be more sensitive to it.

1.4 The Wnt/β-Catenin pathway

1.4.1 Wnt/β-Catenin pathway overview

The Wnt/β-catenin signalization pathway plays an important role in the development of several tissues as well as in their regeneration by stimulating the growth of stem cells and multipotential progenitors [141, 142]. Wnts are secreted extracellular proteins that trigger a wide range of cellular responses upon receptor binding and activation. Most organisms contain multiple Wnt genes, which initiate distinct cellular pathways, namely the canonical Wnt/β-Catenin pathway, the planar cell polarity pathway, or the Wnt/calcium pathway. The cytoplasmic β-Catenin protein is maintained in a cytoplasmic complex that includes APC, Axin, and GSK3β. Phosphorylation by GSK3β causes the degradation of β-catenin unless the Wnt/β-Catenin pathway is activated by Wnt ligands. Wnt proteins initiate the canonical pathway by binding and activating the Frizzled and LRP receptors. Activation of these receptors induces the dissociation of β-catenin from the degradation complex. As its concentration increases, β-Catenin moves to nucleus where displaces the Groucho corepressor from the the TCF/Lef family of transcription factors with which it functions as a co-activator. As a consequence, genes that were previously repressed by the TCF/Groucho complex are now activated by the TCF/β-catenin complex. Many of the TCF/β-Catenin target genes code for Wnt signaling components that are capable of enhancing or antagonizing Wnt pathway activity [141]. Therefore, the Wnt pathway can be regulated by its targets through a feedback loop mechanism [143].

Much of our knowledge of the Wnt pathway was originally derived from studies in *drosophila*. Failure to express either *wingless* or *cut* in the wing margin causes similar phenotypes from which their names were derived: *wingless* and *cut* wing [144-147]. In the wing margin, *cut* is required in a cell-autonomous manner for wingless expression, whereas *wingless* is required for *cut* expression in neighboring cells [146]. Orthologs of *cut* in mammals are Cut homeobox 1 and 2, CUX1 and CUX2. In *Drosophila* imaginable discs, wingless is directly regulated by Cubitus interruptus, Ci, a transcription factor whose regulatory activity is under the control of the hedgehog ligand, Hh [148, 149]. In mammals, there are

four orthologs of Ci: Gli1, Gli2, Gli3 and GliS1. Both Gli1 and Gli2 have been implicated in human cancers [150, 151]. There is evidence to show that the Wnt/β-catenin pathway can be stimulated by Gli transcription factors, whether through activation of Wnt gene expression [152], or indirectly via induction of Snail and downregulation of E-cadherin [153]. Of note, CUX1 (Cut homeobox 1) was also shown to activate Snail gene expression and to cooperate with Snail in the repression of the E-cadherin gene [25].

1.4.2 The Wnt/β-catenin pathway in cancer

Because of its role in cell proliferation, development and differentiation, it is not surprising that the abnormal activation of the Wnt/β-Catenin pathway is frequently observed in cancer.

In colon cancer, the Wnt/β-Catenin pathway is frequently activated following inactivation of the tumour suppressor APC or mutations in the β-Catenin or Axin genes [154]. Recently, another mechanism of Wnt/β-Catenin pathway activation was demonstrated: transcriptional activation of one or several of the Wnt genes leads to the autocrine activation of Frizzled and LRP receptors and the subsequent increase in nuclear β-Catenin [155]. This autocrine activation has been shown in a sizeable proportion (20 – 25%) of breast cancers [155-157], lung cancers [158], neuroblastomas [159], acute myeloid leukemias and myelodysplastic syndromes [160]. Only a few transcription factor so far have been implicated in transcriptional regulation of the Wnt/β-catenin pathway, including the Gli family of factors, CUX1 and ATF3 [52, 146, 148, 149, 161-163].

Given the numerous regulatory steps involved in the activity of the Wnt/β-Catenin pathway, it represents an interesting target for drug development[164]. Indeed, it has recently been shown that small molecule inhibitors of the interaction between TCF4 and β-Catenin blocked the tumorigenic potential of Breast Tumour Initiating Cells derived from a Her2/Neu mouse model of mammary gland tumours. Furthermore, this inhibitor caused an arrest in tumour growth when used in vivo[165].

1.4.3 Activation of the Wnt/β-catenin pathway by CUX1

Transgenic mice expressing the transcription factors CUX1 or Glil under the control of regulatory sequences from the MMTV virus can develop mammary tumours with pathophysiological characteristics similar to that of mammary tumours in MMTV-Wnt1 mice [52, 161]. The molecular basis for the expression of a similar phenotype in these distinct transgenic mouse models remains to be defined. In the case of the MMTV-CUX1 transgenic model, this phenotype was not observed in all tumours but mostly in the adenosquamous carcinomas, which were shown to exhibit high levels of Wnt1, Wnt6, Wnt8b and Wnt10a genes as well as nuclear β-catenin [52].

These findings led us to formulate two hypotheses. First, that MMTV-CUX1 adenosquamous carcinomas represent a model for the autocrine activation of the Wnt/β-catenin pathway. Secondly, that other factors are needed, in addition to CUX1, for the transcriptional activation of this pathway.

To study this, I performed expression profiling on tumours derived from the MMTV-CUX1 mouse model previously described. However, since this model gave rise to tumours of different histological types, importantly solid carcinomas and adenosquamous carcinomas, the profiling was carried out on epithelial cells isolated from these tumours by laser-capture microdissection. Indeed, the proportion of epithelial cells in a solid carcinoma is much greater than that in a an adenosquamous carcinoma, and this method was necessary in order to eliminate any confounding contribution from the neighbouring tissues (connective tissue, infiltrating immune cells, etc.).

2. Methods

Antibodies

The following antibodies were used for Western blotting: anti-Actin (human and mouse), anti-ATR (human) form Santa Cruz Biotechnology (sc-1616, sc-1887); anti-ATM and anti-ATR(mouse) from AbCam (Ab78, Ab2905); anti-ATM(human), anti-p53 from Calbiochem (PC116, OP43); anti-p-ATM from Rockland (200-301-400); Anti-Chk1, p-Chk1, Chk2, p-Chk2 from Cell Signaling (2345, 2344S, 2662, 2661L); anti-53bp1 from BD Biosciences (612522); anti-active β-Catenin clone 8E7 from Millipore (05-665).

For immunofluorescence: anti-γ-H2AX from Cell Signaling (2577); anti-Rad51 from Santa-Cruz (sc-8349); anti-active β-Catenin clone 8E7 from Millipore (05-665); anti-mouse Alexa-488, anti-rabbit Alexa-488, anti-mouse Alexa-534 from Invitrogen.

For flow cytometry experiments: 488-conjugated anti-BrdU from Molecular Probes (A21303).

Cell Culture

All cell lines unless otherwise stated (see below) were maintained in Dulbecco's modified minimum essential medium (DMEM, Wisent) supplemented with 10% Fetal Bovine Serum (Invitrogen) and penicillin-streptomycin (Invitrogen). Cells were grown in humidified incubators at 37^{oC}, 5% CO_2 and atmospheric O_2, except for Mouse Embryonic Fibroblasts (MEF) cells and all cell lines derived from mammary gland tumours of CUX1 transgenic mice, which were grown at 3% O_2.

Hs578t were grown with 5% Fetal Bovine Serum. MCF10A cells were grown in DMEM/F12 media supplemented with 10% fetal bovine serum, penicillin-streptomycin, 5 μg/ml Insuline, 0.25 μg/ml of Hydrocortizone, 10 ng/ml of EGF. HCT116 cells were grown in McCoy's Media supplemented with 10% fetal bovine serum and penicillin-streptomycin.

MEF Cell isolation

Cux1 mutant mice in the albino OF1 outbred strain were obtained from the laboratory of Meinrad Busslinger and were maintained in the OF1 genetic background[8]. Primary MEF cells were prepared from 13.5-day-old embryos. The head, limbs and internal organs were removed, and the body was minced and incubated for 10 min in trypsin. Cells were then washed once in complete medium and seeded in a 100-mm dish. Genotyping was carried out by PCR on DNA extracted from the head and cells from all embryos of the same genotype within each litter were pooled.

siRNA Knockdown

CUX1 knockdown was performed by transfecting cells with a pair of siRNA constructs specific for CUX1 mRNA (5' GAAUCUUCUCGUUUGA-AACUUUGAA & 5' GCUUCAGAGCGAUAAUACACUAUUA) using Lipofectamine2000 (Invitrogen) according to the manufacturer's instructions. Knockdown was performed 7 days and again 2 days prior to performing experiments. In the case of clonogenic survival assays, knockdown was performed 5 days and again 5 hours prior to the experiment.

Retroviral Infection

Retroviruses were produced by transfecting 293VSV cells with either the pREV/TRE empty vector (Clontech) or one encoding p110 CUX1 (CUX1 aa 747–1505, Myc tagged at the amino terminus and hemagglutinin (HA) tagged at the carboxyl terminus). Viruses were applied to cells along with 8ug/ml polybrene and cells were centrifuged at 300g for 1h. Infection were carried on two consecutive days.

Reporter Assays

PCR amplification was performed to obtain fragments of genomic DNA from gene promoters of interest. These fragments were cloned into the luciferase reporter vector, pGL3 (Promega). 40,000 Hs578t cells were plated per well in 12-

well plates. Cells were transfected 24 hours later with 250 ng of reporter DNA and 500ng of p110 expressing effector or empty vector using GeneJuice transfection reagent (Novagen). Each transfection was carried out in triplicate. Cells were lysed 24 hours after transfection and the luciferase activity was measured by addition of luciferin substrate and measurement of light emission over a 30 second period on a BMG FLUOstar Omega microplate reader. Because the internal control plasmid is itself often repressed by CUX1, as a control for transfection efficiency, the purified β-galactosidase protein (Sigma) was included in the transfection mix, and the luciferase activity was then normalized based on β-galactosidase activity as previously described by Howcroft et al.[166].

Top/Fop reporter assays

Cells were plated as per reporter assays (or at 80,000 cells per well for HEK293T cells) and were cotransfected with 0.25 μg of either the TOP8x or FOP8x plasmids (Addgene) and 0.5 μg of either CUX1 p110 or empty vector as effector where indicated, as well as 0.5μg of SFRP1, SFRP2, DKK1 plasmids where indicated. The β-galactosidase protein (Sigma) was included in the transfection mix and the luciferase activity was normalized based on β-galactosidase activity.

Measurement of mRNA Levels

RNA was extracted using TRIzol reagent (Invitrogen) or Qiagen RNeasy Mini kit according to the manufacturer's instructions. cDNA was prepared using Superscript II RNase H-reverse transcriptase kit (Invitrogen) or Qiagen QuantiTect Reverse transcription kit following the manufacturer's instructions. Real time PCR was performed on a Rotor-Gene instrument (Corbett Life Science) using the QuantiTect SYBR Green PCR Kit (Qiagen) and specific primer pairs for each gene (See Supplementary Table 1). mRNA levels of genes of interest are normalized to Hprt1, except in samples isolated from CUX1 mammary gland tumours in which B2MB was used for normalization, due to disruption of the Hprt1 locus by the transgene.

Clonogenic Survival Assay

SiRNA treated MCF7 cells or MEF *Cux1*$^{Z/Z}$ and *Cux1* wild-type cells were exposed to either IR at doses of 1, 2 and 4 Gy, to UV at doses of 2, 5 and 10 J/m^2 or to hydroxyurea at doses of 10μm for 30 minutes. For MCF7, 500 cells were plated in 60 mm dishes in triplicate. For MEFs, 5000 cells were plated. After 10 days incubation, cells were washed with PBS, fixed with 10% phosphate buffered formalin for 10-20 min then stained with 0.1% crystal violet (Acros Organics) in 20% Methanol for 5-10 min. The number of colonies with 50 cells or more was counted.

Immunofluorescence

Cells were plated on glass coverslips and fixed in 4% paraformaldehyde. For γ-H2AX staining, the cell membrane was solubilised in phosphate-buffered saline (PBS) containing 5% FBS and 0.5% Triton X-100. The samples were incubated for 1 h in the solubilising solution containing primary antibodies for γ-H2AX. Secondary detection was done with Alexa Fluor 488- conjugated antibodies (Molecular Probes) and cells were counterstained with DAPI (Molecular Probes). Visualization was done using an Axiovert 200M microscope with an LSM 510 laser module (Zeiss). Images were analyzed using ImageJ64 software.

For Rad51 staining, cells were solubilised in 0.5% Igepal CA-630 and blocked in 10% FBS, 0.1% Igepal for 1h prior to 3h incubation with a primary antibody against Rad51 in blocking solution. Secondary detection and visualization was performed as indicated above.

For β-Catenin staining, ImageJ was used to calculate the mean signal intensity in the nucleus of cells by selecting the nuclear area in the DAPI channel and averaging the signal of the corresponding regions in the green (β-Catenin) channel.

Western Blotting

Nuclear extracts were prepared from cells using a procedure adapted from Lee et al.[167]. Nuclei were obtained by submitting cells to three freeze/thaw cycles in buffer A (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 1 mM dithiothreitol) along with protease inhibitor and phosphatase inhibitor tablets (Roche). Nuclei were lysed in Buffer C (20 mM HEPES [pH 7.9], 25% glycerol, 1.5mM MgCl2, 420 mM NaCl2, 0.2 mM EDTA) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed, and after electrophoretic transfer to polyvinylidene difluoride, membranes were washed in Tris-buffered saline–0.1% Tween 20 (TBS 0.1%T) and blocked in TBS 0.1%T containing 5% milk and 2% BSA. Membranes were probed with antibodies in TBS 0.1%T at room temperature for 1.5 hours for primary antibodies and for 45 min for secondary antibodies. Membranes were exposed on X-Ray films using ECL Western Blotting Substrate (Pierce).

G1/S and G2/M checkpoint assay

Cells were trypsinized, fixed in 75% ethyl alcohol, and stored at 20°C overnight. 50 μl of FBS was then added to each sample. The cells were centrifuged, washed in PBS, and resuspended in 300 μl of PBS containing 200 μg/ml of RNase (Sigma) and 5 μg/ml of propidium iodide (Sigma). Samples were incubated for 15 min at 37°C and analyzed using a FACScan (Becton Dickinson), with doublet discrimination to gate single cells. Cell cycle profiles were obtained with FlowJo software (Tree Star Software).

BrdU incorporation assay

At the indicated times, 5-Bromo-2'-deoxyuridine (BrdU) was added to the culture media at 100mM and incubated for 1 hour. Cells were then trypsinized and fixed in 4% paraformaldehyde. The cell membrane was solubilized in phosphate-buffered saline (PBS) containing 5% FBS and 0.5% Triton X-100. The samples were incubated for 1 h in the solubilizing solution containing a fluorescent BrdU antibody. The cells were then centrifuged, washed in PBS,

stained with propidium iodide and analyzed as for the G1/S and G2/M checkpoint assays, except that BrdU incorporation in S phase cells was scored.

Single Cell Electrophoresis (Comet Assay)

The extent of DNA damage was measured by alkaline lysis comet assays using a modified version of the protocol by Olive et al.[168]. Briefly, microscope slides (Fisher # 12-552) were pre-coated by distributing 900 μl of low-melt agarose (Sigma A9414) and a 3:1 mixture of low-melt agarose and cells in PBS at a concentration of 20,000 cells per ml was distributed on the slide and allowed to solidify before submerging in lysis buffer overnight at 4^{oC}. Slides were then processed and migrated as per Olive et al. and staining using Propidium Iodide. At least 30 pictures of each condition were taken on an Axiovert 200M microscope with an LSM 510 laser module (Zeiss). Comet tail moments were measured using CometScore software (TriTeck Corp).

Cytogenetic Analyses

Karyotyping, metaphase chromosome counts and breakage studies were carried out at the Quebec Leukemia Cell Bank.
(http://www.bclq.org/en/index.html)

Generation of transgenic mice

The p75-CUX1 and p110-CUX1 transgenic mice were generated by site-specific transgenesis into the Hprt locus as described in Cadieux et al.[51]. Each line was backcrossed for at least seven generations with mice of the FVB strain. Two lines of p75 CUX1 transgenic mice were generated; as expected, transgene expression in the FVB genetic background was found to be identical in the two lines. To study tumour burden, we generated cohorts of female mice carrying one copy of the transgene on one chromosome X. As a result of random inactivation of one X chromosome in each cell, the transgene would be expected to be expressed in ~50% of cells in females.

Immunohistochemistry

Formalin-fixed organs were embedded in paraffin and cut in sections of 5 μm and immunohistochemistry staining was done as previously described[50].

Laser-Capture-Microdissection of tumours

Mammary gland tumours from the above-mentioned p110 or p75 CUX1 transgenic mice were embedded in Tissue Freezing Medium (Triangle Biomedical Sciences #TFM-5) and flash frozen in liquid nitrogen. 10μm slices of tissue were prepared in a Microm HM505E Cryostat at ~-30^{oC} on positively charged slides (Fisherbrand Superfrost/Plus #12-550-15). Slides were then stained using a shortened H&E protocol adapted from Ponzo et al [169]. Epithelial cells were then isolated by IR (InfraRed) pulse on an Arcturus XT Laser Capture Microdissection instrument within 30-45 minutes of staining to obtain 100-150 pulses of material of 20-25 μm in diameter. mRNA was isolated from the microdissection caps using the Picopure RNA isolation kit from Arcturus according to the manufacturer's instructions, including the optional DNAse treatment using the Qiagen RNase-Free DNase kit (#79254).

RNA amplification

RNA isolated from microdissected tissue was amplified using the Arcturus RiboAmp HS PLUS RNA amplification kit according to the manufacturer's instructions for 2 rounds of amplification. For RNA isolated from cell lines, 2μg of total RNA obtained using the Arcturus Picopure RNA isolation kit was amplified using the Arcturus RiboAmp PLUS RNA amplification kit according to the manufacturer's instructions for a single round of amplification. Corresponding reference mRNA samples (Stratagene) were amplified using both the 2 round amplification and 1 round amplification method according to the matching experimental samples.

Amplified RNA labeling

Amplified mRNA (aRNA) was labelled using the Arcturus Turbo Labelling Cy5 and Cy3 kits using slightly modified instructions: The labeling reaction was carried out using 5μg of aRNA in a 20μl volume instead of 50μl to increase the dye incorporation rate and the Turbo Blocking buffer provided for subsequent hybridization of sample was not used. Samples and reference samples were labeled with both Cy5 and Cy3 independently.

Expression Array hybridization

Labelled aRNA was hybridized to Agilent's Whole Human Genome Microarray (G4112F), washed and scanned on an Agilent 5μm scanner model G2505B according to the manufacturer's instructions. Dye swapped hybridizations of all samples with references were performed to control for dye hybridization bias and to provide a technical duplicate of each hybridization.

Expression profiling data analysis

Expression profiling array images were processed using the Feature Extraction software from Agilent. The raw data was then processed (Background correction, within-array normalization, between-array normalization) and analyzed using the LIMMA package [170] on the R platform (http://www.R-project.org). Unsupervised hierarchical clustering was performed using the "heatmap.2" function of the "gplots" package, using Euclidean distance to compute the distance matrix.

Public dataset recovery and Meta-analysis

Human breast and lung cancer expression profiling datasets were obtained from the Oncomine database (www.oncomine.org). We retrieved all datasets comprising 90 or more samples and for which expression of at least one of CUX1 and GLIS1 as well as at least 15 out of the 19 Wnt ligand genes were available. Note that Affymetrix datasets contain probes for the CASP transcript at the CUX1 locus. For each dataset, samples were sorted from lowest to highest expression of

Wnt ligand genes using the BreSat algorithm (See Below). The 25% highest Wnt expressing samples were compared to the 25% lowest Wnt expressing samples for *CUX1*, *GLIS1*, *GLI1*, *GLI2* and *GLI3* expression as well as the following markers of epithelial to mesenchymal transition: CDH1, CDH2, SNAI1, SNAI2, OCLN, VIM and TWIST1.

BreSat sorting algorithm

This algorithm is a non-parametric method of ranking samples according to their expression of a given set of genes. Briefly, it computes the sum of the rankings of each sample across the given genes and reorders the samples according to this ranking.

Chromatin Affinity Purification (ChAP)

ChAP was performed on 5 x10^8 Hs578T. The cell nuclei were purified as described in [171], then lysed in RIPA-M buffer (10 mM Tris-HCl pH8, 1 mM EDTA, 0.5 mM EGTA, 150 mM NaCl, 1% Triton X-100, 0.5% DOC, 0.1% SDS, 1 mM PMSF, protease inhibitors) and sonicated on ice to obtain 250- to 800-bp-long DNA fragments. Stably expressed recombinant p110-Tag^2 protein was purified by the Taptag purification method with some modifications [172]. The IgG matrix bound p110-Tag^2/DNA were washed in wash buffer I (20 mM Tris-HCl pH8, 2 mM EDTA, 2 mM EGTA, 150 mM NaCl, 1% NP-40, 0.5% DOC, 0.2% SDS), wash buffer II (20 mM Tris-HCl pH9, 2 mM EDTA, 2 mM EGTA, 500 mM NaCl, 1% NP-40, 0.5% DOC, 0.1% SDS), wash buffer III (50 mM Tris-HCl pH7.5, 2 mM EDTA, 1 mM EGTA, 0.5M LiCl, 1% NP-40, 0.7% DOC,) and then TEV buffer (10mM Tris-HCl (pH8.0), 100mMNaCl, 0.1% TX-100, 0.5mM EDTA, 10% glycerol, 1mM DTT). After TEV protease digestion, the released protein/DNA complexes were purified by affinity chromatography on calmodulin beads in the presence of calcium and then eluted with EGTA. After de-crosslinking, samples were treated with RNase A and Proteinase K. Un-enriched input chromatin was put aside as a control.

Preparation of ChAP purified DNA for Hybridization

ChAP purified chromatin was amplified by the method of Ligation-Mediated PCR as detailed previously [173]. Briefly, ChAPed DNAs and input DNA were blunted, ligated to a unidirectional linker and amplified by PCR for 24 cycles to generate a sufficient amount of DNA. Amplified DNA samples were Cy5 labeled and amplified input controls were Cy3 labeled using Nimblegen's Dual-Color DNA Labeling Kit according to the manufacturer's instructions.

DNA Microarray Hybridization

Labeled samples were hybridized to either NimbleGen's HG17 ENCODE or their HG18 Human Promoter Array Set high density oligonucleotide tiling array (385k probe format) and then washed according to the manufacturer's instructions. Arrays were scanned on an Agilent 5μm scanner model G2505B using customized scan area settings (X: 28, Y:6, Width: 20, Height: 14, values in mm).

ChAP-Microarray Result Analysis

For both array platforms (ENCODE and promoter array), grid alignment, raw signal extraction, peak identification and peak mapping were carried out using the Nimblescan v8.0 software according to the company's instructions. Identified peaks were considered significant with a false discovery rate (FDR) below 0.05, which is considered highly confident. Further analysis of identified binding sites such as distribution and distance to features of interest were carried our using either the R platform for statistical computing (http://www.R-project.org) or scripts written in PERL (Practical Extraction and Report Language, www.perl.org).

ChAP-Microarray Result Validation

Independent ChIP experiments using antibodies specific for endogenous CUX1 were carried out in Hs578t, as previously described[24]. Real-time PCR was used to measure the level of enrichment of genomic target regions in ChIP

DNA vs. the un-enriched input DNA. We selected 25 genes from both the targets identified on the ENCODE array and on the Promoter array set and designed primers specific for the corresponding regions where CUX1 was putatively identified as binding.

ENCODE Binding Sites for c-MYC and E2F1

We used ChIP-chip binding sites for E2F1 and c-Myc downloaded from the website of Dr. Peggy Farnham laboratory at <http://genomics.ucdavis.edu/farnham/suppdata.html>. This dataset contains the binding sites predicted for E2F1, c-MYC and POLR2A (RNA polymerase II) in the ENCODE regions classified by 4 criteria: L1 ($P < 0.0001$ and 98^{th} percentile), L2 ($P < 0.0001$ and 95^{th} percentile), L3 ($P < 0.05$ and 98^{th} percentile) and L4 ($P < 0.05$ and 95^{th} percentile) [72]. Based on the validation of 29 binding sites, Bieda et al. conclude that L1 binding sites are highly reliable, L2 and L3 binding sites are also reliable however based on sparser testing and L4 binding sites are usually artifacts. Binding sites identified with the L1 criteria were used for our analyses. The chromosomal intervals for binding sites predicted for E2F1 and POLR2A belonged to genomic coordinates using hg16, whereas c-MYC binding sites were in hg17. Therefore, the lift-over program found on the online GALAXY platform [174-176] was used and random results were verified using UCSC genome browser to convert hg16 coordinates to those of hg17. There were 1 and 2 binding sites for E2F1 at L1 and L3 respectively (hg16), which could not be mapped to hg17.

Consensus Sequence Analysis

Genomic sequences corresponding to regions of interest (binding sites or other) were obtained using the online GALAXY platform (https://main.g2.bx.psu.edu/). Scripts written in R were used to identify consensus sequences within regions of interest.

Functional Overrepresentation Analysis

Identification of overrepresented gene functions was carried out using the online annotation tool DAVID. Genes that were bound by CUX1 (Targets) were compared with all genes present on the microarray (Background). Overrepresentation of a function depends on the increase in the proportion of genes involved in a given function between CUX1 targets and the background. The P-value is determined using an improved Fisher's exact test from the DAVID software[81, 82].

De Novo Binding Motif Identification

De Novo motif discovery was performed using the DREME (Discriminative DNA Motif Discovery) motif discovery tools form the MEME suite of tools. Comparison with known DNA binding motifs was performed using the TOMTOM algorithm using the JASPAR CORE database as a reference for comparison. (meme.nbcr.net/)[177-179].

DHS and ChromHMM Data Analysis

Data tracks were downloaded from the UCSC's Encode data portal (http://genome.ucsc.edu/ENCODE/). Genomic locations were compared to those of the CUX1 binding sites using scripts written in R. UCSC Accession numbers of the tracks used are: wgEncodeEH000503 (GEO accessions GSM736552 and GSM736634) for the DHS data and wgEncodeEH000786 for the ChromHMM data.

Live-cell Imaging

Cells ($1x10^4$) were plated on 10ug/ml fibronectin coated plates for three hours and then subjected to time lapse-video microscopy for every 5 minutes for a total of 16 hrs. Last forty frames from each condition were subjected to analysis using metamorph software.

Invasion assays

Modified boyden chamber with 8mm pore size were coated with mixture of Matrigel (BD biosciences) with final concentration of 1.25mg/ml and acid purified Rat tail Collagen I (Gibco) with final concentration of 2.5mg/ml for 2 hrs at 37C. Cells ($1x10^6$) were seeded on the bottom of the filters and were incubated at 37°C with 5%CO_2 for 5 hrs to adhere. After 5 hrs, filters were inverted and top compartment was filled with 10%FBS and the bottom compartment were filled with plain DMEM and cells were allowed to invade for 48-72Hrs towards serum. Cells were stained with Calcein-AM (Invitrogen) as per manufacturer's instructions and confocal Z-stacks were acquired every 10um. Cells invading 20um into the matrix and beyond were quantified as percentage of total cells.

3. Results

Rationale

It has been established that CUX1 functions as a transcriptional regulator that can both activate and repress its targets in different contexts. However, the extent of regulation and the biological functions that this regulation plays a role in is not fully established.

The overall goal of my thesis is to study the transcriptional activity of CUX1, both from a global perspective as well as with regards to specific cellular functions.

I investigated the former through the use of genome-scale methods of analysis: Genome-wide location analysis by ChAP-Chip and expression profiling of cells following changes in the levels of CUX1. The findings of this approach have informed us on the general activity of CUX1 in the regulation of its targets, but also provides specific information about a large number of targets. This information was then used by both myself in my subsequent projects and other members of the laboratory in their own projects and may also be useful to other researchers in the future.

Our genome-wide studies of CUX1 regulation have suggested a role for it in the DNA damage response, but the implications that this had for the cells was not established. This lead me to investigate the transcriptional effect of CUX1 on this limited subset of genes and investigate the cellular impacts of CUX1 activity with regards to DDR.

Finally, evidence from a mouse breast cancer model linked CUX1 expression with activation of the Wnt/β-Catenin pathway, but only in a subset of tumours. I again used expression profiling, combined with meta-analysis of public profiling datasets, to understand the transcriptional role of CUX1 and the requirement of other factors in the activation of this pathway.

3.1 Transcriptional regulation by CUX1

3.1.1 Strategy To Identify p110 CUX1 Binding Sites

The overall goal of my thesis was to define the modes of transcriptional regulation by CUX1 and, in particular, determine whether CUX1 can regulate genes at a distance. As detailed in the introduction, previous transcriptional studies and cell-based assays have implicated the p110 CUX1 isoform in transcriptional activation and repression of target genes. Since p110 CUX1 is generated by proteolytic processing, its primary sequence is included in the full-length CUX1 protein sequence. Consequently, all available antibodies that bind to p110 CUX1 also recognize p200 CUX1. Our strategy to identify in vivo binding sites for p110 CUX1 was to isolate chromatin by two different methods. First, we purified chromatin by tandem affinity purification (TAP) using a population of Hs578t cells stably expressing moderate levels of a p110 CUX1 protein with two epitope tags at its C-terminus, p110-Tag^2 (Fig. 2A and B). Chromatin isolated in this manner as well as total chromatin (input) were used in hybridizations on the NimbleGen HG17 ENCODE high density oligonucleotide tiling array. Secondly, binding sites identified in the microarray were then validated by performing independent ChIP in the parental Hs578t cells using CUX1 antibodies, 861 and 1300 (Fig. 2A). Importantly, these cells express endogenous CUX1 proteins only. The strategy of chromatin affinity purification (ChAP) followed by microarray analysis (ChAP-chip) has previously been validated [180], and described in detail [181].

3.1.2 Distribution of CUX1 Binding Sites on the ENCODE Array

Using a stringent false discovery rate (FDR=0.05), 513 CUX1 binding sites were identified on the ENCODE array (Table 1). The recruitment of CUX1 to 23 out of 25 genomic sites (92%) was validated in quantitative-PCR assays using chromatin that was independently obtained from Hs578t cells by immunoprecipitation with CUX1 antibodies (Table 1). 79.6% of probes on the ENCODE array derive from transcribed genomic regions. 70.9% of CUX1 binding sites were located within transcribed regions, indicating a 1.6-fold

enrichment in non-transcribed regions. In comparison, data obtained from ChIP on the ENCODE platform[72] for c-MYC reveals a 1.56 fold enrichment in non-transcribed regions while E2F1 showed a strong enrichment for transcribed regions (Table 2).

Mapping of CUX1 binding sites relative to transcription start sites (TSS) generated a bell-shaped curve of low height around TSS (Fig. 3A). 14.2% of all binding sites overlapped a TSS, and an additional 17% and 16% of binding sites were respectively located in the 4 Kbp region upstream and downstream of a TSS. The number of binding sites gradually declined with increasing distance. Yet, over 6% and 8% of binding sites were situated at more than 40,000 bp upstream or downstream, respectively, from the closest TSS. 53% of CUX1 binding sites are located more than 4,000 bp away from a TSS and approximately 14% of all CUX1 binding sites are situated at more than 40,000 bp from a TSS.

We compared the distribution of CUX1 binding sites with those of 3 randomly generated sets of binding sites, as well as those of c-Myc and E2F1 using the data of Bieda et al., 2006 [72] (Fig. 3B and C). We note that the distributions of randomly generated sets of binding sites exhibited flatter bell-shaped curves around TSS (Supplementary Fig. 1). We conclude that the higher frequency of CUX1 binding sites close to TSS reflects the preferential recruitment of CUX1 to promoter regions. The same cannot be said regarding the binding sites that are located at more than 40 Kbp from TSS, since the same proportions of randomly generated binding sites were located in these regions.

In contrast to CUX1 and c-Myc, the E2F1 transcription factor was found to bind almost exclusively to the region immediately adjacent to TSS. The preference of E2F1 to core promoter regions led the authors to posit that E2F1 is recruited via protein interactions with components of the general transcription machinery [72]. The wider distribution of binding sites observed for CUX1 and c-Myc is also observed for other transcription factors [73, 182, 183] (Supplementary Fig. 2A-C), while other factors show a preference for TSS similarly to E2F1 (Supplementary Fig. 2D-F). Yet other factors show different patterns of binding, such as Pax8, which exhibits preference for non-promoter CpG islands and a

tendency to bind in the 10-100 Kbp range rather than close to the TSS of genes[184].

3.1.3 Binding of CUX1 to Distant Regulatory Elements

We compared the location of CUX1 binding sites that are more than 4 Kb from the nearest TSS to DNAse hypersensitivity mappings and ChromHMM data in human mammary epithelial cells from published datasets. DNAse hypersensitivity sites have been used as markers of regulatory DNA elements such as enhancers, silencers, insulators and locus control regions [185-188]. ChromHMM is a computational method that compiles data from histone modification mappings and integrates them to predict genomic elements such as enhancers[189]. This analysis revealed that respectively 19.2% and 22.1% of distantly located CUX1 binding sites are present within 1 kb of a DNAse hypersensitivity site and of an enhancer predicted (Table 3). Both of these proportions are greater than what is seen for randomly distributed binding sites. However, there was no enrichment of CUX1 binding sites in proximity of insulator elements (Table 3). These results are in agreement with the notion that CUX1 can perform some regulatory functions when binding at a distance from transcription start sites.

3.1.4 Detection of CUX1 Binding Sites and Consensus Binding Motif on Promoter Arrays

Promoter microarrays are useful because they enable one to interrogate easily over 30,000 gene promoters. A limitation is that only a limited amount of promoter sequences can be included for each gene, precluding the detection of far away binding sites that could play a role in transcriptional regulation. Based on the localization of CUX1 binding sites on the ENCODE array, we calculated that between 17.2% to 26.6% of CUX1 binding sites would be identified on commercially available promoter arrays (Table 4A). However, since for many distant CUX1 binding sites another binding site is also present close to the transcription start site, we estimated that between 44.6% to 55.5% of gene targets

would be identified on distinct promoter arrays (Table 4B). In contrast, as E2F1 is targeted to transcription start sites, between 80.4% to 85.8% of E2F1 binding sites would be expected to be identified on a promoter array.

We verified these predictions by performing a ChAP-chip experiment using the Nimblegen promoter microarray. Total chromatin (input) as well as purified chromatin from Hs578t cells expressing p110 CUX1-Tag2 were used in hybridization on the promoter array of NimbleGen. Using a stringent false discovery rate (FDR=0.05), 5828 CUX1 binding sites were identified on 4706 gene promoters (Table 5). The recruitment of CUX1 to 25 out of 25 genomic sites (100%) was validated in quantitative-PCR assays using chromatin that was independently obtained from Hs578t cells by immunoprecipitation with CUX1 antibodies (Table 5). The vast majority of target genes (83.7%) contained only one CUX1 binding site, yet a sizable fraction contained 2 or more binding sites (Table 5).

According to the predictions shown in Table 3B, 44.6% of CUX1 target genes should be identified on the promoter array from Nimblegen. We calculated the proportion of ENCODE genes with a CUX1 binding site that were also identified as putative targets of CUX1 in the promoter array. When we considered all 513 CUX1 binding sites and 445 adjacent ENCODE genes, we found that 92 genes (21%) were identified in the promoter array (Table 4C, second column). When we considered only the 85 ENCODE genes that were regulated in response to changes in CUX1 levels (see below), we found that 27 genes (32%) were identified as putative targets of CUX1 in the promoter array (Table 4C, third column).

The CUX1 consensus binding site, ATCRAT (where R = C or A), was found to be present at 47.2% of the 5828 bound genomic sites (Table 6). This frequency was judged to be significant as the CUX1 consensus binding site was found to be present in only 17.5% of 5828 randomly chosen regions of equal size. Notably, the GC content between bound and unbound regions is practically identical, and thus cannot account for the difference in binding site occurrence (Table 6). Yet, only 8.3% (3633/43778) of the CUX1 consensus sites present on

the array were bound *in vivo*. We conclude that the CUX1 consensus binding site plays a role in the recruitment of CUX1 at specific genomic locations, but the presence of a consensus site is not sufficient.

3.1.5 Identification of Binding Motifs In Genomic Regions Bound by CUX1

We envision that interactions with other transcription factors play an important role in recruiting CUX1 to specific locations. In agreement with this notion, functional analysis revealed distinct sets of cellular functions among gene targets that contain an ATCRAT consensus and those that do not (Table 7A and B). To further test the possibility that CUX1 may interact with other factors, we investigated the presence of binding motifs other than that of CUX1 using the MEME suite of analysis tools (meme.nbcr.net/). We first tested the reliability of the tool by using it to find motifs in the sequences of CUX1 BS in which we had independently determined that they contained the established ATCRAT consensus. As expected, it identified the ATCRAT consensus as the most enriched motif in the set of sequences, by a vast margin (Table 8A, entry 1). We then analyzed binding motifs in the two sets of CUX1 binding sites: those that contained the ATCRAT motif and those that did not. Interestingly, only one common binding motif was found in the two sets, while the rest of the binding motifs were unique to each set (Tables 8A and 8B). These findings support the notion that targeting of CUX1 to specific genomic sites is influenced by protein-protein interactions with other DNA binding proteins.

3.1.6 Regulatory Effects of CUX1 on Putative Targets

To verify the effect of CUX1 on putative targets, we performed expression profiling on three Hs578t cell populations: cells that had been infected with a retrovirus expressing an shRNA against CUX1, cells infected with a retrovirus expressing p110 CUX1, or cells infected with an empty retrovirus. In each case, microarray hybridization were carried in quadruplicate such that a p value could be calculated for each difference in gene expression. Results from expression profiling were validated by repeating the infections and performing RT-qPCR

analysis on 20 genes whose expression went up or down in response to one treatment or the other (Fig. 4A and B). All genes tested in this manner displayed changes in gene expression in the same direction as that observed in the microarray hybridization: genes that were repressed in expression profiling were also repressed when mRNA levels were measured by RT-qPCR. Similar observations were made for genes that were activated. We note, however, that the fold activation or repression calculated by RT-qPCR were not necessarily proportional to the changes observed in microarray hybridization. For example, EEF1A1 and C20ORF44 mRNA were increased respectively 7.3 and 1.8 fold when measured by RT-qPCR, but were increased 1.7 and 1.4 fold in microarray analyses. Some of these differences could be due to the fact that measurements by the two methods were made with RNA prepared from independent experiments. Notwithstanding the differences in magnitude, the effects of CUX1 on gene expression was confirmed for all tested genes.

A total of 445 genes are present on the ENCODE array, and all have a CUX1 binding site located within 213 Kbp of their TSS. Expression profiling results could be matched for 349 of these genes. Using a cut-off of 50% either up or down-regulated and a p value below 0.05, we observed differences in the expression of 26 target genes (7.4%), following changes in CUX1 levels (Table 7A). 20 genes responded to CUX1 knockdown, and 6 genes, to p110 CUX1 overexpression (Table 9A). Among the 26 regulated target genes, 10 genes (38%) were activated and 16 genes (62%) were repressed by CUX1 (Table 9A). Similar proportions of activated and repressed genes were found when a cut-off of 25% change in gene expression was employed (Table 9B). These findings confirm that p110 CUX1 can participate in transcriptional activation or repression depending on promoter context.

Similar results were obtained when we analyzed the expression of putative targets identified on the promoter array. A total of 348 genes, 8.4% of all putative targets for which expression profiling results could be matched, were regulated by CUX1. 287 and 85 genes exhibited regulation in response to CUX1 knockdown or

p110 CUX1 overexpression, respectively. 181 (52%) were up-regulated by CUX1 while 167 (48%) were down-regulated by CUX1.

3.1.7 Effect of Distance on Transcriptional Regulation by CUX1

We noted that CUX1 regulated 7.4% and 8.4% of putative targets from the ENCODE and the promoter arrays, respectively. This result was somewhat surprising since approximately 40% of CUX1 binding sites on the ENCODE array are located more than 40,000 bp away from the closest transcription start site. This observation led us to investigate the relationship between the position of a CUX1 binding site relative to a transcription start site and the probability of a gene to be regulated in response to changes in CUX1 levels. When genes were classified according to the distance between the CUX1 binding site and the transcription start site, we did not observe significant difference in the fraction of targets that were regulated by CUX1 (Fig. 5A and B). However, we observed much variability in the fraction of regulated genes because the number of genes within some distance intervals were very small. Therefore, to increase the sample size, we repeated the analysis this time using a cut-off of 25% either up or down and a p value below 0.05 (Fig. 5C and D). We observed differences in the expression of 62 and 36 genes in response to CUX1 shRNA and CUX1 overexpression, respectively (Table 9B). Again, more genes were found to be regulated by CUX1 using the shRNA approach. Among genes that exhibited regulation by CUX1, 35 genes (41%) were activated by CUX1, and 50 genes (59%) were repressed by CUX1 (Table 9B). The histogram presenting the percentage of regulated genes versus the distance of CUX1 binding sites to TSS shows that essentially the same proportion of genes are regulated whether CUX1 binds close or far away from the TSS (Fig. 5C). Indeed, no statistical difference was observed between genes bound at the TSS and those bound more than 40 Kbp away. We conclude that CUX1 can activate or repress transcription when bound at a distance from a transcription start site.

3.1.8 Effect of Multiple Binding Sites on Transcriptional Regulation by CUX1

The presence of multiple CUX1 binding sites has a modest, yet significant, impact on the probability that a gene is regulated by CUX1. CUX1 regulated 7.9%, 11.2% of genes that contain respectively one or two CUX1 binding sites, respectively (Table 10C).

3.1.9 Effect of Gene Position on Transcriptional Regulation by CUX1

Intuitively, one would assume that a transcription factor is more likely to regulate the closest promoter. Yet, some enhancers will exhibit an effect on a promoter situated on one side, but no effect on the promoter that is on other side on the map. This sort of selectivity between an enhancer and a promoter has been explained by the presence of boundary or insulator elements or by specific interactions between proteins bound at the enhancer and the regulated promoter. Previous studies on CUX1 have all focused on genes that contain a CUX1 binding site within the immediate promoter. To begin to investigate the rules that govern the action of CUX1, we calculated the fraction of different types of CUX1 targets that were regulated in response to changes in CUX1 levels. Three types of genes were analyzed: 1, genes that are the closest to the CUX1 binding site; 2, genes that are further away and in the other direction from the CUX1 binding site; 3, genes that are located further away and are separated by another gene from the CUX1 binding site. For each category, we calculated the percentage of genes that exhibit a 1.25 or 1.5 change in expression following p110 CUX1 overexpression or CUX1 knockdown. Strikingly, essentially similar fractions of genes were regulated whether they were closest to the CUX1 binding site or were located further away in the other direction (Fig. 6, compare 1 and 2). Moreover, the proportion of regulated genes was not significantly lower among genes that belong to the third category (Fig. 6, type 3 genes). We conclude that CUX1 is capable of regulating genes at a distance. Moreover, CUX1 can regulate more than one gene on certain genomic loci.

3.2 CUX1 in the DDR

3.2.1 CUX1 regulates a significant number of genes involved in the DDR

From the transcriptional targets of CUX1 identified in genome-wide location analysis experiments, functional classification of targets revealed an involvement of CUX1 targets in cell cycle checkpoint control (Table 11). We used multiple experimental approaches to investigate the role of CUX1 in transcriptional regulation of 18 putative targets known to be involved in DDR checkpoint signalling (Table 12). Firstly, chromatin immunoprecipitation followed by Real-Time PCR (ChIP-qPCR) was employed to measure CUX1 recruitment to each promoter (Table 12 column 2). Secondly, mRNA expression of DDR target genes was quantified in the context of CUX1 deficiency, either using siRNA-mediated knockdown (Fig. 7A) or by comparing mouse embryonic fibroblast (MEF) cells from wild-type vs. *Cux1*$^{Z/Z}$ mutant mice (Fig. 7B)[8]. A significant transcriptional effect was observed in both assays for 14 out of 18 genes and in one assay for the remaining 4 genes. All genes manifested decreased expression upon CUX1 knockdown, with the exception of CDKN1A (p21), which was upregulated (Fig. 7 & Table 12, columns 3-6). Thirdly, Hs578t cells were infected with a retroviral vector expressing p110 CUX1 and the expression of target genes measured 24 hours later. All genes displayed increased mRNA levels, with the exception of CDKN1A, which was down-regulated (Fig. 7C & Table 12, column 7). Fourthly, we cloned the proximal promoter region of 7 target genes into a reporter construct to verify whether these genomic sequences were sufficient to confer regulation by p110 CUX1 (Fig. 7D & Table 12, column 8). Except for CDKN1A which was repressed, all tested genes were activated by p110 CUX1. The regulatory effects of CUX1 on CDKN1A expression observed using these four assays are in agreement with previous studies reporting the role of CUX1 as a transcriptional repressor of this gene[12, 22, 26, 28, 47, 48, 190]. Altogether, the above data conclusively show that CUX1 binds to the promoters of many DDR genes and regulates their expression.

3.2.2 DDR Signalling is reduced in CUX1 deficient cells

The decrease in DDR gene expression upon CUX1 knockdown suggests that this transcription factor contributes to the maintenance of a transcriptional program required for cellular responses to mutagenic insult. Among CUX1 targets (Table 12) are critical kinases involved in the transmission of DNA damage signals to downstream effectors, specifically ATM/Chk2 and ATR/Chk1, which are rapidly activated by double strand breaks (e.g. IR-induced) and replication stress (e.g. UV-induced) respectively. We therefore investigated whether DDR signalling through these kinases is impaired in the absence of CUX1. Immunoblotting assays were performed to investigate the expression and phosphorylation status of checkpoint kinases following DNA damage. Transfection of MCF7 cells with CUX1-specific siRNA greatly reduced CUX1 expression (Fig. 8A, lanes 1 and 3). IR treatment did not modulate CUX1 expression in MCF7 cells (Fig. 8A, compare lanes 1 and 2, and lanes 3 and 4). Although Chk1 and Chk2 mRNA levels were decreased in CUX1 deficient cells (Fig. 7A and Table 12), protein levels were not affected (Fig 8B and C). This suggests that translational or post-translational mechanisms play an important role in regulating Chk1 and Chk2 protein levels. On the other hand, we noted a decrease in steady-state levels of the ATR and ATM kinases, and of the adaptor protein 53BP1 which is critical for ATM-mediated Chk2 activation in response to DSBs [112] (Fig. 8B and 8C, compare lanes 1 and 2 with 3 and 4). Using phospho-specific antibodies, we also detected a strong reduction in phosphorylation of Chk1 Ser317 following UV-irradiation (Fig. 8C, compare lanes 2 and 4) and an even more striking decrease in phosphorylation of Chk2 Thr68 and ATM Ser1981 after IR (Fig. 8B, compare lanes 2 and 4). These results indicate that both ATM and ATR exhibit reduced activity after DNA damage in CUX1 deficient cells. In addition, p53 accumulation following IR exposure was reduced in siRNA treated cells (Fig. 8D, compare lanes 2 and 4), possibly due to impaired ATM/Chk2 signalling[191]. Importantly, a decrease in ATM and ATR steady-state levels was also observed in *Cux1*$^{Z/Z}$ MEFs (Fig. 8E).

Immunofluorescence microscopy was next employed to evaluate phosphorylation of H2AX (γ-H2AX), well characterized as a very early event mediated by ATR and ATM following exposure to UV and IR, respectively[192]. Consistent with previous findings, diffuse γ-H2AX nuclear staining was evident following treatment with 20J/m^2 of UV, whereas distinct foci were discernable in cells treated with 10 Gy of IR (Fig. 9)[124]. However, in both MEFs and MCF7, knockdown of CUX1 caused a decrease in the proportion of cells showing a positive γ-H2AX signal after UV-irradiation and in the number of γ-H2AX foci per cell after IR (Fig. 9A and 9B). Similar results were obtained following hydroxyurea treatment of cells (Supplementary Fig. 3A).

The above results, taken together, indicate that CUX1 is required for optimal signal transduction downstream of ATM and ATR in response to DNA damage.

3.2.3 CUX1 deficient cells are sensitive to the cytotoxic effects of diverse-acting DNA damaging agents

The above results suggest that knockdown of CUX1 would render cells more sensitive to the cytotoxic effects of diverse genotoxic agents. We tested this using either (i) siRNA-mediated knockdown of CUX1 in MCF-7, or (ii) MEFs from wild-type or *Cux1*$^{Z/Z}$ mutant mice, following exposure to IR or UV. In response to either agent, in both experimental systems, CUX1-deficient cells exhibited significantly decreased clonogenic survival (Fig. 10). Again, similar results were also obtained following exposure to hydroxyurea (Supplementary Fig. 3B). The well established role of CUX1 in cell proliferation [53] cannot account for these differences since colony forming ability in mutagen-treated cells is calculated relative to undamaged cells. Furthermore, in the absence of DNA damage, there was no disparity in the absolute number of colonies between cells expressing more or less CUX1 (Supplementary Fig. 4). We conclude that the ability of cells to survive in the face of genotoxic insult is compromised when CUX1 is either inactivated or its expression reduced.

3.2.4 CUX1 knockdown impacts cell cycle checkpoints

In light of the attenuated ATM/ATR signalling response and reduced viability of CUX1-deficient cells during genotoxic stress, the capacity of such cells to trigger cell cycle checkpoints was investigated. We first evaluated the G1/S checkpoint using flow cytometry to quantify the proportion of cells remaining in G1 24h post-treatment with IR or UV. The assay was performed on cells synchronized by treatment with the mitotic inhibitor nocodazole prior to irradiation, to exclude the possibility of G1 reentry of cells that were in G2/M at the time of irradiation. Treatment of MCF7 cells with 10 Gy IR led to a 21.6% increase in the fraction of cells remaining in G1 (Fig. 11, left panels, from 10.9% to 32.5% G1, nocodazole vs. nocodazole + IR). This increase reflects cells that were prevented from progressing into S phase as a result of the ATM-mediated G1/S checkpoint. Significantly, siRNA-mediated knockdown of CUX1 attenuated the increase in G1 to 13.9% (Fig. 11A, right panels, from 11.8% to 25.7% G1, nocodazole vs. nocodazole + IR). Similarly, following treatment with UV, the increase in the G1 fraction was significantly reduced in siRNA-treated MCF7 cells as compared to the control cells (Fig. 11A, nocodazole vs. nocodazole + UV, results summarized in Fig. 12B). A similar trend was observed when the efficiency of the G1/S checkpoint was compared between wild-type and *Cux1*$^{Z/Z}$ MEF cells following exposure to either IR or UV (Fig. 11C). The above data indicate that RNAi knockdown or genetic inactivation of CUX1 compromises the G1/S checkpoint in cells afflicted with either DSBs or increased replication stress.

The capacity of cells to abrogate DNA replication after irradiation, as controlled by the S phase checkpoint, was also assessed in CUX1 knockdown cells. To this end, we measured incorporation of the deoxyuridine analogue bromodeoxyuridine (BrdU) in wild-type vs. *Cux1*$^{Z/Z}$ MEFs. While both cell types showed a reduction in BrdU incorporation following exposure to 10 Gy of IR, wild-type MEFs showed a more rapid and pronounced decrease in the proportion of cells actively synthesizing DNA following irradiation (Fig. 11D).

Finally, G2/M arrest was evaluated in wild-type vs. *Cux1*$^{Z/Z}$ MEFs, and in siRNA-treated MCF7 cells. The increase in cellular G2 content was measured by

flow cytometry 24 hours after exposure to IR. All irradiated cells exhibited an increase in G2 content although not to the same extent, as illustrated by wild-type vs. *Cux1*$^{Z/Z}$ MEFs (Fig. 12A). While the G2 content of wild-type MEFs increased by 5.2%, 12.9% and 19.9% after exposure to 10Gy, 15Gy and 20Gy, respectively, the G2 content of MEF CUX1$^{z/z}$ increased by only 0.3%, 4.6% and 7.8% under the same conditions (Fig. 12B). G2/M arrest was also significantly reduced in siRNA treated MCF7 cells, although to a less striking extent (Fig. 12C).

Taken together, these results indicate that CUX1 is required for cells to mount a complete DNA damage-induced cell cycle checkpoint response, which is fully consistent with the role of this transcription factor in ATM/ATR regulation documented herein.

3.2.5 CUX1 knockdown causes a decrease in Rad51 focus formation and a delay in the repair of DNA strand breaks

Rad51 focus formation, a well characterized marker of homologous recombination[193], was measured using immunofluorescence. Treatment of MCF7 cells with CUX1 siRNA led to a marked decrease in the proportion of cells displaying 5 or more Rad51 foci after IR (Fig. 13A). Similarly, *Cux1*$^{Z/Z}$ MEFs displayed a strong reduction in cells displaying Rad51 foci as compared with wild-type counterparts (Fig. 13B). These results indicate that CUX1 is required for efficient DNA double strand-break repair by homologous recombination.

This was further demonstrated using the single cell gel electrophoresis assay, commonly known as the comet assay. I measured the disappearance of DNA breaks following exposure to IR, UV and H_2O_2, in MCF7 treated with CUX1 siRNA and in *Cux1*$^{Z/Z}$ MEFs. Judging from the importance of comet tails, DNA breaks persisted significantly longer in CUX1 defective cells (Fig. 13C and D).

Also, *Cux1*$^{Z/Z}$ MEFs cultured in either a 3% or 20% O_2 environment with no additional treatment displayed more damage than the wild-type counterparts, suggesting a higher sensitivity to endogenous DNA damaging agents, eg., reactive oxygen species produced during oxidative respiration (Fig. 13D, rightmost panel).

Interestingly, the baseline level of damage in heterozygous knockout MEFs (*Cux1*$^{wt/z}$), which express half the level of both mRNA and protein as wildtype MEFs (Fig. 14A), also show a higher baseline level of damage, although not as high as full CUX1 knockouts (Fig. 14B). Furthermore, *Cux1*$^{wt/z}$ MEFs show a defect in their ability to repair damage following exposure to H_2O_2 although again, it is not as pronounced as that of *Cux1*$^{z/z}$ MEFs (Fig. 14C). This shows that CUX1 must not only be present for cells to respond to DNA damage properly, but must be present in sufficient amounts.

3.2.6 Genomic instability in CUX1$^{z/z}$ MEF cells

Karyotyping analysis was performed on MEFs derived from *Cux1*$^{Z/Z}$ mutant mice and wild-type littermates. Chromosome counting revealed a greater proportion of tetraploid or near tetraploid cells in CUX1-deficient cells compared with wild-type cells (Fig. 15A). In addition, the number of chromosome breaks, as measured following Giemsa staining, was significantly increased in CUX1-deficient cells (Fig. 15B). The above results indicate that CUX1-deficient MEFs display increased genomic instability.

3.2.7 CUX1 over-expression causes radio-resistance and chemo-resistance in tumour cells

The protective role of CUX1 suggests that its over-expression may render cells even more resistant to DNA damage than cells expressing normal levels. To test this, the p110 isoform of CUX1 was over-expressed in MCF7 breast tumour cells, which resulted in higher survival following exposure to IR, UV and Cisplatin (Fig. 2_16), suggesting that CUX1 can confer radio-resistance and chemo-resistance to tumour cells.

3.3 Activation of the Wnt/β-Catenin pathway by CUX1

3.3.1 Autocrine activation of Wnt/β-catenin pathway in MMTV-CUX1 adenosquamous carcinomas

MMTV-CUX1 transgenic mice were shown to develop mammary tumours of diverse histological types [52]. In particular, nuclear β-catenin expression was observed in adenosquamous carcinomas, but not in solid carcinomas (Fig. 17A, and [52]). These observations suggested that CUX1 plays a causal role in, but is not sufficient by itself for, the upregulation of nuclear β-catenin expression. To investigate the role of CUX1 and other transcription factors in the Wnt/β-catenin pathway, we first established cell lines from the p75-80 adenosquamous carcinoma and the p75-534 solid carcinoma, which respectively exhibit or not nuclear β-catenin expression (Fig. 17A). RT-qPCR analysis indicated that the expression of Wnt1, Wnt2, Wnt4, Wnt6, Wnt7a, Wnt7b, Wnt8b and especially Wnt10a and Wnt10b was much higher in the p75-80 than in the p75-534 cell line (Fig. 17D). The TOP/FOP luciferase reporter assay was employed as a functional assay to monitor the transcriptional activity of TCF/β-catenin complexes in these cell lines. The TOP reporter, which contains 8 TCF binding sites, was activated in the p75-80 but not in the p75-534 cell line, confirming that the β-catenin pathway is activated in the former (Fig. 17B). Transfections in the presence of niclosamide, a Wnt inhibitor that promotes Frizzled1 internalization and downregulates Dvl-2 expression, or cotransfection of plasmids expressing sFRP1, sFRP2 or SOST, antagonists of the interaction between Wnt ligands and their receptors, prevented activation of the TOP reporter in p75-80 cells (Fig. 17C). These results indicate that augmentation of TCF/β-catenin transcriptional activity in p75-80 cells necessitates the activation of the FRP-LRP5/6 receptor complex. Altogether, this analysis demonstrates that increased nuclear expression and activity of β-catenin in p75-80 cells results from the autocrine activation of the Wnt pathway.

3.3.2 CUX1 Is Required for Maximal Expression of the Wnt Genes in Human Tumour Cell Lines

Previous studies have identified a number of human cancer cell lines that display autocrine activation of the Wnt/β-catenin pathway. We performed siRNA-mediated knockdown of CUX1 to verify whether CUX1 is needed for expression of Wnt genes in six of these cell lines, two each from breast (MDA157 and MDA231), ovarian (PA-1 and SKOV3) and lung (H2347 and A427) cancers. In this analysis, we focused our attention on Wnt genes that were identified as transcriptional targets of CUX1 in ChIP-chip and conventional ChIP assays: Wnt 1, 2, 4, 6, 8b and 10a ([24, 52] and unpublished observations). CUX1 knockdown caused a significant downregulation of Wnt genes in the 4 breast and ovarian cell lines, but not in the lung cell lines (Fig. 18). We conclude that CUX1 is required for maximal expression of certain Wnt genes in breast and ovarian cell lines.

3.3.3 Ectopic expression of p110 CUX1 leads to autocrine activation of the Wnt/β-catenin pathway in human tumour cell lines.

In turn, we asked whether ectopic CUX1 expression would activate the Wnt/β-catenin pathway in three cell lines that do not display autocrine activation of this pathway. We observed higher nuclear β-catenin expression in Hs578T, MCF-7 and HEK293 cells stably expressing p110 CUX1 (Fig. 19A). In agreement with these results, co-expression of p110 CUX1 with the TOP and FOP reporter plasmids led to a striking elevation in the TOP/FOP ratio, a readout of TCF/β-catenin transcriptional activity (Fig. 19B). Importantly, co-expression of sFRP1, sFRP2, SOST or DKK1 or transfection in the presence of the niclosamide or IWP-2 inhibitors in HEK293 cells in every case eliminated the stimulatory effect of p110 CUX1 (Fig. 19C). We conclude that ectopic expression of p110 CUX1 can stimulate the Wnt/β-Catenin pathway and it does so through an autocrine loop.

3.3.4 Activation of the Wnt/β-Catenin Pathway in CUX1 Mammary Tumours Is Associated with High Expression of Gli3 and Glis1

Since activation of the β-catenin pathway was observed in a fraction of mammary tumours from MMTV-CUX1 transgenic mice, we reasoned that it may require other transcription factors in addition to CUX1. As an approach to identify such transcription factors, we looked for any factor whose expression was elevated in these mammary tumours. To ensure that changes in gene expression would not be masked by the noise coming from stromal cells, mammary epithelial tumour cells were microdissected prior to performing expression profiling. In total, 17 tumours were analyzed and clustered according to Wnt genes expression (Fig. 20). This procedure revealed three large clusters that displayed low, medium and high Wnt gene expression. Only two genes coding for transcription factors were overexpressed in tumours with high Wnt gene expression: *Glis1* and *Gli3*. Unfortunately, the signals for *Gli1* and *Gli2* did not show any variations in these 17 samples and in many other samples analyzed on the same microarrays; therefore, we cannot make any conclusion regarding the expression of these two genes.

Activation of the Wnt/β-catenin pathway has previously been associated with elevated expression of genes coding for Wnt receptors, markers of stem and progenitor cells, markers of the basal-like subtype and markers of the epithelial to mesenchymal transition (EMT). Among mammary tumours with elevated Wnt gene expression, we observed higher expression of 5 genes typically upregulated in EMT, and lower expression of 2 genes downregulated in EMT (Fig. 2, EMT Up: *Vim*, *Cdh2*, *Snai1*, *Snai2* and *Twist1*; EMT Down: *Cdh1* and *Ocln*). Altogether, expression profiling analysis revealed that elevated Wnt gene expression in mammary tumours from MMTV-CUX1 transgenic mice was associated with higher expression of two GLI family members and correlated well with markers of the EMT phenotype.

3.3.5 Higher Expression of Wnt Genes in Human Breast and Lung Tumours Correlates with Higher *GLI* Gene Expression

Meta-analysis of human cancer expression profiling data was performed to verify whether high Wnt gene expression correlated with that of CUX1 and GLI factors. Human tumour datasets obtained from the Oncomine database were sorted according to Wnt genes expression using the BreSat algorithm. This method linearly orders tumours over an individual gene signature, in this case the Wnt gene signature. As such, this tool allows us to investigate a continuous trend across the data, assessing the relative activation of the Wnt signature across a panel of tumour patients.

The use of this method was validated by using it to rank the tumours obtained from the MMTV-CUX1 mice according to their *Wnt* gene expression. Comparison of this method to that of unsupervised clustering (Fig. 20) showed that both methods identify the same 5 tumours as having the highest expresson of Wnt genes (Supplementary Fig. 5). The BreSat algorithm was thus used to analyze the human datasets as we found that it provides a more effective method of grouping samples by their overall expression profiles.

A heatmap representation of a sample ranking generated with this algorithm is presented in Fig. 21. A cursory visual inspection of the data suggests that CUX1 and GLI factors are expressed at higher levels in breast tumours with high Wnt gene expression (Fig. 21A, top and bottom panels). Indeed, the top 25% of samples ranked according to Wnt expression exhibit significantly higher expression of *CUX1* and *GLI* genes than the bottom 25% of samples (Fig. 21B). Interestingly, a similar correlation was observed in human tumour datasets between *WNT*, *CUX1* and *GLI* gene expression (Table 13). While only a few datasets provide information regarding *CUX1* expression, since Affymetrix microarrays do not have a relevant probe (as discussed in [59]), a clear correlation was observed in 3 out of 3 breast tumour and 2 out of 2 lung tumour datasets between high Wnt and CUX1 gene expression. Moreover, a positive correlation was also observed with GLIS1 in 12 out of 13 datasets (Table 13). Furthermore, higher expression of at least one gene of either GLI1, GLI2 or GLI3 correlated

with high Wnt gene expression in 16 out of 18 datasets (Table 13), with no apparent preference for any of the 3.

A Summary table listing the datasets used for theses analyses as well as the genes for which a significant difference in expression was observed for each dataset are presented as supplementary material (Supplementary Table 2).

3.3.6 Ectopic Expression of Glis1 in a MMTV-CUX1 Tumour Cell line Leads to the Transcriptional Activation of Wnt Genes

Expression profiling analysis suggested that establishment of a "Wnt" phenotype in mammary epithelial cells necessitates high expression of both a CUX1 and a Gli protein. To begin to verify this notion, we turned to the "non-Wnt" and "Wnt" cell lines, p75-534 and p75-80. The p75-534 tumour displayed low expression levels of Wnt as well as *Glis1* mRNA. To verify whether higher Glis1 expression was able to convert p75-534 cells to a "Wnt" phenotype, we ectopically expressed Glis1 using a lentiviral vector and then measured β-catenin protein and Wnt mRNA expression. We observed a 2-fold and a more than 10-fold increase in the expression of 5 and 4 Wnt genes, respectively (Fig. 22A). In agreement with these finding, the steady-state levels of β-catenin protein was significantly increased (Fig. 22B), supporting the idea that the combined expression of GLIS1 genes and CUX1 cooperatively activate the Wnt/β-catenin pathway.

3.3.7 Glis1 and p110 CUX1 Cooperate To Activate the Wnt/β-Catenin Pathway

To further support the cooperation between Glis1 and p110 CUX1, expression vectors for either or both of these factors were introduced into MCF10A cells, which are un-transformed breast epithelial cells derived from breast reduction surgery [194]. Measurements of WNT gene expression revealed that each factor on its own was able to stimulate the expression of several WNT genes, but highest expression was achieved for 8 of 14 measurable WNT genes when both Glis1 and p110 CUX1 were co-expressed (Fig. 23A). In agreement

with these results, the highest levels of the active, non-phosphorylated form of nuclear β-catenin was observed in MCF10A cells co-expressing both Glis1 and p110 CUX1, as measured both by immunofluorescence and immunoblotting (Fig. 23 B, C). In summary, results from 3 assays in the MCF10A tissue culture model system support the notion that Glis1 and p110 CUX1 cooperate to stimulate the Wnt/β-catenin pathway.

3.3.8 Glis1 and p110 CUX1 Cooperate To Stimulate Epithelial-to-Mesenchymal Transition and Invasiveness.

Expression profiling analysis of MMTV-CUX1 mammary gland tumors also revealed that those with elevated Wnt gene expression also displayed higher expression of 5 genes typically upregulated in EMT, and lower expression of 2 genes downregulated in EMT (Fig. 20, EMT Up: *Vim*, *Cdh2*, *Snai1*, *Snai2* and *Twist1*; EMT Down: *Cdh1* and *Ocln*). Coincidentally, we also observed that samples with a high Wnt expression in the Esserman human breast cancer dataset showed significant changes in expression of several markers of the EMT phenotype: *VIM*, *SNAI1* and *TWIST1* genes were elevated whereas CDH1 and OCLN genes were downregulated (Fig. 21 A and C). A similar correlation was observed in the two other breast tumor datasets for which CUX1 expression was available (Table 14).

The migratory and invasiveness properties of MCF10A cells expressing Glis1, p110 CUX1 or the combination, were then tested. Live-cell imaging of these cells showed that expression of Glis1 alone did not results in significant increase in movement speed, directionality of movement or invasiveness of the cells, and while p110 CUX1 overexpression caused a modest increase in those properties, it is the combination of the two factors that lead to significant increases in the motility and invasiveness of the cells (Fig. 24).

3.3.9 CUX1 is required for the invasiveness of cells with an activated Wnt/β-Catenin pathway.

Since Glis1 alone was able to stimulate activation of the Wnt/β-Catenin pathway but not promote motility and invasiveness, we then tested whether CUX1 was required for these properties in the context of cells with a constitutively active Wnt/β-Catenin pathway. We used the human colorectal cancer derived HCT116 cells, which harbor a mutation in the β-Catenin gene preventing the degradation of the protein, as a model. While shRNA mediated knockdown of CUX1 in these cells caused a significant decrease in the levels of its established Wnt ligand targets (Fig. 25A), it did not cause any decrease in the levels of β-Catenin protein or its activity as measured in Top/Fop assays (Fig. 25B and C). However, CUX1 knockdown caused a significant decrease in the invasiveness of these cells (Fig. 25D).

These results suggest that the activation of the Wnt/β-Catenin pathway mediated by Glis1 is insufficient to promote the invasive phenotype associated with EMT and that CUX1 may have additional effects, likely through its established transcriptional regulation of several factors involved in cell motility and invasiveness as described previously[25, 195].

4. Discussion

4.1 Transcriptional regulation by CUX1

4.1.1 Regulation by CUX1 at a distance

Genome-wide location analysis on the ENCODE array revealed that ~25% of CUX1 binding sites are located in the 4-Kbp region upstream and downstream of a TSS, while more than 40% of CUX1 binding sites are situated at more than 40 Kbp from a TSS (Fig. 3). Overall, 7.4% and 8.4% of putative targets on the ENCODE and promoter arrays respectively, exhibited a 1.5-fold change in expression following CUX1 knockdown or p100 CUX1 overexpression (Table 7 and 8). This proportion is within the 1-10% range of potential targets that have been reported to be regulated by other transcription factors [78-80].

Importantly, analysis of the percentage of regulated genes versus the distance of CUX1 binding sites to TSS showed that essentially the same proportion of genes are regulated whether CUX1 binds close or far away from the TSS (Fig. 5A and B). In other words, the probability that a gene is regulated by CUX1 is not affected by the distance between the CUX1 binding site and the TSS. In addition, our results indicate that the position of genes relative to a CUX1 binding site do not determine whether these genes are regulated by CUX1. CUX1 regulated similar percentages of genes whether they were closest to the CUX1 binding site or were located further away in the other direction (Fig. 6, compare 1 and 2). Moreover, CUX1 regulated a surprisingly high proportion of (5.4%) of genes that were separated from its binding site by another gene (Fig. 6). Altogether these results demonstrate that CUX1 can regulate genes at a distance and can regulate more than one gene on certain genomic loci.

The proportion of target genes that were found to be activated or repressed by CUX1, respectively 52% and 48%, is significantly different from what we reported in previous studies on target genes involved in cell cycle progression, cell motility, or the DNA damage response [24, 25, 196]. In each case, a vast majority of genes were found to be activated by p110 CUX1, whether we performed siRNA-mediated knockdown or overexpression of p110 CUX1. One

factor that may explain this could be the functional classes of genes that were studied previously. The functional class of "cell cycle" genes includes mostly genes that stimulate cell cycle progression. Out of 25 cell cycle gene targets identified by ChIP-chip, 22 were activated and 2 were repressed by CUX1 (while only one was not affected) [24]. One of the two repressed genes, $p21^{WAF1/CKI1}$, code for a CDK-inhibitor that blocks cell cycle progression, while the other, CCNH, is involved in transcription and DNA repair. All target genes that were activated play a positive role in cell cycle progression. Similarly, among 19 targets that play a role in DNA damage response, 18 were activated and one was repressed [196]. The repressed gene again was $p21^{WAF1/CKI1}$. Overall, these results are consistent with the notion that CUX1 establishes a transcriptional program that promotes cell cycle progression and at the same time ensures the maintenance of genetic integrity.

4.1.2 CUX1: Required vs. Sufficient for expression

We employed two experimental approaches to examine the transcriptional regulation of genes by CUX1. Expression profiling was performed following shRNA-mediated knockdown of CUX1 or p110 CUX1 overexpression. Among targets identified on the promoter array, 287 genes exhibited a 1.5-fold change in expression following CUX1 knockdown, while 85 genes were regulated in response to p110 CUX1 overexpression. Therefore, more genes were found to be regulated by CUX1 using the shRNA approach. This result can be interpreted to mean that CUX1 is required for optimal expression of many target genes, however, increasing CUX1 expression is not sufficient to modulate the expression of some target genes.

4.1.3 Absence of a requirement for the ATCRAT consensus

The CUX1 consensus binding site, ATCRAT (where R = C or A), was found to be present at 47.2% of the 5828 bound genomic sites (Table 5). We conclude that the presence of a CUX1 consensus binding site contributes to, but is not sufficient for, the recruitment of CUX1 to specific genomic locations. We

envision that interactions with other transcription factors play an important role in recruiting CUX1 to specific locations. In agreement with this notion, functional analysis revealed distinct sets of cellular functions among gene targets that contain an ATCRAT consensus and those that do not (Table 6). We note that functional classes involved in cell cycle were over-represented among target genes that do not contain a consensus CUX1 binding site (Table 6B). In previous studies, CUX1 was shown to interact with E2F factors and cooperate with these factors in the regulation of several cell cycle genes [197, 198]. It is likely that protein-protein interaction with E2F factors reduces the requirement for the presence of a high-affinity binding site for the recruitment of CUX1 on this class of genes.

4.1.4 Implications for other projects

The results of this project, in addition to improving our overall understanding of transcriptional regulation by CUX1, also provide specific information on genes that can be interesting on their own. For instance, a number of additional CUX1 targets identified in the course of this project were added to the original list of DDR related genes of interest identified in earlier ChAP and ChIP experiments[24], including the important ATM and CHK2 kinases. Other targets proved valuable for the studies of other members of the lab[199] and, with the raw results having been made available, it will also be possible for other groups to use these in the context of their own research.

4.2 CUX1 in the DNA damage response

4.2.1 The baseline transcriptional effect of CUX1

In the absence of exogenous DNA damage, CUX1 transcriptionally upregulates various DDR genes including ones encoding the critical signaling kinases ATR and ATM. CUX1 similarly stimulates the expression of (i) DNA damage sensors such as NBS1, TopBP1, and RPA, that directly participate in activation of these kinases [104, 119, 200] and (ii) of the adaptor protein 53BP1 that couples ATM to Chk2 in promoting phosphorylation of the latter [112]. It should be noted that relatively modest reductions in steady-state levels of the aforementioned DDR proteins were observed following either partial or complete abrogation of CUX1 in unstressed cells. However, after exposure to DNA damaging agents, the effects of CUX1 depletion on activation of, and signaling through, both ATM and ATR were much more striking. In addition hallmark ATM/ATR-regulated protective functions including cell cycle checkpoints, survival, homologous recombination and repair of DNA breaks were significantly negatively-impacted following mutagenic insult. This presumably reflects the cooperative effects of CUX1-regulated DDR proteins in triggering a robust ATM/ATR-mediated DDR.

The eukaryotic DDR is classically perceived as being triggered and sustained through post-translational modification of effector proteins imposed after the application of genotoxic stress. However our results provide compelling evidence that prior to sustaining DNA damage, adequate basal DDR protein levels depend on CUX1 transcriptional regulation, and moreover must be in place such that cells can respond efficiently to mutagenic insult. This function of CUX1 (and quite possibly of other transcription factors; see below), can be distinguished from other, purely DNA damage-inducible mechanisms of DDR gene activation, e.g., the SOS response in bacteria [201-204] and yeast [205], or the p53 tumour suppressor pathway in mammalian cells [206]. With this in mind, we speculate that the CUX1-mediated transcriptional response revealed here may be important for full protection against not only exogenous DNA damage, but also highly-genotoxic DNA lesions of endogenous origin such as oxidized bases and strand

breaks generated by cellular free radicals. This notion is supported by the comparative analysis of *Cux1*$^{Z/Z}$ and wild-type MEFs showing a higher number of strand breaks, a greater proportion of tetraploid or near tetraploid cells and an increased number of chromosome breaks in CUX1-deficient MEFs (Fig. 13D, 14A and 14B).

While a multitude of previous investigations have sought to identify transcription factors that are regulated downstream of DDR signal transduction pathways, only a few attempted to reveal such transcription factors whose activity is required prior to the application of genotoxic stress: (i) Expression profiling studies clearly showed a role for E2F family members in the regulation of DNA repair genes involved in homologous recombination, nonhomologous end-joining, and base excision repair [173, 207, 208]. Subsequent chromatin immunoprecipitation studies confirmed that E2F transcription factors directly bind the promoters of DDR genes in the absence of DNA damage [173]; moreover steady-state changes in the expression of DDR genes in Rb-deficient cells was documented [209, 210]. (ii) Stat3$^{-/-}$ MEF cells were shown to exhibit a weaker response and decreased survival following irradiation and evidence from reporter assays suggested that *MDC1*, the gene encoding mediator of DNA damage checkpoint 1, is regulated by STAT3 [211]. (iii) NFκB was shown to bind to the promoter of *ATM* and to be required for optimal expression of this kinase in T cells [212]. (iv) The FoxM1 transcription factor regulates baseline levels of p21 and Chk1, and its overexpression is associated with enhanced checkpoint activity [213]. Although limited in number, the above studies suggest that preparing cells to cope with DNA damage constitutes a critical function of various transcription factors also known to play roles in cell cycle. Consistent with such a notion, we have shown that CUX1, in addition to regulating cell cycle progression and DNA replication in unstressed cells, also contributes to the maintenance of genomic integrity by ensuring that key players in the DDR are present in stoichiometric amounts prior to DNA damage.

4.2.2 Transcriptional regulation by CUX1 following DNA damage

Results from my research and other studies reveal a regulatory loop involving CUX1 and key checkpoint kinases. While we have identified CUX1 as an activator of constitutive DDR gene expression, it should be emphasized that CUX1 itself was shown to be the target of post-translational modification following DNA damage. Specifically, CUX1-derived peptides were identified in two phosphoproteomic studies aimed at identifying proteins phosphorylated after UV or IR [106, 214]. In total, 5 phosphorylation sites were identified at positions 322, 734, 1233, 1312 and 1357. Since none of these sites map within one of the 4 DNA binding domains of CUX1 (CR1, CR2, CR3 and the Cut homeodomain), their phosphorylation would not be expected to inhibit DNA binding. Indeed, CUX1 DNA binding activity is not diminished after DNA damage (Supplementary Fig. 6A, B and D), although the manner in which CUX1 regulates transcription may certainly be altered. For example, in agreement with earlier studies from several groups [12, 22, 26, 28, 47, 48, 190], we observed that CUX1 represses CDKN1A which encodes the p21 cyclin-dependent kinase inhibitor (Fig. 7 and Table 11) in unstressed cells. On the other hand, results using CUX1-deficient cells indicated that it is required for optimal upregulation of CDKN1A following DNA damage (Supplementary Fig. 6E).

We speculate that phosphorylation of CUX1 by checkpoint kinases during periods of genotoxic stress affects its interactions with co-repressors and co-activators, or imparts conformational changes that alter its regulatory properties. The role of CUX1 in the transcriptional program taking place after DNA damage, and the identification of kinases that phosphorylate this transcription factor during genotoxic stress could represent an interesting study for future members of the lab.

4.2.3 A potential direct role of CUX in the DDR and DNA repair

While my work on the role of CUX1 in the DDR was focused on its transcriptional activity and the implications thereof, it is not excluded that the protein may play a more direct role in the mechanistic response to DNA damage.

In particular, the abundance of the p200 isoform in the nucleus and its transient DNA binding activity, as opposed to p110's stable binding, in addition to the fact that it is post-translationally modified after DNA damage, are compatible, for instance, with a potential role for it in the detection of DNA damage and the ensuing signalling cascade, or with an active role in direct repair of DNA lesions. While my own attempts at investigating a non-transcriptional role for CUX1 in the DDR were not successful, other members of the lab have recently obtained data suggesting a direct role for CUX1 in DNA repair, for instance, over-expression of a fragment of CUX1 containing only the CR1 and CR2 domains in cells lead to an faster rate of repair of DNA damage as measured by Comet Assay(Supplementary Fig. 7). This CR1CR2 fragment binds to DNA with a rapid "on/off" rate and may be capable of passive repression through the CCAAT displacement activity but could not transcriptionally activate DDR genes as p110 does. Therefore, this result strongly suggests that p200 CUX1 may play a direct role in DNA repair.

Interestingly, other groups have suggested non-transcriptional roles in DNA Repair for different transcription factors, including p53 and E2F1[215]. For instance, E2F1 is phosphorylated by ATM, localizes to sites of damage and its expression promotes nucleotide excision repair independently of its DNA binding domain and transactivation domain[216, 217].

A plausible model for the overall role of CUX1 in the DDR is that the p110 isoform transcriptionally activates key members of the response pathways, allowing them to accumulate at sufficient levels, and that the p200 isoform then interacts directly with these proteins and cooperates with them directly in the DNA damage response, either at the level of signalling or repair.

4.2.4 A potential role for CUX1 in radio-resistance and chemo-resistance

Since the majority of clinical treatments for cancer rely on DNA damage, on the basis that tumour cells are more sensitive to such damage than normal cells, the protective role of CUX1 has implications for the role it plays in tumour development and clinical management. I have shown that overexpression of the

p110 CUX1 isoform renders tumour cells more resistant to DNA damage (Fig. 16). The increased resistance to IR and Cisplatin is particularly important since those two agents are used in the treatment of tumours in human patients, suggesting that tumours expressing high levels of CUX1 may be more resistant to treatments. Consistent with this notion, increased DDR activity in tumour cells has been associated with radio-resistance in gliomas[218]. Future studies into the potential role of CUX1 in treatment resistance of tumours may be warranted.

4.3 Activation of the Wnt/β-Catenin pathway by CUX1

4.3.1 CUX1 functions through an autocrine mechanism

Strikingly, while there is a single Wingless gene in Drosophila, there are 19 Wnt genes in mammals. We understand that duplication of Wnt genes during evolution led to the association of Wnt coding sequences with a large repertoire of regulatory sequences, a process that ultimately enabled the expression of individual Wnt genes in specific cells and at specific times in order to activate Wnt pathways. Superimposed on the complex transcriptional regulation of Wnt ligands are the combinatorial nature of Wnt receptors and the multiple control steps that can prevent or limit Wnt pathway activation. Delivery of receptors at the cell membrane or secretion of ligands can be inhibited. Moreover, secretion of non-membrane bound receptors can squelch the action of ligands that are present in the extracellular milieu. Altogether, the multiple ways by which Wnt pathways can be activated and inhibited enables precise regulation both in time and space. Yet, perturbations in one or the other control system have been found to cause aberrant activation of the Wnt/ β-catenin pathway in human cancers. The majority of cases were reported in colon cancers and involve inactivating mutations of the APC tumour suppressor or mutations that change the properties of axin or β-catenin itself [154]. In addition, autocrine activation of the Wnt/β-catenin pathway was shown to occur in a sizeable fraction of cancers from multiple organs and tissues. While elevated Wnt gene expression and repression of secreted Frizzled have been reported, we know very little about the transcription factors that play a role in the regulation of these genes.

In this study, I presented evidence from transgenic mouse models, tissue culture systems and human cancer databases implicating CUX1 and GLI family transcription factors in the autocrine activation of the Wnt/β-catenin pathway. A fraction of mammary tumours from MMTV-p110 and MMTV-p75 CUX1 transgenic mice exhibited elevated nuclear β-catenin expression and histo-pathological features similar to that of mammary tumours from MMTV-Wnt1 transgenic mice (Fig. 17A). Using the TOP/FOP reporter system in established tumour cell lines, we showed that TCF/β-catenin transcriptional activity was

elevated in the cell line with higher nuclear β-catenin expression (Fig. 17B). Moreover, ectopic expression of p110 CUX1 led to an increase in nuclear β-catenin levels and TOP/FOP expression ratios in some transformed human cell lines, but noticeably not in the non-transformed MCF10A cell line, (Fig. 19A and B, 23C). Importantly, both in mouse and human cell lines TOP/FOP ratios were drastically reduced following overexpression of secreted proteins or treatment with chemical inhibitors that prevent activation of FZD/LRP receptors by Wnt ligands (Fig. 17C and 19C). These results indicated that CUX1 stimulates TCF/β-catenin activity through an autocrine loop. Indeed, CUX1 was previously shown to bind to the promoter of several Wnt genes and activate their expression [52].

4.3.2 A requirement for cooperation to activate the Wnt/β-Catenin pathway

The fact that only a fraction of mammary tumours from MMTV-CUX1 transgenic mice exhibited elevated nuclear β-catenin expression indicated that CUX1 alone is not sufficient to activate this pathway and provided an opportunity to identify other transcription factors that cooperate with CUX1 in this process. Expression profiling of micro-dissected mammary epithelial tumour cells revealed that Glis1 and Gli3 were overexpressed in mammary tumours with high Wnt expression (Fig. 20). Importantly, meta-analysis of human breast and lung tumour datasets also showed that elevated Wnt gene expression correlated with high levels of both CUX1, GLIS1 and some combination of GLI 1, 2 or 3 (Fig. 21).

Interestingly, Glis1 has been shown to cooperate with Oct3/4, Sox2 and Klf4 (OSK) in the activation of several Wnt gene and in the reprogramming of somatic cells [219]. However, we did not observe upregulation of markers of stem or progenitor cells in mammary tumours with elevated Wnt and Glis1 expression (Supplementary Fig. 7). We conclude that these proteins are not sufficient, and that OSK factors are needed, for the activation of stem cell markers.

4.3.3 A link with epithelial to mesenchymal transition

One functional class of genes whose regulation correlated with that of Wnt, CUX1 and Glis1 are those associated with epithelial-to-mesenchymal

transition (EMT). Both in mouse mammary tumours and in human breast tumour datasets, we observed upregulation of Vim, Cdh2 (N-cadherin), Snai1 and Twist1 and downregulation of Cdh1 (E-cadherin) and Ocln (Fig. 20 and 21), consistent with an epithelial to mesenchymal transition in those tumours.

Further evidence of this was that combined expression of CUX1 and Glis1 in MCF10A cells lead to increases in cell motility and invasiveness, characteristics that are consistent with epithelial to mesenchymal transition (Fig. 24). Importantly, activation of the Wnt/β-Catenin pathway alone in the absence of CUX1 was not sufficient to promote these properties (Fig. 25).

These findings are in agreement with previous studies showing that higher CUX1 expression was associated with higher grade breast tumours and decreased survival [42, 220] and that ectopic CUX1 expression was able to stimulate cell migration and invasion [25, 195].

In summary, these results have identified two transcription factors that play a causal role in the autocrine activation of the Wnt/β-catenin pathway. Yet, there is still much to be discovered regarding the regulation of this pathway. For example, it is not clear why different Wnt genes are regulated in response to changes in CUX1 expression in different cell lines (Fig. 18, 22, 23). It is striking, however, that some elements of transcriptional regulation have been conserved throughout evolution as CUX1 and Glis1 are respectively the orthologs Cut and Ci which regulate wingless in Drosophila. In addition, the results of this study may have long term clinical relevance since it identifies a mechanism of autocrine activation of the Wnt/β-catenin pathway, which is independent of mutations. Given the increasing importance of the pathway as a therapeutic target in human cancers[164, 165], identifying such mechanisms may inform future treatment strategies, which could be particularly interesting if it is associated with a more invasive and EMT-like phenotype.

Summary

The work presented in this thesis provides a global view of transcriptional regulation by CUX1, demonstrates the requirement for that regulation in the DNA damage response and implicates CUX1 in the autocrine activation of the Wnt/β-Catenin pathway in cancer.

1) Transcriptional regulation by CUX1:

- CUX1 can bind both close or at a distance from gene promoters and act both as an activator and a repressor of transcription.
- CUX1 can regulate the expression of genes when binding at a distance.
- The ATCRAT sequence plays a role in recruiting CUX1 to some binding sites but others are independent of it.

2) CUX1 is required for the DNA damage response:

- CUX1 activates and is required for the maintenance of sufficient protein levels of numerous genes involved in the DNA damage response.
- In the absence of CUX1, cells are deficient in cell cycle arrest and repair mechanisms.
- These deficiencies are associated with increased genomic instability.

3) CUX1 activates the Wnt/β-Catenin pathway through an autocrine mechanism in cooperation with the GLI family of transcription factors:

- Some tumours from CUX1 transgenic mice display autocrine activation of the Wnt/β-Catenin pathway.
- This activation is dependent on cooperation with GLI factors.
- This activation is associated with markers and properties of EMT.
- These observations are corroborated by data from human tumour profiling

References

1. Nepveu A: **Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development**. *Gene* 2001, **270**(1-2):1-15.
2. Braun W: **The effect of puncture on the developing wing of several mutants of *Drosophila melanogaster***. *J Exp Zool* 1942, **84**:325-350.
3. Jack J, Dorsett D, Delotto Y, Liu S: **Expression of the cut locus in the Drosophila wing margin is required for cell type specification and is regulated by a distant enhancer**. *Development* 1991, **113**(3):735-747.
4. Liu S, McLeod E, Jack J: **Four distinct regulatory regions of the cut locus and their effect on cell type specification in Drosophila**. *Genetics* 1991, **127**(1):151-159.
5. Neufeld EJ, Skalnik DG, Lievens PM, Orkin SH: **Human CCAAT displacement protein is homologous to the Drosophila homeoprotein, cut**. *Nature Genetics* 1992, **1**(1):50-55.
6. Blochlinger K, Bodmer R, Jack J, Jan LY, Jan YN: **Primary structure and expression of a product from cut, a locus involved in specifying sensory organ identity in Drosophila**. *Nature* 1988, **333**(6174):629-635.
7. Andres V, Nadal-Ginard B, Mahdavi V: **Clox, a mammalian homeobox gene related to Drosophila cut, encodes DNA-binding regulatory proteins differentially expressed during development**. *Development* 1992, **116**(2):321-334.
8. Ellis T, Gambardella L, Horcher M, Tschanz S, Capol J, Bertram P, Jochum W, Barrandon Y, Busslinger M: **The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle**. *Genes Dev* 2001, **15**(17):2307-2319.
9. Heuvel GBV, Bodmer R, McConnell KR, Nagami GT, Igarashi P: **Expression Of a Cut-Related Homeobox Gene In Developing and Polycystic Mouse Kidney**. *Kidney International* 1996, **50**(2):453-461.
10. Quaggin SE, Heuvel GB, Golden K, Bodmer R, Igarashi P: **Primary structure, neural-specific expression, and chromosomal localization of Cux-2, a second murine homeobox gene related to Drosophila cut**. *J Biol Chem* 1996, **271**(37):22624-22634.
11. Gingras H, Cases O, Krasilnikova M, Berube G, Nepveu A: **Biochemical characterization of the mammalian Cux2 protein**. *Gene* 2005, **344**:273-285.
12. Goulet B, Watson P, Poirier M, Leduy L, Berube G, Meterissian S, Jolicoeur P, Nepveu A: **Characterization of a tissue-specific CDP/Cux isoform, p75, activated in breast tumor cells**. *Cancer Res* 2002, **62**(22):6625-6633.
13. Goulet B, Markovic Y, Leduy L, Nepveu A: **Proteolytic processing of cut homeobox 1 by neutrophil elastase in the MV4;11 myeloid leukemia cell line**. *Molecular cancer research : MCR* 2008, **6**(4):644-653.

14. Goulet B, Truscott M, Nepveu A: **A novel proteolytically processed CDP/Cux isoform of 90 kDa is generated by cathepsin L**. *Biol Chem* 2006, **387**(9):1285-1293.
15. Maitra U, Seo J, Lozano MM, Dudley JP: **Differentiation-induced cleavage of Cutl1/CDP generates a novel dominant-negative isoform that regulates mammary gene expression**. *Mol Cell Biol* 2006, **26**(20):7466-7478.
16. Truscott M, Denault JB, Goulet B, Leduy L, Salvesen GS, Nepveu A: **Carboxyl-terminal proteolytic processing of CUX1 by a caspase enables transcriptional activation in proliferating cells**. *J Biol Chem* 2007, **282**(41):30216-30226.
17. Moon NS, Berube G, Nepveu A: **CCAAT displacement activity involves Cut repeats 1 and 2, not the Cut homeodomain**. *J Biol Chem* 2000, **275**(40):31325-31334.
18. Luo W, Skalnik DG: **CCAAT displacement protein competes with multiple transcriptional activators for binding to four sites in the proximal gp91phox promoter**. *J Biol Chem* 1996, **271**(30):18203-18210.
19. Kim EC, Lau JS, Rawlings S, Lee AS: **Positive and negative regulation of the human thymidine kinase promoter mediated by CCAAT binding transcription factors NF-Y/CBF, dbpA, and CDP/cut**. *Cell Growth Differ* 1997, **8**(12):1329-1338.
20. Mailly F, Berube G, Harada R, Mao PL, Phillips S, Nepveu A: **The human cut homeodomain protein can repress gene expression by two distinct mechanisms: active repression and competition for binding site occupancy**. *Mol Cell Biol* 1996, **16**(10):5346-5357.
21. Li S, Moy L, Pittman N, Shue G, Aufiero B, Neufeld EJ, LeLeiko NS, Walsh MJ: **Transcriptional Repression of the Cystic Fibrosis Transmembrane Conductance Regulator Gene, Mediated by CCAAT Displacement Protein/cut Homolog, Is Associated with Histone Deacetylation**. *J Biol Chem* 1999, **274**(12):7803-7815.
22. Nishio H, Walsh MJ: **CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription**. *Proc Natl Acad Sci U S A* 2004, **101**(31):11257-11262.
23. Moon NS, Premdas P, Truscott M, Leduy L, Berube G, Nepveu A: **S Phase-Specific Proteolytic Cleavage Is Required to Activate Stable DNA Binding by the CDP/Cut Homeodomain Protein**. *Mol Cell Biol* 2001, **21**:6332-6345.
24. Harada R, Vadnais C, Sansregret L, Leduy L, Berube G, Robert F, Nepveu A: **Genome-wide location analysis and expression studies reveal a role for p110 CUX1 in the activation of DNA replication genes**. *Nucleic Acids Res* 2008, **36**(1):189-202.
25. Kedinger V, Sansregret L, Harada R, Vadnais C, Cadieux C, Fathers K, Park M, Nepveu A: **p110 CUX1 homeodomain protein stimulates cell migration and invasion in part through a regulatory cascade culminating in the repression of E-cadherin and occludin**. *J Biol Chem* 2009, **284**(40):27701-27711.

26. Coqueret O, Berube G, Nepveu A: **The mammalian Cut homeodomain protein functions as a cell-cycle-dependent transcriptional repressor which downmodulates p21WAF1/CIP1/SDI1 in S phase**. *EMBO J* 1998, **17**(16):4680-4694.
27. Goulet B, Baruch A, Moon NS, Poirier M, Sansregret LL, Erickson A, Bogyo M, Nepveu A: **A Cathepsin L Isoform that Is Devoid of a Signal Peptide Localizes to the Nucleus in S Phase and Processes the CDP/Cux Transcription Factor**. *Mol Cell* 2004, **14**(2):207-219.
28. Santaguida M, Ding Q, Berube G, Truscott M, Whyte P, Nepveu A: **Phosphorylation of the CCAAT displacement protein (CDP)/Cux transcription factor by cyclin A-Cdk1 modulates its DNA binding activity in G(2)**. *J Biol Chem* 2001, **276**(49):45780-45790.
29. Sansregret L, Gallo D, Santaguida M, Leduy L, Harada R, Nepveu A: **Hyperphosphorylation by cyclin B/CDK1 in mitosis resets CUX1 DNA binding clock at each cell cycle**. *The Journal of biological chemistry* 2010, **285**(43):32834-32843.
30. Sansregret L, Nepveu A: **The multiple roles of CUX1: Insights from mouse models and cell-based assays**. *Gene* 2008, **412**(1-2):84-94.
31. Scherer SW, Neufeld EJ, Lievens PM, Orkin SH, Kim J, Tsui LC: **Regional localization of the CCAAT displacement protein gene (CUTL1) to 7q22 by analysis of somatic cell hybrids**. *Genomics* 1993, **15**(3):695-696.
32. Lemieux N, Zhang XX, Dufort D, Nepveu A: **Assignment of the human homologue of the Drosophila Cut homeobox gene (CUTL1) to band 7q22 by fluorescence in situ hybridization**. *Genomics* 1994, **24**(1):191-193.
33. Ozisik YY, Meloni AM, Surti U, Sandberg AA: **Deletion 7q22 in uterine leiomyoma. A cytogenetic review**. *Cancer Genetics & Cytogenetics* 1993, **71**(1):1-6.
34. Fenaux P, Preudhomme C, Lai JL, Morel P, Beuscart R, Bauters F: **Cytogenetics and their prognostic value in de novo acute myeloid leukaemia: a report on 283 cases**. *Br J Haemat* 1989, **73**:61-67.
35. Swansbury GJ, Lawler SD, Alimena G, Arthur D, Berger R, Van Den Berghe H, Bloomfield CD, de la Chappelle A, Dewald G, Garson OM *et al*: **Long-term survival in acute myelogenous leukemia: A second follow-up of the Fourth International Workshop on Chromosomes in Leukemia**. *Cancer Genet Cytogenet* 1994, **73**:1-7.
36. Yunis JJ, Lobell M, Arnesen MA, Oken MM, Mayer MG, Rydell RE, Brunning RD: **Refined chromosome study helps define prognostic subgroups in most patients with primary myelodysplastic syndrome and acute myelogenous leukemia**. *Br J Haemat* 1988, **68**:189-194.
37. Heim S: **Cytogenetic findings in primary and secondary MDS**. *Leukemia Research* 1992, **16**:43-46.
38. Zeng WR, Scherer SW, Koutsilieris M, Huizenga JJ, Filteau F, Tsui LC, Nepveu A: **Loss Of Heterozygosity and Reduced Expression Of the Cutl1 Gene In Uterine Leiomyomas**. *Oncogene* 1997, **14**(19):2355-2365.

39. Zeng WR, Watson P, Lin J, Jothy S, Lidereau R, Park M, Nepveu A: **Refined mapping of the region of loss of heterozygosity on the long arm of chromosome 7 in human breast cancer defines the location of a second tumor suppressor gene at 7q22 in the region of the CUTL1 gene**. *Oncogene* 1999, **18**(11):2015-2021.
40. Neville PJ, Thomas N, Campbell IG: **Loss of heterozygosity at 7q22 and mutation analysis of the CDP gene in human epithelial ovarian tumors**. *Int J Cancer* 2001, **91**(3):345-349.
41. Moon NS, Rong Zeng W, Premdas P, Santaguida M, Berube G, Nepveu A: **Expression of N-terminally truncated isoforms of CDP/CUX is increased in human uterine leiomyomas**. *Int J Cancer* 2002, **100**(4):429-432.
42. Michl P, Ramjaun AR, Pardo OE, Warne PH, Wagner M, Poulsom R, D'Arrigo C, Ryder K, Menke A, Gress T *et al*: **CUTL1 is a target of TGF(beta) signaling that enhances cancer cell motility and invasiveness**. *Cancer Cell* 2005, **7**(6):521-532.
43. Ripka S, Konig A, Buchholz M, Wagner M, Sipos B, Kloppel G, Downward J, Gress T, Michl P: **WNT5A--target of CUTL1 and potent modulator of tumor cell migration and invasion in pancreatic cancer**. *Carcinogenesis* 2007, **28**(6):1178-1187.
44. De Vos J, Thykjaer T, Tarte K, Ensslen M, Raynaud P, Requirand G, Pellet F, Pantesco V, Reme T, Jourdan M *et al*: **Comparison of gene expression profiling between malignant and normal plasma cells with oligonucleotide arrays**. *Oncogene* 2002, **21**(44):6848-6857.
45. Brantley JG, Sharma M, Alcalay NI, Heuvel GBV: **Cux-1 transgenic mice develop glomerulosclerosis and interstitial fibrosis**. *Kidney International* 2003, **63**(4):1240-1248.
46. Ledford AW, Brantley JG, Kemeny G, Foreman TL, Quaggin SE, Igarashi P, Oberhaus SM, Rodova M, Calvet JP, Vanden Heuvel GB: **Deregulated Expression of the Homeobox Gene Cux-1 in Transgenic Mice Results in Downregulation of p27(kip1) Expression during Nephrogenesis, Glomerular Abnormalities, and Multiorgan Hyperplasia**. *Dev Biol* 2002, **245**(1):157-171.
47. Vanden Heuvel GB, Brantley JG, Alcalay NI, Sharma M, Kemeny G, Warolin J, Ledford AW, Pinson DM: **Hepatomegaly in transgenic mice expressing the homeobox gene Cux-1**. *Mol Carcinog* 2005, **43**(1):18-30.
48. Sharma M, Brantley JG, Alcalay NI, Zhou J, Heystek E, Maser RL, Vanden Heuvel GB: **Differential expression of Cux-1 and p21 in polycystic kidneys from Pkd1 null and cpk mice**. *Kidney Int* 2005, **67**(2):432-442.
49. Goulet B, Sansregret L, Leduy L, Bogyo M, Weber E, Chauhan SS, Nepveu A: **Increased expression and activity of nuclear cathepsin L in cancer cells suggests a novel mechanism of cell transformation**. *Molecular Cancer Research* 2007, **5**(9):899-907.
50. Cadieux C, Harada R, Paquet M, Cote O, Trudel M, Nepveu A, Bouchard M: **Polycystic kidneys caused by sustained expression of Cux1 isoform p75**. *J Biol Chem* 2008, **283**(20):13817-13824.

51. Cadieux C, Fournier S, Peterson AC, Bedard C, Bedell BJ, Nepveu A: **Transgenic mice expressing the p75 CCAAT-displacement protein/Cut homeobox isoform develop a myeloproliferative disease-like myeloid leukemia**. *Cancer Res* 2006, **66**(19):9492-9501.
52. Cadieux C, Kedinger V, Yao L, Vadnais C, Drossos M, Paquet M, Nepveu A: **Mouse mammary tumor virus p75 and p110 CUX1 transgenic mice develop mammary tumors of various histologic types**. *Cancer Res* 2009, **69**(18):7188-7197.
53. Sansregret L, Goulet B, Harada R, Wilson B, Leduy L, Bertoglio J, Nepveu A: **The p110 isoform of the CDP/Cux transcription factor accelerates entry into S phase**. *Mol Cell Biol* 2006, **26**(6):2441-2455.
54. Aleksic T, Bechtel M, Krndija D, von Wichert G, Knobel B, Giehl K, Gress TM, Michl P: **CUTL1 promotes tumor cell migration by decreasing proteasome-mediated Src degradation**. *Oncogene* 2007, **26**(40):5939-5949.
55. Ripka S, Riedel J, Neesse A, Griesmann H, Buchholz M, Ellenrieder V, Moeller F, Barth P, Gress TM, Michl P: **Glutamate receptor GRIA3--target of CUX1 and mediator of tumor progression in pancreatic cancer**. *Neoplasia (New York, NY* 2010, **12**(8):659-667.
56. Ripka S, Neesse A, Riedel J, Bug E, Aigner A, Poulsom R, Fulda S, Neoptolemos J, Greenhalf W, Barth P *et al*: **CUX1: target of Akt signalling and mediator of resistance to apoptosis in pancreatic cancer**. *Gut* 2010, **59**(8):1101-1110.
57. Sekharam M, Zhao H, Sun M, Fang Q, Zhang Q, Yuan Z, Dan HC, Boulware D, Cheng JQ, Coppola D: **Insulin-like growth factor 1 receptor enhances invasion and induces resistance to apoptosis of colon cancer cells through the Akt/Bcl-x(L) pathway**. *Cancer Res* 2003, **63**(22):7708-7716.
58. Hilmi C, Larribere L, Giuliano S, Bille K, Ortonne JP, Ballotti R, Bertolotto C: **IGF1 promotes resistance to apoptosis in melanoma cells through an increased expression of BCL2, BCL-X(L), and survivin**. *J Invest Dermatol* 2008, **128**(6):1499-1505.
59. Sansregret L, Vadnais C, Livingstone J, Kwiatkowski N, Awan A, Cadieux C, Leduy L, Hallett MT, Nepveu A: **Cut homeobox 1 causes chromosomal instability by promoting bipolar division after cytokinesis failure**. *Proc Natl Acad Sci USA* 2011, **108**(5):1949-1954.
60. Jenuwein T, Allis CD: **Translating the histone code**. *Science* 2001, **293**(5532):1074-1080.
61. Turner BM: **Cellular memory and the histone code**. *Cell* 2002, **111**(3):285-291.
62. MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ: **Genome-wide transcription factor binding: beyond direct target regulation**. *Trends in genetics : TIG* 2011, **27**(4):141-148.
63. Mantripragada KK, Buckley PG, Diaz de Stahl T, Dumanski JP: **Genomic microarrays in the spotlight**. *Trends in genetics : TIG* 2004, **20**(2):87-94.
64. Bosher JM, Labouesse M: **RNA interference: genetic wand and genetic watchdog**. *Nat Cell Biol* 2000, **2**(2):E31-36.

65. Hofmann A, Nolan GP, Blau HM: **Rapid retroviral delivery of tetracycline-inducible genes in a single autoregulatory cassette [see comments]**. *Proc Natl Acad Sci USA* 1996, **93**(11):5185-5190.
66. Paulus W, Baur I, Boyce FM, Breakefield XO, Reeves SA: **Self-contained, tetracycline-regulated retroviral vector system for gene delivery to mammalian cells**. *J Virol* 1996, **70**(1):62-67.
67. Meyerson M, Gabriel S, Getz G: **Advances in understanding cancer genomes through second-generation sequencing**. *Nature reviews Genetics* 2010, **11**(10):685-696.
68. Gilmour DS, Lis JT: **In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster**. *Mol Cell Biol* 1985, **5**(8):2009-2018.
69. Marzio G, Tyagi M, Gutierrez MI, Giacca M: **HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter**. *Proc Natl Acad Sci USA* 1998, **95**(23):13519-13524.
70. Farnham PJ: **Insights from genomic profiling of transcription factors**. *Nature reviews Genetics* 2009, **10**(9):605-616.
71. Amati B, Land H: **Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death**. *Curr Op Genet Dev* 1994, **4**(1):102-108.
72. Bieda M, Xu X, Singer MA, Green R, Farnham PJ: **Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome**. *Genome Res* 2006, **16**(5):595-605.
73. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ *et al*: **Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs**. *Cell* 2004, **116**(4):499-509.
74. Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA: **Transcriptional Amplification in Tumor Cells with Elevated c-Myc**. *Cell* 2012, **151**(1):56-67.
75. Consortium EP: **The ENCODE (ENCyclopedia Of DNA Elements) Project**. *Science* 2004, **306**(5696):636-640.
76. Peters BA, St Croix B, Sjoblom T, Cummins JM, Silliman N, Ptak J, Saha S, Kinzler KW, Hatzis C, Velculescu VE: **Large-scale identification of novel transcripts in the human genome**. *Genome Res* 2007, **17**(3):287-292.
77. IHGSC: **Finishing the euchromatic sequence of the human genome**. *Nature* 2004, **431**(7011):931-945.
78. Scacheri PC, Davis S, Odom DT, Crawford GE, Perkins S, Halawi MJ, Agarwal SK, Marx SJ, Spiegel AM, Meltzer PS *et al*: **Genome-wide analysis of menin binding provides insights into MEN1 tumorigenesis**. *PLoS genetics* 2006, **2**(4):e51.
79. Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K: **Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells**. *Mol Cell* 2006, **24**(4):593-602.

80. Krig SR, Jin VX, Bieda MC, O'Geen H, Yaswen P, Green R, Farnham PJ: **Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays**. *J Biol Chem* 2007, **282**(13):9703-9712.
81. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nat Protoc* 2009, **4**(1):44-57.
82. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic Acids Res* 2009, **37**(1):1-13.
83. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery**. *Genome biology* 2003, **4**(5):P3.
84. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.
85. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for Gene Set Enrichment Analysis**. *Bioinformatics* 2007, **23**(23):3251-3253.
86. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E *et al*: **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes**. *Nat Genet* 2003, **34**(3):267-273.
87. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC *et al*: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data**. *Nat Genet* 2001, **29**(4):365-371.
88. Rayner TF, Rocca-Serra P, Spellman PT, Causton HC, Farne A, Holloway E, Irizarry RA, Liu J, Maier DS, Miller M *et al*: **A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB**. *BMC Bioinformatics* 2006, **7**:489.
89. Oliver S: **On the MIAME Standards and Central Repositories of Microarray Data**. *Comp Funct Genomics* 2003, **4**(1):1.
90. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository**. *Nucleic Acids Res* 2002, **30**(1):207-210.
91. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M *et al*: **NCBI GEO: archive for functional genomics data sets--update**. *Nucleic Acids Res* 2012.
92. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA *et al*: **NCBI GEO: archive for high-throughput functional genomic data**. *Nucleic Acids Res* 2009, **37**(Database issue):D885-890.

93. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM *et al*: **NCBI GEO: archive for functional genomics data sets--10 years on**. *Nucleic Acids Res* 2011, **39**(Database issue):D1005-1010.
94. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M *et al*: **ArrayExpress update--trends in database growth and links to data analysis tools**. *Nucleic Acids Res* 2012.
95. Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Holloway E *et al*: **ArrayExpress update--an archive of microarray and high-throughput sequencing-based functional genomics experiments**. *Nucleic Acids Res* 2011, **39**(Database issue):D1002-1004.
96. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M *et al*: **ArrayExpress--a public database of microarray experiments and gene expression profiles**. *Nucleic Acids Res* 2007, **35**(Database issue):D747-750.
97. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **ONCOMINE: a cancer microarray database and integrated data-mining platform**. *Neoplasia* 2004, **6**(1):1-6.
98. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, Kulkarni P *et al*: **Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles**. *Neoplasia* 2007, **9**(2):166-180.
99. TCGA: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways**. *Nature* 2008, **455**(7216):1061-1068.
100. TCGA: **Integrated genomic analyses of ovarian carcinoma**. *Nature* 2011, **474**(7353):609-615.
101. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP *et al*: **Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1**. *Cancer Cell* 2010, **17**(1):98-110.
102. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS *et al*: **International network of cancer genome projects**. *Nature* 2010, **464**(7291):993-998.
103. Jennings J, Hudson TJ: **Reflections on the Founding of the International Cancer Genome Consortium**. *Clin Chem* 2012.
104. Lee JH, Paull TT: **ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex**. *Science* 2005, **308**(5721):551-554.
105. Bakkenist CJ, Kastan MB: **DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation**. *Nature* 2003, **421**(6922):499-506.

106. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y *et al*: **ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage**. *Science* 2007, **316**(5828):1160-1166.
107. Hirao A, Kong YY, Matsuoka S, Wakeham A, Ruland J, Yoshida H, Liu D, Elledge SJ, Mak TW: **DNA damage-induced activation of p53 by the checkpoint kinase Chk2**. *Science* 2000, **287**(5459):1824-1827.
108. Matsuoka S, Rotman G, Ogawa A, Shiloh Y, Tamai K, Elledge SJ: **Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro**. *Proc Natl Acad Sci U S A* 2000, **97**(19):10389-10394.
109. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM: **DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139**. *J Biol Chem* 1998, **273**(10):5858-5868.
110. Goldberg M, Stucki M, Falck J, D'Amours D, Rahman D, Pappin D, Bartek J, Jackson SP: **MDC1 is required for the intra-S-phase DNA damage checkpoint**. *Nature* 2003, **421**(6926):952-956.
111. Stucki M, Jackson SP: **gammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes**. *DNA Repair (Amst)* 2006, **5**(5):534-543.
112. Wang B, Matsuoka S, Carpenter PB, Elledge SJ: **53BP1, a mediator of the DNA damage checkpoint**. *Science* 2002, **298**(5597):1435-1438.
113. Cortez D, Wang Y, Qin J, Elledge SJ: **Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks**. *Science* 1999, **286**(5442):1162-1166.
114. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S: **Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints**. *Annu Rev Biochem* 2004, **73**:39-85.
115. Jasin M: **Homologous repair of DNA damage and tumorigenesis: the BRCA connection**. *Oncogene* 2002, **21**(58):8981-8993.
116. Bau DT, Mau YC, Shen CY: **The role of BRCA1 in non-homologous end-joining**. *Cancer Lett* 2006, **240**(1):1-8.
117. Banin S, Moyal L, Shieh S, Taya Y, Anderson CW, Chessa L, Smorodinsky NI, Prives C, Reiss Y, Shiloh Y *et al*: **Enhanced phosphorylation of p53 by ATM in response to DNA damage**. *Science* 1998, **281**(5383):1674-1677.
118. Zhang Y, Xiong Y: **A p53 amino-terminal nuclear export signal inhibited by DNA damage-induced phosphorylation**. *Science* 2001, **292**(5523):1910-1915.
119. Zou L, Elledge SJ: **Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes**. *Science* 2003, **300**(5625):1542-1548.
120. Yang XH, Zou L: **Checkpoint and coordinated cellular responses to DNA damage**. *Results Probl Cell Differ* 2006, **42**:65-92.
121. Nam EA, Cortez D: **ATR signalling: more than meeting at the fork**. *Biochem J* 2011, **436**(3):527-536.
122. Abraham RT: **Cell cycle checkpoint signaling through the ATM and ATR kinases**. *Genes Dev* 2001, **15**(17):2177-2196.

123. Zhao H, Piwnica-Worms H: **ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1**. *Mol Cell Biol* 2001, **21**(13):4129-4139.
124. Marti TM, Hefner E, Feeney L, Natale V, Cleaver JE: **H2AX phosphorylation within the G1 phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks**. *Proc Natl Acad Sci U S A* 2006, **103**(26):9891-9896.
125. Yarden RI, Pardo-Reoyo S, Sgagias M, Cowan KH, Brody LC: **BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage**. *Nat Genet* 2002, **30**(3):285-289.
126. Lakin ND, Hann BC, Jackson SP: **The ataxia-telangiectasia related protein ATR mediates DNA-dependent phosphorylation of p53**. *Oncogene* 1999, **18**(27):3989-3995.
127. Hall-Jackson CA, Cross DA, Morrice N, Smythe C: **ATR is a caffeine-sensitive, DNA-activated protein kinase with a substrate specificity distinct from DNA-PK**. *Oncogene* 1999, **18**(48):6707-6713.
128. Shieh SY, Ahn J, Tamai K, Taya Y, Prives C: **The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites**. *Genes Dev* 2000, **14**(3):289-300.
129. Falck J, Mailand N, Syljuasen RG, Bartek J, Lukas J: **The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis**. *Nature* 2001, **410**(6830):842-847.
130. Molinari M, Mercurio C, Dominguez J, Goubin F, Draetta GF: **Human Cdc25 A inactivation in response to S phase inhibition and its role in preventing premature mitosis**. *EMBO Rep* 2000, **1**(1):71-79.
131. Bartek J, Lukas J: **Mammalian G1- and S-phase checkpoints in response to DNA damage**. *Curr Opin Cell Biol* 2001, **13**(6):738-747.
132. Chehab NH, Malikzay A, Stavridi ES, Halazonetis TD: **Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage**. *Proc Natl Acad Sci U S A* 1999, **96**(24):13777-13782.
133. Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ: **The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases**. *Cell* 1993, **75**(4):805-816.
134. Lin WC, Lin FT, Nevins JR: **Selective induction of E2F1 in response to DNA damage, mediated by ATM-dependent phosphorylation**. *Genes Dev* 2001, **15**(14):1833-1844.
135. Paulovich AG, Hartwell LH: **A checkpoint regulates the rate of progression through S phase in S. cerevisiae in response to DNA damage**. *Cell* 1995, **82**(5):841-847.
136. Kim ST, Xu B, Kastan MB: **Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage**. *Genes Dev* 2002, **16**(5):560-570.
137. Yazdi PT, Wang Y, Zhao S, Patel N, Lee EY, Qin J: **SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint**. *Genes Dev* 2002, **16**(5):571-582.

138. Byun TS, Pacek M, Yee MC, Walter JC, Cimprich KA: **Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint**. *Genes Dev* 2005, **19**(9):1040-1052.
139. Nigg EA: **Mitotic kinases as regulators of cell division and its checkpoints**. *Nat Rev Mol Cell Biol* 2001, **2**(1):21-32.
140. Deckbar D, Jeggo PA, Lobrich M: **Understanding the limitations of radiation-induced cell cycle checkpoints**. *Crit Rev Biochem Mol Biol* 2011, **46**(4):271-283.
141. MacDonald BT, Tamai K, He X: **Wnt/beta-catenin signaling: components, mechanisms, and diseases**. *Dev Cell* 2009, **17**(1):9-26.
142. Nusse R: **Wnt signaling and stem cell control**. *Cell Res* 2008, **18**(5):523-527.
143. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, Morin PJ, Vogelstein B, Kinzler KW: **Identification of c-MYC as a target of the APC pathway**. *Science* 1998, **281**(5382):1509-1512.
144. Couso JP, Bishop SA, Martinez Arias A: **The wingless signalling pathway and the patterning of the wing margin in Drosophila**. *Development* 1994, **120**(3):621-636.
145. de Celis JF, Bray S: **Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the Drosophila wing**. *Development* 1997, **124**(17):3241-3251.
146. Micchelli CA, Rulifson EJ, Blair SS: **The function and regulation of cut expression on the wing margin of Drosophila: Notch, Wingless and a dominant negative role for Delta and Serrate**. *Development* 1997, **124**(8):1485-1495.
147. Neumann CJ, Cohen SM: **A hierarchy of cross-regulation involving Notch, wingless, vestigial and cut organizes the dorsal/ventral axis of the Drosophila wing**. *Development* 1996, **122**(11):3477-3485.
148. Von Ohlen T, Hooper JE: **Hedgehog signaling regulates transcription through Gli/Ci binding sites in the wingless enhancer**. *Mech Dev* 1997, **68**(1-2):149-156.
149. Chen Y, Goodman RH, Smolik SM: **Cubitus interruptus requires Drosophila CREB-binding protein to activate wingless expression in the Drosophila embryo**. *Mol Cell Biol* 2000, **20**(5):1616-1625.
150. Dahmane N, Lee J, Robins P, Heller P, Ruiz i Altaba A: **Activation of the transcription factor Gli1 and the Sonic hedgehog signalling pathway in skin tumours.[Erratum appears in Nature 1997 Dec 4;390(6659):536]**. *Nature* 1997, **389**(6653):876-881.
151. Sheng H, Goich S, Wang A, Grachtchouk M, Lowe L, Mo R, Lin K, de Sauvage FJ, Sasaki H, Hui C-c *et al*: **Dissecting the oncogenic potential of Gli2: deletion of an NH(2)-terminal fragment alters skin tumor phenotype**. *Cancer Res* 2002, **62**(18):5308-5316.
152. Mullor JL, Dahmane N, Sun T, Ruiz i Altaba A: **Wnt signals are targets and mediators of Gli function**. *Curr Biol* 2001, **11**(10):769-773.
153. Li X, Deng W, Lobo-Ruppert SM, Ruppert JM: **Gli1 acts through Snail and E-cadherin to promote nuclear signaling by beta-catenin**. *Oncogene* 2007, **26**(31):4489-4498.

154. Segditsas S, Tomlinson I: **Colorectal cancer and genetic alterations in the Wnt pathway**. *Oncogene* 2006, **25**(57):7531-7537.
155. Bafico A, Liu GZ, Goldin L, Harris V, Aaronson SA: **An autocrine mechanism for constitutive Wnt pathway activation in human cancer cells**. *Cancer Cell* 2004, **6**(5):497-506.
156. Benhaj K, Akcali KC, Ozturk M: **Redundant expression of canonical Wnt ligands in human breast cancer cell lines**. *Oncology Reports* 2006, **15**(3):701-707.
157. Schlange T, Matsuda Y, Lienhard S, Huber A, Hynes NE: **Autocrine WNT signaling contributes to breast cancer cell proliferation via the canonical WNT pathway and EGFR transactivation**. *Breast Cancer Research* 2007, **9**(5).
158. Akiri G, Cherian MM, Vijayakumar S, Liu G, Bafico A, Aaronson SA: **Wnt pathway aberrations including autocrine Wnt activation occur at high frequency in human non-small-cell lung carcinoma**. *Oncogene* 2009, **28**(21):2163-2172.
159. Liu X, Mazanek P, Dam V, Wang Q, Zhao H, Guo R, Jagannathan J, Cnaan A, Maris JM, Hogarty MD: **Deregulated Wnt/beta-catenin program in high-risk neuroblastomas without MYCN amplification**. *Oncogene* 2008, **27**(10):1478-1488.
160. Xu J, Suzuki M, Niwa Y, Hiraga J, Nagasaka T, Ito M, Nakamura S, Tomita A, Abe A, Kiyoi H *et al*: **Clinical significance of nuclear non-phosphorylated beta-catenin in acute myeloid leukaemia and myelodysplastic syndrome**. *Br J Haemat* 2008, **140**(4):394-401.
161. Fiaschi M, Rozell B, Bergstrom A, Toftgard R: **Development of mammary tumors by conditional expression of GLI1**. *Cancer Res* 2009, **69**(11):4810-4817.
162. Grove EA, Tole S, Limon J, Yip L, Ragsdale CW: **The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice**. *Development* 1998, **125**(12):2315-2325.
163. Yan L, Della Coletta L, Powell KL, Shen J, Thames H, Aldaz CM, MacLeod MC: **Activation of the canonical Wnt/beta-catenin pathway in ATF3-induced mammary tumors**. *PloS one* 2011, **6**(1):e16515.
164. Yao H, Ashihara E, Maekawa T: **Targeting the Wnt/beta-catenin signaling pathway in human cancers**. *Expert Opin Ther Targets* 2011, **15**(7):873-887.
165. Hallett RM, Kondratyev MK, Giacomelli AO, Nixon AM, Girgis-Gabardo A, Ilieva D, Hassell JA: **Small molecule antagonists of the Wnt/beta-catenin signaling pathway target breast tumor-initiating cells in a Her2/Neu mouse model of breast cancer**. *PLoS One* 2011, **7**(3):e33976.
166. Howcroft TK, Kirshner SL, Singer DS: **Measure of transient transfection efficiency using beta-galactosidase protein**. *Anal Biochem* 1997, **244**(1):22-27.
167. Lee KA, Bindereif A, Green MR: **A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing**. *Gene Anal Tech* 1988, **5**(2):22-31.

168. Olive PL, Banath JP: **The comet assay: a method to measure DNA damage in individual cells**. *Nat Protoc* 2006, **1**(1):23-29.
169. Ponzo MG, Lesurf R, Petkiewicz S, O'Malley FP, Pinnaduwage D, Andrulis IL, Bull SB, Chughtai N, Zuo D, Souleimanova M *et al*: **Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer**. *Proc Natl Acad Sci U S A* 2009, **106**(31):12903-12908.
170. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments**. *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
171. Weinmann AS, Farnham PJ: **Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation**. *Methods* 2002, **26**(1):37-47.
172. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B: **The tandem affinity purification (TAP) method: a general procedure of protein complex purification**. *Methods* 2001, **24**(3):218-229.
173. Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD: **E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints**. *Genes Dev* 2002, **16**(2):245-256.
174. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J: **Galaxy: a web-based genome analysis tool for experimentalists**. *Curr Protoc Mol Biol* 2010, **Chapter 19**:Unit 19 10 11-21.
175. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J *et al*: **Galaxy: a platform for interactive large-scale genome analysis**. *Genome Res* 2005, **15**(10):1451-1455.
176. Goecks J, Nekrutenko A, Taylor J: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences**. *Genome biology* 2010, **11**(8):R86.
177. Bailey TL: **DREME: motif discovery in transcription factor ChIP-seq data**. *Bioinformatics* 2011, **27**(12):1653-1659.
178. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching**. *Nucleic Acids Res* 2009, **37**(Web Server issue):W202-208.
179. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS: **Quantifying similarity between motifs**. *Genome biology* 2007, **8**(2):R24.
180. Harada R, Berube G, Tamplin OJ, Denis-Larose C, Nepveu A: **DNA-binding specificity of the cut repeats from the human cut-like protein**. *Mol Cell Biol* 1995, **15**(1):129-140.
181. Harada R, Nepveu A: **Chromatin affinity purification**. *Methods in molecular biology (Clifton, NJ)* 2012, **809**(22113280):237-253.
182. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B: **A high-resolution map of active promoters in the human genome**. *Nature* 2005, **436**(7052):876-880.

183. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR *et al*: **Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1**. *Cell* 2005, **122**(1):33-43.
184. Ruiz-Llorente S, de Pau EC, Sastre-Perona A, Montero-Conde C, Gomez-Lopez G, Fagin JA, Valencia A, Pisano DG, Santisteban P: **Genome-wide analysis of Pax8 binding provides new insights into thyroid functions**. *BMC Genomics* 2012, **13**:147.
185. Li Q, Harju S, Peterson KR: **Locus control regions: coming of age at a decade plus**. *Trends Genet* 1999, **15**(10):403-408.
186. Gaszner M, Felsenfeld G: **Insulators: exploiting transcriptional and epigenetic mechanisms**. *Nat Rev Genet* 2006, **7**(9):703-713.
187. Gross DS, Garrard WT: **Nuclease hypersensitive sites in chromatin**. *Annu Rev Biochem* 1988, **57**:159-197.
188. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B *et al*: **The accessible chromatin landscape of the human genome**. *Nature* 2012, **489**(7414):75-82.
189. Ernst J, Kellis M: **ChromHMM: automating chromatin-state discovery and characterization**. *Nature methods* 2012, **9**(3):215-216.
190. Alcalay NI, Sharma M, Vassmer D, Chapman B, Paul B, Zhou J, Brantley JG, Wallace DP, Maser RL, Vanden Heuvel GB: **Acceleration of polycystic kidney disease progression in cpk mice carrying a deletion in the homeodomain protein Cux1**. *Am J Physiol Renal Physiol* 2008, **295**(6):F1725-1734.
191. Shiloh Y: **ATM and related protein kinases: safeguarding genome integrity**. *Nat Rev Cancer* 2003, **3**(3):155-168.
192. Harper JW, Elledge SJ: **The DNA damage response: ten years after**. *Mol Cell* 2007, **28**(5):739-745.
193. Essers J, Houtsmuller AB, van Veelen L, Paulusma C, Nigg AL, Pastink A, Vermeulen W, Hoeijmakers JH, Kanaar R: **Nuclear dynamics of RAD52 group homologous recombination proteins in response to DNA damage**. *Embo J* 2002, **21**(8):2030-2037.
194. Rak JW, Basolo F, Elliott JW, Russo J, Miller FR: **Cell surface glycosylation changes accompanying immortalization and transformation of normal human mammary epithelial cells**. *Cancer letters* 1991, **57**(1):27-36.
195. Kedinger V, Nepveu A: **The roles of CUX1 homeodomain proteins in the establishment of a transcriptional program required for cell migration and invasion**. *Cell Adh Migr* 2010, **4**(3):348-352.
196. Vadnais C, Davoudi S, Afshin M, Harada R, Dudley R, Clermont PL, Drobetsky E, Nepveu A: **CUX1 transcription factor is required for optimal ATM/ATR-mediated responses to DNA damage**. *Nucleic Acids Res* 2012, **40**(10):4483-4495.
197. Seguin L, Liot C, Mzali R, Harada R, Siret A, Nepveu A, Bertoglio J: **CUX1 and E2F1 regulate coordinated expression of the mitotic**

complex genes Ect2, MgcRacGAP, and MKLP1 in S phase. *Mol Cell Biol* 2009, **29**(2):570-581.
198. Truscott M, Harada R, Vadnais C, Robert F, Nepveu A: **p110 CUX1 cooperates with E2F transcription factors in the transcriptional activation of cell cycle-regulated genes**. *Mol Cell Biol* 2008, **28**(10):3127-3138.
199. Sansregret L, Nepveu A: **Gene signatures of genomic instability as prognostic tools for breast cancer**. *Future Oncol* 2011, **7**(5):591-594.
200. Kumagai A, Lee J, Yoo HY, Dunphy WG: **TopBP1 activates the ATR-ATRIP complex**. *Cell* 2006, **124**(5):943-955.
201. Galhardo RS, Hastings PJ, Rosenberg SM: **Mutation as a stress response and the regulation of evolvability**. *Critical Reviews in Biochemistry & Molecular Biology* 2007, **42**(5):399-435.
202. Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T: **The SOS Responses of Prokaryotes to DNA Damage**. In: *DNA Repair and Mutagenesis.* 2nd edn. Washington, DC: ASM Press; 2006: 461-508.
203. Walker GC: **Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli**. *Microbiol Rev* 1984, **48**(1):60-93.
204. Defais M, Fauquet P, Radman M, Errera M: **Ultraviolet reactivation and ultraviolet mutagenesis of lambda in different genetic systems**. *Virology* 1971, **43**(2):495-503.
205. Fu Y, Zhu Y, Zhang K, Yeung M, Durocher D, Xiao W: **Rad6-Rad18 mediates a eukaryotic SOS response by ubiquitinating the 9-1-1 checkpoint clamp**. *Cell* 2008, **133**(4):601-611.
206. Beckerman R, Prives C: **Transcriptional regulation by p53**. *Cold Spring Harbor Perspect Biol* 2010, **2**(8):a000935.
207. Polager S, Kalma Y, Berkovich E, Ginsberg D: **E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis**. *Oncogene* 2002, **21**(3):437-446.
208. DeGregori J, Johnson DG: **Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis**. *Current molecular medicine* 2006, **6**(7):739-748.
209. Liao CC, Tsai CY, Chang WC, Lee WH, Wang JM: **RB.E2F1 complex mediates DNA damage responses through transcriptional regulation of ZBRK1**. *The Journal of biological chemistry* 2010, **285**(43):33134-33143.
210. Ganguly A, Shields CL: **Differential gene expression profile of retinoblastoma compared to normal retina**. *Mol Vis* 2010, **16**:1292-1303.
211. Barry SP, Townsend PA, Knight RA, Scarabelli TM, Latchman DS, Stephanou A: **STAT3 modulates the DNA damage response pathway**. *International Journal of Experimental Pathology* 2010, **91**(6):506-514.
212. De Siervi A, De Luca P, Moiola C, Gueron G, Tongbai R, Chandramouli GVR, Haggerty C, Dzekunova I, Petersen D, Kawasaki E *et al*: **Identification of new Rel/NFkappaB regulatory networks by focused genome location analysis**. *Cell Cycle* 2009, **8**(13):2093-2100.

213. Tan Y, Chen Y, Yu L, Zhu H, Meng X, Huang X, Meng L, Ding M, Wang Z, Shan L: **Two-fold elevation of expression of FoxM1 transcription factor in mouse embryonic fibroblasts enhances cell cycle checkpoint activity by stimulating p21 and Chk1 transcription**. *Cell Prolif* 2010, **43**(5):494-504.
214. Stokes MP, Rush J, Macneill J, Ren JM, Sprott K, Nardone J, Yang V, Beausoleil SA, Gygi SP, Livingstone M *et al*: **Profiling of UV-induced ATM/ATR signaling pathways**. *Proc Natl Acad Sci U S A* 2007, **104**(50):19855-19860.
215. Velez-Cruz R, Johnson DG: **E2F1 and p53 Transcription Factors as Accessory Factors for Nucleotide Excision Repair**. *Int J Mol Sci* 2012, **13**(10):13554-13568.
216. Liu K, Lin FT, Ruppert JM, Lin WC: **Regulation of E2F1 by BRCT domain-containing protein TopBP1**. *Mol Cell Biol* 2003, **23**(9):3287-3304.
217. Guo R, Chen J, Zhu F, Biswas AK, Berton TR, Mitchell DL, Johnson DG: **E2F1 localizes to sites of UV-induced DNA damage to enhance nucleotide excision repair**. *J Biol Chem* 2010, **285**(25):19308-19315.
218. Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD, Rich JN: **Glioma stem cells promote radioresistance by preferential activation of the DNA damage response**. *Nature* 2006, **444**(7120):756-760.
219. Maekawa M, Yamaguchi K, Nakamura T, Shibukawa R, Kodanaka I, Ichisaka T, Kawamura Y, Mochizuki H, Goshima N, Yamanaka S: **Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1**. *Nature* 2011, **474**(7350):225-229.
220. Michl P, Downward J: **CUTL1: a key mediator of TGFbeta-induced tumor invasion**. *Cell Cycle* 2006, **5**(2):132-134.

TABLES

Table 1 – CUX1 Binding Sites on the ENCODE Array

Number and average width of CUX1 binding sites identified on the NimbleGen HG17 ENCODE array using chromatin purified from Hs578t cells. Also shown are the number and percentage of binding sites that were validated in an independent an ChIP experiment. The validation rate is also shown independently for sites that contained the ATCRAT consensus sequence as well as for sites that did not.

Table 1 – CUX1 Binding Sites on the ENCODE Array

# of binding sites	513
Average site width (bp)	553
Sites tested in qPCR	25
Validation rate	92%
Validation rate (with consensus)	100%
Validation rate (no consensus)	90%

Table 2 – Distribution of CUX1, Myc and E2F1 Binding Sites in Transcribed and Non-Transcribed Regions

Number of CUX1, C-Myc and E2F1binding sites in transcribed and un-transcribed regions. Also indicated are the fold enrichments in transcribed regions. P Values are calculated using a Fisher's exact test.

Table 2 – Distribution of CUX1, Myc and E2F1 Binding Sites in Transcribed and Non-Transcribed Regions

	Encode platform	CUX1	c-MYC	E2F1
Number of binding sites		513	172	204
Non-transcribed regions	20.4%	28.1%	28.5%	5.9%
Transcribed regions	79.6%	70.9%	71.5%	94.1%
Enrichment in un-transcribed regions		1.61	1.56	0.24
P Value		0.0018	0.1333	<0.0001

Table 3 – A Fraction of CUX1 Binding Sites Locate Close to Enhancer Elements and DHS Sites

Percentages of CUX1 binding sites located more than 4 Kbps away from a TSS but within 1 Kbp of the indicated type of genomic element. Percentages are shown for a set of randomly generated binding sites of the same size distribution as CUX1. P Value is calculated using a Fisher's exact test. See Methods for information on the datasets used.

Table 3 – A Fraction of CUX1 Binding Sites Locate Close to Enhancer Elements and DHS Sites

Type	CUX1	Random	Fold Difference	P Value
DHS	19.2%	12.9%	1.49	0.0109
Enhancers	22.1%	15.2%	1.45	0.0100
Insulators	4.43%	4.40%	1.01	1.0000

Table 4 – Binding Sites and Target Genes Predicted To Be Identified in Promoter Arrays

A Percentages of CUX1, C-Myc and E2F1 binding sites that were identified on the ENCODE array and that are located within the boundaries of promoters on various promoter array platforms.

B Percentages of CUX1, C-Myc and E2F1 target genes that are identified on the ENCODE array and whose binding site are located within the boundaries of promoters on various promoter array platforms.

C The second column shows the number of genes on the ENCODE. The third column shown the number and percentage of these genes that were also identified on the Nimblegen promoter array.

Table 4 – Binding Sites and Target Genes Predicted To Be Identified in Promoter Arrays

A

Platform	Promoter array boundaries	% of Binding sites predicted in promoter array		
		CUX1	C-Myc	E2F1
Nimblegen	-3.5kb to + 0.75kb	17.2%	26.8%	80.4%
Agilent	-5.5kb to + 2.5kb	23.4%	34.3%	84.3%
Affymetrix	-7.5kb to + 2.45kb	26.6%	34.9%	85.8%

B

Platform	Promoter array boundaries	% of Target genes predicted in promoter array		
		CUX1	C-Myc	E2F1
Nimblegen	-3.5kb to + 0.75kb	44.6%	36.0%	90.2%
Agilent	-5.5kb to + 2.5kb	57.1%	45.9%	92.2%
Affymetrix	-7.5kb to + 2.45kb	58.5%	48.3%	92.2%

C

	Total	Identified on Nimblegen Promoter Array
All Genes on the ENCODE Array	445	92 (21%)
Regulated Genes on the ENCODE Array (1.25)	85	27 (32%)
Regulated Genes on the ENCODE Array (1.5)	26	8 (31%)

Table 5 – CUX1 Binding Sites on the Promoter Array

Columns 1 and 2: Number and average width of CUX1 binding sites identified on the NimbleGen HG18 Human Promoter Array using chromatin purified from Hs578t cells. Also shown are the number and percentage of binding sites that were validated in an independent an ChIP experiment.

Columns 3 and 4: Number of genes with the indicated number of CUX1 binding site. Note that validation was performed on 14 and 11 sites that contained or not an ATCRAT motif. All of them were validated.

Table 5 – CUX1 Binding Sites on the Promoter Array

Genes on array	20593	Number of sites/gene	Number of genes
CUX1 Binding sites	5828	1	3942
Genes bound by CUX1	4706	2	643
Average Site Width (bp)	513	3	90
Sites tested in qPCR	25	4	23
Validation rate	100%	5+	8

Table 6 –CUX1 Consensus Binding Sites and Bound Genomic Regions

Columns 2-4, occurrence of the CUX1 consensus binding site, ATCRAT (where R = C or A), within the 5828 genomic regions bound by CUX1 on the promoter array. To calculate the p value, an equal number of randomly chosen regions of equal width was searched for the presence of the CUX1 consensus binding site: ***: $p<0.001$. Column 5 shows the GC content of bound and unbound regions.

Table 6 –CUX1 Consensus Binding Sites and Bound Genomic Regions

	Regions	Regions with consensus	% with consensus	GC Content
Bound Regions	5828	2749	47.2% ***	47.3%
Unbound Regions	5828	1020	17.5%	47.0%

Table 7A – Functions of CUX1 Target Genes That Contain a Consensus CUX1 Binding Site

Ten most over-represented biological functions of CUX1 target genes from the promoter array which contain a consensus CUX1 binding sequence (ATCRAT). Overrepresentation is determined using the online DAVID tool (see methods).

Table 7B – Functions of CUX1 Target Genes That Do Not Contain a Consensus CUX1 Binding Site

Ten most over-represented biological functions of CUX1 target genes from the promoter array which do not contain a consensus CUX1 binding sequence (ATCRAT). Overrepresentation is determined using the online DAVID tool.

Table 7A – Functions of CUX1 Target Genes That Contain a Consensus CUX1 Binding Site

Functional Term	Fold Enrichment	P Value
Macromolecular complex assembly	1.39	3.1E-04
Microtubule cytoskeleton organization	1.88	6.1E-04
Cytoskeleton organization	1.47	6.6E-04
Response to DNA damage stimulus	1.51	7.2E-04
Negative regulation of programmed cell death	1.49	1.2E-03
Anti-apoptosis	1.68	1.2E-03
Cellular response to stress	1.38	1.4E-03
Cellular macromolecule catabolic process	1.32	1.5E-03
Protein localization	1.28	1.7E-03
Translational elongation	1.97	2.3E-03

Table 7B – Functions of CUX1 Target Genes That Do Not Contain a Consensus CUX1 Binding Site

Functional Term	Fold Enrichment	P Value
Ribonucleoprotein complex biogenesis	2.11	3.3E-06
Translation	1.73	1.1E-05
RNA processing	1.55	1.1E-05
Cell cycle	1.41	5.4E-05
Mitotic cell cycle	1.62	6.1E-05
Ribosome biogenesis	2.17	7.4E-05
Nuclear mRNA splicing, via spliceosome	2.02	8.0E-05
Cell cycle process	1.45	1.9E-04
Establishment of protein localization	1.38	1.9E-04
Translational elongation	2.19	3.0E-04

Table 8 – Identification of DNA Motifs in CUX1 Binding Sites

A 10 Most enriched DNA motifs found in CUX1 binding sites that contain the ATCRAT CUX1 consensus. Proteins with DNA binding motifs highly similar to the consensus are listed in the rightmost column. K=G/T, M=A/C, R=A/G, Y=C/T, S=C/G, W=A/T, B=C/G/T, V=A/C/G, H=A/C/T, D=A/G/T

B 10 Most enriched DNA motifs found in CUX1 binding sites that do not contain the ATCRAT CUX1 consensus. Proteins with DNA binding motifs highly similar to the consensus are listed in the rightmost column.

Table 8 – Identification of DNA Motifs in CUX1 Binding Sites

A

Motif	Reverse Complement	E-Value	Transcription factors
ATCRAT	ATYGAT	3.5E-735	Cux1, Pbx1
GGGYGGGR	YCCCRCCC	4.8E-35	Klf4, Klf7, Sp1, Sp4, Zfp281, Zfp740, Egr1
AAATAHW	WDTATTT	1.9E-27	-
CTBCCTS	SAGGVAG	6.30E-26	Spi1, Stat3, Fev, Sfpi1
CWCCDCC	GGHGGWG	6.60E-23	-
DRGGAAA	TTTCCYH	6.20E-21	-
BSTGTGTG	CACACASV	1.20E-20	-
RGAGAAR	YTTCTCY	2.60E-14	-
ACRCWG	CWGYGT	3.70E-14	-
RAAACAAA	TTTGTTTY	1.90E-11	Sox11, Sox4, Foxd3, Foxi1

B

Motif	Reverse Complement	E-Value	Transcription factors
DTATTTW	WAAATAH	3.80E-35	-
CYCCRCCC	GGGYGGRG	4.60E-34	Klf4, Klf7, Sp1, Sp4, Zfp281, Zfp740, Egr1
CAYTTCY	RGAARTG	1.50E-26	Gabpa, Stat1
CACASAS	STSTGTG	3.20E-23	Runx1
DGGAAA	TTTCCH	5.00E-22	Stat1, Nfatc2, Rela, Rel, Fev
CCRCCDCC	GGHGGYGG	6.40E-19	-
GSAGAGR	YCTCTSC	3.90E-17	-
CHGCAGC	GCTGCDG	1.30E-16	Myf, Ascl2
CATTTWM	KWAAATG	2.90E-26	-
DTTTCTS	SAGAAAH	1.70E-13	-

Table 9 – Genes on the ENCODE Array That are Regulated in Response to CUX1 Overexpression or CUX1 Knockdown

A Number and percentage of genes on the ENCODE platform that exhibit a 1.5 fold change in expression following p110 CUX1 overexpression or CUX1 knockdown. "Upregulated by CUX1" are genes whose expression is increased following p110 CUX1 and/or decreased following CUX1 knockdown. Conversely, "Downregulated by CUX1" are genes whose expression is decreased following p110 CUX1 and/or increased following CUX1 knockdown.

B Number and percentage of target genes on the ENCODE platform that exhibit a 1.25 fold change in expression following p110 CUX1 overexpression or CUX1 knockdown. Genes were analyzed as in **A**.

Table 9 – Genes on the ENCODE Array That are Regulated in Response to CUX1 Overexpression or CUX1 Knockdown

A

Effect of CUX1 on 327 Putative Target Genes	CUX1 Overexpression or Knockdown	CUX1 Overexpression	CUX1 Knockdown
Up- or Downregulated	26 (7.4%)	6 (1.7%)	20 (5.7%)
Upregulated	10 (2.9%)	3 (0.9%)	7 (2.0%)
Downregulated	16 (4.6%)	3 (0.9%)	13 (3.7%)

B

Effect of CUX1 on 327 Putative Target Genes	CUX1 Overexpression or Knockdown	CUX1 Overexpression	CUX1 Knockdown
Up- or Downregulated	85 (24.4%)	36 (10.3%)	62 (17.8%)
Upregulated	35 (10.0%)	18 (5.2%)	24 (10.9%)
Downregulated	50 (14.3%)	18 (5.2%)	38 (6.9%)

Table 10 – Number of Genes on the Promoter Array That Are Regulated in Response to CUX1 Overexpression or CUX1 Knockdown

A Number and percentage of all genes and CUX1 target genes on the promoter array that exhibit a 1.5 fold change in expression following p110 CUX1 overexpression or CUX1 knockdown. "Upregulated by CUX1" are genes whose expression is increased following p110 CUX1 and/or decreased following CUX1 knockdown. Conversely, "Downregulated by CUX1" are genes whose expression is decreased following p110 CUX1 and/or increased following CUX1 knockdown. The total number of genes on the array and target genes were 17586 and 4140, respectively.

B Number and percentage of all genes and CUX1 target genes on the promoter array that exhibit a 1.25 fold change in expression following p110 CUX1 overexpression or CUX1 knockdown. Genes were analyzed as in A.

C Effect of the number of CUX1 binding sites on the probability that target genes exhibit a change in expression following p110 CUX1 overexpression or CUX1 knockdown, depending on the number of CUX1 binding site present in their promoter region. **: $P<0.01$, ***:$P<0.001$ on a Fisher's exact test vs. genes whose promoter have only 1 CUX1 binding site.

Table 10 – Number of Genes on the Promoter Array That Are Regulated in Response to CUX1 Overexpression or CUX1 Knockdown

A

Effect of CUX1 on All Genes and 4140 Putative Targets	CUX1 Overexpression or Knockdown		CUX1 Overexpression	CUX1 Knockdown
Gene List	All Genes	Target Genes	Target Genes	Target Genes
Up- or Downregulated	1231 7.0%	347 8.4%	85 2.1%	287 6.9%
Upregulated	591 3.4%	181 4.4%	28 0.7%	169 4.1%
Downregulated	640 3.6%	167 4.0%	57 1.4%	118 2.8%

B

Effect of CUX1 on All Genes and 4140 Putative Targets	CUX1 Overexpression or Knockdown		CUX1 Overexpression	CUX1 Knockdown
Gene List	All Genes	Target Genes	Target Genes	Target Genes
Up- or Downregulated	4880 27.7%	1437 34.7%	568 13.7%	1083 26.1%
Upregulated	2290 13.0%	696 16.8%	261 6.3%	546 13.2%
Downregulated	2590 14.7%	744 17.9%	307 7.4%	537 13.0%

C

Number of sites/target	Number of Targets	Targets with Profiling Data	1.5 fold change	1.25 fold change
Any #	4706	4140	347 8.4%	1437 34.7%
1	3942	3527	278 7.9%	1182 33.5%
2+	643	613	** 69 11.2%	*** 255 41.6%

Table 11 – Over-Represented Biological Functions of CUX1 Targets

A single list of putative targets of CUX1 was compiled from 8 individual ChIP-chip experiments from cell lines overexpressing p110 CUX1. Genes that were bound by CUX1 (Targets) were compared with all genes present on the microarray (Background) by using a web-based functional annotation tool, DAVID. Overrepresentation of a function depends on the increase in the proportion of genes involved in a given function between CUX1 targets and the background. The P Value is determined using an improved Fisher's exact test from the DAVID software. The top 20 significantly over-represented functions are shown.

Table 11 – Over-Represented Biological Functions of CUX1 Targets

Function	Background	Targets	P Value
Mitotic cell cycle	2.6%	4.9%	2.12E-10
DNA replication and chromosome cycle	1.5%	3.0%	3.46E-08
Cell cycle	5.6%	8.3%	4.18E-08
Cell proliferation	8.5%	11.6%	8.83E-08
M phase of mitotic cell cycle	1.0%	2.3%	9.66E-08
Mitosis	1.0%	2.2%	1.65E-07
M phase	1.3%	2.6%	3.57E-07
DNA replication	1.1%	2.3%	1.44E-06
Nuclear division	1.3%	2.5%	1.90E-06
S phase of mitotic cell cycle	1.2%	2.3%	2.24E-06
DNA metabolism	4.2%	6.3%	2.37E-06
Cell growth and/or maintenance	30.3%	34.0%	7.61E-05
Intracellular transport	4.5%	6.1%	0.00044
Nucleosome assembly	0.5%	1.1%	0.00049
Protein metabolism	19.9%	22.6%	0.00061
DNA dependent DNA replication	0.6%	1.2%	0.00074
Cell cycle checkpoint	0.3%	0.7%	0.00079
Regulation of cell cycle	3.1%	4.3%	0.00090
Small GTPase mediated signal transduction	1.9%	2.8%	0.00110
Protein folding	1.0%	1.7%	0.00186

Table 12 – Transcriptional Targets of CUX1 Involved in DDR

Table shows the validation of regulation of DDR genes by CUX1. Column 2 shows enrichment of CUX1 at each gene's promoter. Column 3 shows mRNA levels of DDR genes following siRNA knockdown. Column 4 shows mRNA levels in *Cux1*$^{Z/Z}$ MEFs relative to wild-type littermates. Column 5 shows mRNA levels following Overexpression of p110 CUX1 by retroviral infection. All mRNA levels are shown normalized to HPRT1; GAPDH, ACTB and UBC are shown as additional housekeeping genes that are unaffected by CUX1 levels. Column 6 shows level of activation of reporter constructs by p110 CUX1.

* Indicates p Value <0.05. ** <0.01. *** <0.001.

Table 12 – Transcriptional Targets of CUX1 Involved in DDR

Symbol	ChIP-qPCR	siRNA		MEF CUX1$^{z/z}$		Overexpression		Reporter	
ATM	5.1	0.74	*	0.61	***	2.9	**		
ATR	3.9	0.56	***	0.56	***	2.3	***	4.7	*
BRCA1	2.4	0.40	**	0.63	***	4.5	***		
CCNG1	4.9	0.71	*	0.68	***	3.5	***	2.6	*
CDKN1A	2.7	1.32	***	1.32	*	0.5	*	0.2	***
CHK1	3.2	0.58	***	0.57	***	4.4	***	3.7	**
CHK2	2.1	0.68	**	0.79	***	2.6	***		
FANCD2	2	0.76	**	0.71	***	2.6	***		
MDM2	5.3	0.93		0.57	***	3	***	1.7	**
NBS1	2.2	0.56	***	0.97		2.6	**		
RAD17	2.8	0.50	***	1.10		2.2	**		
RPA1	2.2	0.89		0.65	***	3	***		
RPA2	3.7	0.61	***	0.87	*	5.5	***		
RPA3	19.4	0.65	***	0.86	**	3.2	***	5.7	*
SMC1L1	3.1	0.82	*	0.57	***	2	***		
TOPBP1	3.3	0.52	***	0.45	***	3.6	**		
TP53	6	0.67	*	0.72	*	2.4	*	9	**
TP53BP1	2.5	0.65	*	0.82	*	3.3	***		
GAPDH	0.91	1.06		0.94		1.02			
ACTB	1.22	1.00		1.01		1.01			
UBC	0.84	0.96		1.08		1.05			

Table 13 – Correlation Between *WNT*, *CUX1* and *GLI* Gene Expression in Human Tumour Datasets

The table shows the number of human tumour datasets retrieved from Oncomine in which the top 25% of samples ranked according to Wnt gene expression display significantly higher expression of *CUX1*, *CTSL2*, *GLI1*, *GLI2*, *GLI3* or *GLIS1* than the bottom 25% of samples. "Gli genes" refers to any combination of *GLI1*, *GLI2* and/or *GLI3*.

Table 13 – Correlation Between *WNT*, *CUX1* and *GLI* Gene Expression in Human Tumour Datasets

Gene	Breast datasets	Lung datasets	All organs
CUX1	3 of 3 (100%)	2 of 2 (100%)	5 of 5 (100%)
Glis1	7 of 8 (88%)	5 of 5 (100%)	12 of 13 (92%)
1 Gli gene	8 of 9 (89%)	8 of 9 (89%)	16 of 18 (89%)
2 Gli genes	5 of 9 (56%)	6 of 9 (67%)	11 of 18 (61%)
3 Gli genes	1 of 9 (11%)	3 of 9 (33%)	4 of 18 (22%)

Table 14 – Correlation Between *WNT* Gene Expression and EMT Markers in Human Tumour Datasets

The table shows for each dataset the number of EMT markers whose expression is differentially expressed between the top and bottom 25% of samples ranked according to *WNT* gene expression.

Table 14 – Correlation Between *WNT* Gene Expression and EMT Markers in Human Tumour Datasets

Dataset	Differentially Expressed EMT Genes
Esserman Breast	5 of 7 (71%)
Gluck Breast	5 of 7 (71%)
TCGA Breast	6 of 7 (86%)

Supplementary Table 1 – Primers for qPCR Analysis

All primers were designed using the MacVector software and verified for specificity across the genome of interest by using the In Silico PCR function of the UCSC Genome Browser website (http://genome.ucsc.edu/).

Supplementary Table 1 - Primers for qPCR analysis

Human cDNA	Forward primer	Reverse Primer
ATM	CAGGCGAAAAGAATCTGGGG	GCACAAAGTAGGGTGGGAAAGC
ATR	TGAAAGGGCATTCCAAAGCG	CAATAGATAACGGCAGTCCTGTCAC
BRCA1	GCAGAGAGTCAGACCCTTCAATGG	GCCCAGGTTTCAAGTTTCCTTTTC
CCNG1	GACAAGCCTGAGAAGGTAAACTGTG	GTGTGACTCCTCCAATAGCATTTTC
CDKN1A	ATGGTGGCAGTAGAGGCTATGGAC	TGGAAGGTGTTTGGGGTCAGAC
CHEK1	CAGGTCTTTCCTTATGGGATACCAG	TGGGGTGCCAAGTAACTGACTATTC
CHEK2	GCTATTGGTTCAGCAAGAGAGGC	TCAGGCGTTTATTCCCCACC
CUX1	TGAACGACCCCAACAATGTGG	GGCTTTTGCTGATACGCTCG
FANCD2	CGACTGAAACAGGGAGAACACAGC	GCACACTGGAAACTGGGAATGC
Glis1	AAGATGTGGTGTCCAGCGGC	GGGCTCCTTCAGGTGTCTGTGTAG
HPRT1	AACACTTCGTGGGGTCCTTTTC	CTTTGCTTTCCTTGGTCAGGC
MDM2	CCATTGAACCTTGTGTGATTTG	GGCAGGGCTTATTCCTTTTC
NBS1	AATGGAACAGTGAGGAATGGAGG	CAAGATTTGGAAGGTGAGAGTGATG
RAD17	TAGACCCTGACAGCGGAGATGAAG	CACTGGCACTATTCTGACTCAAAGG
RPA1	TTTCAGCCCAGTAACCAGTCTTTC	TCCCAATCACATCCACGGAG
RPA2	GCCTGTTTTCATTTCCCACTTTGG	TTCTTTTCCTCTGCCCCTGGAG
RPA3	AAATCTGCCCCATAGACACCCG	TTTTCCAGCCTCCCTACGAAGC
SMC1L1	GGAAACCCTTAGCAAACCTGCC	GCCCTTGAACCAGACACTATCCAC
TOPBP1	GTGCTTCATCGCTCCTACCTTG	TGCTCCCTCAACAATGCCAG
TP53	CTACCTCCCGCCATAAAAAACTC	CCCACAACAAAACACCAGTGC
TP53BP1	TGGCAACCCCGTGAAAATC	CCACCACATCAAATACCCCTAAAG
Wnt8b	TCTTCTCCCTTTCCTTTCCCAG	GGTGTTCAAGTTAGCACTCCAGAGC
Wnt8a	GAACAGCCACAACACATCCAGG	TGACCTCAGTTTTCCTCTCTTCCAC
Wnt7b	TTGAGGCACTTGGCTGCTCC	GGTAAAAGGACCAGGACAGGGC
Wnt7a	TTCACCTACGCCATCATTGCCG	TTCTCTTTGTCGCAGCCACAGTCG
Wnt2b	CGCTCTGGAGATTTGAAGGGAG	GCACACCAAGAAGTATCGGGAAG
Wnt2	TGCTGACAGAGAGGTTTCCAGAGC	CTTGAAGGGTGATGGGCATTAG
Wnt16	GCAAGTAACCACTCCATCCAGC	CAGGTAAGGGAAACTCTCAGGTCC
Wnt11	GGACATCAAAGGAAACCGACAAG	GCCCTGAAAGGTCAAGTCTGTATC
GAPDH	TCCACCACCCTGTTGCTGTA	ACCACAGTCCATGCCATCAC
ACTB	GAAGTCCCTTGCCATCCTAAAAG	ATGCTATCACCTCCCCTGTGTG
UBC	CAAGACAAGGAAGGCATCCCTC	ATCCCACCTCTAAGACGGAGCAC
Mouse cDNA	**Forward primer**	**Reverse Primer**
ATM	CAGCAGCACCTCTGATTCTTACAAC	ACCTTAGCCTTAGGACCTGACTGG
ATR	AGGACACTCCAAAGCACCACTG	GCAGCCCTGTTACTCTATTTCGG
BRCA1	CTGATGACCTGCTGGATGATGTTG	TGGAGTCGCTCTCTTCTGACGATT
CCNG1	AAGTCTTTCTGCCACTCTGACCTG	CCTTGATTTGAATGCTTCCTGG
CDKN1A	CAGTGTTGAATACCGTGGGTGTC	CGTGAGACGCTTACAATCTGAGTG
CHEK1	AACAGGGCTTTCCTTGTGGG	GGCATTGGTAAGATTTGTCCGC
CHEK2	CAGGAAAAGAACTCGGTGACCC	ATTGGAGCGTGGAAGGAAGC
CUX1	CAGCGCTTATTTGGGGAGACC	TGGAACCAGTTGATGACGGTG
FANCD2	GGACCTTGTAGAACTTCAGGGAATC	CAGGATGCTTTGTTGCTATCAGC
HPRT1	TTGGGCTTACCTCACTGCTTTC	ATCGCTAATCACGACGCTGG
MDM2	TGGAACCCCCAGGAAGAGTG	AACCACCAGGAACCACGGAG
NBS1	CAAGTGAAGGTTGAAAAGCAGGAG	TTGACCACGACTCTGTCTGAAGTTC
RAD17	GGGACACTTTACTTGTTCAGCCTG	TGTATTCCAATCACCACCGAGG
RPA1	GAGTGTGCTTTTCCAGTGGTGG	AACAGACCCTTTGATTCCAGGC
RPA2	AACTCTGGTCATCAGGCTTATTGG	GCAGGCAAGTGCTTTACTACTGTG
RPA3	GCTGGAAAAGATTCATCCCAC	ACAGAGCAGGGTCGCCTTG
SMC1L1	AGTGTGCTTCAGCGGATTGC	TCTGTTCAAATGCCTGCTTGG
TOPBP1	ATGCTGGAATCCCCTCCTTG	GTGGTTGTGTATGGCTCACTTCAG
TRP53	TGAGGTTCGTGTTTGTGCCTG	GGGTGAAATACTCTCCATCAAGTGG
TRP53BP1	GGGTTTTCTCATTTGGTGGTGAC	GTTGGGTGGGTTCATTTGTGTAAC
Wnt1	CCAGCAGCAAAACCCTACATTCTC	TACACAGTGATGAGGAGGCAGGAC
Wnt2	CAGATTCCAACAACCCAGAAAGTC	TCACCAAGGATGCTATCAACAGC
Wnt3	CAGAAAGAGGAGATAAAATGGGGG	CACAACAGGAAAGAAGTGGCTGAC
Wnt3a	TGCTGTTGAGGCAATGGTCAC	TTGTGGCAGATGGGCTGTATG
Wnt5b	AAAGACGGTTCTGTCACCTGCTAC	CCTGTCATTCTCAAAGCCTCCC
Wnt6	CTCCTGGAAACCTTTAGCACACC	AGTTCACACAGACGCCTGACAAC
Wnt7a	CAGGGGTTATTTCCAGGTATCTGC	CAGGCATCTGAGTTTCACCAGC
Wnt7b	AGATTTACTGGAGACCCCACGG	CACAGAAGAAGGACAAAACCCAAG

Wnt9a	GCCCACTAATGAATGCCACG	TCCCCAGAACTGAAAGAGATGAAG
Wnt10b	GCTTCTCCTTCCGTTCAGTTGTTG	TATCCATTCCCACCCTTCCTGC
Wnt11	CAATCTGCCCCCAAATCTCTG	TCTGTCCTCACCCTTGACCAAC
GAPDH	TCCACCACCCTGTTGCTGTA	ACCACAGTCCATGCCATCAC
ACTB	TATTGGCAACGAGCGGTTCC	GGCATAGAGGTCTTTACGGATGTC
UBC	CGCACCCTGTCAGACTACAACATC	CCCAAGAACAAGCACAAGGAGG
Human genomic	**Forward primer**	**Reverse Primer**
ATM	TGGGCTCTGGAATCATACGGC	TAACGCTCACGAGTGCTCACCAC
ATR	ACTCTGCTGTTCTCAAGCCTGG	TCGCTGGACTGTCATAGTTCTCAC
BRCA1	CAGAAAGAGCCAAGCGTCTC	CAATAAGCCGCAACTGGAAG
CCNG1	GATTGGCCGGACTTCTCAC	GGGACTCGTAGGCAAGAGGA
CDKN1A	CAGACAACTCACTCGTCAAATCCTC	AATCTCCCTACACCCTACACTCACC
CHEK1	GCCGCCGACATTCAGA	CATGCCCTCCCTCACTAATC
CHEK2	CCCCTGAGACTGAGGTTCTTGG	CCACCCTAACTTTTCCACGGC
FANCD2	GTGGAGCAATGGTCGTAGTCTCTC	ATGAGGAAGCCAAGGTTCGG
MDM2	GGTGCCTGTCGGGTCA	ACTGCAGTTTCGGAACGTGT
NBS1	ACCTGGTGGTTGGAAAAGGAAC	CAAACGCACGAAACTACATAACTGG
RAD17	GAGAAACGACCCGAAATGCTC	TGGTGGATGCCTCACTCCTTAG
RPA1	CCATCCTTACTTGACCCTTGTCTG	CATTGGAGAGTTGAAATAGCCTGG
RPA2	CGGCGTGCTCAGGTTC	TCATAGGCAAGAGGGCGTAG
RPA3	GGCAAGAGGGAAGGCGAGACA	GGCGGGAGTCGGCACT
SMC1L1	CAACAAACACTTGCTCCTGCG	GCATAATCCTGCTGTGACTTCTGTC
TOPBP1	GCAGCAGAAGAGCGGAAATGTC	GCAAAGCCCTAAACCCAAAGG
TP53	CAGCCCGAACGCAAAG	CTTGTCATGGCGACTGTCC
TP53BP1	GCCCGCCACTCAAGAAATCC	TTCACGCCCTCTCAAGGTCC
GAPDH	TTTTCCCTCTTCTTGACTCACCC	GTGCCTTTCATTCCATCCAGC
ACTB	AAGGCAACTTTCGGAACGGC	CCAAAACTCTCCCTCCTCCTCTTC
UBC	AAAAGAGGCGGAAACCCCAC	CTCCCTGTTGGCATCAAGTAGG

Supplementary Table 2 – Public Datasets Used for Meta-Analysis

Table shows all datasets obtained from Oncomine with sample number (N). "Y" indicates a gene is significantly upregulated in the top 25% samples with highest Wnt expression compared to the lowest 25% in the corresponding dataset. "N" indicates it is not significantly upregulated. Significance implies $P < 0.05$ on a Welch corrected Student's T test.

Supplementary Table 2 – Public datasets used for Meta-Analysis

Dataset information		Genes with significant change				
Dataset name	Sample number	CUX1	GLI1	GLI2	GLI3	GLIS1
Bonnefoi_Breast	161	-	Y	N	N	Y
Esserman_Breast	120	Y	Y	Y	N	Y
Gluck_Breast	149	Y	Y	Y	N	Y
Hatzis_Breast	508	-	Y	Y	Y	-
Kao_Breast	327	-	Y	N	Y	Y
Lu_Breast	130	-	N	N	N	N
Pawitan_Breast	159	-	N	N	Y	Y
TCGA_Breast	529	Y	Y	Y	N	Y
vandeVijver_Breast	295	-	Y	Y	N	Y
Bhattacharjee_Lung	186	-	N	Y	Y	-
Bild_Lung	111	Y	N	N	N	Y
Bittner_Lung	101	-	Y	N	N	Y
Director's Challenge_Lung	443	-	Y	Y	Y	-
Hou_Lung	92	-	N	Y	Y	Y
Kim_Lung	138	-	N	Y	N	Y
Raponi_Lung	130	-	Y	N	Y	-
TCGA_Lung	165	Y	Y	Y	Y	Y
Zhu_Lung	90	-	Y	Y	Y	-

FIGURES

Figure 1 - CUX1 Isoforms Exhibit Distinct DNA Binding and Transcriptional Properties

The p150, p110, p90 and the p80 isoforms are generated by proteolytic processing, while the p75 isoform is the product of a mRNA that is initiated at a transcription start site within intron 20. Proteolytic processing by cathepsin L takes place at the end of the G1 phase in normal cells, but is constitutive in many transformed cells. Proteolytic processing by a caspase is observed in proliferating cells and is not associated with apoptosis. The DNA binding properties are determined by the combination of DNA binding domains present within the isoform. While the full-length isoform only represses transcription, the shorter isoforms p110, p90, p80 and p75 can function as transcriptional repressors or activators depending on the promoter.

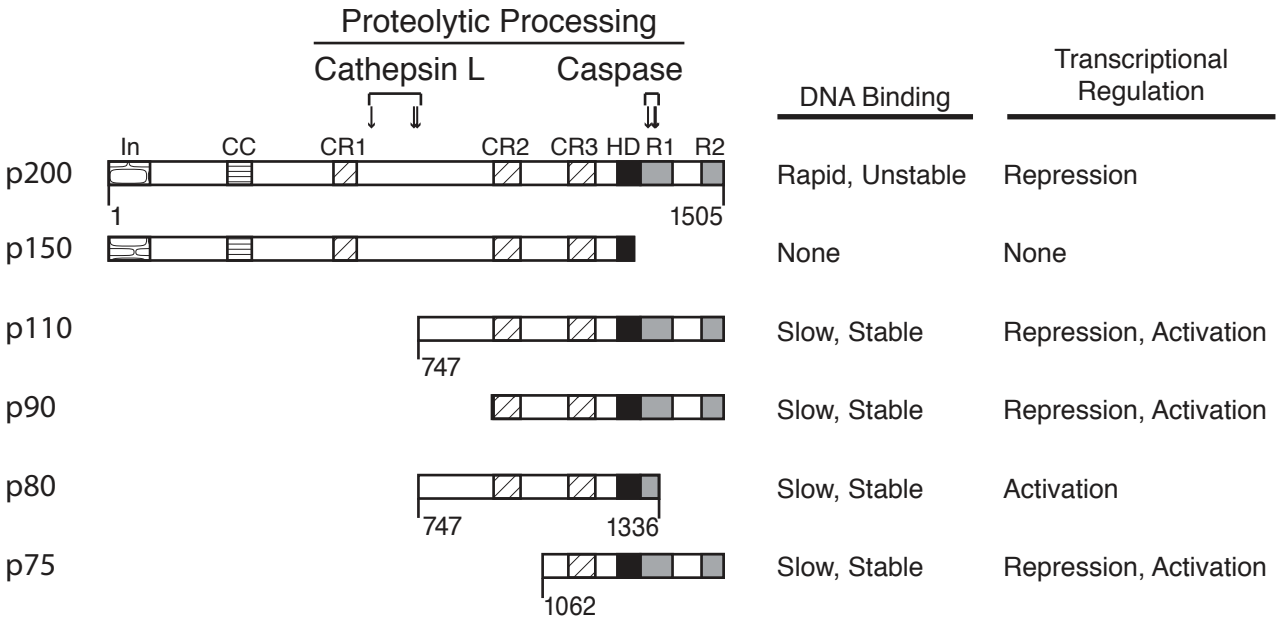

Proteolytic Processing
Cathepsin L
Caspase
DNA Binding
Transcriptional Regulation
In
CC
CR1
CR2
CR3
HD
R1
R2
p200
1
1505
Rapid, Unstable
Repression
p150
None
None
p110
747
Slow, Stable
Repression, Activation
p90
Slow, Stable
Repression, Activation
p80
747
1336
Slow, Stable
Activation
p75
1062
Slow, Stable
Repression, Activation

Figure 1

Figure 2 – Expression of CUX1 Recombinant Protein

A Schematic representation of CUX1 proteins with some of the functional domains: ID, inhibitory domain; CC, coiled-coil; CR1, CR2 and CR3, Cut repeat 1, 2 and 3; HD homeodomain; CBD, calmodulin binding domain; Prot A, protein A. The regions recognized by the 861 and 1300 antibodies are shown.

B Hs578t cells were infected with a retroviral vector to establish a population of cells stably expressing a recombinant p110 CUX1 protein with two tags at its C-terminus, p110 CUX1-Tag^2. A population stably carrying the empty vector was used as a control. Nuclear extracts were prepared from each population of cells and analyzed by Western blot using the 861 and 1300 CUX1 antibodies.

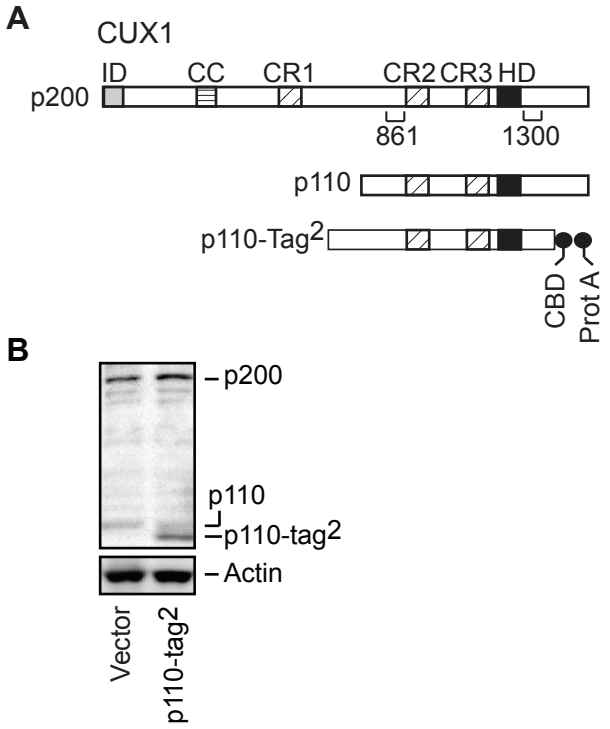

A
CUX1
ID
CC
CR1
CR2
CR3
HD
p200
861
1300
p110
p110-Tag2
CBD
Prot A
B
p200
p110
p110-tag2
Actin
Vector
p110-tag2

Figure 2

Figure 3 – Distribution of CUX1, C-Myc and E2F1 Binding Sites Relative to Transcription Start Sites

A Percentage of CUX1 binding sites located at various distances from the closest transcription start site. The "0" column indicates genes where the CUX1 binding site overlaps the start site.

B Location of C-Myc binding sites as per A.

C Location of E2F1 binding sites as per A.

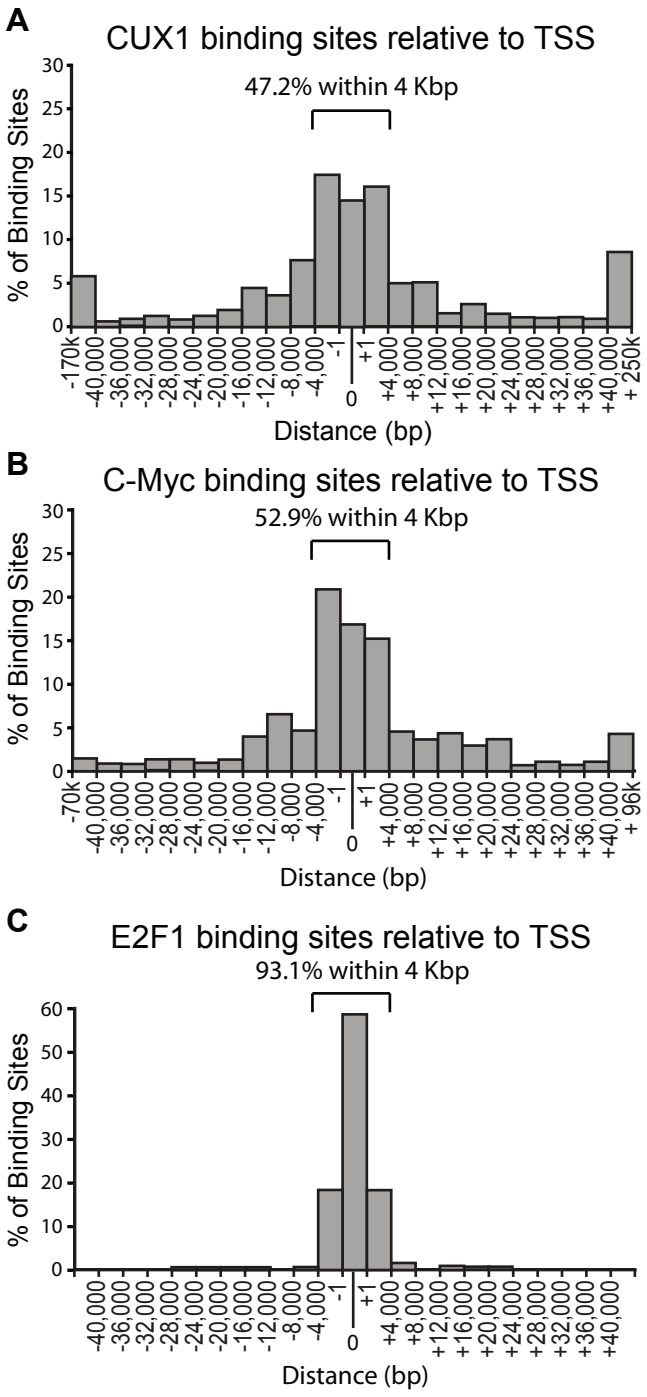

A
CUX1 binding sites relative to TSS
47.2% within 4 Kbp
% of Binding Sites
Distance (bp)
B
C-Myc binding sites relative to TSS
52.9% within 4 Kbp
% of Binding Sites
Distance (bp)
C
E2F1 binding sites relative to TSS
93.1% within 4 Kbp
% of Binding Sites
Distance (bp)

Figure 3

Figure 4 – Overexpression and Knockdown of CUX1 and Array Validation

A Hs578t cells were infected with a lentiviral vector expressing p110 CUX1 or nothing (vector). RNA and proteins were purified 48 hours post-infection. mRNA expression was measured by quantitative real-time PCR using primer pairs specific for CUX1 and for each target. Expected up or down indicates regulation that was observed by expression profiling.

B Hs578t cells were infected with a lentiviral vector expressing CUX1 shRNA or a scrambled RNA. RNA and proteins were purified 5 days after infection. (B) CUX1 expression was analyzed by and RT-qPCR and immunoblotting.

C RNA levels of the indicated genes were measured by RT-qPCR in cells treated as in **A**. Expected up or down indicates regulation that was observed by expression profiling.

D RNA levels of the indicated genes were measured by RT-qPCR in cells treated as in **B**. Expected up or down indicates regulation that was observed by expression profiling. * $p<0.05$, *** $p<0.001$ on a Student's T test.

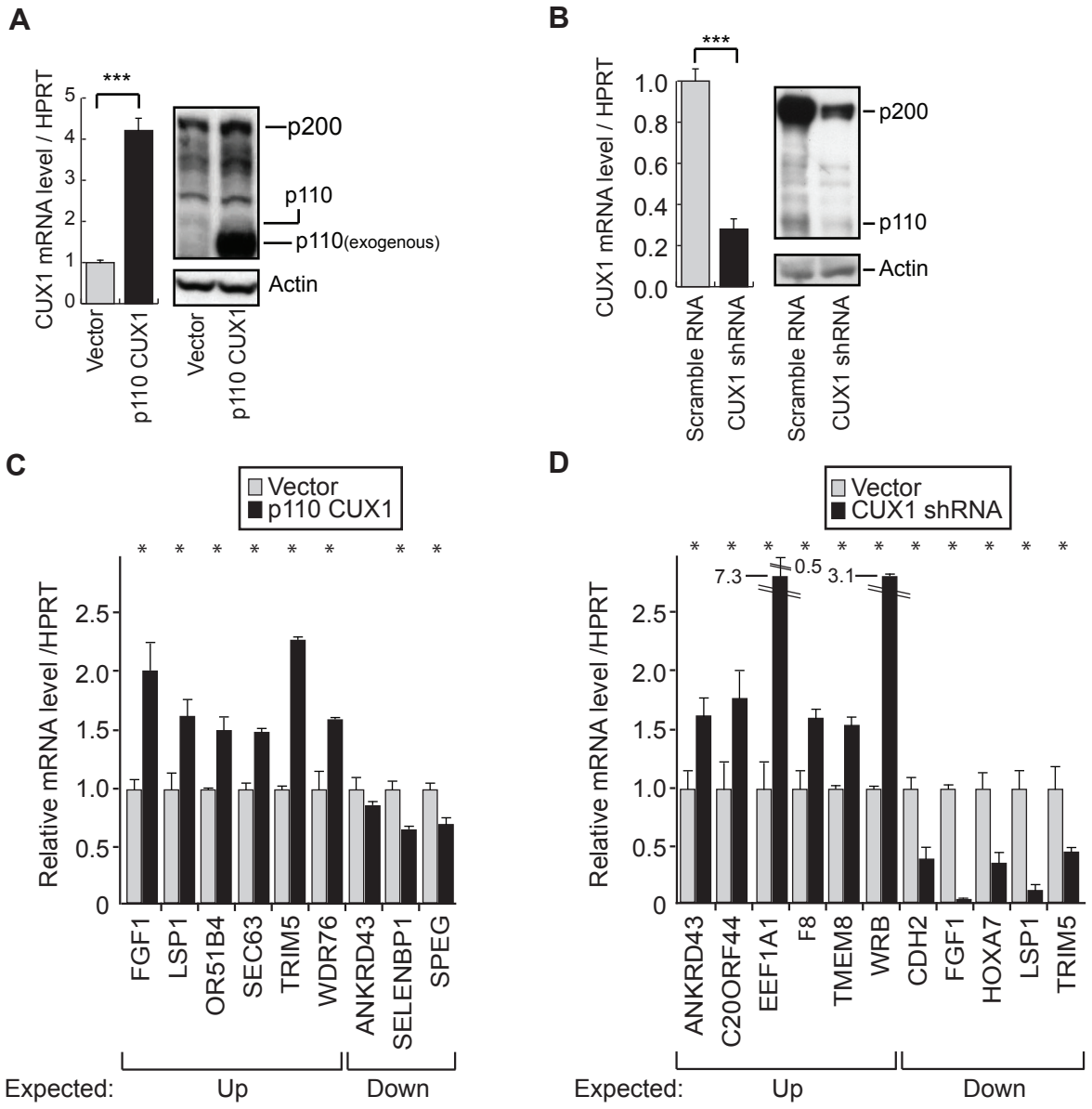

A
CUX1 mRNA level / HPRT
***
Vector
p110 CUX1
p200
p110
p110(exogenous)
Actin
B
CUX1 mRNA level / HPRT
***
Scramble RNA
CUX1 shRNA
p200
p110
Actin
C
Vector
p110 CUX1
Relative mRNA level /HPRT
FGF1
LSP1
OR51B4
SEC63
TRIM5
WDR76
ANKRD43
SELENBP1
SPEG
Expected:
Up
Down
D
Vector
CUX1 shRNA
7.3
0.5
3.1
Relative mRNA level /HPRT
ANKRD43
C20ORF44
EEF1A1
F8
TMEM8
WRB
CDH2
FGF1
HOXA7
LSP1
TRIM5
Expected:
Up
Down

Figure 4

Figure 5 – Effect of Distance on Regulation by CUX1

A Genes from the ENCODE array have been organized according to the distance between their transcription start site (TSS) and the closest CUX1 binding site. The "0" column indicates genes where the CUX1 binding site overlaps the start site. The histogram shows, for each interval of distance, the percentage of genes that exhibit a 1.5 fold change in expression following p110 CUX1 overexpression or CUX1 knockdown. The total number of genes within each interval is indicated within each column.

B As in A, except that the regulation by CUX1 is expressed as either activation by CUX1 or repression by CUX1

C As in A , but with a threshold of 1.25 fold change in expression.

D As in B, but with a threshold of 1.25 fold change in expression.

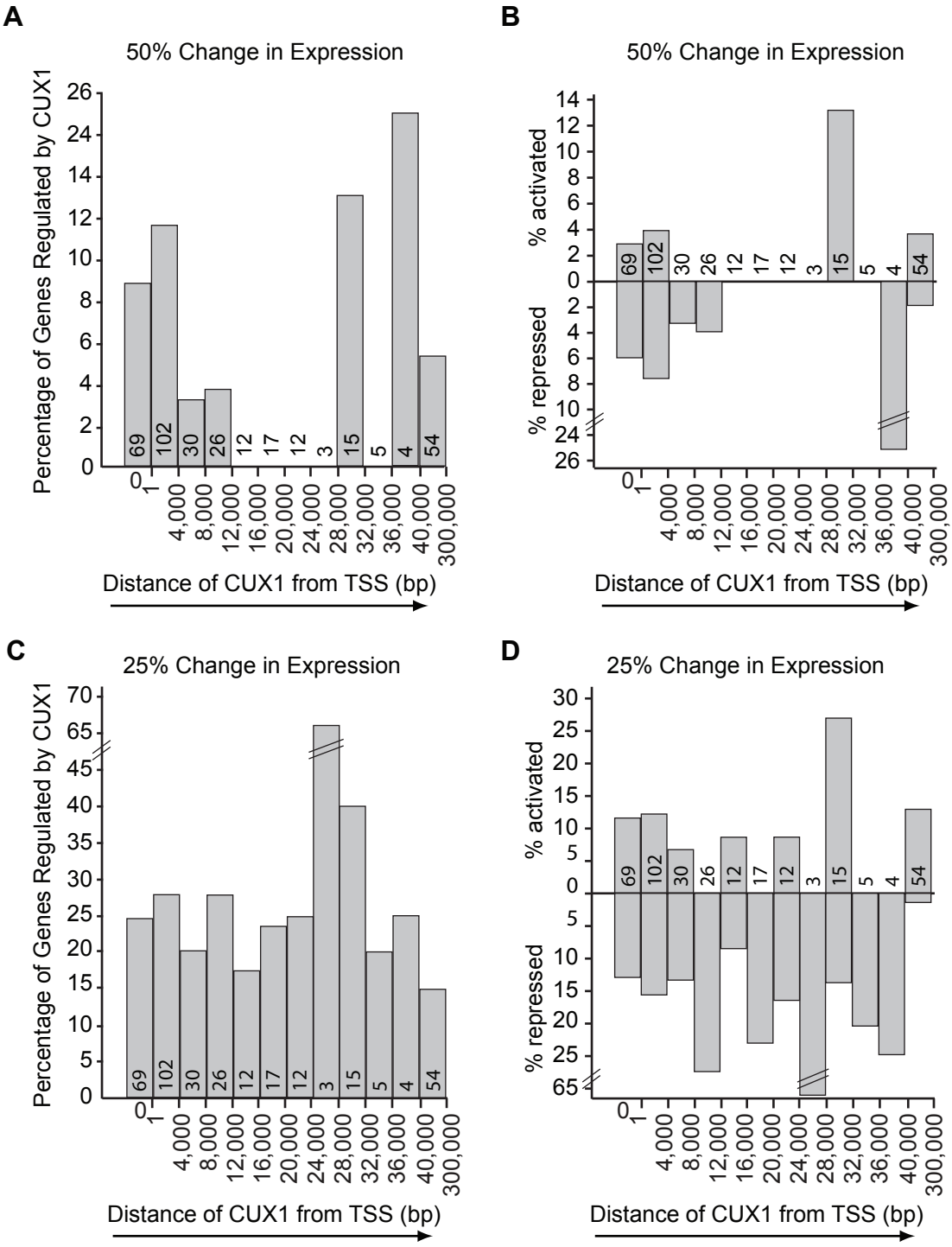
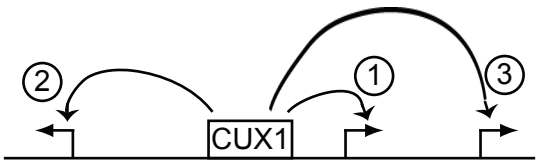

A
50% Change in Expression
Percentage of Genes Regulated by CUX1
Distance of CUX1 from TSS (bp)
B
50% Change in Expression
% activated
% repressed
Distance of CUX1 from TSS (bp)
C
25% Change in Expression
Percentage of Genes Regulated by CUX1
Distance of CUX1 from TSS (bp)
D
25% Change in Expression
% activated
% repressed
Distance of CUX1 from TSS (bp)
69
102
30
26
12
17
12
3
15
5
4
54
0
1
4,000
8,000
12,000
16,000
20,000
24,000
28,000
32,000
36,000
40,000
300,000

Figure 5

Figure 6 – Relationship Between Gene Position and Regulation by CUX1. Three types of situations are depicted in the diagram. 1, genes that are the closest to the CUX1 binding site; 2, genes that are further away and in the other direction from the CUX1 binding site; 3, genes that are located further away and are separated by another gene from the CUX1 binding site. For each category, the table shows the percentage of genes that exhibit a 1.25 or 1.5 change in expression following p110 CUX1 overexpression or CUX1 knockdown.



	①	②	③
1.25 fold	24.8%	21.1%	24.7%
1.5 fold	7.8%	7.9%	5.4%

Figure 6

Figure 7 – Transcriptional Regulation of DDR Genes by CUX1.

A MCF7 cells were transfected with CUX1-specific siRNA. Top panel: mRNA and protein levels of CUX1 are shown following knockdown. Bottom panel: mRNA levels of DDR gene targets are shown. All mRNA levels are normalized to HPRT1. The values are the mean of three measurements and error bars represent standard deviation. * $p < 0.05$ on a student's T Test.

B Mouse Embryonic Fibroblasts (MEFs) were obtained from *Cux1* knockout embryos ($cux1^{Z/Z}$) and wild-type littermates. Levels of CUX1 and DDR gene targets are shown as in A; * in top panel corresponds to non-specific band recognized by CUX1 antibody.

C Hs578T cells were infected with a retrovirus expressing p110 CUX1 or with the empty vector. RNA was prepared 24h post-infection and levels of DDR target genes was measured by real-time PCR and normalized to HPRT1. The values are the mean of three measurements and error bars represent standard deviation.

D The promoter regions of target genes were cloned into a luciferase reporter plasmid. Hs578T cells were transfected with each reporter plasmid along with a vector expressing p110 CUX1 or with an empty vector. The values are the mean of three measurements and error bars represent standard deviation.

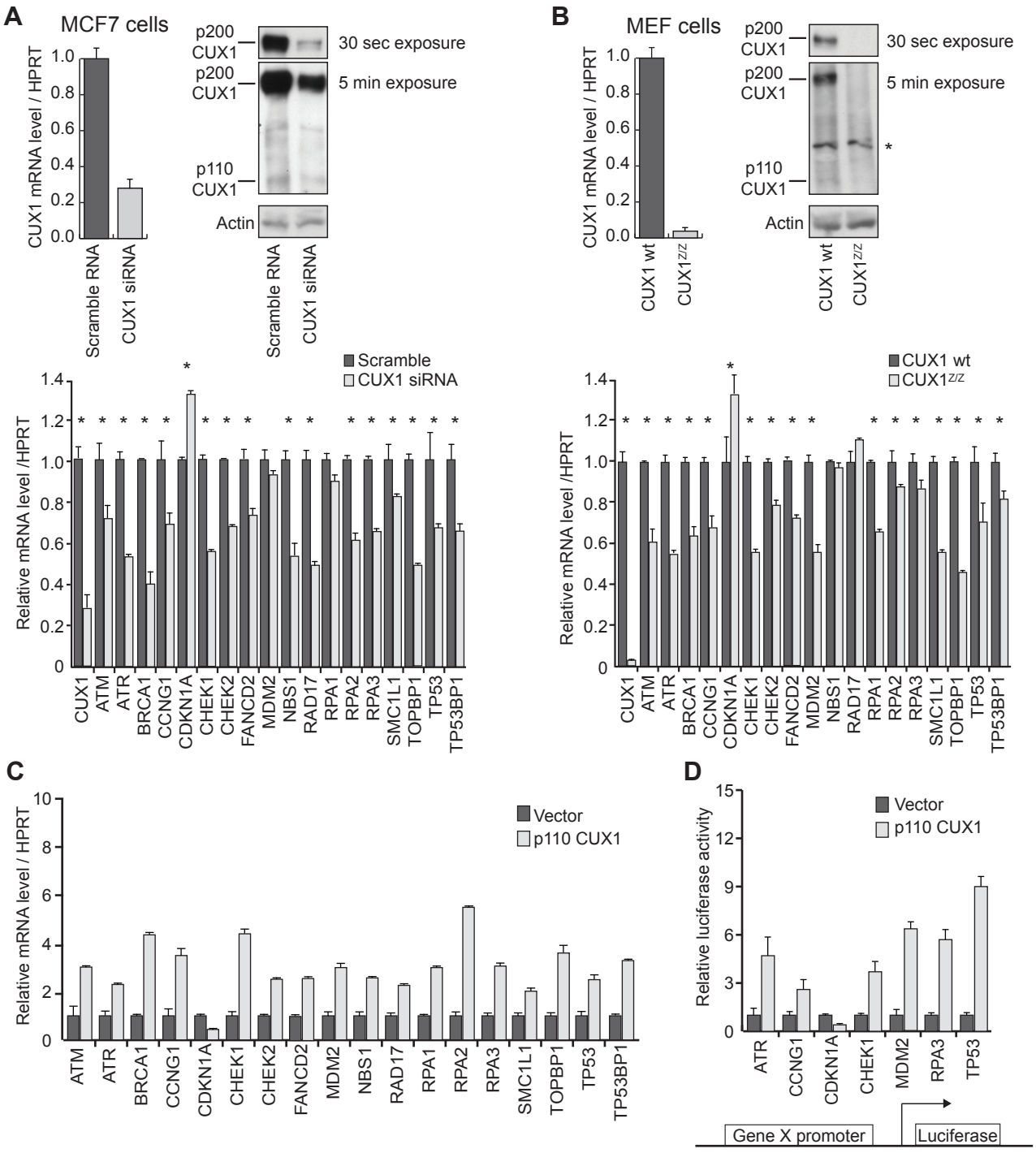

A
MCF7 cells
CUX1 mRNA level / HPRT
Scramble RNA
CUX1 siRNA
p200 CUX1
p110 CUX1
30 sec exposure
5 min exposure
Actin
Scramble
CUX1 siRNA
Relative mRNA level /HPRT
CUX1
ATM
ATR
BRCA1
CCNG1
CDKN1A
CHEK1
CHEK2
FANCD2
MDM2
NBS1
RAD17
RPA1
RPA2
RPA3
SMC1L1
TOPBP1
TP53
TP53BP1
B
MEF cells
CUX1 wt
CUX1Z/Z
C
Vector
p110 CUX1
Relative mRNA level / HPRT
D
Relative luciferase activity
Gene X promoter
Luciferase

Figure 7

Figure 8 – Effect of CUX1 Knockdown on DDR Signalling Kinases, Partners and Substrates.

A, **B**, **C** and **D** MCF7 Cells were transfected with CUX1-specific siRNA. Nuclear extract were prepared and analyzed by immunoblotting.

A Cells were exposed to 10 Gy of IR, incubated for 1h prior to harvest and the extracts were immunoblotted for CUX1 to assess knockdown.

B Cells were treated as in A and immunoblotted for ATM, p-ATM, 53BP1, Chk2 and p-Chk2. Actin was used to control for equal loading.

C Cells were exposed to 20 Js of UV, incubated for 2h prior to harvest and the extracts were immunoblotted for ATR, Chk1 and p-Chk2.

D Cells were exposed to 10 Gy of IR, incubated for 6h prior to harvest and the extracts were immunoblotted for p53 and CUX1. Actin was used to control for equal loading.

E Nuclear extracts from *Cux1*$^{Z/Z}$ and wild-type MEFs were immunoblotted for ATM or ATR following exposure to 10 Gy of IR (left panel) or 20Js of UV (right panel), respectively.

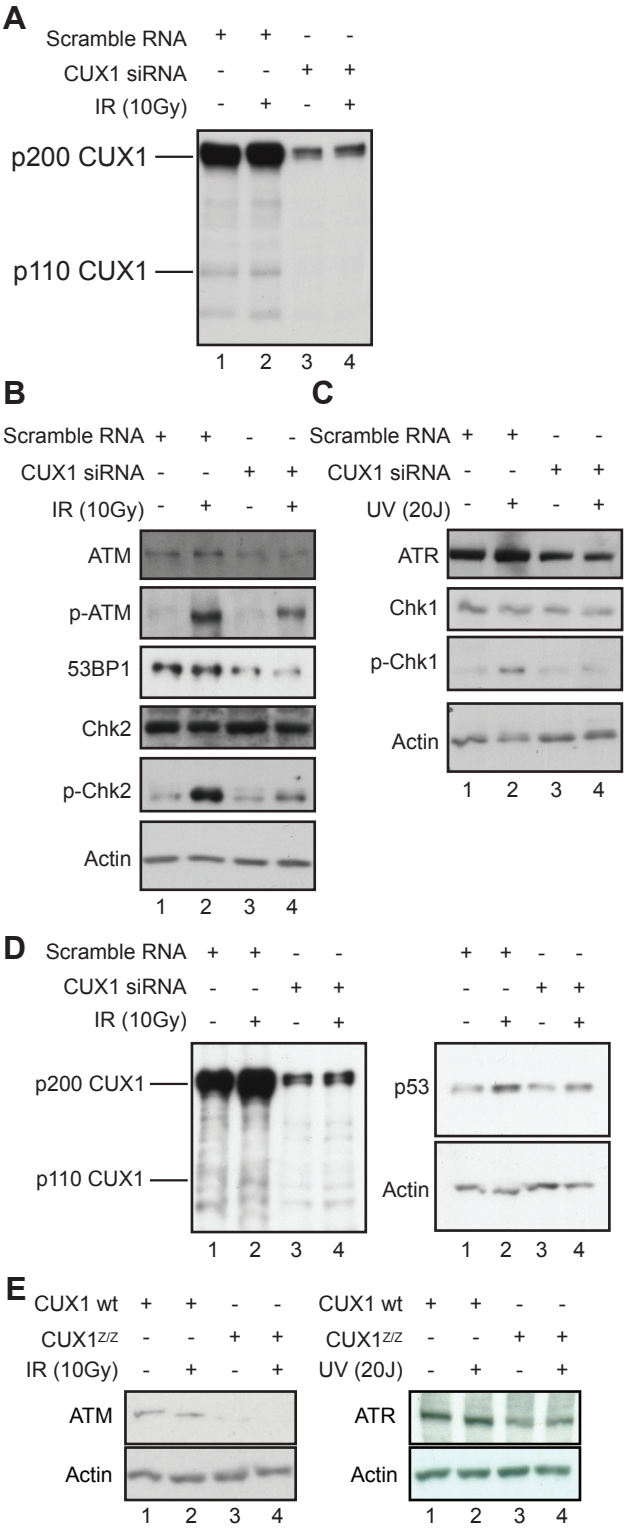

A
Scramble RNA + + - -
CUX1 siRNA - - + +
IR (10Gy) - + - +
p200 CUX1
p110 CUX1
1 2 3 4
B
Scramble RNA + + - -
CUX1 siRNA - - + +
IR (10Gy) - + - +
ATM
p-ATM
53BP1
Chk2
p-Chk2
Actin
1 2 3 4
C
Scramble RNA + + - -
CUX1 siRNA - - + +
UV (20J) - + - +
ATR
Chk1
p-Chk1
Actin
1 2 3 4
D
Scramble RNA + + - -
CUX1 siRNA - - + +
IR (10Gy) - + - +
p200 CUX1
p110 CUX1
1 2 3 4
+ + - -
- - + +
- + - +
p53
Actin
1 2 3 4
E
CUX1 wt + + - -
CUX1Z/Z - - + +
IR (10Gy) - + - +
ATM
Actin
1 2 3 4
CUX1 wt + + - -
CUX1Z/Z - - + +
UV (20J) - + - +
ATR
Actin
1 2 3 4

Figure 8

Figure 9 – γ-H2AX Staining in Cells After DNA damage.

A MCF7 cells were transfected with CUX1-specific siRNA. Cells were fixed and stained by immunofluorescence for γ-H2AX. The proportion of γ-H2AX positive cells was counted after treatment with UV. For cells treated with IR, the number of γ-H2AX foci per nuclei was counted. Representative images from experiments at 4h (UV) and 1h (IR) are shown next to each graph. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; Fisher's exact test for UV treatment or a student's T test for IR treatment. **B** MEF cells from *Cux1*$^{Z/Z}$ knockout embryos and wild-type littermates were exposed to DNA damage. Cells were treated and counted as in A. Representative images from experiments at 4h (UV) and 1h (IR) are shown next to each graph.

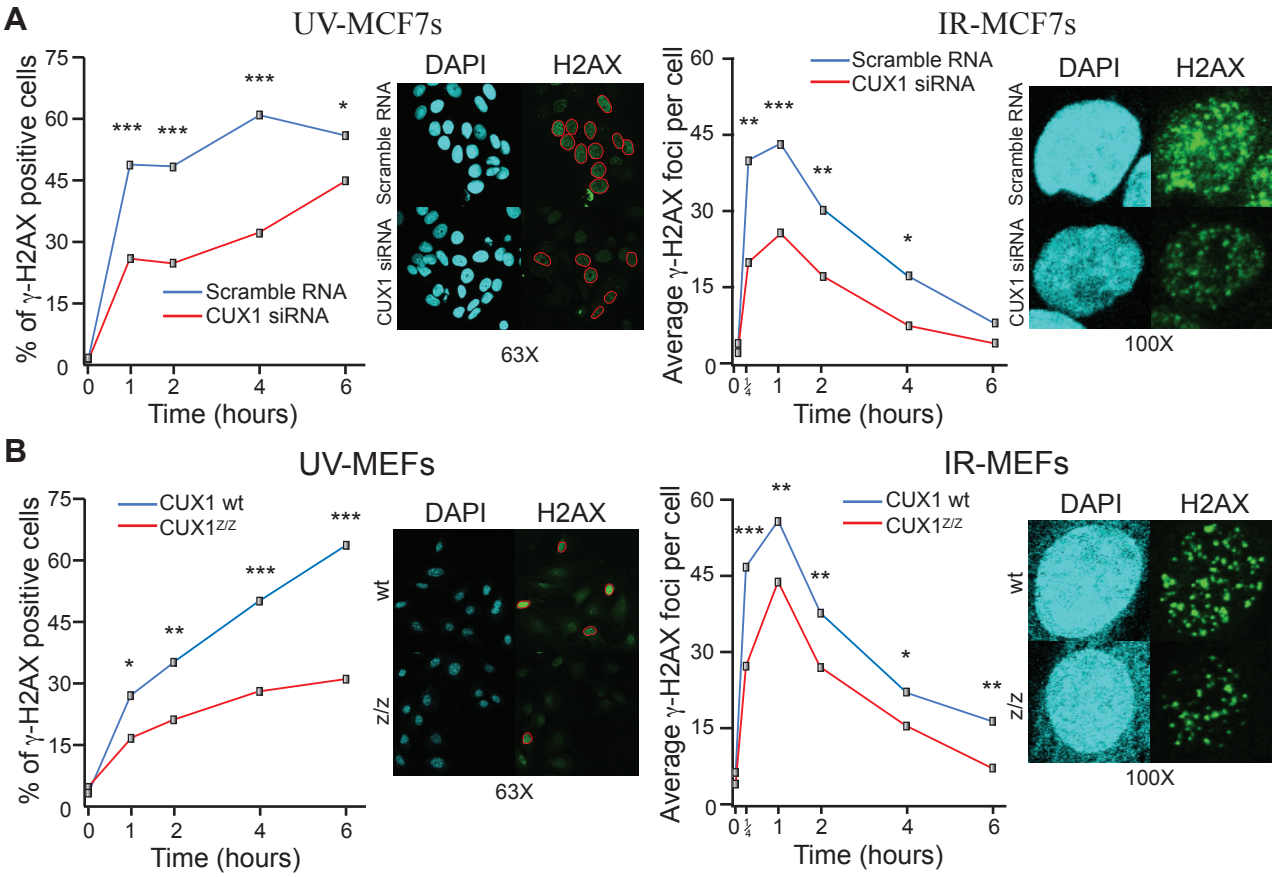

A
UV-MCF7s
IR-MCF7s
% of γ-H2AX positive cells
Scramble RNA
CUX1 siRNA
Time (hours)
DAPI
H2AX
63X
Average γ-H2AX foci per cell
100X
B
UV-MEFs
IR-MEFs
CUX1 wt
CUX1Z/Z
wt
z/z

Figure 9

Figure 10 – Effect of CUX1 Knockdown on Survival Following DNA damage.

A MCF7 cells were transfected with CUX1-specific siRNA prior to exposure to DNA damage. 500 cells were plated in triplicate and incubated for 10 days. Clones were fixed, stained and counted and the cloning efficiency of unexposed cells was set to 1. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; student's T test.

B MEF cells from *Cux1*$^{Z/Z}$ knockout embryos and wild-type littermates were exposed to DNA damage. 5000 cells were plated in triplicate and incubated for 10 days. Clones were fixed, stained and counted and cloning efficiency of unexposed cells was set to 1.

C Representative images from experiments with 1Gy IR in MEF cells (left) and 5J UV in MCF7 cells (right).

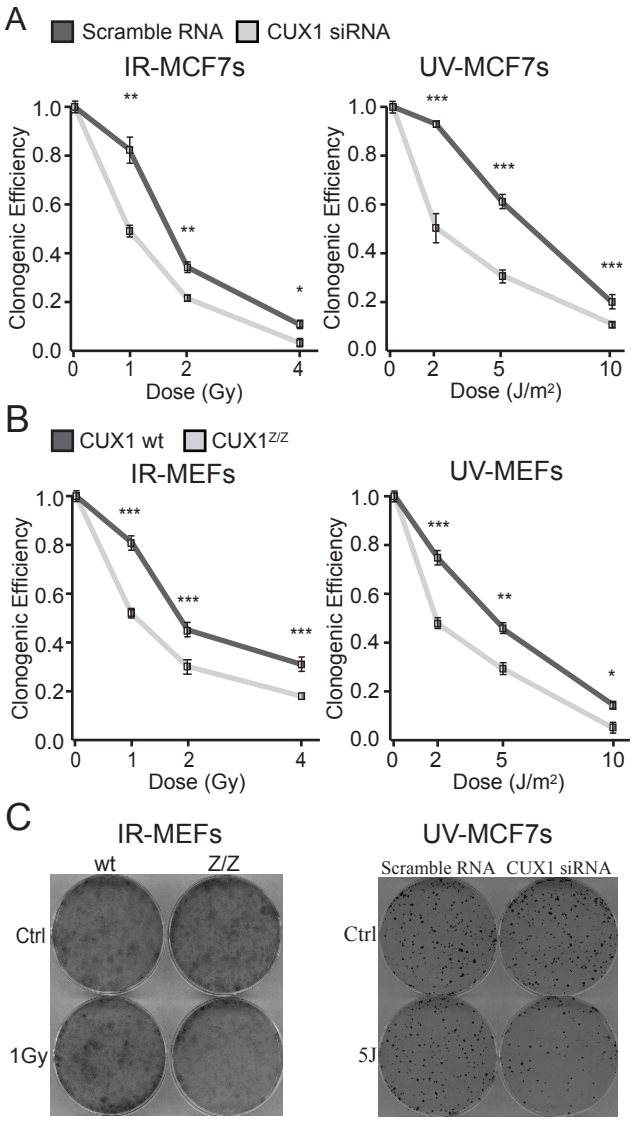

A
Scramble RNA
CUX1 siRNA
IR-MCF7s
UV-MCF7s
Clonogenic Efficiency
Dose (Gy)
Dose (J/m2)
B
CUX1 wt
CUX1Z/Z
IR-MEFs
UV-MEFs
C
IR-MEFs
UV-MCF7s
wt
Z/Z
Scramble RNA
CUX1 siRNA
Ctrl
1Gy
5J

Figure 10

Figure 11 – Effect of CUX1 Knockdown on G1/S and S Phase Arrest Following DNA damage.

A MCF7 cells were transfected with CUX1-specific siRNA and exposed to either 1 μm Nocodazole, Nocodazole + 10 Gy of IR or Nocodazole + 20 Js of UV. Cells were fixed with ethanol 24 hours after exposure, stained with Propidium Iodine (PI) and analyzed for cell cycle distribution by flow cytometry.

B A histogram of the increase in G1 content after IR and UV in MCF7 cells. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; student's T test.

C A histogram of the increase in G1 content after IR and UV in $Cux1^{Z/Z}$ and $Cux1^{wt}$ MEF cells treated and analyzed as in A.

D $Cux1^{Z/Z}$ and *Cux1* wild-type MEF cells were exposed to 10 Gy IR. 1 to 4 hours post exposure, the cells were labeled with BrdU for 1h before fixation with 4% PFA. BrdU incorporation was measured by flow cytometry.

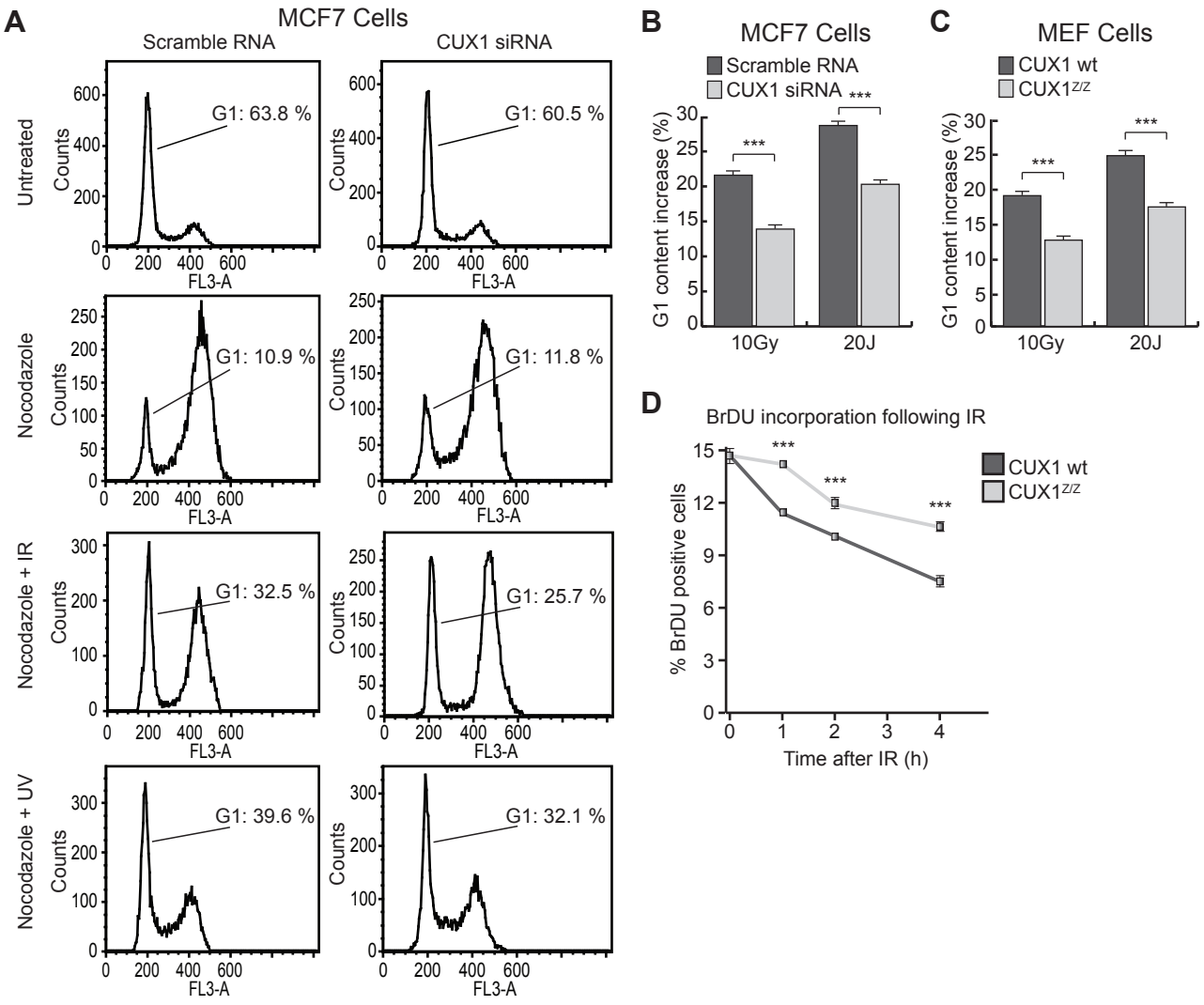

A
MCF7 Cells
Scramble RNA
CUX1 siRNA
Untreated
Nocodazole
Nocodazole + IR
Nocodazole + UV
Counts
FL3-A
G1: 63.8 %
G1: 60.5 %
G1: 10.9 %
G1: 11.8 %
G1: 32.5 %
G1: 25.7 %
G1: 39.6 %
G1: 32.1 %
B
MCF7 Cells
Scramble RNA
CUX1 siRNA
G1 content increase (%)
10Gy
20J
C
MEF Cells
CUX1 wt
CUX1Z/Z
D
BrDU incorporation following IR
% BrDU positive cells
Time after IR (h)

Figure 11

Figure 12 – Effect of CUX1 Knockdown on G2/M Arrest Following IR Exposure.

A *Cux1*$^{Z/Z}$ and *Cux1* wild-type MEF cells were exposed to 10 Gy IR. Cells were fixed with ethanol 24 hours after exposure, stained with Propidium Iodine (PI) and analyzed for cell cycle distribution by flow cytometry.

B A histogram of the increase in G2 content after IR in MEF cells. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; student's T test.

C A histogram of the increase in G2 content after IR in MCF7 cells treated with either CUX1 siRNA or scrambled siRNA.

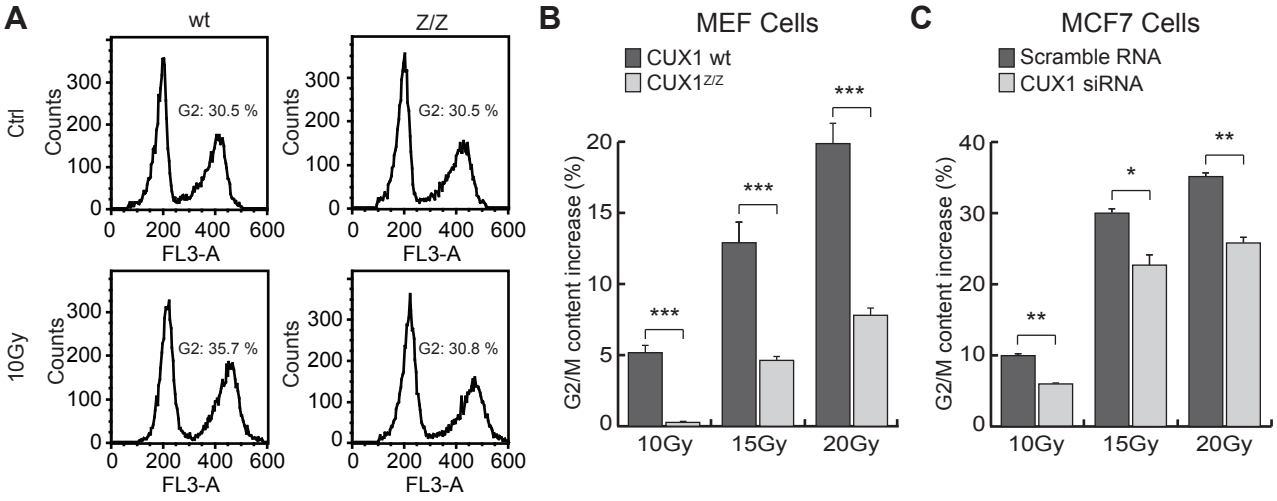

A
wt
Z/Z
Ctrl
10Gy
Counts
FL3-A
G2: 30.5 %
G2: 30.5 %
G2: 35.7 %
G2: 30.8 %
B
MEF Cells
CUX1 wt
CUX1Z/Z
G2/M content increase (%)
10Gy
15Gy
20Gy
C
MCF7 Cells
Scramble RNA
CUX1 siRNA

Figure 12

Figure 13 – Effect of CUX1 Knockdown on Rad51 Focus Formation and DNA Damage Repair.

A MCF7 cells were transfected with CUX1-specific siRNA before exposure to 5 Gy of IR. Cells were fixed 1hr after exposure and stained by immunofluorescence for Rad51 foci. The proportion of cells displaying 5 or more foci is shown. Non-irradiated cells are shown as controls. ***: P Value <0.001; student's T test.

B *Cux1*$^{Z/Z}$ and wild-type MEFs were irradiated or mock-irradiated with 5 Gy of IR. Cells were fixed 1hr later and stained by immunofluorescence for Rad51 foci. The proportion of cells displaying 5 or more foci is shown.

C MCF7 cells were transfected with CUX1-specific siRNA and then exposed to 2 Gy IR, 5J UV or 10mm H_2O_2 for 30 min. At the indicated times, cells were collected and strand breaks quantified by Alkaline Single Cell Gel Electrophoresis (Comet Assay). Comet tail moments were scored for at least 30 cells per conditions. Error bars represent standard error. * Indicates p Value <0.05, ** <0.01, *** <0.001; student's T test.

D *Cux1*$^{Z/Z}$ and wild-type MEFs were exposed to IR, UV or H_2O_2.and DNA breaks were quantified as in C. In addition, comet tail moments were measured in MEFs maintained at 3% and 20% oxygen with no further treatment.

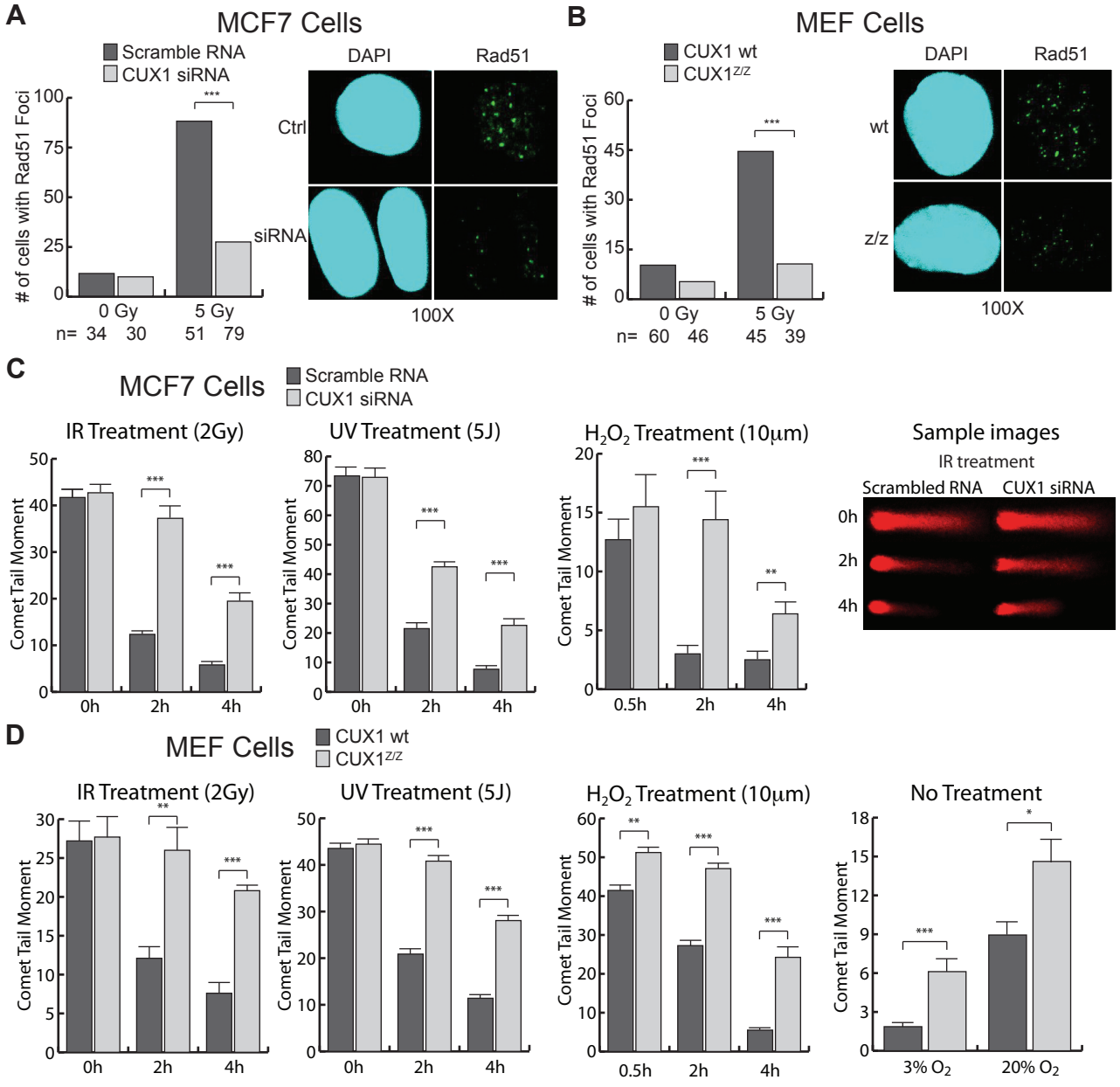

A
MCF7 Cells
Scramble RNA
CUX1 siRNA
# of cells with Rad51 Foci
0 Gy
5 Gy
n= 34 30 51 79
DAPI
Rad51
Ctrl
siRNA
100X
B
MEF Cells
CUX1 wt
CUX1Z/Z
# of cells with Rad51 Foci
0 Gy
5 Gy
n= 60 46 45 39
DAPI
Rad51
wt
z/z
100X
C
MCF7 Cells
Scramble RNA
CUX1 siRNA
IR Treatment (2Gy)
UV Treatment (5J)
H2O2 Treatment (10μm)
Sample images
IR treatment
Scrambled RNA
CUX1 siRNA
Comet Tail Moment
0h
2h
4h
0.5h
D
MEF Cells
CUX1 wt
CUX1Z/Z
IR Treatment (2Gy)
UV Treatment (5J)
H2O2 Treatment (10μm)
No Treatment
3% O2
20% O2

Figure 13

Figure 14 – Mouse Cells Lacking one Copy of CUX1 Suffer Increased Endogenous DNA damage and Show Impaired Damage Response.

A Mouse Embryonic Fibroblasts (MEFs) were obtained from *Cux1* knockout embryos ($Cux1^{Z/Z}$), wild-type and heterozygous ($Cux1^{wt/z}$) littermates. mRNA and protein levels of CUX1 are shown. p110 CUX1 is shown separately on a higher Acrylamide percentage gel to increase resolution.

B MEF cells isolated from wild-type, heterozygous knockout or homozygous knockout mice from the same litter were grown in 3% O_2. Cells were collected and strand breaks quantified by Comet Assay on at least 50 cells per condition.* p <0.05, ** p<0.01, *** p<0.001; student's T test. Representative images of Comet assay are shown on the right.

C MEF cells isolated as in B were exposed to 10μm H^2O^2 for 20 minutes at 4^{oC} and allowed to recover for the indicated time. Cells were collected and strand breaks quantified by Comet Assay on at least 30 cells per condition.* p <0.05, ** p<0.01, *** p<0.001; student's T test.

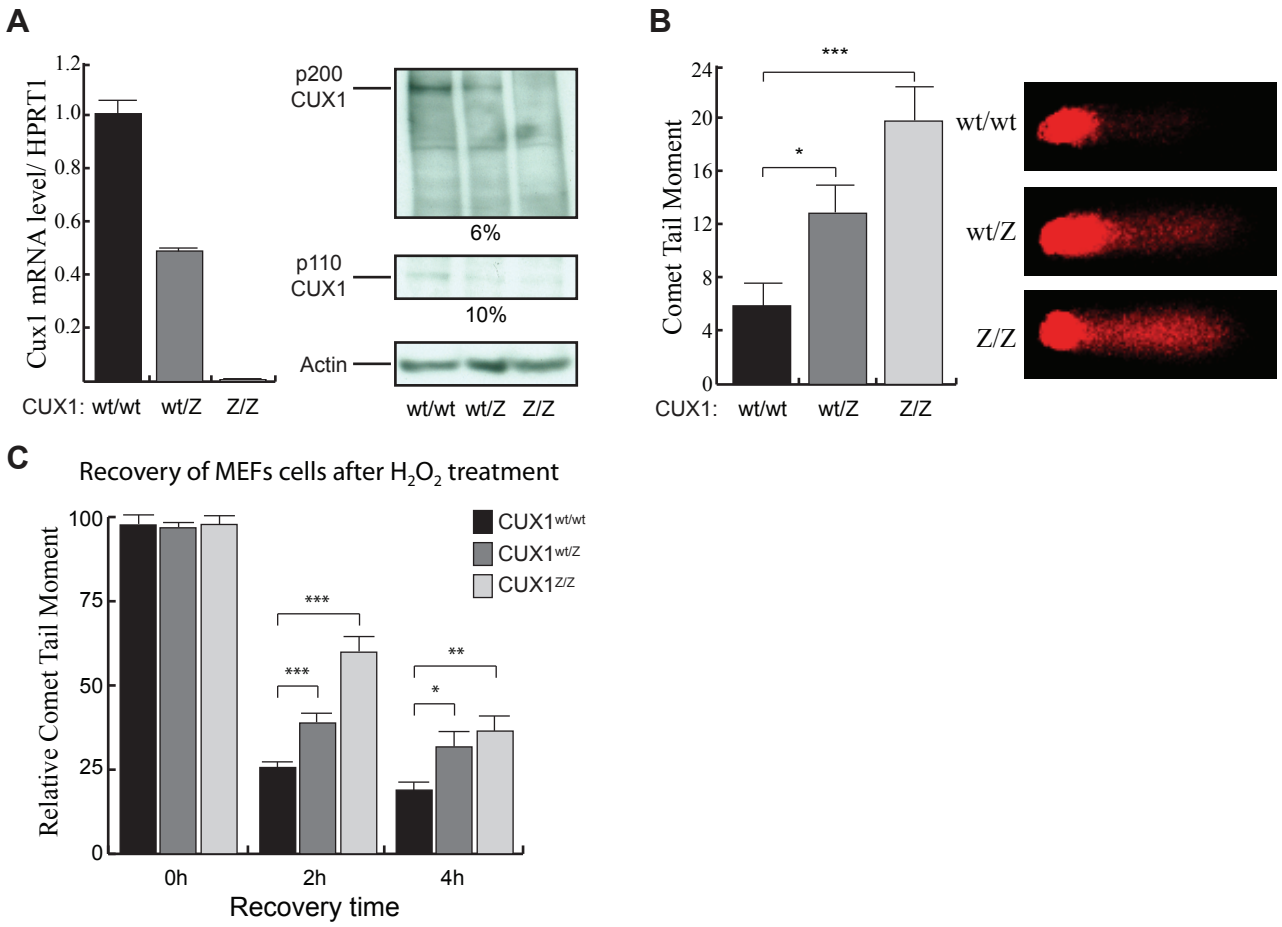

A
Cux1 mRNA level/ HPRT1
1.2
1.0
0.8
0.6
0.4
0.2
CUX1: wt/wt wt/Z Z/Z
p200 CUX1
6%
p110 CUX1
10%
Actin
wt/wt wt/Z Z/Z
B
Comet Tail Moment
24
20
16
12
8
4
0
***
*
CUX1: wt/wt wt/Z Z/Z
wt/wt
wt/Z
Z/Z
C
Recovery of MEFs cells after H_2O_2 treatment
Relative Comet Tail Moment
100
75
50
25
0
CUX1wt/wt
CUX1wt/Z
CUX1Z/Z
***
***
**
*
0h
2h
4h
Recovery time

Figure 14

Figure 15 – *Cux1*$^{Z/Z}$ MEF Cells Exhibit Genomic Instability.

A The number of metaphase chromosomes per cell was counted in *Cux1*$^{Z/Z}$ and wild-type MEFs. Percentages of diploid/near diploid cells and tetraploid/near tetraploid cells in each population is shown. * $p < 0.05$; Fisher's exact test.

B Cells were stained with Giemsa and the number of chromosome breaks per individual cell was counted for *Cux1*$^{Z/Z}$ and wild-type MEFs. The average number of breaks per cell is shown. * $p < 0.05$, ** < 0.01; student's T test.

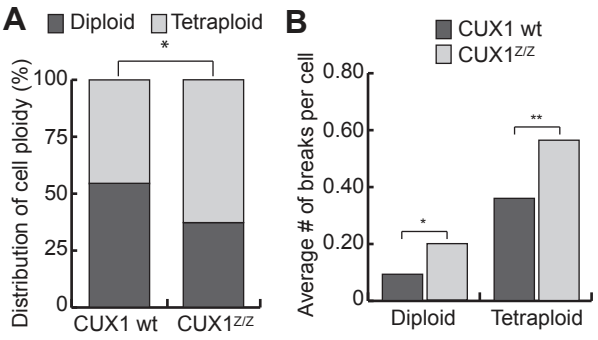

A
Diploid
Tetraploid
*
Distribution of cell ploidy (%)
100
75
50
25
0
CUX1 wt
CUX1Z/Z
B
CUX1 wt
CUX1Z/Z
Average # of breaks per cell
0.80
0.60
0.40
0.20
0
*
**
Diploid
Tetraploid

Figure 15

Figure 16 – p110 CUX1 Over-expression Increases Tumour Cell Resistance to DNA Damaging Agents.

A MCF7 cells stably over-expressing p110 CUX1 or an empty vector were exposed to DNA damage. 500 cells were plated and incubated for 10 days. Clones were fixed, stained and counted and the cloning efficiency of unexposed cells was set to 1. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; student's T test.

B Western blot for CUX1 on cells used in **A**. Actin is used as a loading control.

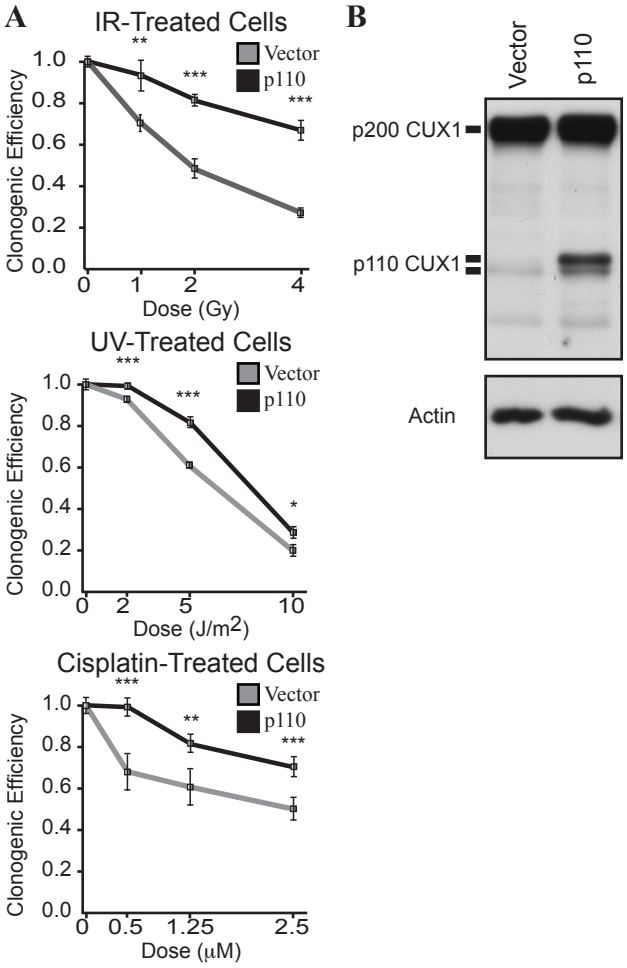

A
IR-Treated Cells
Vector
p110
Clonogenic Efficiency
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
4
Dose (Gy)
**
***
***
UV-Treated Cells
Vector
p110
Clonogenic Efficiency
0
2
5
10
Dose (J/m2)
***
***
*
Cisplatin-Treated Cells
Vector
p110
Clonogenic Efficiency
0
0.5
1.25
2.5
Dose (μM)
***
**
***
B
Vector
p110
p200 CUX1
p110 CUX1
Actin

Figure 16

Figure 17 – Autocrine Activation of the Wnt/β-Catenin Pathway in Some Mammary Tumours from MMTV-CUX1 Transgenic Mice.

A Immunohistochemical staining for β-catenin was performed in two solid mammary gland tumours and two adenosquamous carcinomas from MMTV-CUX1 transgenic mice. In subsequent experiments, the adenosquamous p75-80 and solid p75-534 tumours and their derived cell lines served as models for "Wnt" and "non-Wnt" tumours

B The TOP/FOP luciferase reporter assay was performed in cell lines established from the "Wnt" p75-80 and "non-Wnt" p75-534 mammary tumours.

C Cells of the "Wnt" p75-80 line were cotransfected with two plasmids as follows: a TOP or FOP luciferase reporter and a vector expressing either sFRP1, sFRP2 or SOST or an empty vector. Where indicated, niclosamide or the carrier (no inhibitor) was added to the cells at the time of transfection.

D mRNA levels of Wnt ligand genes were measured by RT-qPCR in the p75-80 and p75-534 mammary tumour cell lines.

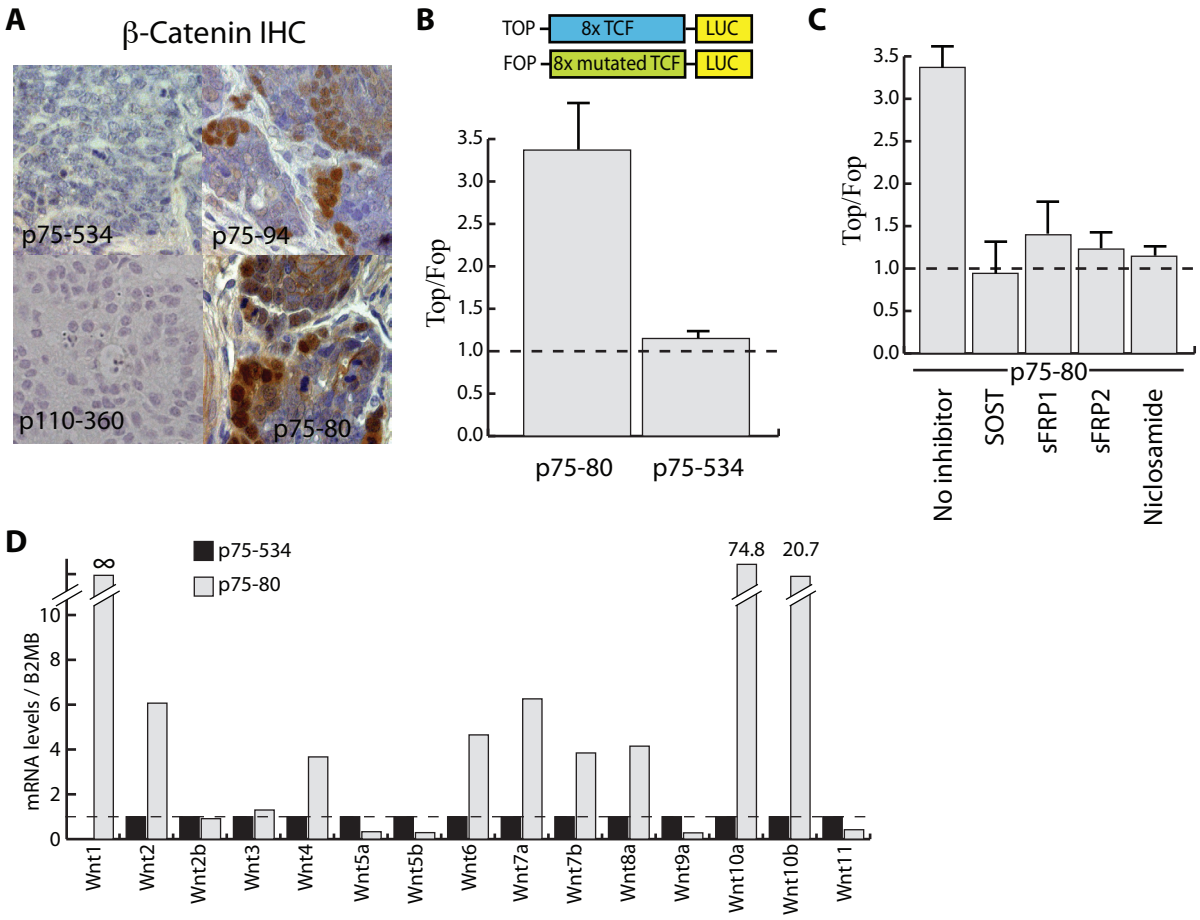

A
β-Catenin IHC
p75-534
p75-94
p110-360
p75-80
B
TOP
8x TCF
LUC
FOP
8x mutated TCF
LUC
Top/Fop
p75-80
p75-534
C
Top/Fop
p75-80
No inhibitor
SOST
sFRP1
sFRP2
Niclosamide
D
p75-534
p75-80
mRNA levels / B2MB
74.8
20.7
Wnt1
Wnt2
Wnt2b
Wnt3
Wnt4
Wnt5a
Wnt5b
Wnt6
Wnt7a
Wnt7b
Wnt8a
Wnt9a
Wnt10a
Wnt10b
Wnt11

Figure 17

Figure 18 – CUX1 Is Required for Maximal Expression of Wnt Genes in Human Tumour Cell Lines.

CUX1 specific or scrambled siRNA were transfected in a panel of 6 human cancer cell lines from breast (**A**), ovarian (**B**) and lung (**C**) cancers. 3 days later total mRNA was isolated and quantitative RT-PCR was performed using GAPDH mRNA as a control. The values represent fold difference in mRNA expression between cells treated with CUX1 or scrambled siRNA. (Und: undetected).

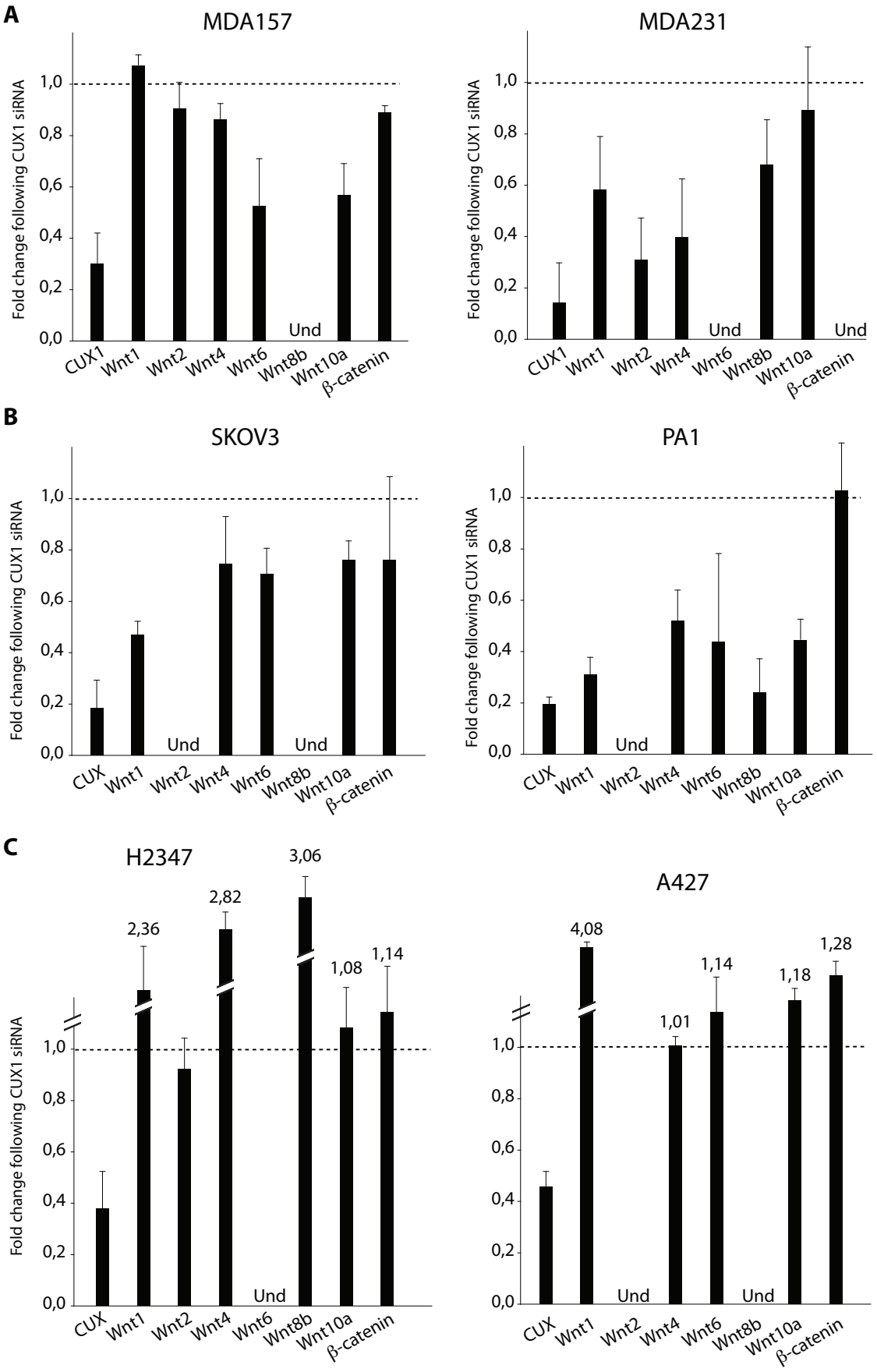

A
MDA157
MDA231
Fold change following CUX1 siRNA
1,0
0,8
0,6
0,4
0,2
0,0
Und
CUX1
Wnt1
Wnt2
Wnt4
Wnt6
Wnt8b
Wnt10a
β-catenin
B
SKOV3
PA1
CUX
C
H2347
2,36
2,82
3,06
1,08
1,14
A427
4,08
1,01
1,14
1,18
1,28

Figure 18

Figure 19 – Ectopic Expression of P110 CUX1 Leads to Autocrine Activation of the Wnt/β-Catenin Pathway in Human Tumour Cell Lines.

A Hs578T, MCF-7 and HEK293 cells were infected with retroviruses to establish cells stably expressing p110 CUX1 or not (vector). Nuclear and cytoplasmic extracts were analyzed by immunoblotting with the indicated antibodies.

B Hs578T, MCF-7 and HEK293 cells were cotransfected with a TOP (wild-type) or FOP (mutant) luciferase reporter and a plasmid vector expressing p110 CUX1 or an empty vector. 36 hours after transfection, whole cell extracts were prepared and processed to measure luciferase activity. The standard deviation of 3 transfections is shown and the results are expressed as TOP over FOP normalized to β-galactosidase activity from an internal control. A schematic representation of the reporter construct is shown in Fig. 17B.

C HEK293T cells were cotransfected with three plasmids as follows: a TOP or FOP luciferase reporter, a vector expressing p110 CUX1 or an empty vector, and a vector expressing either sFRP1, sFRP2, SOST or DKK1 or an empty vector. Where indicated, niclosamide, IWP-2 or the carrier was added to the cells at the time of transfection.

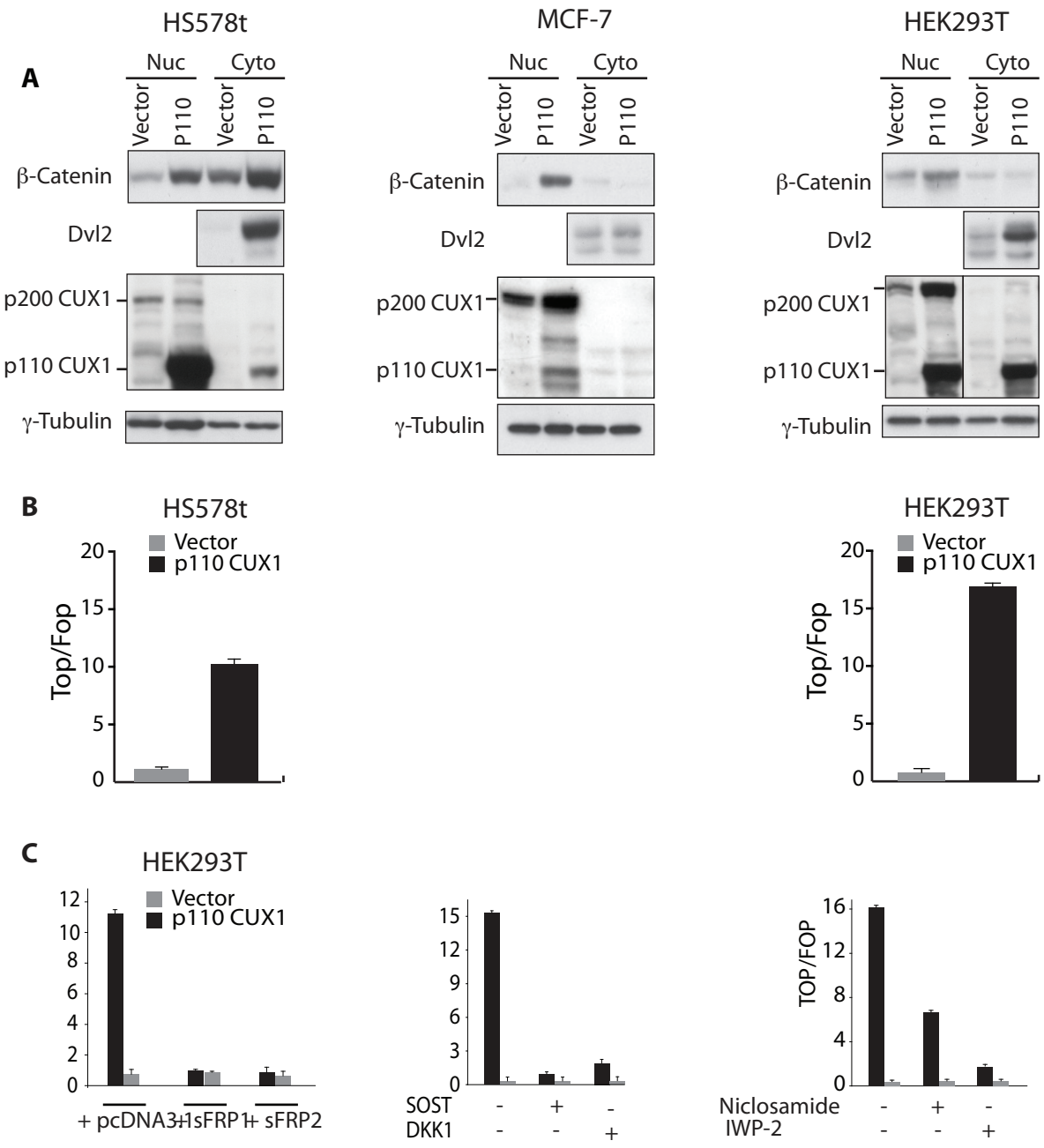

A
HS578t
MCF-7
HEK293T
Nuc
Cyto
Vector
P110
β-Catenin
Dvl2
p200 CUX1
p110 CUX1
γ-Tubulin
B
HS578t
HEK293T
Vector
p110 CUX1
Top/Fop
C
HEK293T
Vector
p110 CUX1
+ pcDNA3
+ sFRP1
+ sFRP2
SOST
DKK1
Niclosamide
IWP-2
TOP/FOP

Figure 19

Figure 20 – Activation of the Wnt/β-Catenin Pathway in MMTV-CUX1 Mammary Tumours Is Associated with High Expression of Glis1 and Gli3. Expression profiling was performed on microdissected mammary epithelial tumour cells from MMTV-CUX1 transgenic mice. The figure shows a heatmap of mammary tumours clustered according to Wnt gene expression. Expression of Glis1 and Gli3 in each tumour is shown below. Note that the probes for Gli1 and Gli2 on the microarray did not reveal any variation among samples and therefore were not considered.

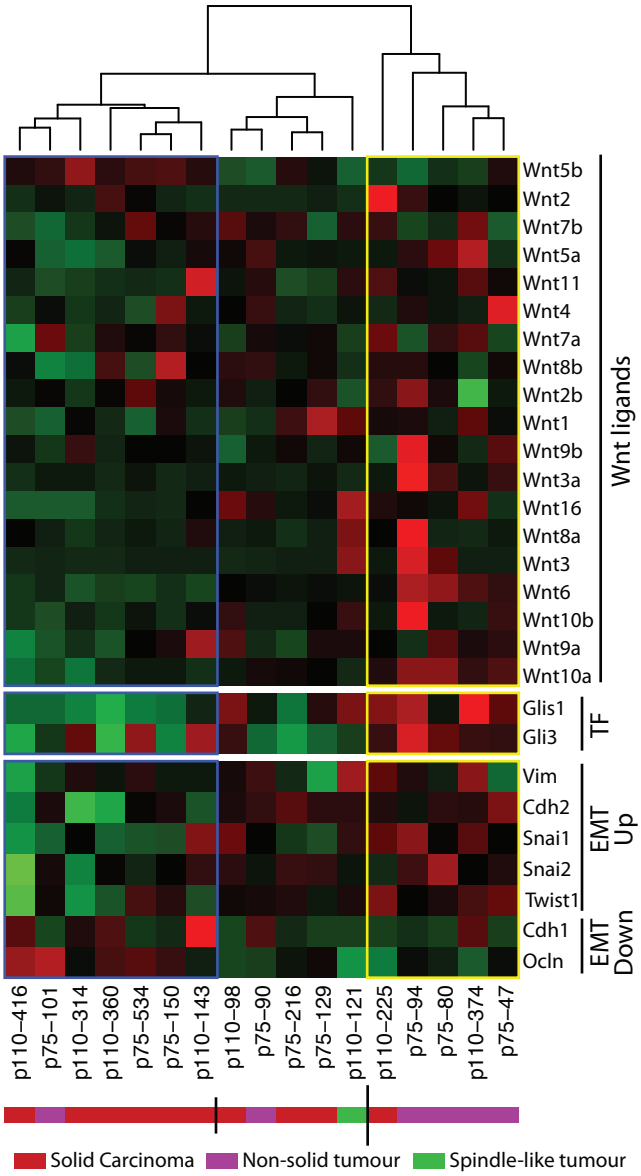

Wnt5b
Wnt2
Wnt7b
Wnt5a
Wnt11
Wnt4
Wnt7a
Wnt8b
Wnt2b
Wnt1
Wnt9b
Wnt3a
Wnt16
Wnt8a
Wnt3
Wnt6
Wnt10b
Wnt9a
Wnt10a
Wnt ligands
Glis1
Gli3
TF
Vim
Cdh2
Snai1
Snai2
Twist1
EMT Up
Cdh1
Ocln
EMT Down
p110–416
p75–101
p110–314
p110–360
p75–534
p75–150
p110–143
p110–98
p75–90
p75–216
p75–129
p110–121
p110–225
p75–94
p75–80
p110–374
p75–47
Solid Carcinoma
Non-solid tumour
Spindle-like tumour

Figure 20

Figure 21 – Correlation Between *WNT, CUX1* and *GLI* Gene Expression in a Human Breast Tumour Dataset.

A Heatmap of a human breast tumour dataset sorted according to Wnt genes expression using the BreSat algorithm. *CUX1, GLI1, GLI2* and *GLIS1* expression in each tumour is shown below.

B *CUX1, GLI1, GLI2* and *GLIS1* expression in the top 25% and bottom 25% samples sorted according to Wnt genes expression. * indicates $p < 0.05$, ** < 0.01, *** < 0.001 on a Welch-corrected student's T test.

C *CDH1, OCLN, SNAI1, TWIST1, VIM* expression in the top 25% and bottom 25% samples sorted according to Wnt genes expression. * indicates $p < 0.05$, ** < 0.01, *** < 0.001 on a Welch-corrected student's T test.

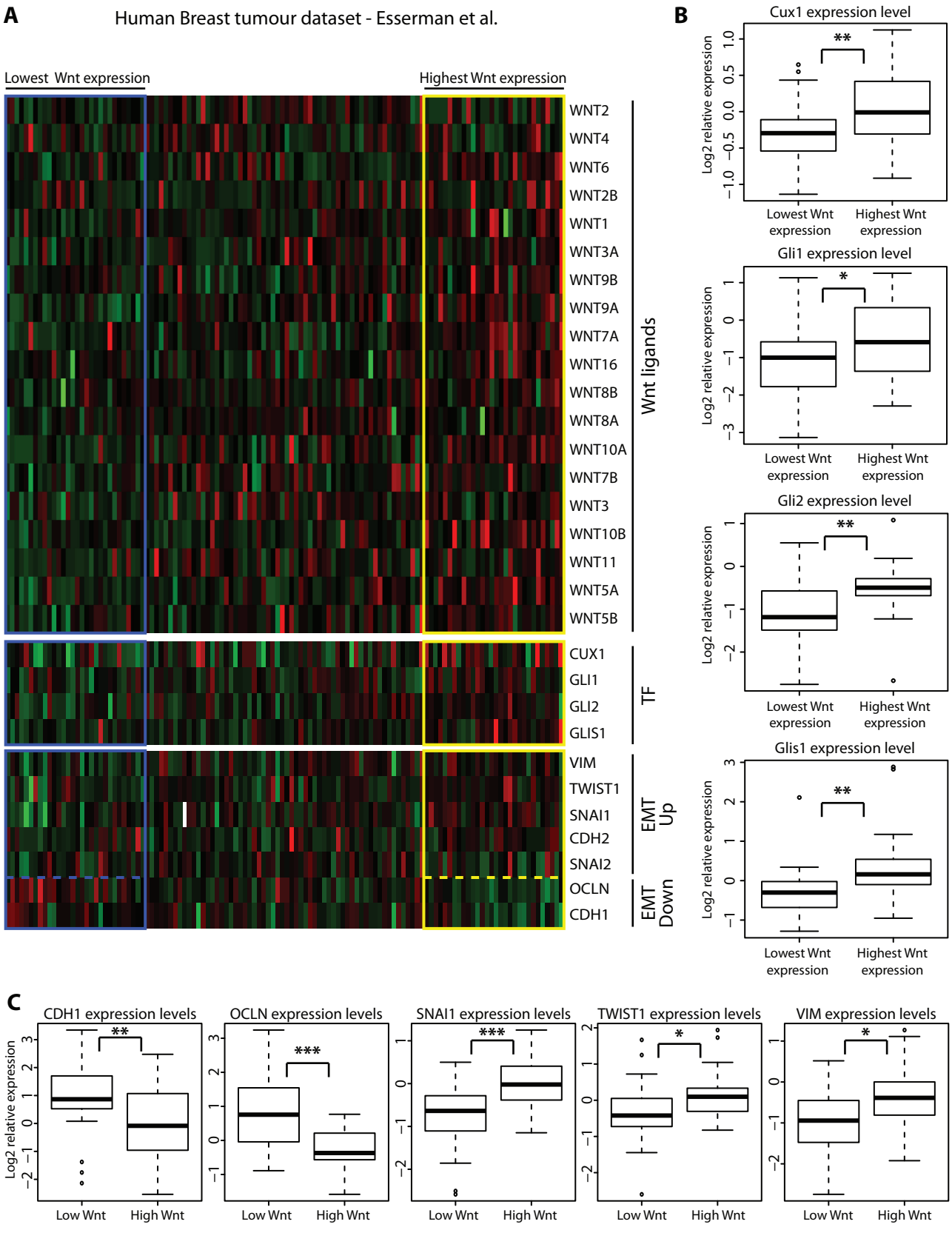

A
Human Breast tumour dataset - Esserman et al.
Lowest Wnt expression
Highest Wnt expression
WNT2
WNT4
WNT6
WNT2B
WNT1
WNT3A
WNT9B
WNT9A
WNT7A
WNT16
WNT8B
WNT8A
WNT10A
WNT7B
WNT3
WNT10B
WNT11
WNT5A
WNT5B
Wnt ligands
CUX1
GLI1
GLI2
GLIS1
TF
VIM
TWIST1
SNAI1
CDH2
SNAI2
EMT Up
OCLN
CDH1
EMT Down
B
Cux1 expression level
Gli1 expression level
Gli2 expression level
Glis1 expression level
Log2 relative expression
Lowest Wnt expression
Highest Wnt expression
C
CDH1 expression levels
OCLN expression levels
SNAI1 expression levels
TWIST1 expression levels
VIM expression levels
Log2 relative expression
Low Wnt
High Wnt

Figure 21

Figure 22 Ectopic Expression of GLIS1 in a MMTV-CUX1 Tumour Cell line Leads to the Transcriptional Activation of Wnt Genes

A Cells of the "non-Wnt" p75-534 tumour cell line were infected with a lentivirus expressing GLIS1 or nothing (vector). Two days later, β-Catenin expression was analyzed by immunoblotting

B Cells were treated as in A and Wnt mRNA levels were measured by RT-qPCR.

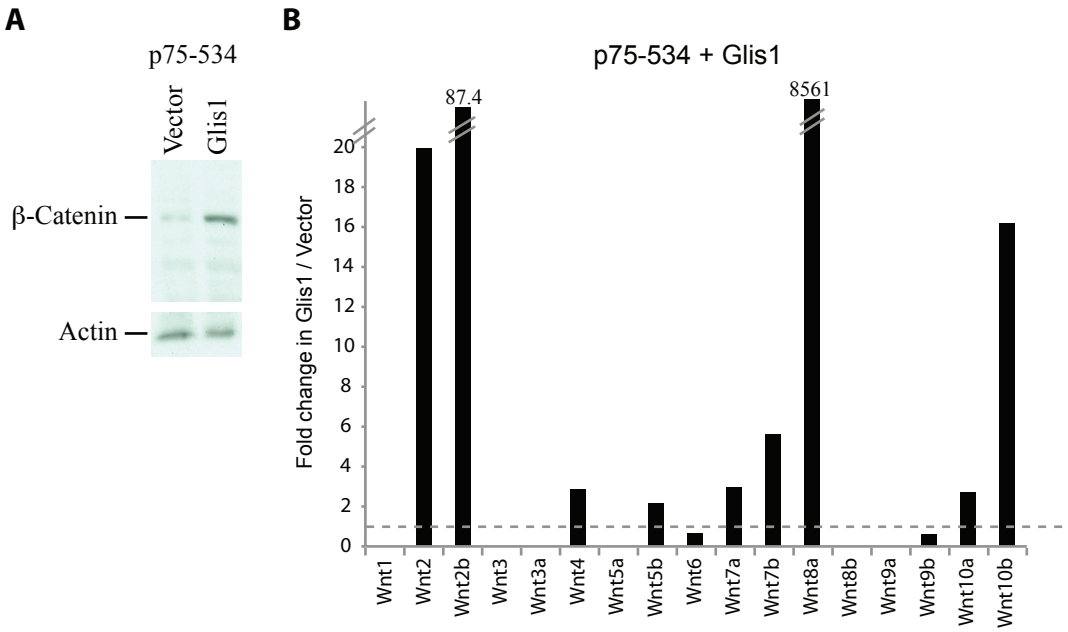

A
p75-534
Vector
Glis1
β-Catenin
Actin
B
p75-534 + Glis1
87.4
8561
Fold change in Glis1 / Vector
0
2
4
6
8
10
12
14
16
18
20
Wnt1
Wnt2
Wnt2b
Wnt3
Wnt3a
Wnt4
Wnt5a
Wnt5b
Wnt6
Wnt7a
Wnt7b
Wnt8a
Wnt8b
Wnt9a
Wnt9b
Wnt10a
Wnt10b

Figure 22

Figure 23 – GLIS1 and p110 CUX1 Cooperate To Activate of the Wnt/β-Catenin Pathway.

MCF10A cells were infected with retroviruses expressing nothing (vector), GLIS1, p110 CUX1, or both and stable populations were established.

A Wnt mRNA levels were measured RT-qPCR.

B β-Catenin protein levels were measured by Immunofluorescence using antibodies specific for active, non-phosphorylated form of the protein. Histogram shows the mean nuclear signal for β-Catenin for at least 50 cells per condition. A representative image of each condition is shown below.

C β-Catenin protein levels were measured by immunoblotting using antibodies specific for active, non-phosphorylated form of the protein. γ-Tubulin is shown as a loading control.

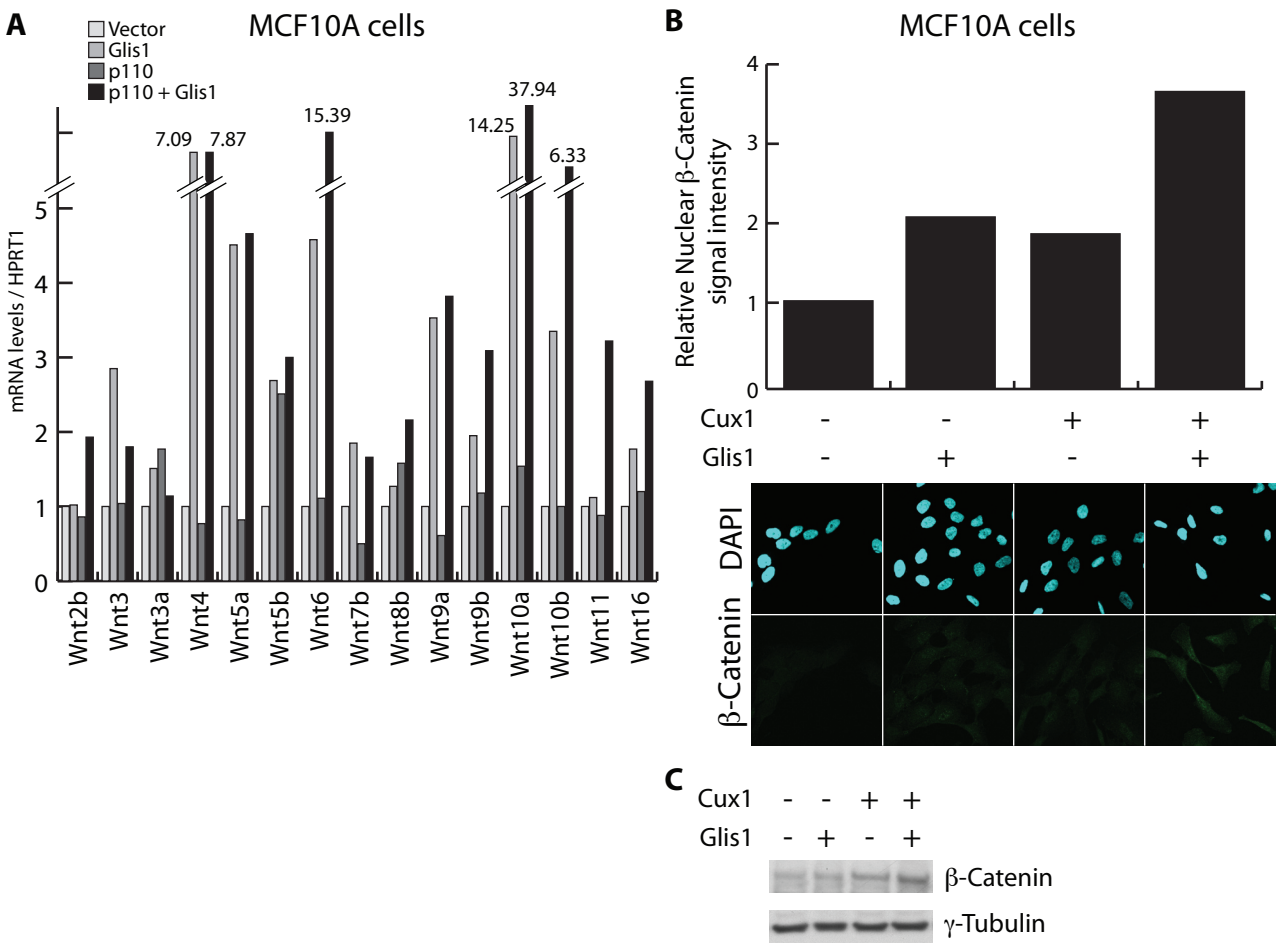

A
MCF10A cells
Vector
Glis1
p110
p110 + Glis1
mRNA levels / HPRT1
7.09
7.87
15.39
14.25
37.94
6.33
Wnt2b
Wnt3
Wnt3a
Wnt4
Wnt5a
Wnt5b
Wnt6
Wnt7b
Wnt8b
Wnt9a
Wnt9b
Wnt10a
Wnt10b
Wnt11
Wnt16
B
MCF10A cells
Relative Nuclear β-Catenin signal intensity
Cux1
Glis1
DAPI
β-Catenin
C
Cux1
Glis1
β-Catenin
γ-Tubulin

Figure 23

Figure 24 – Co-expression of Glis1 and p110 CUX1 increases motility and invasiveness of cells.

A MCF10A cells overexpressing Glis1, p110 CUX1 or both were filmed by live cell imaging and their movement speed was measured. The average of 3 independent experiments is shown with error bars representing standard deviation. ** P Value <0.01, *** <0.001 on 3 combined Welch corrected T tests using the Fisher method.

B Cells were treated as in A and distance from origin was measured

C MCF10A cells overexpressing Glis1, p110 CUX1 were subjected to inverted Boyden chamber assays. Histogram shows the average of 3 independent experiments.

D Images from a representative invasion experiment **C**.

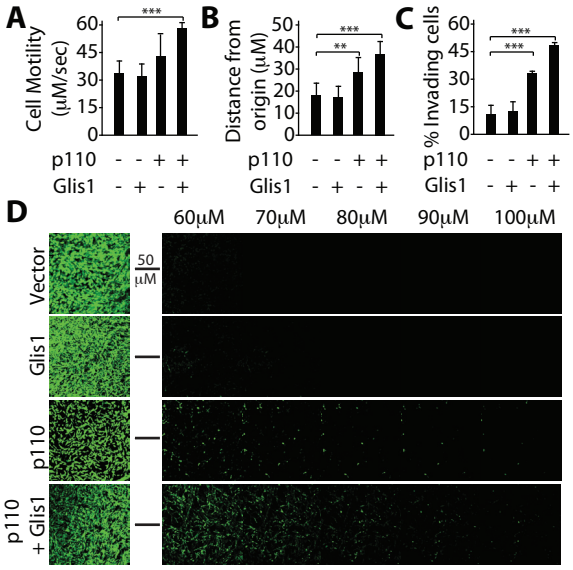

A
Cell Motility (μM/sec)
60
45
30
15
0
***
p110 - - + +
Glis1 - + - +
B
Distance from origin (μM)
50
40
30
20
10
0
***
**
p110 - - + +
Glis1 - + - +
C
% Invading cells
60
45
30
15
0
***
***
p110 - - + +
Glis1 - + - +
D
60μM
70μM
80μM
90μM
100μM
Vector
Glis1
p110
p110 + Glis1
50 μM

Figure 24

Figure 25 – CUX1 is required for the invasiveness of HCT116 cells independently of β-Catenin activity.

A HCT116 cells stably expressing a Doxycycline-inducible CUX1 specific shRNA were treated with Dox or vehicle control for 6 days. mRNA was extracted and Wnt gene expression was measured by RT-qPCR.

B Nuclear extracts of cells treated as in **A** were immunobloted for active β-Catenin.

C Cells treated as in **A** were cotransfected with a TOP (wild-type) or FOP (mutant) luciferase reporter. Luciferase activity was measured 36 hours after transfection.

D Cells treated as in **A** were subjected to inverted Boyden chamber assays. Histogram shows the average of 3 independent experiments. Images from a representative invasion experiment are shown to the right. ** P Value <0.01, *** <0.001 on 3 combined Welch corrected T tests using the Fisher method.

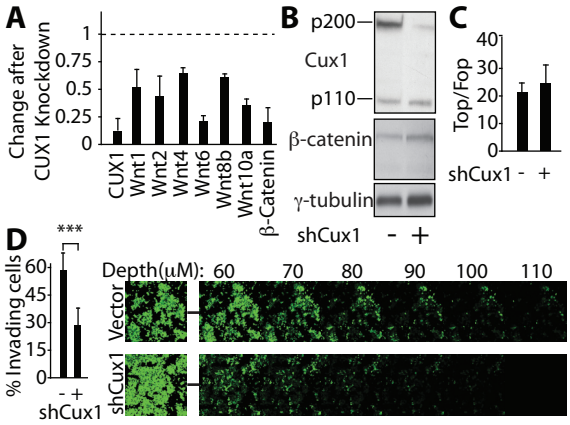
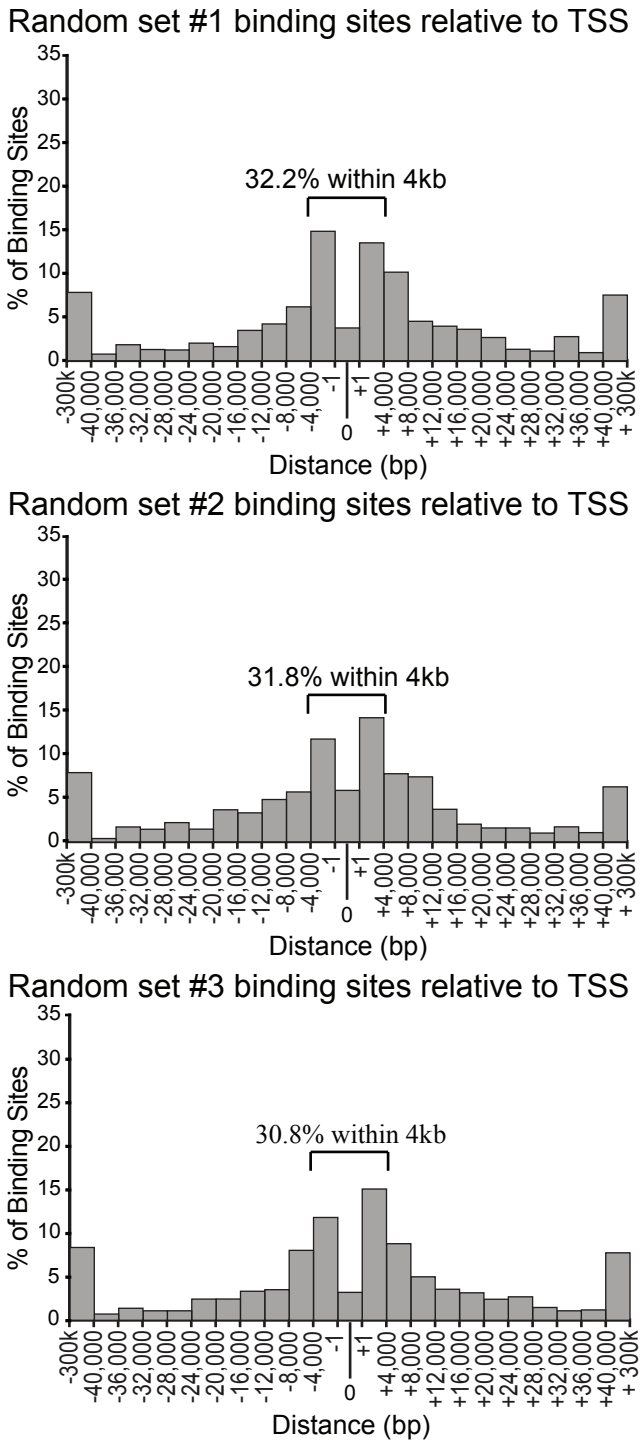

A
Change after CUX1 Knockdown
1
0.75
0.5
0.25
0
CUX1
Wnt1
Wnt2
Wnt4
Wnt6
Wnt8b
Wnt10a
β-Catenin
B
p200
Cux1
p110
β-catenin
γ-tubulin
shCux1 - +
C
Top/Fop
40
30
20
10
0
shCux1 - +
D
***
% Invading cells
60
45
30
15
0
- +
shCux1
Depth(μM): 60 70 80 90 100 110
Vector
shCux1

Figure 25

Supplementary Figure 1 – Distribution of 3 Random Sets of Binding Sites Relative to Transcription Start Sites.

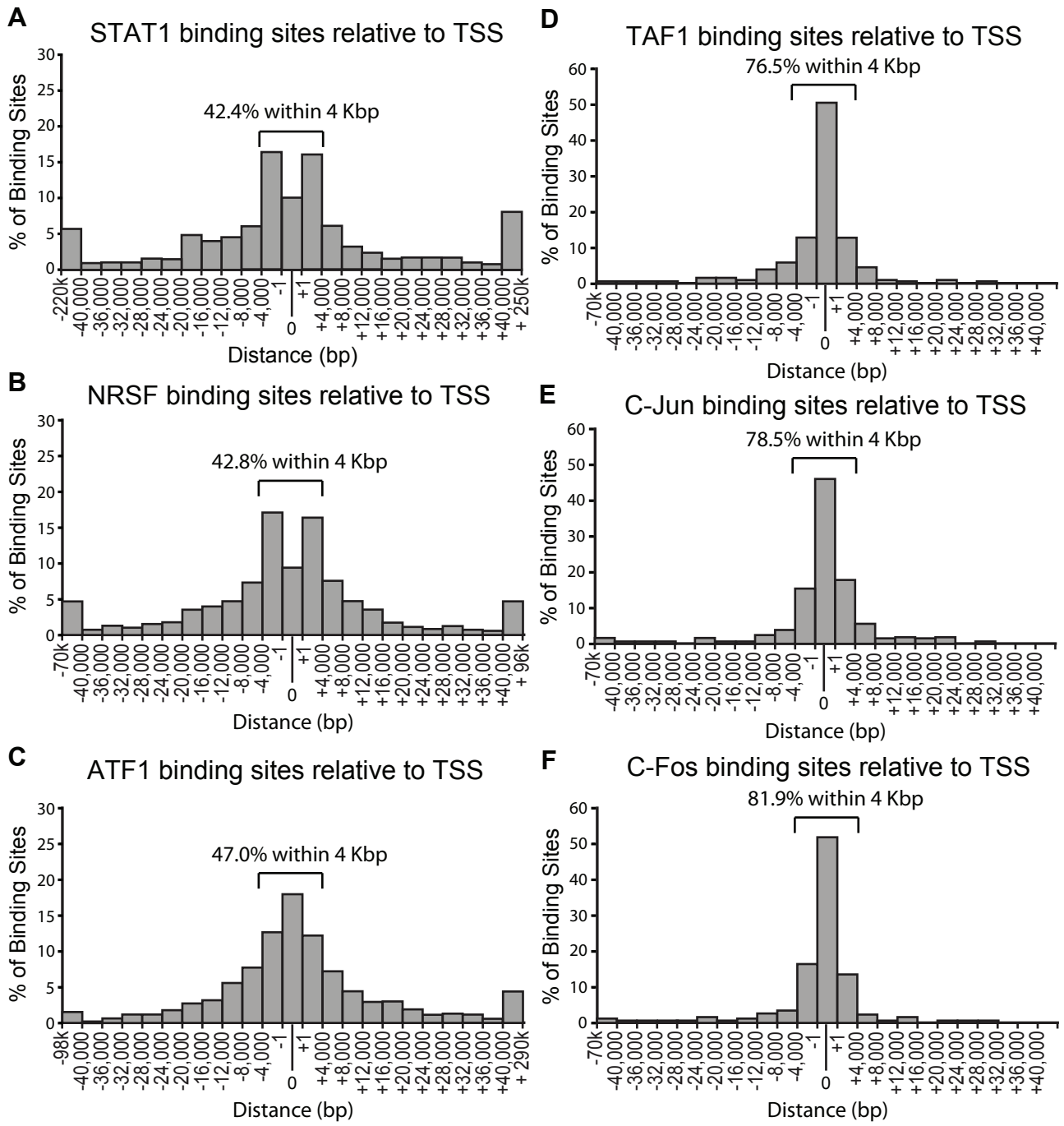
Percentage of Randomly generated binding sites located at various distances from the closest transcription start site. The "0" column indicates genes where the binding site overlaps the start site. 3 sets of 513 randomly located sites having the same distribution of sizes as CUX1 binding sites were generated.



Supplementary Figure 1

Supplementary Figure 2 – Distribution of binding sites Relative to Transcription Start Sites

Percentage of binding sites located at various distances from the closest transcription start site for 6 different transcription factors. The "0" column indicates genes where the binding site overlaps the start site.

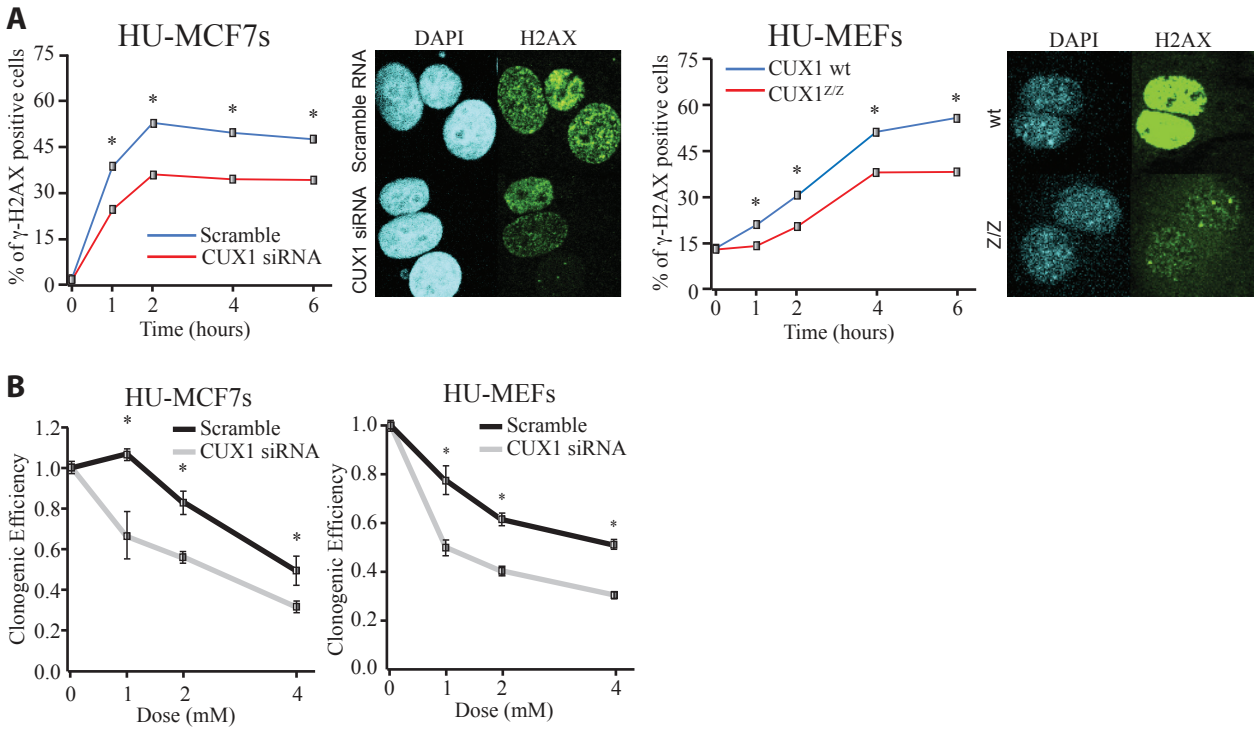


Supplementary Figure 2

Supplementary Figure 3 – γ-H2AX Signalling and Survival Following Hydroxyurea Treatment.

A MCF7 cells transfected with CUX1-specific siRNA or Scramble siRNA or MEF cells from *Cux1*$^{Z/Z}$ knockout embryos and wild-type littermates were exposed to Hydroxyurea (HU). Cells were fixed and stained by immunofluorescence for γ-H2AX and the proportion of γ-H2AX positive cells was counted after treatment. * $p < 0.05$ on a Fisher's exact test.

B Cells as in A were exposed to HU. 500 MCF7 cells or 5000 MEF cells were plated in triplicate and incubated for 10 days. Clones were fixed, stained and counted and the cloning efficiency of unexposed cells was set to 1. * $p < 0.05$ on a student's T test.

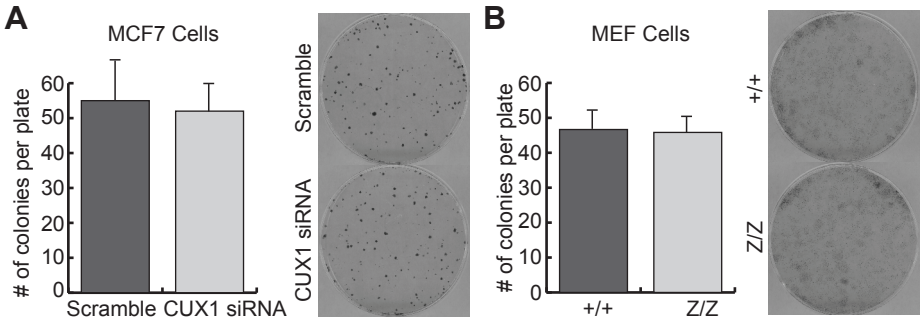


Supplementary Figure 3

Supplementary Figure 4 - Clonogenic Efficiency of MCF7s and MEFs.

A 500 MCF7 cells treated with either a CUX1 specific siRNA or a scrambled control were plated for 10 days prior to fixation and staining.

B 5000 MEF cells from *Cux1*$^{z/z}$ or *Cux1* wild-type littermates were plated for 10 days prior to fixation and staining.



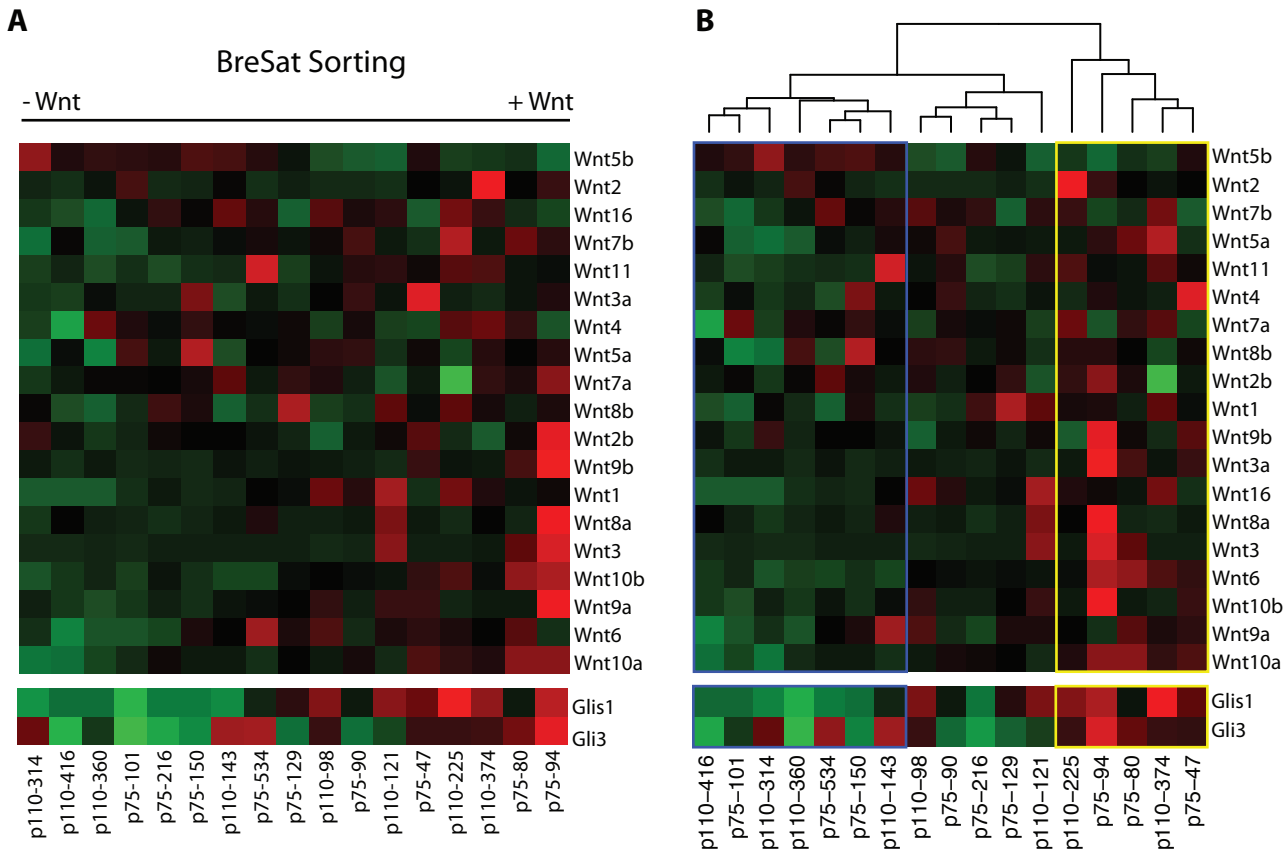
Supplementary Figure 4

Supplementary Figure 5 – BreSat Sorting vs. Clustering of CUX1 Mammary Tumours

Comparison of the heatmap representations of mammary tumours ranked according to Wnt expression using the BreSat algorithm vs. hierarchical clustering according to the same set of genes.

A BreSat Sorting.

B Unsupervised hierarchical clustering using Euclidean distance.



Supplementary Figure 5

Supplementary Figure 6 – CUX1 DNA Binding Activity and Effect of CUX1 Knockdown on p21 mRNA Expression After DNA Damage.

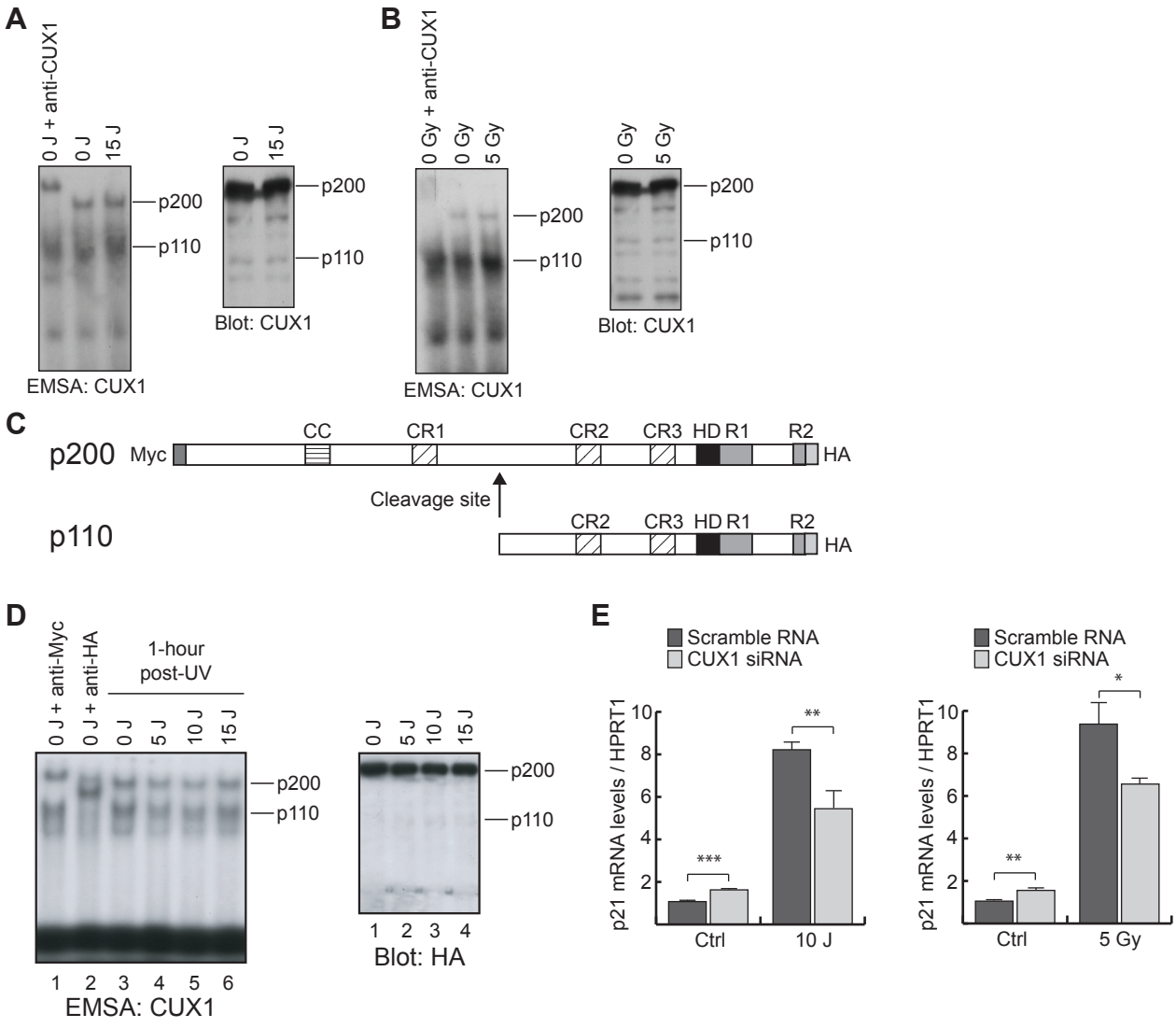
A Nuclear protein extracts were prepared from HS578T cells exposed to 15 J UV or left untreated. Left: DNA binding by endogenous CUX1 was assessed by Electrophoretic Mobility Shift Assay (EMSA) using double-stranded oligonucleotides containing a CUX1 consensus binding site. Right: Western blot showing equal levels CUX1 protein.

B Nuclear protein extracts were prepared from HS578T cells exposed to 5 Gy IR or left untreated. Left: DNA binding by endogenous CUX1 was assessed by EMSA as in A. Right: Western blot showing equal levels CUX1 protein.

C Schematic representation of recombinant CUX1 protein is shown. The full length p200 isoform is tagged at both the N-terminus (Myc tag) and C-Terminus (HA tag). p200 is cleaved at the indicated site into the p110 isoform.

D Nuclear protein extracts were prepared from NIH 3T3 cells stably expressing the p200 CUX1 recombinant protein described in C. Left: DNA binding by recombinant CUX1 was assessed by EMSA as in A. Right: Western blot showing equal levels of recombinant CUX1 proteins.

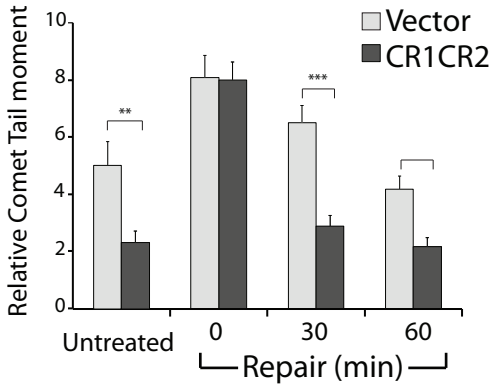
E MCF7 cells were transfected with CUX1-specific siRNA or scrambled control. Left: Cells were then exposed to 10 J of UV or left untreated. Right: Cells were exposed to 5 Gy IR or left untreated. p21 mRNA levels were measured by qPCR and normalized to HPRT1. *: P value <0.05, ** < 0.01, *** <0.001.



Supplementary Figure 6

Supplementary Figure 7 – Overexpression of the CR1CR2 Fragment of CUX1 Is Sufficient To Accelerate the Repair of DNA Stand Breaks.

DLD1 cells stably carrying a vector expressing nothing (Vector) or a CUX1 peptide encompassing the Cut repeats 1 and 2 fused to a nuclear localization signal (CR1CR2-NLS) were exposed to 10 μm H_2O_2 for 20 min on ice. Cells were collected after the indicated repair period in H_2O_2-free medium. Each bar represents the average of at least 30 comets and processed for the Comet assay. **: $p<0.01$, ***: $p<0.001$



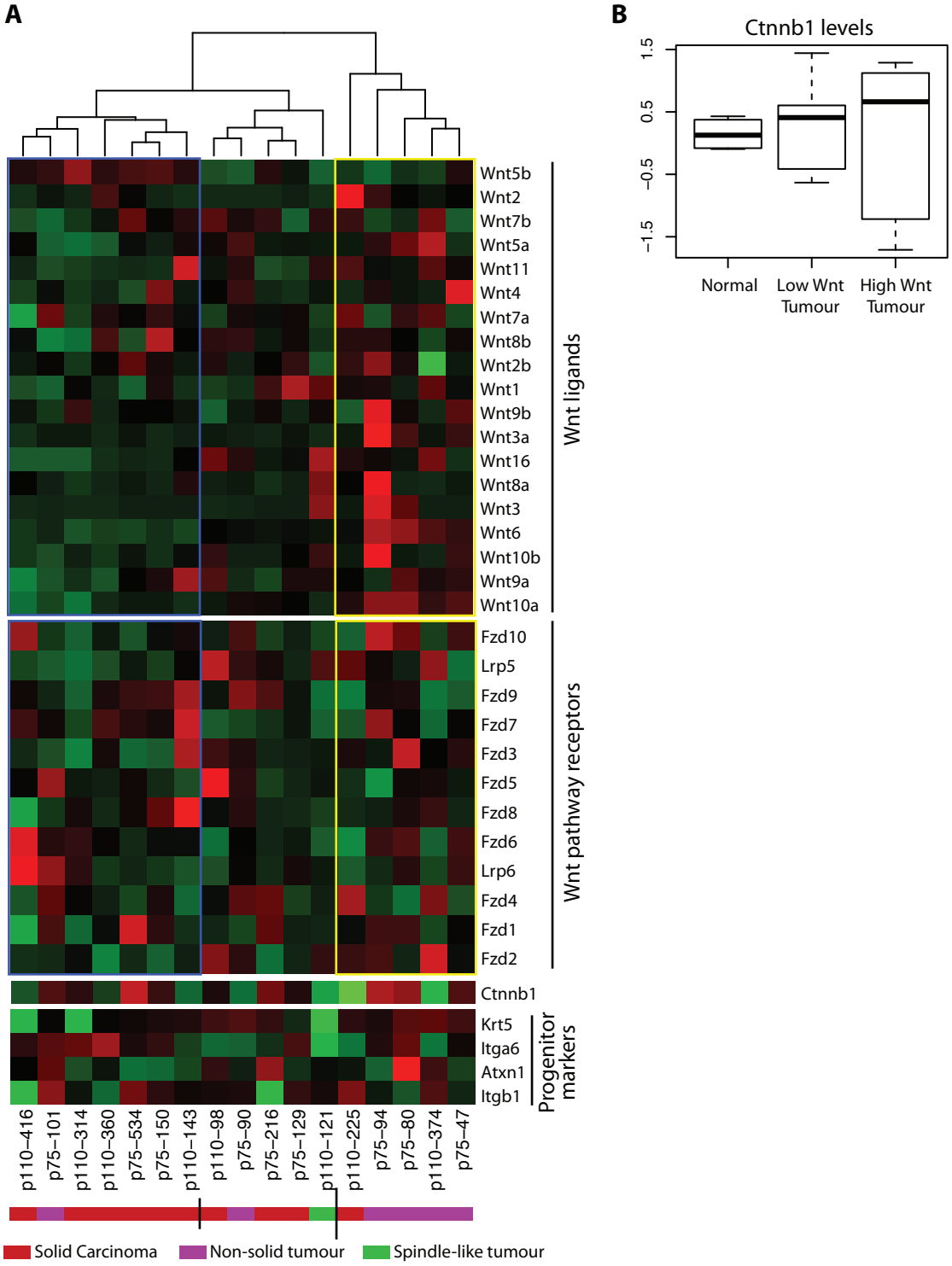
Unpublished results from Z. Ramdzan, Hulea L. and Vadnais C.

Supplementary Figure 7

Supplementary Figure 8 - Activation of the Wnt/β-Catenin Pathway in MMTV-CUX1 Mammary Tumours is Independent of Frzd Receptor and β-Catenin Transcript Levels.

A Heatmap of mammary tumours clustered according to Wnt gene expression, with corresponding expression of Frzd receptor genes, β-Catenin and markers of progenitor cells.

B Boxplot representation of β-Catenin expression in normal mammary epithelial cells, "non-Wnt" tumours and "Wnt" tumours.



Supplementary Figure 8