

Transcriptome Based Bioinformatic Analysis of a Unique Ovarian Cancer Model

Samuel Nicholas Smillie

Department of Human Genetics

McGill University

Montreal, Quebec, Canada

March 2013

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

Master of Science

© Samuel Nicholas Smillie, 2013

ACKNOWLEDGMENTS

I would like to express my gratitude towards those who have supported my pursuit of this degree, most importantly my supervisor, Dr. Patricia Tonin. Patricia, your dedication and passion have served to inspire me during moments of doubt. Your support and confidence made this journey possible and enabled me to grow as a person and scientist. Your patience continues to amaze me. Thank you.

Many thanks to the members of Dr. Tonin's research group, past and present. Dr. Neal Cody, Dr. Michael Quinn, Dr. Ashley Birch, Dr. Paulina Wojnarowicz, Dr. Karen Gambaro, Suzanna Arcand, Nancy Hamel, Lena Dolman, and Moria Belanger, you have made my time at McGill wonderful and have all contributed to this thesis in your own ways. The excellent research performed by Dr. Neal Cody and Dr. Michael Quinn laid the foundation for this project for which I am extremely grateful. Special thanks to Dr. Karen Gambaro; you have been a wonderful friend and mentor. I cannot imagine my time here without your positive attitude and selfless nature. You are the best.

I would also like to acknowledge members of Dr. Anne-Marie Mes-Masson's research team, Jason Madore, Michael Quinn, Guergana Tchakarska and Liliane Meunier. A special thanks to Michael for his guidance and Liliane for contributing the *CP* immunohistochemistry results to this thesis.

Many thanks to the members of my supervisory committee Dr. Peter Siegel and Dr. Nada Jabado who helped guide my project as well as to the Research Institute of the McGill University Health Centre and the Department of Medicine at McGill University for financial support.

Lastly, I would like to thank my family for their love as well as all the members of the Department of Human Genetics whom I am fortunate to call friends.

CONTRIBUTION OF AUTHORS

The candidate completed the majority of the research presented and all aspects of the project were under the supervision of Dr. Patricia Tonin. Suzanna Arcand extracted the gene expression data file GEOD-18520 from the ArrayExpress Archive (<http://www.ebi.ac.uk/arrayexpress/>). The candidate conducted comparative transcriptome analyses, gene annotation of probe sets and hierarchical clustering. Bioinformatic analyses were performed by the candidate under the advice of Dr. Michael Quinn. The candidate under the guidance of Dr. Karen Gambaro performed western blotting. Immunohistochemistry was performed by Liliane Muenier of Dr. Anne-Marie Mes-Masson's research team. All statistical analyses were performed by the candidate.

TABLE OF CONTENTS

Page	Section	Title
1		Title page
2		Acknowledgements
5		Table of contents
9		Abstract – English
11		Résumé – French
13		Abbreviations
14	1	INTRODUCTION
14	1.1	Ovarian cancer clinical presentation
16	1.2	Epithelial ovarian cancer (EOC) classification
17	1.3	Suspected origins of development
18	1.4	Models to study EOC
18	1.5	OV-90: a unique EOC cell line
20	1.6	Transcriptome analyses in the study of EOC
21	1.7	Transcriptome analyses of the OV-90 hybrid cell lines
23	1.8	Project hypothesis and objectives
24	2	MATERIALS AND METHODS
24	2.1	Derivation of gene expression data for comparative transcriptome analysis
25	2.2	Hierarchical cluster analysis
25	2.3	Derivation of candidate gene list from comparative transcriptome analysis
26	2.4	Gene annotation of probe sets
26	2.5	Bioinformatics analyses using software tools
27	2.6	Survey of the candidates identified by comparative transcriptome analysis
28	2.7	Western blot analysis
29	2.8	Immunohistochemistry analysis

31	3	RESULTS
31	3.1	Hierarchical clustering
32	3.2	Comparative transcriptome analysis
33	3.3	Differentially expressed gene categories
34	3.4	Bioinformatics analysis of differentially expressed genes using computational tools
36	3.5	Interrogation of TCGA analysis of HGSC
38	3.6	Protein expression analysis of a candidate gene and correlation with clinical parameters
40	4	DISCUSSION
40	4.1	Validation of the hybrids as a model for the study of HGSC
46	4.2	Gene candidates identified in the TCGA study on HGSC
49	4.3	Implicated molecular pathways in HGSC
51	4.4	Gene candidates implicated in G-protein coupled receptor signaling
60	4.5	The role of ceruloplasmin in HGSC
63	4.6	Conclusion
65	4.7	Future Directions
71		FIGURES
80		TABLES
89		REFERENCES
103		APPENDICES
103		Appendix I – DAVID functional annotation clustering
114		Appendix II – IPA biological functions
119		Appendix III – IPA canonical pathways
122		Appendix IV – TCGA expression profiles

LIST OF FIGURES

Page	Figure	Title
71	1	Hierarchical clustering of #18520 dataset
72	2	Comparative transcriptome analysis selection criteria
73	3	Candidate gene expression categories
74	4	<i>CP</i> mRNA and protein expression in OV-90neo and hybrids
75	5	<i>CP</i> mRNA expression in a series of HGSC and normal reference tissues
77	6	Representative CP immunohistochemistry staining in HGSC and FT samples
78	7	Kaplain Meir analysis of CP in HGSC

LIST OF TABLES

Page	Table	Title
80	1	Candidate gene list
83	2	DAVID gene functional classification
84	3	IPA top 20 biological functions
85	4	IPA canonical pathways
86	5	IPA molecular networks
87	6	TCGA regional copy number alterations
88	7	Clinical characteristics of the patient cohort

ABSTRACT (ENGLISH)

Ovarian cancer is a heterogeneous and deadly malignancy where both an understanding of the disease biology and effective therapies are lacking. Efforts to characterize the disease are challenged by the genetically and clinically distinct manifestations of the disease currently defined by histopathology. The serous histopathological subtype represents the majority of cases and fatalities due to the disease and is the focus of this thesis. Using a unique series of ovarian cancer cell lines (OV-90neo, RH-5, RH-6 and RH-10) representative of aspects of the high-grade serous subtype, we examined the genes differentially expressed at the mRNA level between the tumorigenic OV-90neo and the non-tumorigenic genetically modified ‘hybrid’ (RH-5, RH-6 and RH-10) cell lines, within a series of publicly available high-grade serous tumor samples and ovarian surface epithelium cytobrushings. Selecting for those genes that also differed in their expression levels between the high-grade serous tumors and the ovarian surface epithelium, we identified a small, focused, candidate gene list that captured aspects of high-grade serous biology and processes related to the tumorigenicity of the disease. Using established bioinformatics programs, we identified biological conditions, processes, pathways and functions associated with this candidate list. These results supported the relevance of the hybrid cell lines to high-grade serous ovarian cancer and identified gene candidates and molecular pathways for future research. Our research group had previously implicated the candidate gene ceruloplasmin (*CP*) in the disease, leading us to characterize its protein expression in a large series of clinically annotated high-grade serous

tumor samples. The results identified a statistically significant relationship between protein expression of *CP* and patient progression-free survival, supporting further research into the role of this gene in ovarian cancer. We have demonstrated the utility of the hybrid cell lines in investigating ovarian cancer biology and suggest that they may serve as an effective model for the study of this disease.

RÉSUMÉ (FRANÇAIS)

Le cancer de l’ovaire est une maladie hétérogène et mortelle pour laquelle la compréhension des mécanismes biologiques ainsi que des thérapies efficaces manquent. Les efforts mis en œuvre pour caractériser la maladie sont entravés notamment par une disparité clinique et génétique des manifestations de la maladie, qui est elle-même définie par des critères histologiques. Le sous-type histologique séreux représente la majorité des cas et des décès dus à la maladie, et sera le sujet central de cette thèse. En se servant d’une série de lignées cellulaires de cancer de l’ovaire (OV-90neo, RH-5, RH-6 and RH-10), représentatives du sous-type séreux de haut grade, nous avons examiné les gènes différentiellement exprimés au niveau de l’ARN messager entre la lignée cellulaire tumorigénique OV-90neo et les lignées « hybrides » génétiquement modifiées non-tumorigéniques (RH-5, RH-6 and RH-10) dans une série d’échantillons de tumeurs séreuses de haut grade et de cellules épithéliales de l’ovaire prélevées à la cytobrosse. provenant d’une étude indépendante et accessible au public. En sélectionnant les gènes montrant aussi une expression différentielle entre les tumeurs séreuses de haut grade et les cellules épithéliales de surface de l’ovaire, nous avons identifié une liste de gènes candidats, réduite et ciblée, qui capture les aspects biologiques du sous-type séreux de haut grade ainsi que les processus liés à la tumorigénicité de la maladie. En utilisant des programmes bio-informatiques, nous avons identifié les conditions, processus, voies de signalisation et fonctions biologiques associés à cette liste de candidats. Ces résultats ont permis de confirmer la validité des lignées cellulaires hybrides comme modèle pour l’étude du cancer de l’ovaire de type séreux de haut grade et d’identifier des gènes

candidats et des voies de signalisation pour les recherches futures. Notre groupe de recherche avait précédemment impliqué le gène candidate ceruloplasmin (*CP*) dans la maladie, nous menant ainsi à caractériser l'expression de la protéine dans une large série d'échantillons de tumeurs ovariennes séreuses de haut grade annotées pour l'information cliniques. Les résultats identifièrent une corrélation statistiquement significative entre l'expression de la protéine CP et la survie des patientes sans progression de la maladie, encourageant des recherches plus approfondies sur le rôle de ce gène dans le cancer de l'ovaire. Nous avons démontré l'utilité des lignées cellulaires hybrides pour l'examen de la biologie du cancer de l'ovaire et proposons ces cellules comme modèle efficace pour étudier la maladie.

ABBREVIATIONS

OC	ovarian cancer
EOC	epithelial ovarian cancer
HGSC	high-grade serous ovarian cancer
OSE	ovarian surface epithelium
FT	fallopian tube
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
n	number
MAS5.0	Microarray Suite 5.0
DAVID	Database for Annotation, Visualization and Integrated Discovery
IPA	Ingenuity Pathway Analysis
TCGA	The Cancer Genome Atlas
CGH	comparative genomic hybridization
SDS	sodium dodecyl sulphate
HRP	horseradish peroxidase
TMA	tissue microarray
μm	micrometer
mm	millimeter
SPSS	Statistical Package for the Social Sciences
T	tumorigenic
N	non-tumorigenic
mRNA	messenger ribonucleic acid
kDa	kilodalton
CI	confidence interval
DNA	deoxyribonucleic acid
HPV	human papillomavirus

1. INTRODUCTION

1.1 Ovarian cancer clinical presentation

Ovarian cancer (OC) is the most deadly gynecological malignancy. It is the 7th most common cancer among Canadian women and the 5th most lethal. The 5-year relative survival ratio of 42% and an overall survival of approximately 30% illustrate the inability to effectively manage the disease (Canadian Cancer Statistics, 2011). In Canada, 2,600 women are estimated to be diagnosed with OC in 2011 while 1,750 are predicted to die from the disease (Canadian Cancer Statistics, 2011). The estimated lifetime risk of developing OC is 1 in 69 and the lifetime risk of dying from the disease is 1 in 92 (Canadian Cancer Statistics, 2011). The majority of cases occur in older women with median and mean ages at diagnosis of 60 and 59, respectively (Cannistra 2004; Kerlikowske *et al.*, 1994). The disease burden is not unique to Canada as most of the Western world suffers from similar rates of incidence and mortality. This can be seen in the estimated 22,280 new OC cases and 15,460 attributed deaths in the United States of America (Siegel *et al.*, 2012).

Two primary issues contribute to the poor outcome of affected women. The first challenge is that the majority of women are diagnosed with disease that has already metastasized beyond the ovary (Siegel *et al.*, 2012). This is the result of vague and common disease symptoms in conjunction with an absence of reliable screening methods (Ozols 2002; Seidman *et al.*, 2004; Stirling *et al.*, 2005). If this problem was overcome and women were diagnosed while their disease was still confined to the ovary, it is postulated that 5-year survival rates

would likely more closely reflect the 90% survival observed in stage one disease (Buys *et al.*, 2011).

The second challenge is that in the front-line use of platinum and taxane-based therapies, only half of patients exhibit complete responses, and those with advanced stages will eventually relapse (Pliarchopoulou & Pectasides, 2010; Vaughan *et al.*, 2011). With each relapse, patients are less likely to respond to treatment as drug resistance often develops (Yap *et al.*, 2009). While a number of targeted therapies have been developed recently, biomarkers will be needed to identify which patients will benefit from a given chemotherapeutic agent (Yap *et al.*, 2009). This along with the identification of new targets makes furthering the understanding of OC biology paramount in efforts to improve outcome for women affected by the disease (Vaughan *et al.*, 2011).

A number of factors can influence an individual's risk of developing OC. The strongest risk factor outside of increased age (Yancik, 1992) is a family history of breast or ovarian cancer, where most of the risk is the result of inherited mutations in the BRCA1/BRCA2 genes (Sueblinvong and Carney, 2009). However, only 5-10% of cases are hereditary and the majority of cases occur in patients without a family history of either disease (Lynch *et al.*, 1993; Risch, 1998). Factors that lower a woman's risk of developing OC include the number of full term pregnancies, prolonged use of oral contraceptives, breast-feeding, oophorectomy and tubal ligations (Tortolero-Luna and Mitchell, 1995; Hennessy *et al.*, 2009; Berek *et al.*, 2010). It is thought that these reproductive events can alter hormone levels affecting disease risk (Salehi *et al.*, 2008) however, the majority of risk is attributed to factors yet unknown.

1.2 Epithelial ovarian cancer classification

Complicating efforts to understand OC, the disease is very heterogeneous. ‘Ovarian cancer’ is an umbrella term for a range of tumors localized to the ovary. Tumors of an epithelial origin, termed epithelial ovarian cancer (EOC) comprise the majority (90%) of OC cases with tumors of stromal and germ cell origin each representing approximately 5% of all cases (Auersperg *et al.*, 2001; Ozols *et al.*, 2005). EOC is also heterogeneous. In addition to being graded based on the degree of cellular differentiation and staged by localization of the disease to the ovaries or lack thereof, it is also categorized based on histopathology (Benedet *et al.*, 2000). The major histopathological subtypes are defined as serous, endometrioid, mucinous, clear cell and undifferentiated adenocarcinomas (Seidman *et al.*, 2004). The histopathological subtypes display marked differences in morphology, molecular genetic abnormalities, response to therapy and potentially, site of origin (Shih and Kurman, 2004; Zorn *et al.*, 2005; Kobel *et al.*, 2008; Vaughan *et al.*, 2011). It has been proposed that the subtypes be considered as separate diseases in research as well as clinical settings (Kobel *et al.*, 2008). As cancers of the high-grade serous histopathological (HGSC) subtype represent the largest proportion of EOC cases (Seidman *et al.*, 2004), our group has focused on understanding the molecular genetics involved in this disease in the hopes of improving the diagnoses and treatment of affected women. While 5-year survival rates for patients are influenced by stage at diagnoses, response to first-line therapy and the amount of residual disease following cytoreductive surgery, a

considerable amount of variation in outcome is due to heterogeneity within HGSC itself (Kobel *et al.*, 2008; Berns and Bowtell, 2012; Hanrahan *et al.*, 2012).

1.3 Suspected origins of development

It has been widely accepted that EOCs arise from the single layer of epithelial cells lining the outer edge or “surface” of the ovary (ovarian surface epithelium – OSE) (Godwin *et al.*, 1992; Feeley and Wells, 2001). However recent studies have implicated secretory epithelial cells of the distal fallopian tube (FT) as a potential origin (Lee *et al.*, 2007; Crum *et al.*, 2007). This question has yet to be fully answered, as evidence in support of both theories exists (Hunn and Rodriguez, 2012). The activation of “protective” molecular pathways in OSE as a result of oral contraceptive use (Rodriguez *et al.*, 1998; Rodriguez *et al.*, 2002), the identification of premalignant dysplasia in the OSE (Scully, 1995) combined with coordinate tumor suppressor loss (Yang *et al.*, 2002; Roland *et al.*, 2003), and detection of non-malignant to malignant OSE transition in early grade cancers (Plaxe *et al.*, 1990) all provide support to the OSE as a site of origin. Evidence in support of the FT hypothesis includes increased risk for FT cancers in women with *BRCA1/BRCA2* mutations, indices of dysplasia in distal FT cells among healthy women, coincident involvement of FT lesions or carcinomas in EOCs, and similarities between FT and EOC histology (Auersperg *et al.*, 2001; Medeiros *et al.*, 2006; Lee *et al.*, 2007; Crum *et al.* 2007; Roh *et al.*, 2010; Chivukula *et al.*, 2011). It is possible that EOC could develop from epithelial cells of the ovarian surface or FT and neither possibility can be dismissed at this time.

1.4 Models to study EOC

Researchers have applied a variety of models to investigate the biology of EOC. EOC tumors represent an ideal source of genetic material because they reflect the disease as it manifests in the host. In the case of EOC where tumor masses are generally abundant (Eisenkop *et al.*, 2003), they provide the added benefit of being comprised of many tumor cells allowing for sufficient assay material. Researchers have successfully used tumor samples to identify mutations observed in the disease, genomic anomalies, and transcriptome profiles (Ouellet *et al.*, 2006; Wiegand *et al.*, 2010; TCGA, 2011; Wojnarowicz *et al.*, 2012). An alternative to examining the molecular genetic features of tumor specimens is the use of EOC cell lines. These include short-term cultures of tumor cells, long-term passages of tumor cells and oncogene transformed OSE cells (Garson *et al.*, 2005). While the transcriptome of cell lines can be affected by culture conditions (Leung *et al.*, 2001, Zorn *et al.*, 2003), they provide numerous benefits. For example, long-term passage cell lines derived from malignant ovarian tumor samples can be examined for the purpose of characterizing their molecular genetic properties, *in vitro* growth characteristics (anchorage independent growth, spheroid formation, wound healing), and tumorigenic potential in mouse tumor xenograft models (Garson *et al.*, 2005). They can also be genetically manipulated and examined for changes in phenotype.

1.5 OV-90: a unique EOC cell line

OV-90 is a long-term passage cell line derived from the ascites of a chemotherapy naïve patient (Provencher *et al.*, 2000). It exhibits many of the

somatic genetic features, such as somatic *TP53* mutation and chromosomal anomalies characteristic of over 90% of HGSC cases (Provencher *et al.*, 2000). OV-90 is able to grow in anchorage independent conditions in soft agar, forms three-dimensional spheroids with tight-junctions in hanging drop cultures, and is capable of forming tumors in mouse xenograft tumor models in *nude* and *scid* mice at both sub-cutaneous and intraperitoneal sites (Provencher *et al.*, 2000; Cody *et al.*, 2007; Tonin *et al.*, unpublished). Interestingly, OV-90 exhibits an aggressive disease course similar to that observed in the patient from which it was derived, with tumors forming rapidly in mouse xenograft tumor models and ascites present in the peritoneal cavity (Provencher *et al.*, 2000). Cell lines representative of HGSC from chemotherapy naïve patients are very rare, particularly those that develop tumors in mouse xenograft models (Ouellet *et al.*, 2005, Létourneau *et al.*, 2012). These features suggest that OV-90 may be a suitable cell line for which to study HGSC biology.

With the goal of identifying potential EOC tumor suppressor genes on the short (p) arm of chromosome 3, Dr. Tonin's group genetically manipulated a clone of OV-90 - OV-90neo - by introducing fragments of chromosome 3 in an attempt to complement its 3p hemizyosity and generate a series of cell lines with altered phenotypes (Cody *et al.*, 2007; Cody *et al.*, 2008). Three of the resulting OV-90neo cell line 'hybrids,' RH-5, RH-6 and RH-10, that were derived from the experiment were characterized. They exhibited a loss of anchorage independent growth, inability to form spheroids, and loss of tumorigenicity when assessed in mouse xenograft models (Cody *et al.*, 2007). These hybrid cell lines provide a

unique opportunity to investigate molecular mechanisms associated with tumor suppression.

1.6 Transcriptome analyses in the study of EOC

Transcriptome analyses are a powerful technique to investigate the biology of EOC. Expression profiling has revealed differences between tumors of different histopathological subtypes, varying malignancy and examined putative sites of disease origin (Ono *et al.*, 2000; Welsh *et al.*, 2001; Hibbs *et al.*, 2004; Santin *et al.*, 2004; Bignottie *et al.*, 2006). Transcriptome analyses have also examined gene expression differences between EOC tumors and normal reference tissues in comparative analyses to identify molecular pathways deregulated as a consequence of the disease (Hibbs *et al.*, 2004; Lu *et al.*, 2004; Shridhar *et al.*, 2001; Santin *et al.*, 2004; Welsh *et al.*, 2001; Ono *et al.*, 2000; Bonome *et al.*, 2005; Birch *et al.*, 2008; Wojnarowicz *et al.*, 2008). One of the challenges arising from these comparative analyses is inter and intra-tumor heterogeneity that contributes to large candidate lists that are difficult to reproduce (Le Page *et al.*, 2004; Tinker *et al.*, 2006). One strategy used in cancers other than EOC has been to narrow the focus of transcriptome analyses by incorporating *in vitro* cell line models which bear important phenotypes of the disease in question (Segal *et al.*, 2005). Our research group has characterized the long-term passage EOC cell line OV-90 to determine if it can serve in a similar role as an appropriate model for HGSC biology (Provencher *et al.*, 2000; Ouellet *et al.*, 2005; Létourneau *et al.*, 2012).

In the case of comparative analyses, the selection of normal disease-free reference material requires consideration as the choice of cell type and growth conditions impacts gene expression profiles identified in tumor models examined (Auersperg *et al.*, 2001; Zorn *et al.*, 2003). The selection and use of normal references has evolved along with our understanding of molecular genetics, EOC biology, and tissue availability. The relative accessibility of OSE cell primary cultures and the ease of obtaining sufficient quantities of RNA make them an appealing choice (Zorn *et al.*, 2003). Unfortunately, these cells, as well as exogenously immortalized OSE cells, are affected by culture conditions and in the case of the latter; the immortalization processes itself (Leung *et al.*, 2001, Zorn *et al.*, 2003). Alternatively, whole ovary can be used with the benefit of avoiding culture conditions and providing adequate RNA with the drawback of being largely comprised of stromal cells that do not reflect the OSE biology (Zorn *et al.*, 2003). OSE cytobrushings capture only the epithelial cells surrounding the ovary but often do not yield sufficient RNA for transcriptome analysis without undergoing amplification cycles (Bonome *et al.*, 2005). In the case of normal FT samples, alternate tissue preparation post-surgery is required due to a mixture of cells beyond the secretory epithelial cells of suspected OC origin (Marquez *et al.*, 2005). As a result, these samples are still difficult to acquire. To provide the most amount of information, EOC researchers often incorporate combinations of models in their studies.

1.7 Transcriptome analyses of the OV-90 hybrid cell lines

A comparative transcriptome analysis of the parental OV-90neo cell line and the hybrids RH-5, RH-6 and RH-10, identified 1204 probe sets as differentially expressed between OV-90neo and the hybrids (Cody *et al.*, 2007). A review of the reprogrammed genes represented by the 1204 probe sets indicated their implication in OC (Cody *et al.*, 2007). To further explore the genes identified as reprogrammed in the hybrids, the 1204 probe sets identified as differentially expressed between OV-90neo and the hybrids (Cody *et al.*, 2007) were examined within a series of 17 OSE primary cultures and 17 HGSC samples (Quinn *et al.*, 2009). This revealed that a number of the genes de-regulated in the hybrids were also differentially expressed in HGSC samples when compared to primary cultures of OSE (Quinn *et al.*, 2009). Nearly 70% of these de-regulated genes exhibited patterns of expression in the hybrids which, based on the phenotype of tumorigenicity, suggested the hybrids were more ‘normal’ compared to OV-90neo, as these genes displayed patterns of expression in the hybrids similar to that observed in normal OSE cells (Quinn *et al.*, 2009). To verify these expression patterns, 30 genes were investigated by RT-PCR in the 17 OSE and HGSC samples (Quinn *et al.*, 2009). Several of these genes have been previously described in OC, indicating the relevance of the OV-90 hybrid cell lines to the study of HGSC biology (Quinn *et al.*, 2009). Limitations of this comparative transcriptome analysis included the small number of HGSC samples investigated, the integration of gene expression data from an early generation, lower density expression microarray and the use of OSE short-term cultures in the comparative analysis (Quinn *et al.*, 2009).

1.8 Project hypothesis and objectives

The objective of my thesis is to address the limitations of the aforementioned studies and test the hypothesis that the OV-90 and hybrid cell lines comprise a suitable model for the study of HGSC. To this end, I have extracted gene expression data associated with the genes de-regulated in the OV-90 hybrids from a larger and independently derived gene expression data set of HGSC samples and uncultured OSE cytobrushings (Bonome *et al.*, 2005; <http://www.ebi.ac.uk/arrayexpress/>). I then performed a comparative analysis of gene expression profiles to identify those genes exhibiting consistent differences in expression between HGSC and OSE samples from this pre-defined gene list. To aid the interpretation of the resulting candidate list, I have utilized bioinformatic resources. These tools integrate vast stores of biological information associated with known genes, allowing researchers to upload gene lists of interest and interrogate them for relationships with biological functions, pathways, molecular networks, diseases and disorders (Huang *et al.*, 2009; Jimenez-Marin *et al.*, 2009; Pitteri *et al.*, 2009; Helleman *et al.*, 2010). By identifying pathways or processes associated with the genes transcriptionally altered as a result of the abrogation of tumorigenicity in the OV-90 derived hybrids, insight may be gained into molecular mechanisms important in the disease biology identifying biomarkers or therapeutic targets. Having applied these programs and a survey of the literature to the genes identified in our comparative transcriptome analysis, I described the associated features of each gene and the pathways implicated. With the assistance of Dr. Mes-Masson's group, I have also characterized the protein expression of a candidate gene in a

series of HGSC samples and examined the relationship between protein expression and patient outcome.

The hypothesis of this research is that further characterization of the genes de-regulated in the hybrids has the potential to identify important disease pathways and provide additional evidence of their utility as a model of HGSC (Seitz *et al.*, 2006; Klebig *et al.*, 2005; Stronach *et al.*, 2003).

The aims of this thesis were to: 1) examine the genes represented by the 1204 probe sets differentially expressed in the OV-90 derived hybrids (RH-5, RH-6, RH-10) in a larger, independently derived and publically available series of HGSC (n=53) and OSE cytobrushings (n=10) samples and 2) acquire information on implicated genes through bioinformatics using established programs to inform disease biology and identify areas for future research.

2 MATERIALS AND METHODS

2.1 Derivation of gene expression data for comparative transcriptome analysis

The gene expression data file GEOD-18520 in MAS5.0 format (hereafter referred to as #18520) was obtained from the ArrayExpress Archive (<http://www.ebi.ac.uk/arrayexpress/>). This file contains data derived from 10 OSE cytobrushings and 53 laser captured microdissected late stage HGSC tumors that were assayed using the Affymetrix GeneChip® U133 Plus 2.0 microarray as previously described (Bonome *et al.*, 2005). Data was normalized as previously described (Cody *et al.*, 2007). This GeneChip® contains 54,613 probe sets that map to 47,000 transcripts representing 39,500 genes (www.affymetrix.com). The

gene expression data file containing normalized data for OV-90neo, RH-5, RH-6 and RH-10 cell lines used to identify the 1204 probe sets differentially expressed three-fold between OV-90neo and the hybrids was described previously (Cody *et al.*, 2007; Quinn *et al.*, 2009). This MAS5.0 formatted data was also assayed using the Affymetrix GeneChip[®] U133 Plus 2.0 microarray allowing for matched probe set comparisons of gene expression between datasets.

2.2 Hierarchical clustering analysis

Hierarchical clustering was performed using selected normalized gene expression data derived from the 53 HGSC and 10 OSE samples (#18520). Only expression data corresponding to the 1204 probe sets differentially expressed between OV-90neo and the hybrids (RH-5, RH-6 and RH-10) were used in this analysis (Cody *et al.*, 2007). The Multiple Experiment Viewer (Saeed *et al.*, 2006) was used to perform the unsupervised clustering under default settings with Pearson correlation as the distance metric and average linkage clustering as the linkage method.

2.3 Derivation of candidate gene list from comparative transcriptome analysis

The expression data for the 1204 probe sets differentially expressed between OV-90neo and the hybrids were extracted from the gene expression data for the 10 OSE cytobrushings and 53 HGSC samples present in the #18520 dataset. Comparing the expression of each of the 53 HGSC samples to the mean of the OSE cytobrushings (n=10), we selected for probe sets where greater than

75% of the HGSC samples displayed at least a three-fold change in the same direction relative to OSE mean. This was an arbitrary cut-off selected to capture expression patterns that were representative of the majority of HGSC samples. The selected probe sets and corresponding genes represent the result of the first thesis objective.

2.4 Gene annotation of probe sets

Probe sets were annotated using NetAffxTM Analysis Center Batch Query version 32 (www.affymetrix.com/analysis/index.affx), and supplemented by the Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.7, Huang *et al.*, 2009; Huang *et al.*, 2009*) and Ingenuity Pathway Analysis (IPA, Ingenuity® Systems, build number 124019, www.ingenuity.com) bioinformatics programs.

2.5 Bioinformatics analyses using software tools

DAVID (Huang *et al.*, 2009; Huang *et al.*, 2009*) and IPA (Ingenuity® Systems, www.ingenuity.com) bioinformatics software tools were used to identify gene-associated molecular functions that were over-represented in the gene list generated by comparative analyses of the HGSC and OSE gene expression data. In DAVID, the Gene Functional Classification Tool was used to group genes with functional similarity via probe set entry under default settings with background selected for the probe sets represented on the Affymetrix U133 Plus 2.0 GeneChip®. Identifying the microarray platform used to generate the candidate list allows the software to evaluate the enrichment of related genes or associated

annotations among only the genes that could potentially be selected, as opposed to all genes in the human genome including those not represented on the given microarray. Functional Annotation Clustering was also performed using the same parameters.

An Ingenuity Pathway Core Analysis (Ingenuity® Systems, www.ingenuity.com) was performed on the candidate list with default setting and background set to the probe sets represented on the Affymetrix U133 Plus 2.0 GeneChip®.

2.6 Survey of the candidates identified by comparative transcriptome analysis

Expression data for 506 HGSC samples (relative to normal whole FT samples) as reported by The Cancer Genome Atlas Research Network (TCGA, 2011), was accessed through the TCGA Data Portal (tcga-data.nci.nih.gov/tcga/). This data was generated using the Custom Agilent 244K Gene Expression Microarray (TCGA, 2011). The log₂ tumor/normal ratios were extracted for gene candidates from the TCGA Data Portal. No information was available for *AKRIC2*, *FAM155A*, *CFC1*, *GPR133*, *TWIST2* and *LOC100128893*. We used this data to construct histograms of the relative gene expression levels of our candidates across the 506 HGSC samples surveyed by the TCGA. These patterns were then compared to those observed in the #18520 dataset used in our comparative transcriptome analysis. We examined how representative the expression patterns observed in our study are to a larger sample of HGSCs, as well as what effect the use of whole FT as normal reference may have on

expression patterns. Patterns observed in the TCGA samples (2011) were considered consistent with the #18520 dataset if at least two-thirds (358) of the HGSC samples displayed the same pattern of over or under-expression (relative to normal reference) observed in the #18520 dataset. Genes where greater than two-thirds of the samples had the opposite pattern were considered inconsistent and genes that failed to meet either designation were labeled as having no clear profile.

In addition to the expression data gathered from the TCGA Data Portal, we also examined our candidates in light of mutation and copy number alteration data presented in their study (TCGA, 2011). Somatic mutations detected via whole genome exome sequencing in 316 HGSC surveyed by the TCGA were extracted along with regions of focal or regional copy number alterations identified in the 489 HGSC genomes analyzed for copy number changes via Agilent 1M Human Genome CGH (comparative genomic hybridization) Microarray as reported in TCGA (2011). Somatic mutation information was used to examine the frequency of mutation in our gene candidates. Copy number alterations defined as focal or regional (affecting whole chromosome arms) in the HGSC genomes analyzed were used to examine copy number events that map to our gene candidates (TCGA, 2011).

2.7 Western blot analysis

Western blotting was performed on the OV-90 and hybrid cell lines. The mouse monoclonal antibody against Ceruloplasmin (LF-MA0159) was purchased from Young in Frontier co. (Seoul, Korea). Approximately equal amounts of total

protein extracted from OV-90, OV-90neo, RH-5, RH-6 and RH-10 were loaded on a SDS-polyacrylamide gel, and following electrophoresis, transferred onto a nitrocellulose membrane using standard techniques. Membranes were blocked with 5% milk and probed with anti-ceruloplasmin (LF-MA0159, Young in Frontier co., Seoul, Korea) at a dilution of 1:100. The primary antibody was detected with a conjugated HRP secondary antibody and visualized by enhanced chemiluminescence. Equal protein loading was confirmed by reprobing the membrane with anti-actin.

2.8 Immunohistochemistry analysis

With the assistance of Dr. Mes-Masson's research group, protein expression of the candidate in HGSC was described using a tissue microarray (TMA) of HGSC and normal FT samples as per established protocols (Le Page *et al.*, 2012). The TMA used contained 0.6 mm diameter cores of 260 formalin fixed paraffin embedded HGSC tumor samples and 11 normal FT samples as previously described (Le Page *et al.*, 2012). It was sectioned at 4 μ m and stained using the BenchMark XT automated stainer (Ventana Medical System Inc.). The optimal concentration for anti-CP was determined by serial dilution. With the automated stainer, antigen retrieval was carried out using Cell Conditioning 1 (Ventana Medical System Inc.; #950–124) for 30 minutes. Prediluted primary CP antibody (1 in 100) was applied to the TMA and incubated for 60 minutes. The UltraView DAB detection kit was used for primary antibody capture (Ventana Medical System Inc.; #760–091). Counterstaining was performed with hematoxylin

(Ventana Medical System Inc.; #760–2021). All sections were digitally scanned using a 20x 0.75 NA objective with a resolution of 0.3225 μm .

The scanned HGSC TMA was visualized and the epithelial tissue component scored according to the staining intensity of the cellular membrane and cytoplasm (0 for absence, 1 for weak, 2 for moderate) (Le Page *et al.*, 2012). In cores where staining was of variable intensity, the higher intensity was reported. Two observers independently scored the results. Differences between observers greater than 1 unit per core were re-evaluated to reach inter-observer concordance. Inter-rater correlation was greater than 75% and an average of the two observer scores was used in subsequent analyses.

Clinical data such as tumor grade, stage, progression-free survival and overall survival was available for 196 samples represented on the TMA (Le Page *et al.*, 2012). Progression-free survival was defined as time from surgery and first progression based on scan imaging and blood CA125 level, while overall survival was defined as time from surgery to death from OC.

The Pearson correlation test (two-tailed) was used to estimate the correlation between CP protein expression scores and the clinical measures of progression-free survival and overall survival intervals. Kaplan-Meier curve analyses and the log-rank test were used to measure significant differences. Receiver operative characteristic curves failed to determine a threshold value for CP that optimized sensitivity and specificity for patient progression so the median CP score was used. Univariate Cox proportional hazards modeling was used to estimate the hazards ratio for CP expression with respect to progression-free and

overall survival. All statistical analyses were performed using the Statistical Package for the Social Sciences software version 11.0 (SPSS, Inc.).

Ethical approval for the collection and banking of tissue specimens as well as criteria used in assessing specimen quality and clinical correlate measures were as previously described (Le Page *et al.*, 2012).

3 RESULTS

3.1 Hierarchical clustering

To assess the disease relevance of the 1204 probe sets and 843 represented genes previously identified as differentially expressed between OV-90neo and the hybrids (Cody *et al.*, 2007), their expression levels in the 53 HGSC and 10 OSE samples of the #18520 dataset were examined. An unsupervised hierarchical clustering analysis was performed on the 53 HGSC and 10 OSE samples using expression data from only the 1204 probe sets. The experiment revealed the ability of the genes represented by these probe sets to cluster HGSC samples separately from OSE samples (Figure 1). While it can be observed that OSE samples are clustered in a separate branch from the HGSC samples, a number of genes exhibited similar expression patterns in both OSE and HGSC samples. Within the separate groups of OSE and HGSC samples, subclusters of samples resulted from differences in gene expression levels among samples within these larger groups. Some genes also illustrated expression levels in a subset of OSE samples that were more consistent with those seen in HGSC samples and vice versa. While these expression patterns illustrate the

heterogeneity of the gene expression levels across OSE and HGSC samples for the 1204 probe sets (and 843 represented genes) identified as differentially expressed between OV-90neo and the hybrids (Cody *et al.*, 2007), the most important observation is the ability of these genes to cluster OSE and HGSC samples into separate groups in an unsupervised analysis. This suggests the hybrid cell lines may represent an aspect of HGSC biology, particularly at the level of the transcriptome.

3.2 Comparative transcriptome analysis

Among the genes differentially expressed as a result of potential transcriptional reprogramming in the OV-90 hybrids, we sought to identify those that were also differentially expressed between the HGSC and OSE samples of the #18520 dataset. A schematic of the filtering process starting from the approximately 54,000 probe sets on the Affymetrix U133 Plus 2.0 GeneChip[®], down to the 106 probe sets that comprise our candidate list can be seen in Figure 2. It outlines the selection of the 1204 probe sets previously defined by their association with genes exhibiting at least a 3-fold difference in expression between the tumorigenic OV-90neo cell line and the genetically modified, non-tumorigenic hybrids, RH-5, RH-6 and RH-10 (Cody *et al.*, 2007). From these 1204 selected probe sets, we identified 106 probes that exhibited at least a 3-fold difference in expression value between >75% of the HGSC samples relative to the OSE mean (Table 1). These arbitrary selection criteria were chosen to identify genes with clear and consistent expression differences between HGSC and OSE samples with the aim of identifying a candidate list of genes representative of the

disease. The 106 probe sets and the 92 genes they represent define our candidate list of genes differentially expressed in both the OV-90 hybrid cell lines and the HGSC/OSE samples examined. This list reflects the accomplishment of our first objective and the remainder of this thesis addresses the second objective of characterizing these genes and their relevance of HGSC biology.

3.3 Differentially expressed gene categories

Within the 92 genes of our candidate list, four expression patterns can be used to categorize genes based on their relative levels of expression across OV-90neo, the hybrid cell lines, HGSC and OSE samples. The branches of the flowchart presented in Figure 2 terminate into these four categories and present the number of candidate genes in each group. These categories were developed to provide a visual way of interpreting the expression patterns seen in these four sample groups and how the patterns can be related to the phenotype of tumorigenicity as it is observed in these samples.

Figure 3 presents histogram representations of the expression patterns across cell lines and tissue samples for each of the four categories with annotation of sample tumorigenic potential where (T) indicates tumorigenic potential and (N) represents non-tumorigenic samples. Category one genes have higher expression levels in the tumorigenic (T) OV90 and tumor samples than in the non-tumorigenic (N) hybrids and normal samples. In category three, genes have higher expression levels in the hybrids and normal (N) samples compared to OV90 and tumor (T) samples. Both category one and three genes illustrate expression patterns where the relative level of expression is correlated to the tumorigenic

potential of the sample across both the cell lines and the HGSC and OSE samples. These genes are represented by 46% of the 106 probe sets examined and their histograms have been shaded purple in Figure 3.

Category two and four genes do not show the same consistent association with the phenotype of tumorigenicity across both cell lines and tissue samples (green histograms in Figure 3). For example, category two genes display higher levels of expression in OV90 (T) compared to the hybrids (N) but lower levels in the tumors (T) compared to normal samples (N). In category four, the pattern is reversed with lower expression in OV90 (T) compared to the hybrids (N) but higher expression in the tumors (T) compared to normal tissue (N).

3.4 Bioinformatics analysis of differentially expressed genes using computational tools

To determine whether certain biological processes were enriched within the 92 genes of our candidate list, we applied two computational methods using DAVID (Huang *et al.*, 2009, Huang *et al.*, 2009*). With platform background set to the Affymetrix U133 Plus 2.0 GeneChip[®], all genes represented by the 106 probe sets were recognized by the program with the exception of *LSAMP*. Gene Functional Classification yielded one group of functionally related genes with a significant enrichment score of 2.04 ($p=0.00912$) (Table 2). These nine genes had associated terms related to cell signaling functions such as glycoprotein, transmembrane region, integral-to-membrane, G-protein coupled receptor, signal peptide, disulfide bond, transducer and others. Functional Annotation Clustering revealed seven clusters with significant enrichment scores (Appendix I).

Consistent with the results of Gene Functional Classification, the terms associated with the top scoring Functional Annotation Cluster describe glycoprotein signaling while others can be summarized as the following: limb development, mammary gland development, oxio-reductase activity, protein processing, extra-cellular signal response and innate immune response.

To further explore potential biological networks associated with the 92 differentially expressed genes, we also applied IPA (Ingenuity® Systems, www.ingenuity.com), as this resource provides similar but more detailed bioinformatics outputs. IPA recognized 89 of the 92 genes represented by the 106 probes of our candidate list, but the genes *GATA6*, *LGR4* and *NRG4* were not recognized by the program's annotation database. The results of a Core Analysis identified a number of biological functions significantly associated with the candidate genes identified in our study (Table 3, complete list in Appendix II). The most significant of these associated functions included carcinoma, cancer, adenocarcinoma and tumorigenesis ($p < 5.08 \times 10^{-8}$). Other significantly associated functions more specific to OC included reproductive system disorder, gynecological disorder and invasion of ovarian cancer cell lines. Also present in the top 20 biological functions were those associated with cancer phenotypes such as invasion of cells, proliferation of cells, colony formation of cells, and cell movement.

Canonical Pathways analysis revealed several statistically significant signaling pathways associated with our candidate list (Table 4, complete list in Appendix III). The five most significantly implicated pathways ($p < 0.0134$) include G-protein coupled receptor signaling, C21-steroid hormone metabolism,

androgen and estrogen metabolism, cAMP-mediated signaling and clathrin-mediated endocytosis signaling. Ovarian cancer signaling was also a statistically significant pathway containing the gene candidates *FGF9*, *FZD7* and *PRKAR2B* ($p=0.0398$).

Eight molecular networks associated with our candidate list were identified in the network analysis output (Table 5), with top functions including cancer and related processes such as cellular movement, DNA replication, recombination and repair, as well as cellular growth and proliferation.

3.5 Interrogation of TCGA analysis of HGSC

Given the scale and comprehensive nature of the study on HGSC by The Cancer Genome Atlas Research Network (TCGA, 2011), we sought to examine our 92 gene candidates with respect to their gene expression, mutation analyses, and copy number aberration results.

Gene expression data was available for 86 of the 92 candidates. This data was presented as log2 of the HGSC/whole FT ratio. Histograms of the gene expression values for the 506 HGSC samples assayed were created to allow the visual inspection of expression patterns (Appendix IV). We compared the expression patterns of our gene candidates in the #18520 dataset to these results. Using the criteria that at least two-thirds of the TCGA HGSC samples ($n>357$) must show the same pattern of expression (either higher or lower expression in HGSCs compared to normal reference) observed in the #18520 dataset to be considered consistent, we found that 62 of the 86 (72%) genes examined had expression patterns consistent between datasets. If greater than two-thirds of the

HGSC samples had an opposite expression pattern to that seen in the #18520 dataset, they were considered to have an inconsistent expression pattern. Only *STS*, *F2R*, *FGF9*, *NPL*, and *TNNT3* had inconsistent patterns of expression between the two datasets. A larger number of genes (n=19) failed to meet either cut-off and had no clear trend towards either consistent over, or under-expression in the HGSC samples relative to the FT reference. The TCGA group identified a 193-gene signature predictive of outcome however there was no overlap between this list and the 92 genes identified by our analysis (TCGA, 2011).

Of the 18,500 genes screened for somatic mutations in 316 HGSC samples, the TCGA reported 19,356 instances of somatic mutation (TCGA, 2011). Forty-nine of the 92 genes identified in our analysis had intragenic mutations with a total of 108 instances of somatic mutation among the 316 samples examined. The majority of mutations were classified as single nucleotide variants (n=94) with only eight cases of indels variants. Of our gene candidates, the most frequently mutated gene was *DMD* (n=9) followed by *NES* (n=6) and *AOX1*, *BCHE*, *LAMA4*, *PAPPA*, *TNXB*, *CGNL1*, *PRICKLE2*, and *GPR133* each mutated in four samples. *ALDH1A1*, *DPP4*, *F2R*, *SLC26A4*, *TBX3*, *SLC16A4*, and *PCDH17* were mutated in three samples each, while *MECOM*, *ME1*, *RGS4*, *VIM*, *FZD7*, *SORBS2*, *FAM134B* and *ARHGAP18* were mutated in only two samples each. The remaining gene candidates reported as somatically mutated were only observed in one HGSC sample. The frequency of mutation for our gene candidates is consistent with what would be expected by chance, given the baseline mutation rate observed in this study (TCGA, 2011).

Copy number alterations identified in the 489 HGSC genomes analyzed were categorized as either regional aberrations spanning chromosome arms, or focal copy number changes in the TCGA study (TCGA, 2011). Of the 92 genes identified in our analysis, 29 were located on chromosome arms that exhibited statistically significant recurrent gains in copy number (TCGA, 2011). Conversely, 47 gene candidates were located on chromosome arms that exhibited losses in copy number (Table 6). Five gene candidates mapped within the 63 regions identified as focally amplified (*EPCAM*, *MECOM*, *AKRIC2*, *TSPAN8*, *LYZ*). Nine candidate genes mapped within one of the 50 regions of focal deletion (*SLC16A4*, *SNCA*, *CFI*, *CAMK2D*, *F2R*, *LAMA4*, *TNNT3*, *FGF9*, *AXL*).

The gene expression results of the TCGA study support the notion that the expression patterns observed for our candidate genes are representative of HGSC. Mutation and copy number analysis results for our 92 candidates are consistent with the high frequency of somatic mutation and genomic instability characteristic of the disease biology.

3.6 Protein expression analysis of a candidate gene and correlation with clinical parameters

In an effort to expand our knowledge of HGSC biology, we sought to characterize a gene candidate by examining its protein expression in HGSC. As seen in Figure 4 A, ceruloplasmin (*CP*) is a gene with robust mRNA expression in OV-90neo, but expressed at much lower levels in the hybrid cell lines (Cody *et al.*, 2007; Quinn *et al.*, 2009). The repeated observation of elevated *CP* mRNA levels in HGSC tumors relative to normal controls (Figure 5) along with antibody

availability made *CP* a strong candidate for further analysis (Quinn *et al.*, 2009; Quinn *et al.*, 2009; TCGA, 2011; Tonin *et al.*, unpublished; Axela, www.axelabiosensors.com). To confirm antibody specificity, western blot analysis was performed using protein extracts from OV-90, OV-90neo and the hybrid cell lines. CP was detected at approximately 132 kDa as expected in OV-90 and OV-90neo (Figure 4 B). Consistent with the low mRNA levels observed previously, no protein expression was detected in RH-5, RH-6 or RH-10 (Cody *et al.*, 2007; Quinn *et al.*, 2009).

We utilized immunohistochemistry to assay the CP protein expression of 196 HGSC tumor samples contained on a TMA for which corresponding patient overall and progression-free survival data was available (Table 7). Also present on the TMA were 11 normal FT samples. After eliminating samples with insufficient material within the core to score, 165 HGSC and 8 FT samples were examined for CP expression. All eight FT samples examined displayed positive staining, indicating the presence of CP. Staining in the HGSC tissues was scored in the epithelial compartment as either absent, weak or moderate (0, 1 or 2) for descriptive purposes. No samples displayed intense levels of staining despite the high levels of mRNA expression noted in HGSC samples previously by our group (Quinn *et al.*, 2009; Quinn *et al.*, 2009; TCGA, 2011; Tonin *et al.*, unpublished; Axela, www.axelabiosensors.com). Examples of staining can be seen in Figure 6. Of the 165 HGSC samples, 113 (68%) displayed some degree of staining with the majority of samples (53%) falling into staining categories of 0.5 and 1.0 (average intensity scores).

To determine whether CP protein expression is associated with progression-free or overall survival in patients with HGSC, we performed a Kaplan-Meier and Cox proportional hazards model test on the 165 HGSC samples for which staining and clinical correlate information was available. Kaplan-Meier curves demonstrated no significant association between CP protein expression and overall survival in this cohort (Figure 7 A). However, CP expression was found to significantly associate with increased progression-free survival ($p=0.019$; log rank=5.535) as demonstrated in Figure 7 B. The mean progression-free survival interval for patients with detectable CP expression was 42.9 months compared to 19.7 months for patients lacking CP expression.

Univariate Cox regression analysis with CP expression as a categorical variable (on, off) reflected the relationship between CP protein expression and increased progression-free survival. Detectable CP protein was associated with a hazards risk of 0.641 for progression-free survival indicating the better outcome for patients with CP expression (95% CI 0.438-0.939, $p=0.022$). No significant association was observed between CP protein expression and overall survival. Spearman's correlation test revealed no correlation between CP protein expression and standard prognostic variables indicating CP protein as independent indicator of progression-free survival.

4 DISCUSSION

4.1 Validation of the hybrids as a model for the study of HGSC

Since the development of the hybrids we have performed research to characterize these cell lines, and determine if they are representative of HGSC

biology (Cody *et al.*, 2007). The initial transcriptome analysis of the hybrids revealed that despite the drastic change in phenotype between OV-90neo and the hybrids RH-5, RH-6 and RH-10, there were few significant differences in gene expression between them as assessed by two-way comparative analyses (91-95% correlated) (Cody *et al.*, 2007). However, among the 1204 probe sets differentially expressed at least three-fold between OV-90neo and the non-tumorigenic hybrids, several EOC related genes including *CAVI*, *DAB2*, *SFRP1*, *CDH1*, *EVI-1*, *CP* and others were represented (Table 2 in Cody *et al.*, 2007). An overlap of pathways containing these genes, combined with the coordinated transcriptional reprogramming observed in all three of the hybrids as a consequence of tumorigenic suppression and alteration of growth characteristics, suggested that molecular networks characteristic of EOC were modulated as a result of chromosome 3 fragment transfer (Cody *et al.*, 2007).

Subsequent research examined the 1204 probe sets and 843 corresponding genes differentially expressed between OV-90neo and the hybrids in a comparative transcriptome analysis of normal OSE primary cultures and HGSC tumors (Quinn *et al.*, 2009). Selecting a subset of the 1204 probe sets also differentially expressed between 17 HGSC and 17 OSE primary cultures resulted in a candidate list of 374 probe sets. Of these, 70% displayed expression patterns that when examined in both the OV-90neo/hybrid cell lines and the OSE and HGSC samples, related to the tumorigenicity of the samples (Quinn *et al.*, 2009). For example, a gene with higher expression in tumorigenic OV-90neo compared to the non-tumorigenic hybrids had a higher expression in the HGSC samples compared to the OSE cells. A number of genes represented by these 374 probe

sets had been previously implicated in EOC. Examples included *TACSTD1*, *SCNN1A*, *CDH1*, *FLH2*, *AXL*, *ELF3* and *DAB2* (Quinn *et al.*, 2009). Other genes such as *GREM1*, while not implicated in EOC, had been described in other cancer types, or in the cases of *EVI-1*, *INHBA* and *FSTL1*, are involved in cancer-associated processes such as TGF-*beta* signaling (Quinn *et al.*, 2009). These findings further indicated that the reprogramming of OV-90neo that occurred as a result of chromosome 3-fragment transfer affected molecular pathways relevant to HGSC.

The present study has expanded on the aforementioned research to test the hypothesis that further characterization of the genes de-regulated in the hybrid cell lines can identify important disease pathways and provide additional evidence of their utility as a model of HGSC. In our comparative transcriptome analysis, a larger independently derived series of HGSC samples and uncultured OSE cytobrushings as the normal reference provided an increased sample size and eliminated potential effects due to cell culture (Cody *et al.*, 2008, Ferley *et al.*, 2008). This dataset also had the advantage of being captured on the same U133 Plus 2.0 Affymetrix GeneChip[®] used to assay the OV-90neo - hybrid cell line transcriptomes allowing all probe sets to be compared directly with no information lost in the comparative transcriptome analysis. By incorporating bioinformatic resources we were able to identify molecular pathways, processes and biological functions statistically enriched in our candidate list. Our use of immunohistochemistry to investigate a candidate gene's protein expression in a series of clinically annotated HGSC samples provided information on both the gene's protein expression in HGSC, as well as it's potential as a disease

biomarker. The following discussion of results highlights the evidence in support of our previous findings and illustrates that the hybrids are a useful model for the study of HGSC biology.

Prior to performing our comparative transcriptome analysis, we examined the 1204 probe sets differentially expressed between OV-90neo and the hybrids within the #18520 HGSC dataset (Cody *et al.*, 2007). Unsupervised hierarchical clustering analysis produced interesting results (Figure 1). The ability of the 1204 probe sets to separate OSE and HGSC samples into different groups suggests that the genes represented by these probe sets capture a transcriptional profile representative of the disease. While this experiment did not address what role these genes play in the disease with respect to function, it illustrated the ability of these genes identified on the basis of a change in phenotype (loss of tumorigenic potential in the hybrids) to separate an independent series of HGSC and OSE samples into groups defined by the phenotype of tumorigenicity.

The hierarchical clustering experiment examined the relationship between all 1204 probe sets differentially expressed in the hybrids (Cody *et al.*, 2007) and their expression levels in the HGSC and OSE samples. However we wanted to focus on the genes among this list that were also differentially expressed between HGSC and OSE samples. This required the removal of genes that while differentially expressed between OV-90neo and the hybrids, displayed relatively similar expression levels between HGSC and OSE samples. The criteria we applied to identify our candidate list aimed to identify genes that displayed consistently different levels of expression between OSE cytobrushings and HGSC tumor samples. The candidate list that resulted from this comparative

transcriptome analysis included 106 probe sets capturing 92 known and hypothetical genes. Of these 92 genes, 34 were also identified in the Quinn *et al.*, (2009) study, despite the use of different microarray platforms, tumor samples, normal reference and methods used to define differential expression. The initial characterization of the #18520 dataset performed by Bonome *et al.*, (2005) identified 3605 probe sets significantly differentially expressed between the HGSC tumors and OSE cytobrushings. While their analysis involved a statistical measure of differential expression, more than half of our gene candidates (47), were present on this list. The results of the TCGA study comparison discussed later in this section provide further support to the accuracy with which our selection criteria have captured genes consistently transcriptionally perturbed in HGSC. The categorization of genes based on their relative mRNA expression levels across OV-90neo, the hybrids, OSE cytobrushings and HGSC tumor samples allowed us to make some important observations.

Figures 2 and 3, outline the non-uniform distribution of candidate genes across the four expression categories. It is interesting that while a similar number of probe sets were downregulated versus upregulated in the hybrids compared to OV-90neo (674 vs 530), over 81% of the 106 candidate probe sets (categories two and three) were downregulated in HGSC samples compared to OSE cytobrushings. This observation is not unique to our study as Quinn *et al.* noted that 61% of their candidate genes displayed lower expression levels in HGSC samples compared to OSE primary cultures (2009). In examining the link between expression levels and the phenotype of tumorigenicity, categories 1 and 3 comprise genes where expression patterns align with tumorigenicity in both the

cell lines and human tissues assayed. However, these categories represent only 45% of the gene candidates. This result is interesting given the ability of the 1204 probe sets differentially expressed in the hybrids to separate OSE and HGSC samples in a hierarchical clustering analysis. It is important to recall that while the hybrids are no longer able to form tumors in nude and *scid* mice, they still bear a number of the hallmarks of cancer. They have retained the complex genomic background of the parental OV-90 cell line including a mutant *TP53* and remain able to grow indefinitely in culture though several growth characteristics have been altered. The transcriptional reprogramming of these genes may represent pathways frequently perturbed in the development of the disease with the expression levels themselves dependent on the genetic background of the cells in question. In addition, it is critical to remember that the hybrids are based on a clone of a single EOC cell line inherently limiting the ability to capture all aspects representative of the disease. Thus, while the genes included in categories 2 and 4 may show expression patterns that do not correlate with tumorigenicity between the hybrid model and the HGSC/OSE samples, it is important to not discount the relevance of these genes to HGSC biology. The results of our bioinformatics analyses discussed below highlight this point.

As the top six biological functions associated with our candidate list by IPA are carcinoma, cancer, adenocarcinoma, tumorigenesis, reproductive system disorder and gynecological disorder, it appeared that the genes captured by our analysis were associated not only with cancer but the organ systems implicated in OC specifically (Table 3). Other statistically significant biological functions implicating either OC or gynecological malignancies included endometrial cancer,

genital tumor, invasion of ovarian cancer cell lines, uterine cancer, serous ovarian carcinoma process, ovarian tumor, and ovarian cancer. In addition to other cancer types implicated, a number of cancer related processes are described such as proliferation of cells, genetic disorder, colony formation of cells, cell movement, migration of cancer cells, metastasis, migration of cells, invasion of tumor cell lines, migration of endothelial cells, movement of tumor cell lines, proliferation of epithelia cells, and many others.

The molecular networks described by IPA provided a similar perspective with functions corresponding to the seven networks identified including terms such as cancer, cellular movement, gene expression, DNA replication, recombination and repair, cellular growth and proliferation, cell-to-cell signaling and interaction, cell death, and small molecule biochemistry. Three of the seven networks have cancer as a top related function.

Lastly, ovarian cancer signaling is one of 14 canonical pathways significantly enriched within our candidate list. Together, these results provide strong evidence to support the use of the hybrids as a tool in the study of EOC as well as continued investigation of the genes identified in this study. Given the association of the transcriptionally modified genes with the abrogation of tumorigenicity in the hybrids, it is possible that the genes may be directly involved in such pathways in HGSC.

4.2 Gene candidates identified in the TCGA study on HGSC

The importance of reproducibility to the confirmation of research findings led us to examine our candidates within the comprehensive study on HGSC

performed by the TCGA (2011). The most useful information provided by their study for this project was the mRNA expression data, because not only was a different technology platform used to assay gene expression, but FT samples were also used as a normal reference (TCGA, 2011). By examining the expression of the genes identified in our study in this data, we were able to assess how well our selection criteria achieved the goal of identifying genes with consistent expression patterns in HGSC. Furthermore we were able to examine the impact of using OSE cytobrushings as a normal control as opposed to FT. An examination of the histograms constructed to display the expression levels (log base-2 of the tumor/normal ratio) for the 86 genes with available data revealed very good concordance with the expression patterns observed in our study (Appendix IV). Of the 11 category one genes with higher expression in HGSC than OSE, seven also had higher expression in the HGSC samples examined by TCGA when compared to FT (TCGA, 2011). Similarly 22 of the 29 category three genes display the same pattern of higher expression in HGSC compared to normal reference. This concordance is also noted in category two and four genes where of the 46 genes with expression data available, 33 have expression patterns consistent with the lower expression in HGSC observed in our study. These results provide evidence to suggest the genes identified in our study are representative of patterns seen in HGSC generally.

Given the success of our selection criteria in identifying genes with consistent patterns of differential expression in HGSC samples relative to normal reference tissues, it is not surprising that none of the genes identified in this study were present in the 193-gene signature predictive of outcome described by the

TCGA (2011). The variation in expression needed for genes to be capable of stratifying patients across any parameter (clinical or otherwise) was partially precluded by our goal - as reflected in our selection criteria - of identifying genes with consistent expression patterns across HGSC.

An examination of the somatic mutation data generated by the TCGA study (2011) revealed that over half of our gene candidates were mutated in at least one of the 316 HGSC samples examined. Seventy-six of our 92 genes map to chromosome arms that experienced statistically significant losses or gains of copy number while 14 gene candidates were mapped to regions that exhibited statistically significant focal amplifications or deletions. The large number of candidates somatically mutated at frequencies consistent with the baseline mutation rate suggests they are not being targeted for deactivation by somatic mutation. However, in examining the nine genes identified as significantly mutated, several genes display fewer than ten instances of mutation, illustrating the impact that mutation distribution modeling has on the identification of significantly mutated genes (TCGA, 2011).

In the case of copy number events, the majority of our gene candidates are located on chromosome arms that exhibited frequent and non-random copy number alterations however with 30 chromosome arms identified as experiencing these copy number changes, these results make it difficult to conclude individual genes are targeted by copy number changes (TCGA, 2011). Conversely, stronger evidence for important roles in the disease can be made for the five and nine gene candidates that map to regions of focal amplification and deletion respectively. Of the amplified candidates, *EPCAM* and *MECOM* are noted to have therapeutic

antagonists available (TCGA, 2011). As the authors suggest, genes identified in these focal regions of copy number alteration may be suitable targets for future drug development (TCGA, 2011).

Overall, the mRNA expression data provided by the TCGA study was supportive of our findings and while most of our gene candidates did not appear to be targeted by somatic mutations or copy number alterations (TCGA, 2011), it remains possible that these events, while uncommon at a single locus, represent the consistent perturbation of a molecular pathway or biological function common to our candidates and critical to the disease.

4.3 Implicated molecular pathways in HGSC

In addition to providing insight into the disease relevance of the hybrids and the genes captured in our comparative transcriptome analysis, bioinformatics resources have also identified molecular functions and pathways associated with the candidate genes. DAVID Functional Gene Classification (Table 2) identified a statistically significant enrichment of genes associated with cell signaling as described by annotation terms such as G-protein coupled signaling, transmembrane, and transducer. Functional Annotation Clustering supported this result, with glycoprotein signaling as the top scoring term.

Interestingly, the top scoring Canonical Pathway identified by IPA as over represented in our candidate list was G-protein coupled receptor signaling. An examination of the four next highest ranked signaling pathways and associated genes (C21-steroid hormone metabolism, androgen and estrogen metabolism, cAMP-mediated signaling and clathrin-mediated endocytosis signaling) suggested

that these top five canonical pathways likely represent two broader signaling processes. As cAMP-mediated signaling propagates signals from membrane-bound receptors including G-protein coupled receptors (Marinissen and Gutkind, 2001) and clathrin-mediated endocytosis facilitates the recycling of G-protein coupled receptors from the membrane back into the cytoplasm (Ferguson, 2001), it is not surprising to see an extensive overlap in gene candidates associated with these three canonical pathways (Table 4). It appears that G-protein coupled receptors can even continue to signal to cAMP after being internalized via clathrin-mediated endocytosis (Calebiro *et al.*, 2010). Similar to the three aforementioned pathways, C21-steroid hormone metabolism, and androgen and estrogen metabolism are highly related pathways with the first two of the three associated genes - *AKR1C1*, *HSD17B2* and *STS* - in common. Interestingly, the genes implicated in the ovarian cancer signaling canonical pathway include *FGF9*, *FZD7* and *PRKAR2B*. These genes were also represented in the three related pathways of G-protein coupled receptor signaling, cAMP-mediated signaling and clathrin-mediated endocytosis.

The association of our candidate list with G-protein coupled receptor signaling along with related canonical pathways prompted further investigation into the role of this pathway and implicated gene candidates in EOC biology.

G-protein coupled receptors have come to be recognized for their role in cancer relatively recently (Dorsam and Gutkind, 2007). The involvement of these receptors in cell proliferation and metastasis is complex and involves crosstalk between a number of growth factor receptors and signaling pathways (Lappano and Maggiolini, 2011). As the largest family of cell-surface markers involved in

signal transmission (Dorsam and Gutkind, 2007), there is potential for novel drug development, particularly given the therapeutic efficacy of current drugs directed at G-protein coupled receptors (Lappano and Maggiolini, 2011). In EOC, a large portion of G-protein coupled receptor signaling research has focused on the role of lysophosphatidic acid (LPA), a phospholipid with growth-factor-like activity that functions in a G-protein dependent manner (Mills and Moolenaar, 2003). Present in the ascites fluid often produced in EOC, LPA acts as a potent mitogen (Xu *et al.*, 1995) with the ability to promote metastasis, transactivate EGFR signaling (Lappano and Maggiolini, 2011) and stimulate *VEGF* (Hu *et al.*, 2001). Unfortunately, the complex nature of the transduction networks mediated by G-protein coupled signaling (Marinissen and Gutkind, 2001) and the limited number of anti-cancer agents antagonistic to G-protein coupled signaling available at this time leave much to be done before these pathways can be harnessed for clinical benefit (Lappano and Maggiolini, 2012). Despite these challenges, promising results have been reported such as the development of LPA antagonist analogues capable of reducing breast cancer xenograft tumor size and vascularization in nude mice (Zhang *et al.*, 2009). An examination of our gene candidates implicated in these signaling pathways provides further insight into the role G-protein coupled signaling may play in EOC biology.

4.4 Gene candidates implicated in G-protein coupled receptor signaling

F2R also known as *PARI* is involved in the interrelated G-protein coupled receptor signaling, cAMP-mediated signaling and clathrin-mediated endocytosis canonical pathways identified by IPA. The gene encodes a protease-

activated G-protein coupled receptor activated by the proteolytic cleavage of the N-terminal region via thrombin (Vu *et al.*, 1991). Characterization of *F2R* mRNA levels in EOC revealed expression in samples of low-malignant potential and invasive malignant disease contrasted by a lack of expression in OSE (Grisaur-Gronovsky *et al.*, 2005). Breast cancer research identified the ability of activated *F2R* to promote tumor growth and invasion in mouse tumor xenograft models (Boire *et al.*, 2005). Similar findings were noted in EOC when *F2R* inhibition via allosteric receptor blocking agents was shown to inhibit angiogenesis, ascites formation, invasion and metastasis in nude mouse xenograft models of EOC (Agarwal *et al.*, 2008). The matrix metalloprotease MMP1, was identified as an upstream activator of *F2R* with further research identifying the pro-angiogenic effects of *F2R* signaling as mediated by the release of angiogenic factors (Agarwal *et al.*, 2008; Agarwal *et al.*, 2010). These findings were supported by the identification of the established pro-angiogenic ligand LPA as an upstream agonist (Wang *et al.*, 2011). Interestingly, the oncogene-like effects of *F2R* are consistent with the increased expression observed in the HGSC samples assayed by the TCGA (2011), but not with the decreased expression noted in the HGSC samples of the #18520 dataset. This discrepancy may reflect the use of different HGSC samples, normal reference tissue or gene expression platforms.

Modulating the role of G-protein coupled receptors such as *F2R* are regulators of G-protein signaling proteins like *RGS2* and *RGS4*. These proteins act as negative regulators of G-protein mediated signaling by accelerating the deactivation of G-proteins (Hurst *et al.*, 2009). Regulators of G-protein coupled receptors have been shown to play an important role in attenuating the LPA-

mediated proliferative signaling prevalent in EOC cells (Hurst *et al.*, 2008). Research into effectors of migration and invasion in breast cancer has revealed RGS4 acts to inhibit the formation of lamellipodia required for these processes by attenuating G-protein coupled receptor signaling (Xie *et al.*, 2009). Both *RGS2* and *RGS4* had decreased expression in the HGSC samples present in the #18520 dataset while only *RGS2* had decreased expression in the TCGA study (2011). *RGS4* had no clear profile in the TCGA samples with about equal numbers of samples showing either decreased or increased expression relative to normal FT reference (TCGA, 2011).

In addition to the role of *F2R* in angiogenesis, G-protein coupled receptor signaling appears to affect this process via the renin-angiotensin system (Suganuma *et al.*, 2005). The angiotensin II type 1 receptor is encoded by the *AGTR1* gene and in addition to its role in blood pressure, fluid and electrolyte balance, has recently become implicated for a role in promoting cancer progression in multiple cancer types (George *et al.*, 2010). Beyond the correlation between *AGTR1* expression and malignant potential in a series of EOC samples of varying malignancy, receptor blockage was shown to reduce dissemination and angiogenesis in mouse xenografts (Suganuma *et al.*, 2005). In a series of 67 EOC patients, *AGTR1* staining was performed with the majority of cases showing positive staining and a statistically significantly worse overall and progression-free survival compared to negative staining patients tumor samples (Ino *et al.*, 2006). *AGTR1* staining also correlated with increased tumor microvessel density and VEGF expression, providing a potential mechanism of action (Ino *et al.*,

2006). Interestingly, this gene had lower expression levels in the HGSC samples of both #18520 and TCGA (2011) data sets relative to normal controls.

G-protein coupled signaling receptors include the frizzled receptor family. This family of receptors is responsible for initiating the canonical Wnt signaling pathway, which plays a critical role in multiple cancer types (King *et al.*, 2011). Over-expression of *FZD7* in colorectal cancers has been noted in comparison to normal tissues and corresponds to decreased patient overall survival (Ueno *et al.*, 2009). Over-expression has also been noted in triple-negative breast cancer and Wilm's tumors (Yang *et al.*, 2011; Pode-Shakeed *et al.*, 2011). *FZD7* gene expression inhibition (knockdown in *in vitro* and *in vivo* models of colorectal cancer resulted in decreased cancer cell viability, invasion and metastasis (Ueno *et al.*, 2009), while knockdown in triple-negative breast cancer cell lines had reduced proliferation, invasion, colony formation and tumorigenicity in xenograft models of the disease (Yang *et al.*, 2011). Anti-FZD7 antibodies have resulted in Wilm's tumor cell death, decreased proliferation and graft survival *in vivo* (Pode-Shakeed *et al.*, 2011). Despite *FZD7* being implicated in the ovarian cancer signaling canonical pathway in addition to G-protein coupled receptor signaling by IPA Core Analysis, it has received little attention in relation to EOC biology. This gene had decreased expression in the #18520 HGSC samples but no clear expression pattern in the TCGA study (2011).

Further downstream of frizzled Wnt receptors such as *FZD7* lies the Ca^{2+} signaling cascade where *CAMK2D* encodes a subunit of the calcium/calmodulin-dependent kinase 2 (CAMKII) (Rodriguez-Mora *et al.*, 2005). This Wnt/ Ca^{2+} transducer has been implicated for its involvement in a number of cancers

including EOC. Prostate cancer research has shown that *CAMK2D* along with other CAMK2 subunit encoding genes is expressed in prostate cancer cells and that CAMK2 activity facilitates the activation of Akt promoting apoptosis resistance and cellular growth (Rokhlin *et al.*, 2007), while colorectal cancer research has identified CAMK2 mRNA and protein to be increased in the transition from adenoma to adenocarcinoma (Hennig *et al.*, 2011). Endogenous inhibitors of CAMK2 termed ‘hCAMK2N α/β ’ have been shown to inhibit colorectal cancer cell growth (Zhang *et al.*, 2001; Wang *et al.*, 2008) as well as EOC cell growth with a decrease in tumorigenicity *in vivo* as a result of decreased Akt activity (Ma *et al.*, 2009). *CAMK2D* was also demonstrated to potentiate the effect of cisplatin (Arora *et al.*, 2010) while single nucleotide polymorphisms in the gene appear to modify risk for EOC (Permuth-Wey *et al.*, 2011). *CAMK2D* displayed decreased expression in the HGSC samples of both the #18520 and TCGA (2011) datasets.

PRKAR2B, encodes a regulator of cAMP signaling that despite limited investigation in most cancers has been noted as down-regulated at least 5-fold in primary cultures of HGSC compared to primary cultures of OSE (Santin *et al.*, 2004). However others have noted up-regulation in HGSC tumors compared with OSE cells (Baranova *et al.*, 2006) suggesting expression levels may be impacted by culture conditions. In the #18520 dataset HGSC samples had decreased *PRKAR2B* expression but showed no clear pattern in the TCGA data (2011).

Further upstream in the G-protein coupled receptor signaling cascade lies *PTHLH* which encodes the largely paracrine acting parathyroid-hormone related protein (PTHrP) (McCauley & Martin, 2012). Originally discovered for its role in

malignancy-associated hypercalcemia (Wysolmerski & Broadus, 1994), it was investigated in gynecological malignancies per a potential link with human papilloma virus (HPV) where approximately 25% of EOCs had positive staining for the protein despite an absence of association with HPV (MacKenzie *et al.*, 1994). An expanded series of EOC tumors including multiple histopathological subtypes confirmed positive staining for PTHrP in the majority of HGSC, all clear cell and half of endometrioid adenocarcinomas of the ovary, while mucinous and cystadenomas lacked staining (Fukuniski *et al.*, 1994). Research into the mechanisms of *PTHLH* regulation identified the ability of TGF- β 1 to stimulate PTHrP production in an EOC cell line (Yasui *et al.*, 1997). While a promising target, much remains to be learned about the complex *PTHLH* signaling relationships (McClauley and Martin, 2012) as demonstrated by positive feedback loops with TGF- β 1 in the context of epithelial-mesenchymal transition, the involvement of EGF, VEGF and extracellular signal-regulated kinases in this process (Ardura *et al.*, 2010). The relationship between *PTHLH*, TGF- β signaling and epithelial-mesenchymal transition is particularly interesting given the up-regulation of *PTHLH* expression in the hybrids, the implication of TGF- β signaling in the previous comparative transcriptome analyses performed on the hybrids (Quinn *et al.*, 2009) and evidence suggesting the hybrids may have transitioned to a more mesenchymal phenotype as a result of chromosome 3 fragment transfer (Cody *et al.*, 2007). *PTHLH* had decreased expression in the #18520 dataset HGSC samples but no clear profile in the TCGA data (2011).

The *FGF9* gene encodes fibroblast growth factor 9, which was first reported as mutated in colorectal and endometrial carcinomas (Abdel-Rahman *et al.*, 2008). Comparative transcriptome analyses aimed at identifying genes involved in Wnt/ β catenin dysregulation in endometrioid EOC (a common event, ~40%) identified *FGF9* as up-regulated 6-fold in endometrioid EOC samples with Wnt/ β catenin dysregulation compared to those with intact Wnt signaling (Hendrix *et al.*, 2006). Subsequent research demonstrated *FGF9* was regulated by upstream Wnt signals and possessed mitogenic effects including the promotion of invasion and anchorage independent growth (Hendrix *et al.*, 2006). *FGF9* also appears to play a role in the normal ovary with evidence suggesting it promotes the production of progesterone in healthy rat granulosa cells (Drummond *et al.*, 2007). An analysis of late stage, HGSC samples revealed an upregulation of Wnt targets *AXIN2* and *FGF9* across all 16 tumors with four tumors demonstrating over a 1000-fold increase when compared to an immortalized normal OSE cell line (Schmid *et al.*, 2011). The expression data from the TCGA supports these findings (2011) but in the #18520 dataset, *FGF9* had decreased expression compared to OSE cytobrushings.

Platelet-derived growth factor is a dimeric growth factor composed of combinations of alpha and/or beta chains with each chain encoded by a unique gene (Starksen *et al.*, 1987). *PDGF β* encodes the beta chain of this growth factor whose expression in EOC has been correlated to malignancy when assessed by immunohistochemistry in EOC samples of varying malignancy and histopathological subtype (Henriksen *et al.*, 1993). An examination of HGSC

samples via fluorescent immunohistochemistry revealed the majority of samples were positive for PDGF α and PDGF β with a number of samples also positive for activated forms of the receptors (Apte *et al.*, 2004). PDGFR β and PDGFR α inhibition in EOC cell lines has been shown to result in a reduced number of tumor vascular endothelial cells, decreased tumor weight, ascitic fluid mass and increased survival in peritoneal xenograft mouse models (Matei *et al.*, 2004; Machida *et al.*, 2005). PDGFR inhibition in combination with paclitaxel in nude mice bearing intraperitoneal injected EOC cell lines resulted in the apoptosis of tumor associated endothelial cells leading to reduced vessel density and tumor proliferation (Apte *et al.*, 2004). Follow-up research by Matei *et al.* (2006) elucidated the ability of PDGF-PDGFR autocrine signaling to promote progression in EOC. A phase II clinical trial of PDGFR inhibitor imatinib mesylate in combination with docetaxel in advanced, platinum-resistance EOC and peritoneal cancers demonstrated no clear benefit over docetaxel alone, however the patient group studied had been heavily pre-treated (Matei *et al.*, 2008). Surprisingly, PDGF β had decreased mRNA expression in the #18520 HGSC samples, the HGSC samples assessed by Santin *et al.*, (2004) and no clear pattern was observed in the TCGA expression data (2011).

DAB2 encodes a well established tumor suppressor gene first identified as down-regulated in EOC compared to normal ovarian tissues with particularly reduced expression in the serous histopathological subtype (Mok *et al.*, 1998). An examination of premalignant regions bordering EOC tumor and normal tissue identified dysplastic morphology in combination with loss of *DAB2* expression

and a weakened basement membrane, suggesting *DAB2* loss of expression may be an early event in EOC development (Yang *et al.*, 2002). The consistent down-regulation of *DAB2* in HGSC is illustrated by the 5-fold down-regulation in primary cultures of HGSCs compared to primary cultures of OSE (Santin *et al.*, 2004). These results are supported by the reduced expression noted in the #18520 dataset HGSC samples as well as the TCGA tumors (2011).

GPR133 encodes a lesser-known member of the G-protein coupled receptor family first identified at the mRNA level in human pituitary and putamen tissues (Vanti *et al.*, 2003). Despite no known ligand for this receptor, research has confirmed the traditional G-protein coupled nature of its downstream signaling (Bohnekamp & Schoneberg, 2011; Gupte *et al.*, 2012). However much of its biology remains to be elucidated, including its role in EOC. While it displayed decreased expression in the #18520 HGSC samples, no gene expression information was available in the TCGA study (2011).

The above review of gene candidates recognized involved in G-protein coupled receptor signaling and related processes in our candidate list reveals not only an extensive body of research investigating the role of many of these candidates in EOC biology, but illustrates a complex and inter-related network of cellular signals and processes controlling hallmark cancer features such as proliferation, migration, invasion, metastasis and angiogenesis (Apte *et al.*, 2004; Xie *et al.*, 2009, Boire *et al.*, 2005; Agarwal *et al.*, 2008; Lappano and Maggiolini, 2011; Suganuma *et al.*, 2005). Many of the questions regarding the early stages of HGSC disease development remain unanswered and our understanding of the pathway crosstalk remains limited. As others have noted, this

large family of cell-surface signaling receptors represents a promising avenue for therapy development (Lappano and Maggiolini, 2012). As reprogramming of OV-90 appears to have affected the expression of many G-protein coupled receptor signaling genes, we suggest that the hybrids may be suitable for the continued investigation of these pathways and their role in EOC biology.

4.5 The role of ceruloplasmin in HGSC

CP encodes a glucoprotein ferroxidase that transports the majority of the copper found in the bloodstream (Varela *et al.* 1997). Synthesized in the liver as well as by tumor cells, it behaves as an acute-phase reactant protein meaning its concentration increases in plasma in response to inflammation (Varela *et al.*, 1997). The repeated observation of elevated serum levels of this protein in various types of solid malignancies (Linder *et al.*, 1981) led to the investigation of its utility as a diagnostic serum biomarker. This revealed a sensitivity and specificity of approximately 80% (Varela *et al.*, 1997), making it a poor diagnostic marker when compared to current tools (Nossov *et al.*, 2008). At the mRNA level, it has been identified as upregulated in EOC across multiple histological subtypes as well as in early and late stages of the disease (Shirdhar *et al.*, 2001; Hough *et al.*, 2001; Lu *et al.*, 2004; Bignotti *et al.*, 2006). *CP* was also recognized as a useful biomarker in identifying EOC primary tumors as the site of origin in cases of metastatic carcinomas where the primary site is unknown (Buckhaults *et al.*, 2003). *CP* was noted as overexpressed at the protein level in the ascites fluid of women with EOC (Gortzak-Uzan *et al.*, 2007) however very little is known about its protein expression in tumors. Immunohistochemistry in a small series (n=20)

of HGSC samples confirmed overexpression in 40% of tumors compared to normal OSE tissues, which showed no expression (Lee *et al.*, 2004). Consistent with these results, CP protein expression was noted to be increased in a pool of malignant EOC samples compared to benign disease cases when assessed using mass spectrometry (Waldemarson *et al.*, 2012). Increased levels of CP glycopeptide sialylation were also reported in the serum of EOC patients compared to healthy controls, suggesting this protein may experience aberrant post-translational modifications in the disease (Shetty *et al.*, 2012). Interestingly, a recent EOC case study suggested that the use of copper lowering agents in patients might partially re-sensitize cancer cells to platinum chemotherapy, although these findings need to be confirmed in a larger study (Fu *et al.*, 2012).

Consistent with the increased *CP* mRNA levels in EOC noted in the aforementioned studies, research by our own group has confirmed this result. *CP* was a gene markedly downregulated in the OV-90 derived hybrids (Figure 4 A) rendered non-tumorigenic as a result of chromosome 3 fragment transfer (Cody *et al.*, 2007). It was shown to have significantly higher mRNA expression in the series of 17 HGSC samples compared with primary cultures of OSE used in previous comparative transcriptome analyses (Quinn *et al.*, 2009). Moreover, *CP* expression was increased in a series of 79 *TP53* mutation positive HGSC samples analyzed by our group (Figure 5 B) using a chemiluminescence-based gene expression array (Quinn *et al.*, 2009; Tonin *et al.*, unpublished, Axela, www.axelabiosensors.com). In our study, *CP* was observed as a gene associated with the top biological function ‘carcinoma’ by IPA Core Analysis, where it also presented in the top ranked network (cancer, cellular movement, gene expression).

Furthermore, *CP* mRNA expression was much higher in the HGSC samples of the #18520 dataset than in the OSE normal reference (Figure 5 A). While research done by other groups has largely focused on either serum levels or mRNA expression in HGSC, we were interested in characterizing CP protein expression in both our cell lines and a larger series of HGSC tumor samples than previously examined (Lee *et al.*, 2004). The results of western blot analysis demonstrated that CP protein expression corresponded to mRNA levels in the OV-90 derived hybrids with OV-90neo displaying robust CP protein expression while no protein was detectable in any of the three non-tumorigenic hybrids (Figure 4). Immunohistochemistry staining on the 165 HGSC samples analyzed revealed the majority of samples had detectable protein expression in the epithelial compartment (68%). However the level of staining intensity observed was difficult to quantify with the majority of positive samples displaying low level staining. Of the FT samples scored, 100% (n=8) stained positive for CP. This was a particularly intriguing result given that FT has yet to be examined for CP protein expression and the absence of staining in OSE previously reported (Lee *et al.*, 2004).

The association of CP protein expression with progression-free survival suggests this protein may be a useful biomarker to help stratify HGSC patients. The increased progression-free interval and 0.641 hazards ratio associated with patient samples where CP expression is detected may be useful in the clinical management of patients. Having only examined progression-free survival and overall survival, it is difficult to speculate as to why CP expression correlates with

longer progression-free survival. Additional research is needed to shed light on this question.

Given the strong protein expression and mRNA levels we observed in OV-90neo, high mRNA levels in HGSC noted generally, and the presence of a subset of the *TP53* mutation positive HGSC samples previously identified as overexpressing *CP* mRNA (Quinn *et al.*, 2009; Tonin *et al.*, unpublished, Axela, www.axelabiosensors.com), we had expectations of intense CP staining in the HGSC samples interrogated on the TMA. Accordingly, we were surprised to note moderate staining as the intensity score upper limit with overall expression favoring low intensity levels. There are several explanations for this observation including the possibility that the majority of produced CP is secreted as in the case of the liver (Harned *et al.*, 2012). It is also possible that protein levels of *CP* do not closely correlate to mRNA levels but rather variation in protein expression is due to post-transcriptional and post-translational regulation (de Sousa Abreu *et al.*, 2009). The replication of our findings using alternative samples and antibodies would support further investigation of the relationship between *CP* mRNA and protein expression in HGSC.

4.6 Conclusion

This thesis has outlined compelling support for the relevance of the hybrids as a model to study HGSC biology. The initial characterization of OV-90 and the hybrid cell lines (RH-5, RH-6 and RH-10) (Provencher *et al.*, 2000, Cody *et al.*, 2007) prompted the comparative transcriptome analysis of these cell lines with HGSC samples and OSE primary cultures performed by Quinn *et al.* (2009).

The identification of genes previously implicated in EOC as transcriptionally altered both between the hybrids and OV-90neo as well as between OSE primary cultures and HGSC samples, suggested the hybrids experienced a transcriptional programming as a consequence of chromosome 3 fragment transfer that relates to HGSC biology. Addressing some of the limitations of the previous comparative analysis (Quinn *et al.*, 2009), we incorporated a larger, publically available and independently derived dataset of OSE cytobrushings and HGSC samples (#18520) captured on the same GeneChip[®] microarray used to assay OV-90neo and the hybrids. Starting from the 1204 probe sets identified as differentially expressed three-fold between OV-90neo and the hybrid cell lines (Cody *et al.*, 2007), we identified a candidate list of 92 genes differentially expressed between HGSC and OSE samples in the publically available dataset (#18520). Incorporating established bioinformatics programs such as DAVID and IPA, we identified both biological processes and molecular pathways reflective of EOC. The most significantly associated canonical pathways implicate the G-protein coupled receptor signaling family with numerous gene members transcriptionally modified in both the hybrids and HGSC. Many of these gene family members or related genes have been investigated previously in EOC however others have only been examined in the context of other cancer types or in a non-cancer context. We believe this signaling network and those members identified here represent important candidates for EOC research. Among the 92 genes captured in our analysis, ceruloplasmin (*CP*) represents a candidate noted as overexpressed in OV-90neo relative to the hybrids as well as in HGSC samples relative to normal OSE cytobrushings. This is a gene that has been identified previously as

overexpressed at the mRNA level in HGSC samples both in our research and that of others. With its protein expression investigated previously in only a small sample of HGSC samples, we sought to examine its expression in a larger series of HGSC samples for which we had some indices of clinical outcome. Surprisingly, we did not observe the robust levels of protein expression in the samples examined by immunohistochemistry that were expected based on the mRNA expression levels previously observed. This observation, along with the indication that the presence of detectible *CP* protein expression is correlated with a statistically significant increase in progression-free survival, warrants the further investigation of the gene in HGSC. We suggest that the hybrids should continue to be characterized for their utility as a tool for the study of HGSC biology and that additional progress into our understanding of the fundamentals of the disease will translate into better outcomes for the women affected.

4.7 Future Directions

To expand further upon these findings, some limitations of our experiments can be addressed. Microarray analyses are an established technique used to assess gene expression genome wide, however there are some trade-offs in this method of assaying gene expression. While over 35, 000 expressed sequences are captured by the Affymetrix U133 Plus 2.0 GeneChip, detected transcripts are limited to those with corresponding probe sets on the microarray. Techniques such as RNA-sequencing capture a less biased picture of mRNA transcripts with greater insight into alternatively spliced transcripts. However, these features of microarray gene expression capture can also be favorable, as they reduce the

‘noise’ associated with the capture of alternative transcripts of unknown relevance and provide for easier inter-study comparisons. In this study, the use of the U133 Plus 2.0 GeneChip in both the transcriptome characterization of the hybrid cell lines as well as the #18520 dataset was a major strength of our analysis. Future research on the hybrids and OC should integrate more and higher density expression datasets as they become available. Increased efforts on behalf of journals and authors to make expression data publically available will result in more comprehensive studies.

The use of bioinformatics resources to characterize our gene candidates greatly assisted the identification of molecular pathways and biological processes involved in OC. The overlap of results obtained from the established programs used in our analysis as well as supporting literature lent confidence to the program results, but highlight a limitation of this approach. The databases used to construct gene-gene relationships and identify molecular pathways and functions associated with a given gene list are based on a priori knowledge; the same knowledge presented in the literature. So while bioinformatics programs greatly assist researchers in characterizing a large candidate list, they are still limited by our current knowledge. Fortunately, as we continue to expand our knowledge of molecular biology and genetics, these resources will only improve. Future work using the hybrids should take advantage of new features available in IPA including the ability to analyze gene expression data. This feature allows the program to go beyond identifying molecular pathways associated with a gene list and identify whether implicated pathways are likely to be activated or repressed

based on the gene expression data, providing greater insight into the biology of the hybrids and OC.

We utilized immunohistochemistry to provide a descriptive assessment of the protein expression of ceruloplasmin in a large series of HGSC samples. The overall survival and progression-free survival interval data for specimens present on the tissue microarray used allowed us to examine putative relationships between CP protein expression and these clinical outcomes. We found there to be a statistically significant relationship between CP protein expression and progression free survival. These novel findings are very interesting, and we suggest that future research should expand this experiment to include more samples and specimens of varying histopathological subtypes. While we had sufficient power to detect the significance of the relationship between CP protein expression and progression free survival, this relationship was only detected after dichotomizing the staining intensity scoring into 'on' and 'off.' This result may reflect the difficulty of quantifying protein expression via immunohistochemistry or support the use of a larger series of HGSC samples in the future. It would also be beneficial to include a larger series of FT samples to confirm the high percentage of FT samples (100%), we observed as positive for CP protein expression as this too is a novel finding.

The OV-90neo derived hybrids are the focus of this study, however by using it to help characterize HGSC biology, we are limited by the clonal nature of their origin. With a genetic background specific to one individual and subjected to culture conditions, we risk ignoring a number of genes important to HGSC biology but not differentially expressed in the hybrids. This sacrifice in the pursuit

of a narrowed focus centered on a specific phenotype (tumorigenicity) is apparent when we consider all the genes differentially expressed within the HGSC and OSE samples of the #18520 dataset. Using the same criteria requiring greater than 3-fold differences in expression across 75% of the HGSC samples relative to the OSE mean, we identified 2072 probe sets corresponding to 1504 unique genes. This is a much larger gene list than the 92 characterized in this study and likely includes additional genes and pathways of relevance to HGSC biology. However, by using the hybrids as a ‘filter’ to reduce the size of the candidate list and focus on the phenotype of tumorigenicity, our candidate list could be examined in greater detail than would be possible with a larger list, and was further enriched for processes relevant to HGSC biology as observed in Table 3. The development of additional cell lines representative of HGSC biology will help address the limitations associated with the use of clonal cell lines and aid our ability to study this disease.

Unlike the previous transcriptome analysis performed by our group on the hybrids (Quinn *et al.*, 2009), we used OSE cytobrushings as a normal control to avoid transcriptional changes that would reflect culture conditions rather than OSE biology. However, the distal FT is also considered to be a potential origin of EOC (Lee *et al.*, 2007; Crum *et al.*, 2007). The difficulty in obtaining these samples prevented their inclusion as a normal reference in our comparative transcriptome analysis. However, we tried to address how their use may affect the expression profiles identified in our study by examining the expression data presented by the TCGA (2011). As described in the discussion, this comparison indicated the choice of OSE or FT did not affect the expression patterns presented

here in the majority of cases. If samples become available to our group, they will be included in future analyses.

In our use of the #18520 dataset (Bonome *et al.*, 2005), we had a limited knowledge of the samples, lacking clinical information pertaining to patients from which samples were obtained. This made it difficult to answer questions regarding sample heterogeneity and methodological biases. For example, in reviewing the expression of a number of gene candidates in the OSE samples present in the #18520 dataset, the OSE samples frequently appeared to separate into two groups, one with lower expression levels, and the other with higher levels. These groups were comprised of the same samples in most cases where the effect is observed. While we used mean expression values across all 10 samples in our comparative analysis, the question of why these samples appeared to represent two separate groups of samples rather than a uniformly distributed group remains unknown. So while the use of independent datasets provided otherwise unavailable OSE cytobrushing expression data, it also prevented further data analyses. These issues could be reduced by larger sample sizes and additional publically available expression data.

We performed a comparative transcriptome analysis using fold-changes as the basis for selection criteria. For a project that focused in part on the use of bioinformatics resources this may seem a less desirable approach than the use of statistical cutoffs. However, in the original design of this project we intended to validate a number of candidate's mRNA expression levels via semi-quantitative RT-PCR. For this reason and the experience of our lab regarding the sensitivity of this technique, we selected the aforementioned criteria such that validation would

be achievable (Arcand *et al.*, 2005). Furthermore, we were interested in identifying genes whose expression patterns would be amenable to replication and did not want differences that, while statistically significant due to potential outliers, were not necessarily representative of the majority of HGSC cases. To address this potential weakness in our study design, we did perform a student's t-test to examine which of the 1204 probe sets differentially expressed in the hybrid model were statistically differentially expressed in the HGSC/OSE #18520 dataset. The resulting list (when p set to <0.05) was comprised of 643 probe sets of which 105 of the 106 probe sets characterized in this study were present (*F2R* excluded). This reflects positively on our methods of analysis and we suggest that the use of complimentary analyses is an aspect of our study that should be included in future research.

The design of this project centered on the phenotype of tumorigenicity as we were interested in revealing more information regarding the pathways and molecular mechanisms associated with the aggressive biology of HGSC. However, the transfer of chromosome 3 fragments that resulted in the derivation of the non-tumorigenic OV-90 derived hybrids did not alter the underlying genetic background of the parental cell line. These hybrids still possess *TP53* mutations as well as significant chromosomal anomalies and the ability to grow in culture indefinitely (Provencher *et al.*, 2000, Tonin *et al.*, unpublished). This project used the hybrids as a normal reference for the comparative transcriptome analysis of OV-90neo, however, the OV-90 derived hybrids could also be used as a tool to study more indolent forms of the disease. The findings of Quinn *et al.*, (2009), support this application as of 19 genes upregulated in the hybrids examined for

expression by RT-PCR in a series of benign serous cystadenomas, 16 had detectable expression. Eighteen of these genes were also detected as expressed in the non-tumorigenic TOV-81D serous adenocarcinoma cell line (Quinn *et al.*, 2009; Provencher *et al.*, 2000) suggesting the hybrids may have expression profiles and *in vivo* characteristics similar to more indolent forms of OC. It would also be useful to determine if the hybrids - while incapable of forming tumors in xenograft mouse models - remain viable and continue to display an immortalized phenotype *in vivo*, as well as what, if any effect *in vivo* growth conditions (such as those examined for OV-90 and other OC cell lines developed by our group) would have on gene expression levels (Cody *et al.*, 2007; Cody *et al.*, 2008).

The hybrids should continue to be examined through the use of new techniques and sample types as advancing fields such as proteomics will benefit from well characterized, disease representative cell lines in the same way these disease models have aided transcriptome analyses.

FIGURES

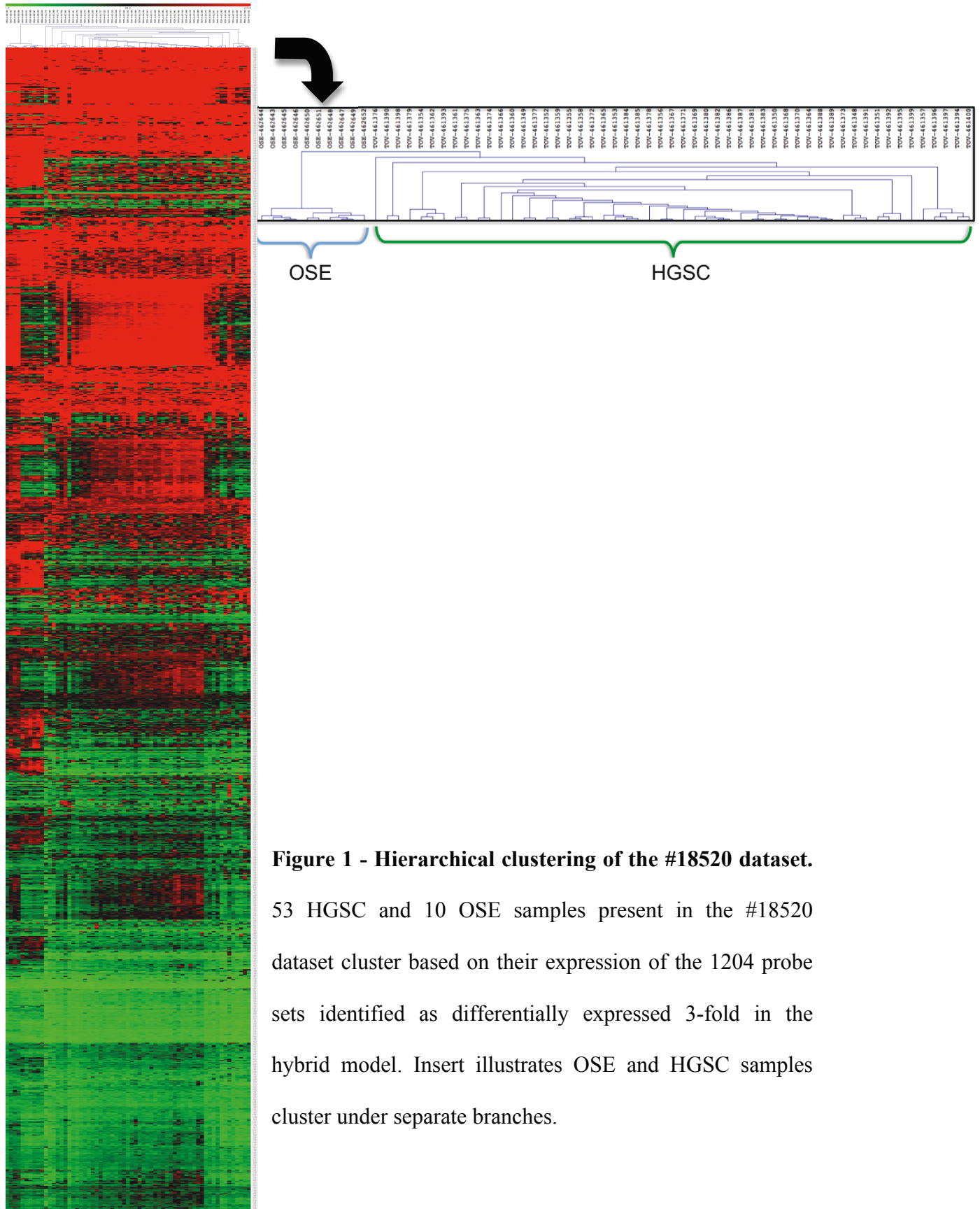


Figure 1 - Hierarchical clustering of the #18520 dataset.

53 HGSC and 10 OSE samples present in the #18520 dataset cluster based on their expression of the 1204 probe sets identified as differentially expressed 3-fold in the hybrid model. Insert illustrates OSE and HGSC samples cluster under separate branches.

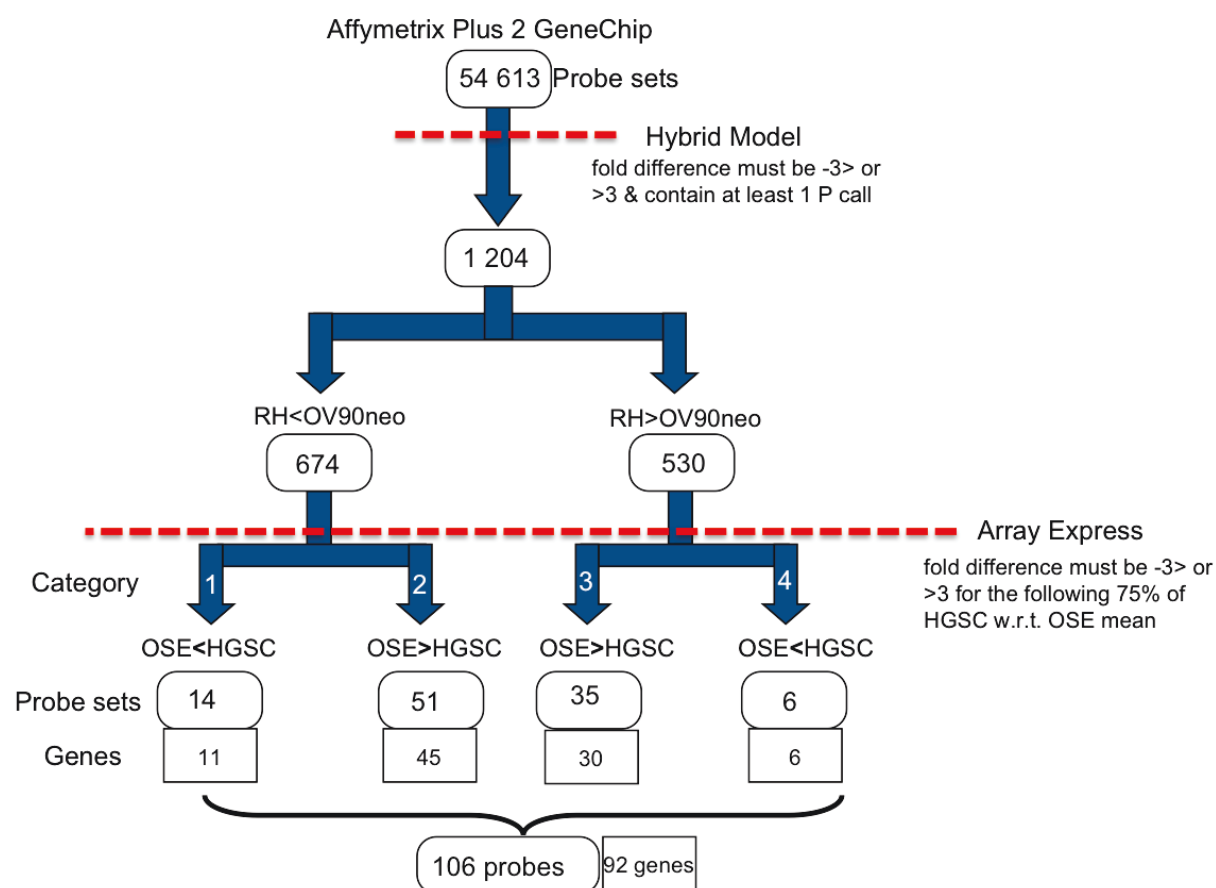


Figure 2 - Comparative transcriptome analysis selection criteria. Flowchart illustrating the selection criteria applied first in OV-90neo and the hybrids, and then in the #18520 dataset to reduce the candidate list from over 54 000 probe sets to 106. These 106 probe sets correspond to 92 unique genes of hypothesized relevance to the biology of HGSC.

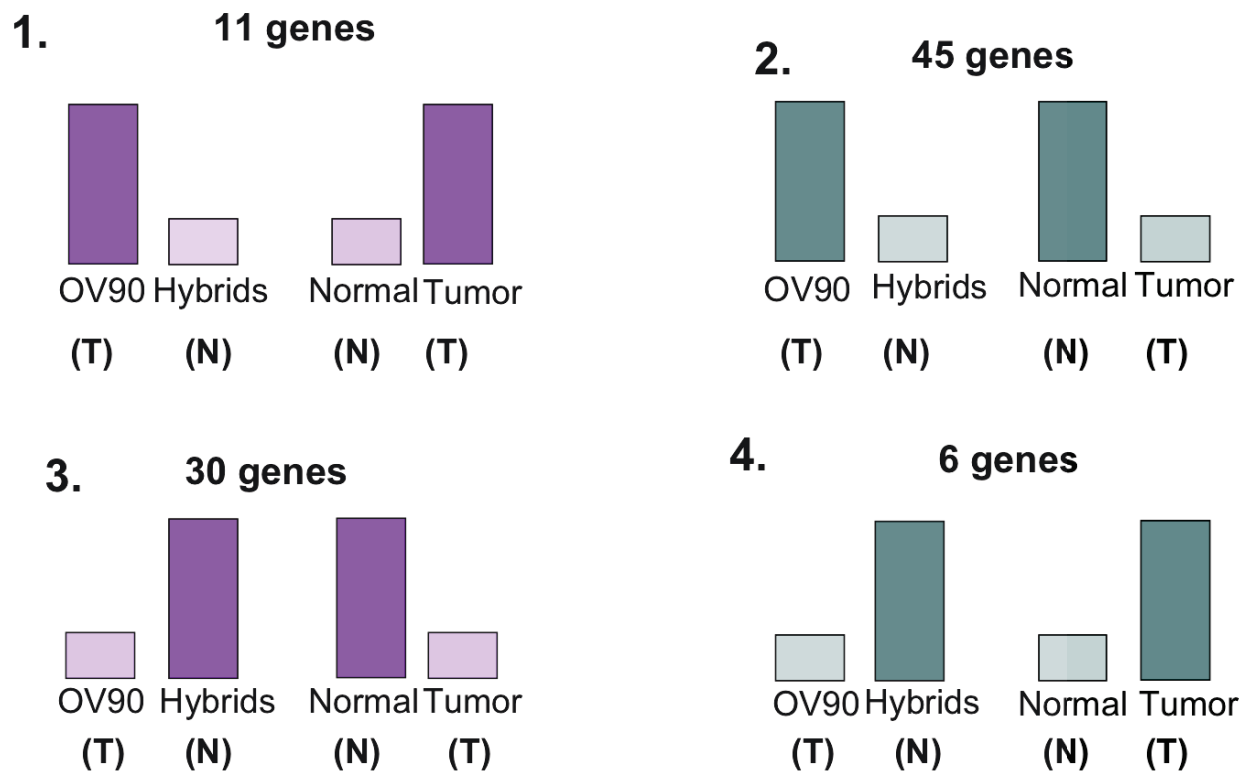


Figure 3 - Candidate gene expression categories. Representation of the four expression patterns noted for genes across the samples examined in this study. Darker shading corresponds to relatively higher expression level and lighter shading lower expression. Categories 1 and 3 are colored similarly to indicate their genes exhibit patterns that are consistent with the phenotype of tumorigenicity across samples. Conversely, category 2 and 4 genes have patterns that do not follow the phenotype of tumorigenicity.

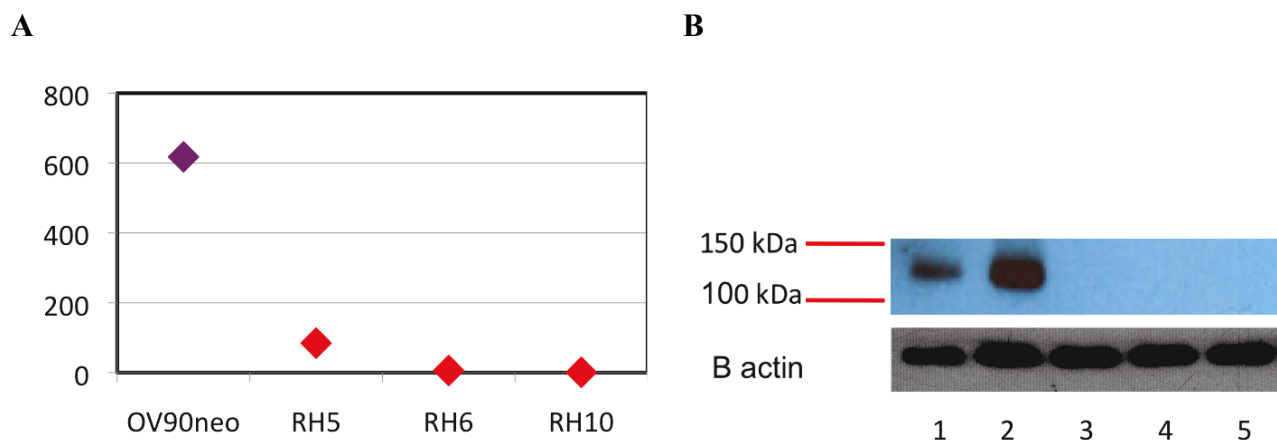


Figure 4 – CP mRNA and protein expression in OV-90neo and hybrids. .

A – mRNA expression levels of CP in OV-90neo, RH-5, RH-6 and RH-10.

B - Western blot analysis of total protein extracts from OV-90, OV-90neo, RH-5, RH-6 and RH-10 (lanes 1-5 respectively) hybridized with anti-ceruloplasmin (diluted at 1:100, LF-MA0159, Young in Frontier). CP is detected at the expected size of approximately 132kDa in OV-90 and OV-90neo.

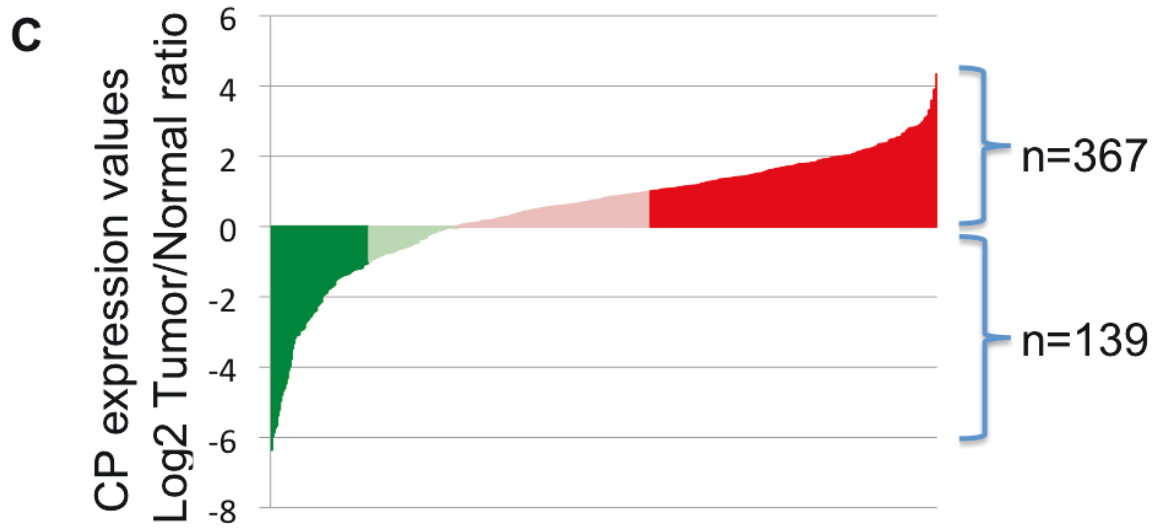
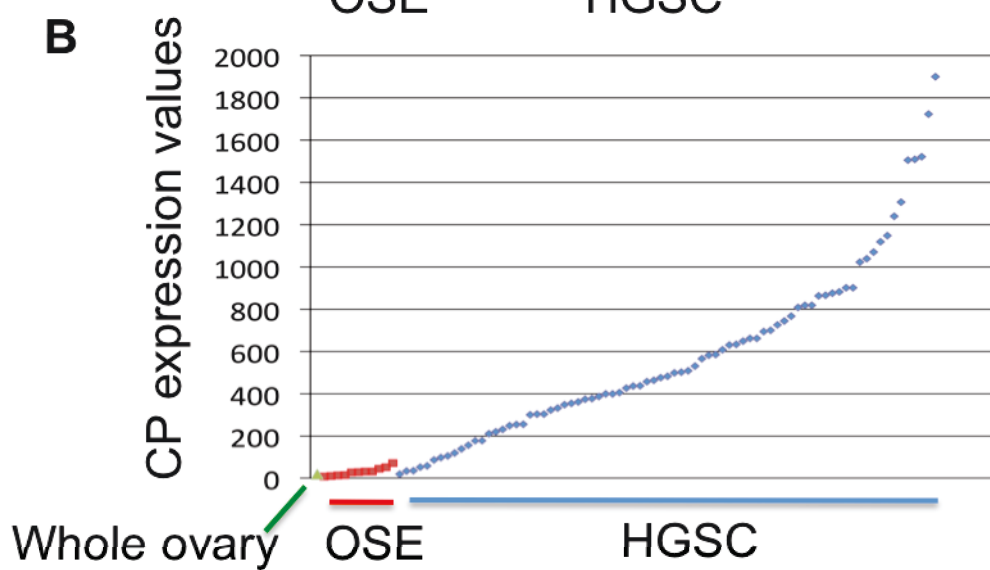
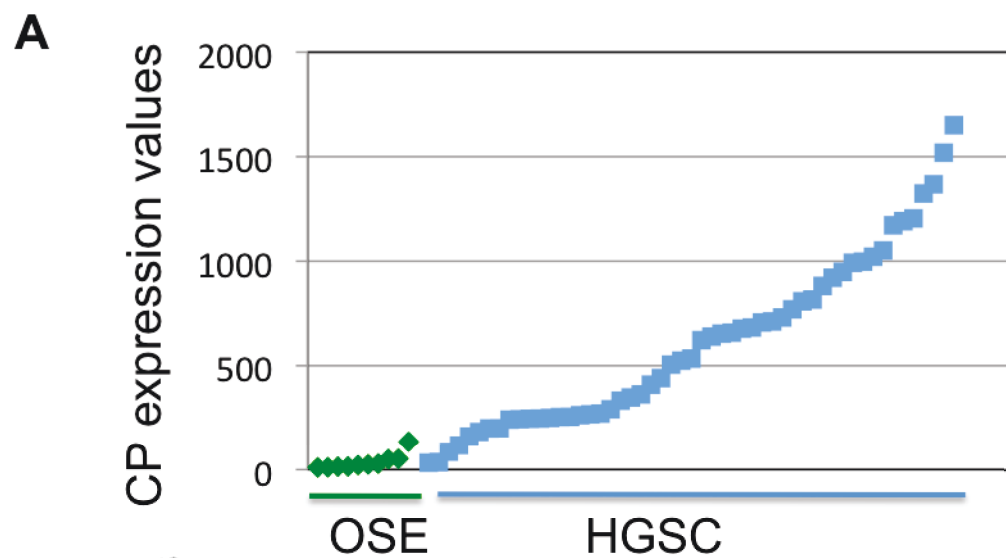


Figure 5 – CP mRNA expression in a series of HGSC and normal reference tissues. **A** - Expression in the OSE cytobrushings and HGSC samples of the #18520 dataset. **B** - Expression in whole ovary, OSE primary cultures, and *TP53* mutation positive HGSC samples (Quinn *et al.*, 2009, Tonin *et al.*, unpublished; Axela, www.axelabiosensors.com). **C** – Log2 of the tumor/normal ratio from HGSC and fallopian tube samples (TCGA, 2011).

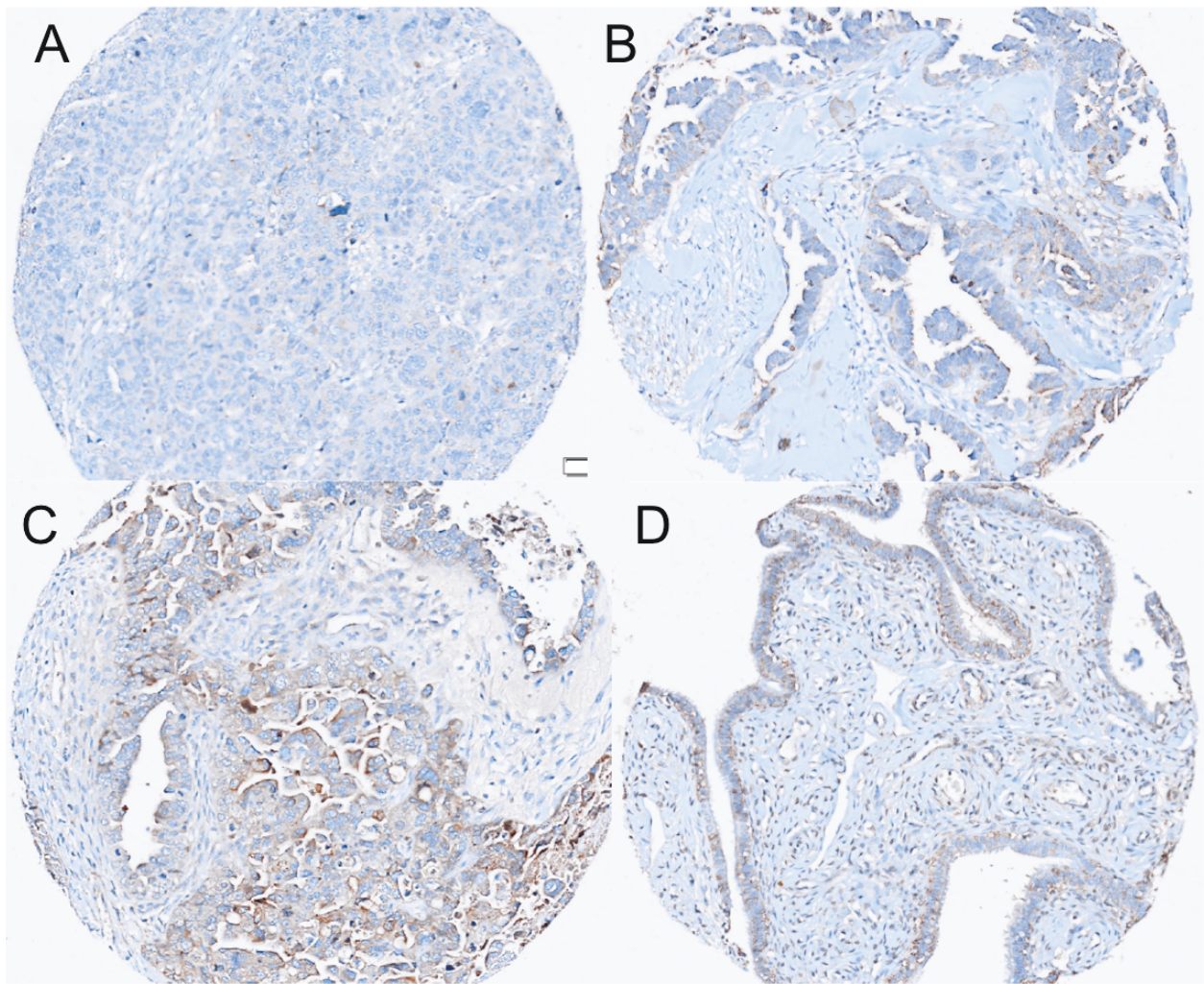
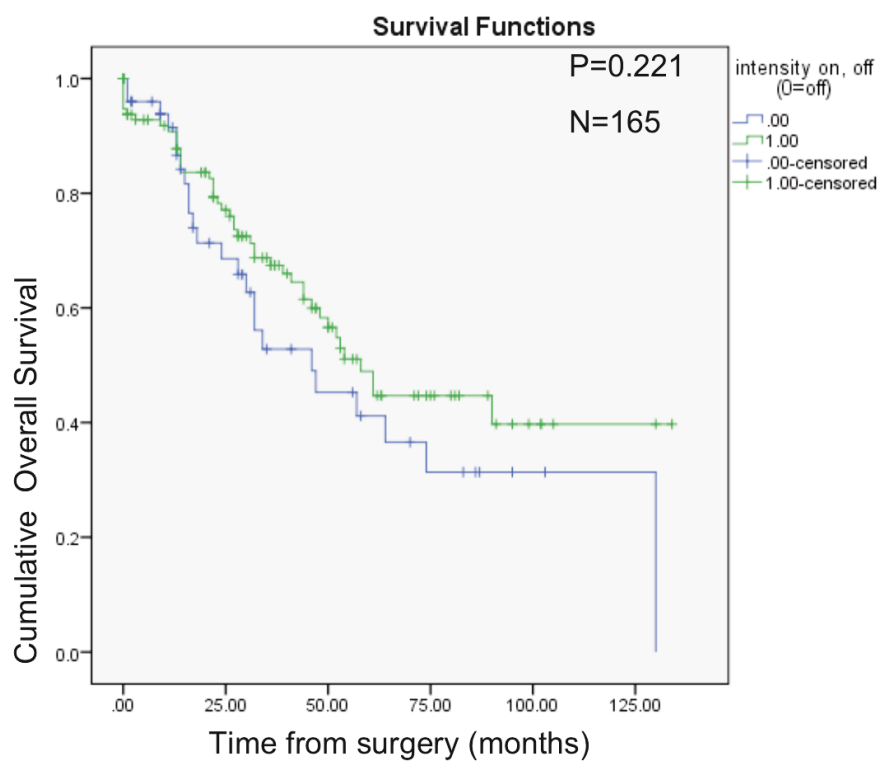


Figure 6 - Representative CP immunohistochemistry staining in HGSC and FT samples. Images of TMA cores assessed for CP staining. Sections A-C illustrate negative (A), weak (B) and moderate (C) staining of CP in HGSC samples. No intense staining was noted. Section D is representative of the positive staining noted in all 8 fallopian tube samples. Images were obtained from the OlyVIA viewer (Olympus America Inc., Center Valley, PA, USA).

A



B

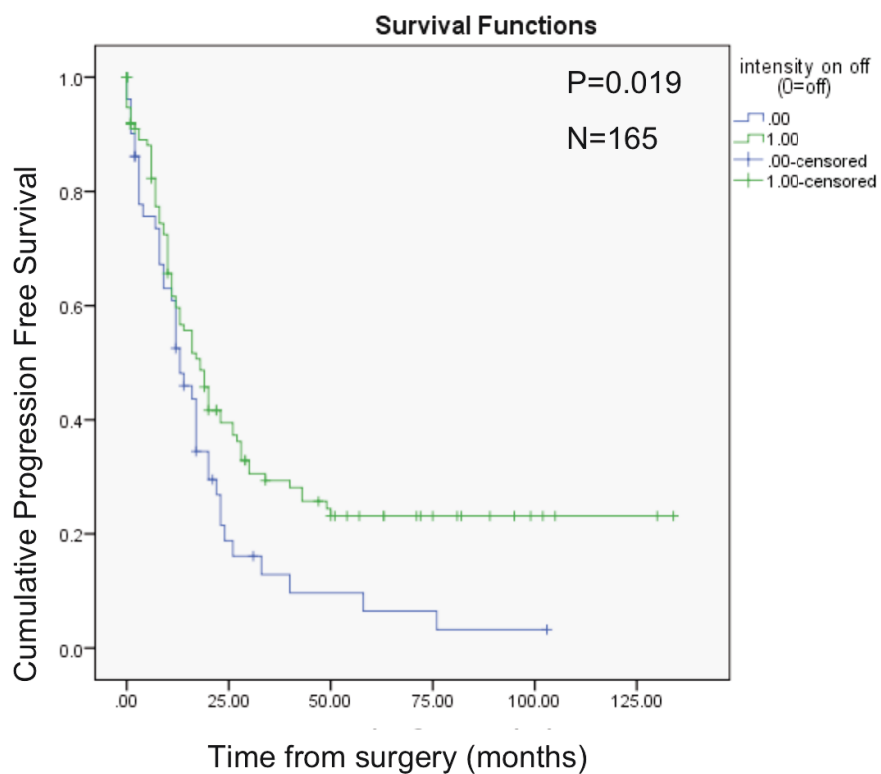


Figure 7 – Kaplan-Meier analysis of CP and survival rates in HGSC.

A – Overall survival curve. **B** – Progression free survival curve. Significance (p) is indicated by log rank.

TABLES

Table 1 – Candidate gene list

Probe Set ID	Gene Symbol	Gene Title	Entrez Gene	Alignments	Chromosomal Location	Expression Category	In Hybrids vs OV90	In HGSC vs OSE
1558034_s_at	CP	ceruloplasmin (ferroxidase)	1356	chr3:148925268-148939829 (-) // 99.74 // q25.1	chr3q23-q25	1	down regulated	up regulated
201839_s_at	EPCAM	epithelial cell adhesion molecule	4072	chr2:47596466-47614157 (+) // 85.8 // p21	chr2p21	1	regulated down	regulated up
203768_s_at	STS	steroid sulfatase (microsomal), isozyme S	412	chrX:7137496-7272851 (+) // 88.85 // p22.31	chrXp22.32	1	regulated down	regulated up
204846_at	CP	ceruloplasmin (ferroxidase)	1356	chr3:148891375-148939579 (-) // 99.94 // q24	chr3q23-q25	1	regulated down	regulated up
205363_at	BBOX1	butyrobetaine (gamma), 2-oxoglutarate dioxygenase (gamma-butyrobetaine hydroxylase) 1	8424	chr11:27062997-27149354 (+) // 92.56 // p14.2	chr11p14.2	1	down regulated	up regulated
210827_s_at	ELF3	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	1999	chr1:201979723-201985131 (+) // 100.0 // q32.1	chr1q32.2	1	regulated down	regulated up
216470_x_at	PRSS1 /// PRSS2 /// PRSS3	protease, serine, 1 (trypsin 1) /// protease, serine, 2 (trypsin 2) /// protease, serine, 3	5644 /// 5645 /// 5646	chr7:142469328-142471794 (+) // 92.75 // q34 /// chr9:33796651-33799178 (+) // 88.59 // p13.3	chr7q32-qter[7q34] /// chr7q34 /// chr9p11.2	1	down regulated	up regulated
221884_at	MECOM	MDS1 and EVI1 complex locus	2122	chr3:168801294-168833223 (-) // 97.42 // q26.2	chr3q24-q28	1	regulated down	regulated up
225645_at	EHF	Ets homologous factor	26298	chr11:34682443-34684826 (+) // 91.5 // p13	chr11p12	1	regulated down	regulated up
226420_at	MECOM	MDS1 and EVI1 complex locus	2122	chr3:168801294-168833223 (-) // 97.42 // q26.2	chr3q24-q28	1	regulated down	regulated up
227253_at	CP	ceruloplasmin (ferroxidase)	1356	chr8:92171298-92172198 (+) // 81.96 // q21.3	chr3q23-q25	1	regulated down	regulated up
227475_at	FOXQ1	forkhead box Q1	94234	chr6:1314110-1314993 (+) // 92.71 // p25.3	chr6p25	1	regulated down	regulated up
227725_at	ST6GALNAC1	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylglucosaminide alpha-2,6-sialyltransferase 1	55808	chr17:74620837-74639839 (-) // 88.11 // q25.1	chr17q25.1	1	down regulated	up regulated
228360_at	LYPD6B	LY6/PLAUR domain containing 6B	130576	chr2:149895251-150071776 (+) // 98.88 // q23.1	chr2q23.1-q23.2	1	regulated down	regulated up
1554485_s_at	TMEM37	transmembrane protein 37	140738	chr2:120194280-120195652 (+) // 99.06 // q14.2	chr2q14.2	2	regulated down	regulated down
1555564_a_at	CFI	complement factor I	3426	chr4:110661853-110723196 (-) // 96.0 // q25	chr4q25	2	regulated down	regulated down
1555745_a_at	LYZ	lysozyme	4069	chr12:69742188-69746999 (+) // 96.49 // q15	chr12q15	2	regulated down	regulated down
1559214_at	NRG4	neuregulin 4	---	chr15:76233277-76304785 (-) // ??? // q24.2	chr15q24.2	2	regulated down	regulated down
1563075_s_at	---	---	---	chr7:138736749-138737554 (-) // 17.05 // q34	chr7q34	2	regulated down	regulated down
202388_at	RGS2	regulator of G-protein signaling 2, 24kDa	5997	chr1:192778170-192781403 (+) // 97.55 // q31.2	chr1q31	2	regulated down	regulated down
203680_at	PRKAR2B	protein kinase, cAMP-dependent, regulatory, type II, beta	5577	chr7:106685177-106801864 (+) // 98.59 // q22.3	chr7q22	2	regulated down	regulated down
203717_at	DPP4	dipeptidyl-peptidase 4	1803	chr2:162848758-162930567 (-) // 99.79 // q24.2	chr2q24.3	2	regulated down	regulated down
203824_at	TSPAN8	tetraspanin 8	7103	chr12:71518881-71551565 (-) // 95.38 // q21.1	chr12q14.1-q21.1	2	regulated down	regulated down
203881_s_at	DMD	dystrophin	1756	chrX:31137344-33146545 (-) // 99.25 // p21.2	chrXp21.2	2	regulated down	regulated down
204041_at	MAOB	monoamine oxidase B	4129	chrX:43625857-43741622 (-) // 93.01 // p11.3	chrXp11.23	2	regulated down	regulated down
204058_at	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic	4199	chr6:83920111-84140787 (-) // 96.8 // q14.2	chr6q12	2	regulated down	regulated down
204337_at	RGS4	regulator of G-protein signaling 4	5999	chr1:163039149-163046592 (+) // 95.3 // q23.3	chr1q23.3	2	regulated down	regulated down
204818_at	HSD17B2	hydroxysteroid (17-beta) dehydrogenase 2	3294	chr16:82068862-82132137 (+) // 99.93 // q23.3	chr16q24.1-q24.2	2	regulated down	regulated down
205158_at	RNASE4	ribonuclease, RNase A family, 4	6038	chr14:21167516-21168492 (+) // 86.95 // q11.2	chr14q11.1	2	regulated down	regulated down
205234_at	SLC16A4	solute carrier family 16, member 4 (monocarboxylic acid transporter 5)	9122	chr1:110905504-110933636 (-) // 85.45 // p13.3	chr1p13.3	2	regulated down	regulated down
205357_s_at	AGTR1	angiotensin II receptor, type 1	185	chr3:148415633-148460788 (+) // 98.11 // q24	chr3q21-q25	2	regulated down	regulated down
205433_at	BCHE	butyrylcholinesterase	590	chr3:165490692-165555250 (-) // 99.88 // q26.1	chr3q26.1-q26.2	2	regulated down	regulated down
205466_s_at	HS3ST1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	9957	chr4:11400442-11401739 (-) // 99.39 // p15.33	chr4p16	2	regulated down	regulated down
206167_s_at	ARHGAP6	Rho GTPase activating protein 6	395	chrX:11155666-11683821 (-) // 96.13 // p22.2	chrXp22.3	2	regulated down	regulated down
206529_x_at	SLC26A4	solute carrier family 26, member 4	5172	chr7:107301079-107358250 (+) // 91.74 // q22.3	chr7q31	2	regulated down	regulated down
207761_s_at	METTL7A	methyltransferase like 7A	25840	chr12:51318801-51326288 (+) // 89.79 // q13.12	chr12q13.12	2	regulated down	regulated down
209699_x_at	AKR1C2	aldo-keto reductase family 1, member C2 (dihydrodiol dehydrogenase 2; bile acid binding protein; 3-alpha hydroxysteroid dehydrogenase, type III)	1646	chr10:5005619-5020158 (+) // 97.21 // p15.1 /// chr10:5031964-5046050 (-) // 99.84 // p15.1	chr10p15-p14	2	down regulated	down regulated

209829_at	FAM65B	family with sequence similarity 65, member B	9750	chr6:24804512-24911195 (-) // 91.06 // p22.3	chr6p22.3-p21.32	2	regulated down	regulated down
212224_at	ALDH1A1	aldehyde dehydrogenase 1 family, member A1	216	chr9:75515586-75567971 (-) // 97.58 // q21.13	chr9q21.13	2	regulated down	regulated down
213397_x_at	RNASE4	ribonuclease, RNase A family, 4	6038	chr14:21168264-21168753 (+) // 73.55 // q11.2	chr14q11.1	2	regulated down	regulated down
213800_at	CFH	complement factor H	3075	chr1:196621161-196659370 (+) // 80.8 // q31.3	chr1q32	2	regulated down	regulated down
213802_at	PRSS12	protease, serine, 12	---	chr4:119201193-119202261 (-) // 97.45 // q26	chr4q26	2	regulated down	regulated down
214825_at	FAM155A	family with sequence similarity 155, member A	728215	chr13:107822317-108519091 (-) // 91.03 // q33.3 chr1:196788934-196801288 (+) // 99.33 // q31.3	chr13q33.3	2	regulated	regulated
215388_s_at	CFH /// CFHR1	complement factor H /// complement factor H-related 1	3075 /// 3078	/// chr1:196712580-196716603 (+) // 59.48 // q31.3	chr1q32	2	down regulated	down regulated
218532_s_at	FAM134B	family with sequence similarity 134, member B	54463	chr5:16473165-16475344 (-) // 97.14 // p15.1	chr5p15.1	2	down regulated	down regulated
219093_at	PID1	phosphotyrosine interaction domain containing 1	55022	chr2:22988710-230020682 (-) // 82.49 // q36.3	chr2q36.3	2	down regulated	down regulated
219263_at	RNF128	ring finger protein 128	79589	chrX:106028304-106040219 (+) // 95.56 // q22.3	chrXq22.3	2	regulated down	regulated down
219355_at	CXorf57	chromosome X open reading frame 57	55086	chrX:105855886-105922667 (+) // 99.93 // q22.3 chr2:131279011-131285566 (+) // 96.22 // q21.1	chrXq22.3	2	regulated	regulated
223753_s_at	CFC1 /// CFC1B	cripto, FRL-1, cryptic family 1 /// cripto, FRL-1, cryptic family 1B	55997 /// 653275	/// chr2:131350351-131356906 (-) // 96.45 // q21.1	chr2q21.1	2	down regulated	down regulated
225166_at	ARHGAP18	Rho GTPase activating protein 18	93663	chr6:129897289-130031347 (-) // 93.94 // q22.33	chr6q22.33	2	down regulated	down regulated
225171_at	ARHGAP18	Rho GTPase activating protein 18	93663	chr6:129897289-130031347 (-) // 93.94 // q22.33	chr6q22.33	2	down regulated	down regulated
225817_at	CGNL1	cingulin-like 1	84952	chr15:57744336-57842915 (+) // 85.99 // q21.3	chr15q21.3	2	regulated down	regulated down
226492_at	SEMA6D	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	80031	chr15:48061683-48066420 (+) // 97.4 // q21.1	chr15q21.1	2	regulated down	regulated down
226534_at	KITLG	KIT ligand	4254	chr12:88886569-88890031 (-) // 91.26 // q21.32	chr12q22	2	regulated down	regulated down
227061_at	---	---	---	chr3:112315643-112316945 (-) // 88.88 // q13.2	chr3q13.2	2	regulated	regulated
227561_at	DDR2	discoidin domain receptor tyrosine kinase 2	4921	chr1:162752803-162754116 (+) // 99.32 // q23.3 /// chr10:38637601-38638921 (+) // 96.66 // p11.1	chr1q23.3	2	down regulated	down regulated
228218_at	LSAMP	limbic system-associated membrane protein	4045	chr3:115523823-115526937 (-) // 96.38 // q13.31	chr3q13.2-q21	2	regulated down	regulated down
228863_at	PCDH17	protocadherin 17	27253	chr13:58302855-58303444 (+) // 89.47 // q21.1	chr13q21.1	2	regulated down	regulated down
230006_s_at	SVIP	small VCP/p97-interacting protein	258010	chr11:22842156-22842625 (-) // 79.14 // p14.3	chr11p14.2	2	down regulated	down regulated
230869_at	FAM155A	family with sequence similarity 155, member A	728215	chr13:107820887-107821421 (-) // 89.6 // q33.3	chr13q33.3	2	regulated down	regulated down
231042_s_at	CAMK2D	calcium/calmodulin-dependant protein kinase II delta	---	chr4:114374777-114375242 (+) // 76.14 // q26 chr12:131555397-131626010 (+) // 98.39 // q24.33	chr4q26	2	regulated down	regulated down
232267_at	GPR133	G protein-coupled receptor 133	283383	q24.33	chr12q24.33	2	regulated down	regulated down
234650_at	LGR4	leucine-rich repeat-containing G protein-coupled receptor 4	---	chr11:27440397-27443381 (-) // 50.58 // p14.1	chr11p14.1	2	down regulated	down regulated
236118_at	LOC100128893	hypothetical protein LOC100128893	1E+08	chr18:19746856-19747678 (-) // 94.87 // q11.2	chr18q11.2	2	regulated down	regulated down
240385_at	GATA6	GATA binding protein 6	---	chr18:19773571-19774031 (+) // 100.0 // q11.2	chr18q11.2	2	regulated	regulated
1554127_s_at	MSRB3	methionine sulfoxide reductase B3	253827	chr12:65672521-65857168 (+) // 94.28 // q14.3	chr12q14.3	3	up regulated	regulated down
201116_s_at	CPE	carboxypeptidase E	1363	chr4:166300093-166419697 (+) // 95.17 // q32.3	chr4q32.3	3	up regulated	regulated down
201278_at	DAB2	disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)	1601	chr5:39371775-39425331 (-) // 99.31 // p13.1	chr5p13	3	up regulated	regulated down
201426_s_at	VIM	vimentin	7431	chr10:17270929-17279591 (+) // 94.16 // p13	chr10p13	3	up regulated	regulated down
201473_at	JUNB	jun B proto-oncogene	3726	chr19:12902332-12904129 (+) // 98.27 // p13.2	chr19p13.2	3	up regulated	regulated down
202035_s_at	SFRP1	secreted frizzled-related protein 1	6422	chr8:41119480-41166980 (-) // 98.73 // p11.21	chr8p12-p11.1	3	up regulated	regulated down
202036_s_at	SFRP1	secreted frizzled-related protein 1	6422	chr8:41121846-41166971 (-) // 97.76 // p11.21	chr8p12-p11.1	3	up regulated	regulated down
202037_s_at	SFRP1	secreted frizzled-related protein 1	6422	chr8:41119478-41166980 (-) // 98.81 // p11.21	chr8p12-p11.1	3	up regulated	regulated down
202202_s_at	LAMA4	laminin, alpha 4	3910	chr6:112430076-112575802 (-) // 99.87 // q21	chr6q21	3	up regulated	regulated down
202686_s_at	AXL	AXL receptor tyrosine kinase	558	chr19:41724822-41767670 (+) // 91.33 // q13.2	chr19q13.1	3	up regulated	regulated down
203706_s_at	FZD7	frizzled homolog 7 (Drosophila)	8324	chr2:202899309-202903160 (+) // 100.0 // q33.1	chr2q33	3	up regulated	regulated down
203951_at	CNN1	calponin 1, basic, smooth muscle	1264	chr19:11649664-11661247 (+) // 99.2 // p13.2	chr19p13.2-p13.1	3	up regulated	regulated down
203989_x_at	F2R	coagulation factor II (thrombin) receptor	2149	chr5:76011750-76031298 (+) // 87.03 // q13.3	chr5q13	3	up regulated	regulated down
204200_s_at	PDGFB	platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)	5155	chr22:39619718-39640990 (-) // 87.82 // q13.1	chr22q12.3-q13.1 22q13.1	3	up regulated	regulated down
204457_s_at	GAS1	growth arrest-specific 1	2619	chr9:89559277-89562104 (-) // 91.58 // q21.33	chr9q21.3-q22	3	up regulated	regulated down
204466_s_at	SNCA	synuclein, alpha (non A4 component of amyloid precursor)	6622	chr4:90646311-90758350 (-) // 77.0 // q22.1	chr4q21	3	up regulated	regulated down
204467_s_at	SNCA	synuclein, alpha (non A4 component of amyloid precursor)	6622	chr4:90646704-90756846 (-) // 82.63 // q22.1	chr4q21	3	up regulated	regulated down
205083_at	AOX1	aldehyde oxidase 1	316	chr2:201450537-201536214 (+) // 93.09 // q33.1	chr2q33	3	up regulated	regulated down
205100_at	GFP2	glutamine-fructose-6-phosphate transaminase 2	9945	chr5:179727699-179780315 (-) // 98.51 // q35.3	chr5q34-q35	3	up regulated	regulated down

206404_at	FGF9	fibroblast growth factor 9 (glia-activating factor)	2254	chr13:22245874-22276184 (+) // 93.66 // q12.11	chr13q11-q12	3	up regulated	regulated down
208609_s_at	TNXB	tenascin XB	7148	chr6:32009125-32065972 (-) // 99.95 // p21.33	chr6p21.3	3	up regulated	regulated down
210299_s_at	FHL1	four and a half LIM domains 1	2273	chrX:135252084-135293103 (+) // 97.7 // q26.3	chrXq26	3	up regulated	regulated down
210355_at	PTH1H	parathyroid hormone-like hormone	5744	chr12:28111023-28123873 (-) // 97.19 // p11.22	chr12p12.1-p11.2	3	up regulated	regulated down
211756_at	PTH1H	parathyroid hormone-like hormone	5744	chr12:28115262-28123001 (-) // 79.0 // p11.22	chr12p12.1-p11.2	3	up regulated	regulated down
212097_at	CAV1	caveolin 1, caveolae protein, 22kDa	857	chr7:116199521-116201233 (+) // 94.53 // q31.2	chr7q31.1	3	up regulated	regulated down
223405_at	NPL	N-acetylneuraminate pyruvate lyase (dihydropicolinate synthase)	80896	chr1:182758928-182799519 (+) // 96.31 // q25.3	chr1q25	3	up regulated	regulated down
224940_s_at	PAPPA	pregnancy-associated plasma protein A, pappalysin 1	5069	chr9:119160611-119164153 (+) // 88.82 // q33.1	chr9q33.2	3	up regulated	regulated down
224941_at	PAPPA	pregnancy-associated plasma protein A, pappalysin 1	5069	chr9:119160611-119164153 (+) // 88.82 // q33.1	chr9q33.2	3	up regulated	regulated down
225544_at	TBX3	T-box 3	6926	chr12:115108059-115121567 (-) // 98.23 // q24.21	chr12q24.1	3	up regulated	regulated down
225728_at	SORBS2	sorbin and SH3 domain containing 2	8470	chr4:186506597-186508184 (-) // 86.55 // q35.1	chr4q35.1	3	up regulated	regulated down
225968_at	PRICKLE2	prickle homolog 2 (Drosophila)	166336	chr3:64079526-64253655 (-) // ??? // p14.1	chr3p14.1	3	up regulated	regulated down
227070_at	GLT8D2	glycosyltransferase 8 domain containing 2	83468	chr12:104382761-104443921 (-) // 98.74 // q23.3	chr12q	3	up regulated	regulated down
227399_at	VGLL3	vestigial like 3 (Drosophila)	389136	chr3:86989786-86991590 (-) // 96.5 // p12.1	chr3p12.1	3	up regulated	regulated down
227607_at	STAMBPL1	STAM binding protein-like 1	57559	chr10:90672857-90683244 (+) // 92.46 // q23.31	chr10q23.31	3	up regulated	regulated down
229404_at	TWIST2	twist homolog 2 (Drosophila)	117581	chr2:239757162-239832230 (+) // 50.77 // q37.3	chr2q37.3	3	up regulated	regulated down
1569986_x_at	TNNT3	troponin T type 3 (skeletal, fast)	7140	chr11:1940932-1959932 (+) // 73.79 // p15.5	chr11p15.5	4	up regulated	regulated up
202575_at	CRABP2	cellular retinoic acid binding protein 2	1382	chr1:156669409-156675375 (-) // 95.87 // q23.1	chr1q21.3	4	up regulated	regulated up
204447_at	ProSAPiP1	ProSAPiP1 protein	9762	chr20:3143272-3149207 (-) // 95.64 // p13	chr20p13	4	up regulated	regulated up
212909_at	LYPD1	LY6/PLAUR domain containing 1	116372	chr2:133402367-133427833 (-) // 98.81 // q21.2	chr2q21.2	4	up regulated	regulated up
218678_at	NES	nestin	10763	chr1:156638557-156647199 (-) // 97.21 // q23.1	chr1q23.1	4	up regulated	regulated up
227261_at	KLF12	Kruppel-like factor 12	11278	chr13:74260154-74261896 (-) // 98.41 // q22.1	chr13q22	4	up regulated	regulated up

Table 2 – DAVID gene functional classification

Gene Functional Classification		
Gene Group 1	Enrichment Score - 2.04	
Probe set(s)	Gene Symbol	Gene Name
203824_at	TSPAN 8	tetraspanin 8
201839_s_at	EPCAM	epithelial cell adhesion molecule
230869_at, 214825_at	FAM155A	family with sequence similarity 155, member A
232267_at	GPR133	G protein-coupled receptor 133
1559214_at	NRG4	neuregulin 4
227070_at	GLT8D2	glycosyltransferase 8 domain containing 2
226492_at	SEMA6D	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D
203706_s_at	FZD7	frizzled homolog 7 (Drosophila)
228863_at	PCDH17	protocadherin 17

Table 3 – IPA top 20 biological functions

Function Annotation	p-value	Molecules	# Molecules
carcinoma	9.13E-11	AGTR1, AKR1C1/AKR1C2, ALDH1A1, ARHGAP18, AXL, CAV1, CFH, CFI, CNN1, CP, CRABP2, DAB2, DDR2, DMD, DPP4, EHF, EPCAM, F2R, FAM134B, FHL1, FZD7, GAS1, GLT8D2, HSD17B2, JUNB, LAMA4, LYZ, MECOM, NES, PAPP, PDGFB, PRSS12, RGS2, RGS4, SFRP1, SLC16A4, STS, TBX3, TNNT3, TNXB, TSPAN8, TWIST2, VIM	43
cancer	2.13E-08	AGTR1, AKR1C1/AKR1C2, ALDH1A1, ARHGAP18, AXL, CAV1, CFH, CFI, CNN1, CP, CRABP2, DAB2, DDR2, DMD, DPP4, EHF, EPCAM, F2R, FAM134B, FHL1, FZD7, GAS1, GLT8D2, HSD17B2, JUNB, LAMA4, LYZ, MECOM, NES, PAPP, PDGFB, PRSS12, PTHLH, RGS2, RGS4, SFRP1, SLC16A4, SORBS2, STS, TBX3, TNNT3, TNXB, TSPAN8, TWIST2, VIM	45
adenocarcinoma	3.75E-08	CAV1, CFH, CNN1, CP, DAB2, DDR2, DPP4, EPCAM, F2R, FHL1, GAS1, GLT8D2, SFRP1, TBX3, TNXB, TWIST2, VIM	17
tumorigenesis	5.07E-08	AGTR1, AKR1C1/AKR1C2, ALDH1A1, ARHGAP18, AXL, CAV1, CFH, CFI, CNN1, CP, CRABP2, DAB2, DDR2, DMD, DPP4, EHF, EPCAM, F2R, FAM134B, FHL1, FZD7, GAS1, GLT8D2, HSD17B2, JUNB, KITLG, LAMA4, LYZ, MECOM, NES, PAPP, PDGFB, PRSS1/PRSS3, PRSS12, PTHLH, RGS2, RGS4, SFRP1, SLC16A4, SORBS2, STS, TBX3, TNNT3, TNXB, TSPAN8, TWIST2, VIM	47
reproductive system disorder	7.82E-08	AGTR1, ALDH1A1, CAV1, CFH, CNN1, CP, CRABP2, DAB2, DDR2, DPP4, EPCAM, F2R, FHL1, GLT8D2, HS3ST1, HSD17B2, JUNB, LAMA4, MECOM, SFRP1, SORBS2, STS, TBX3, TNXB, TWIST2, VIM	26
gynecological disorder	2.34E-07	ALDH1A1, CAV1, CFH, CNN1, CP, CRABP2, DAB2, DDR2, DPP4, EPCAM, F2R, FHL1, GLT8D2, HS3ST1, MECOM, SFRP1, SORBS2, TBX3, TNXB, TWIST2, VIM	21
cardiovascular disorder	1.37E-06	AGTR1, ALDH1A1, ARHGAP6, AXL, BCHE, CAMK2D, CAV1, CFC1/CFC1B, CGNL1, DDR2, DMD, DPP4, F2R, FAM65B, FHL1, JUNB, KITLG, KLF12, LAMA4, LSAMP, LYPD6B, MAOB, MECOM, MSRB3, NPL, PAPP, PDGFB, PID1, PRICKLE2, RGS2, RGS4, SNCA, SORBS2, TBX3, VIM	35
invasion of cells	7.58E-06	AXL, CAV1, DAB2, DPP4, ELF3, EPCAM, F2R, JUNB, KITLG, NES, PAPP, RGS4, SFRP1, VIM	14
metastatic colorectal cancer	1.10E-05	AKR1C1/AKR1C2, CAV1, CRABP2, DMD, F2R, FZD7, PAPP, RGS4	8
endometrial cancer	1.29E-05	CAV1, CNN1, DDR2, FHL1, GLT8D2, TBX3, TNXB, TWIST2	8
genital tumor	1.36E-05	AGTR1, ALDH1A1, AXL, CAV1, CFH, CFI, CP, DAB2, DPP4, EHF, EPCAM, F2R, KITLG, LYZ, MECOM, NES, SFRP1, VIM	18
invasion of ovarian cancer cell lines	1.49E-05	AXL, CAV1, DPP4, PAPP	4
uterine cancer	1.78E-05	ALDH1A1, CAV1, CFH, CNN1, CRABP2, DDR2, FHL1, GLT8D2, SFRP1, SORBS2, TBX3, TNXB, TWIST2	13
proliferation of cells	1.86E-05	AGTR1, AXL, CAMK2D, CAV1, CNN1, DAB2, DDR2, DPP4, EHF, ELF3, EPCAM, F2R, FGF9, FHL1, FZD7, GAS1, JUNB, KITLG, LAMA4, MECOM, PAPP, PDGFB, PRKAR2B, PTHLH, RGS2, RGS4, RNF128, SFRP1, SNCA, STS, TBX3, VIM	32
development of forelimb	2.33E-05	CRABP2, GAS1, MECOM, TBX3	4
tissue development	4.06E-05	AGTR1, ALDH1A1, AXL, BCHE, CAV1, CRABP2, DAB2, DMD, FGF9, FOXQ1, GAS1, HSD17B2, JUNB, KITLG, LAMA4, MECOM, PAPP, PDGFB, PRICKLE2, PRKAR2B, PTHLH, RGS2, RGS4, SFRP1, SLC26A4, SNCA, STS, TBX3	28
genetic disorder	4.84E-05	AGTR1, AKR1C1/AKR1C2, ALDH1A1, ARHGAP18, ARHGAP6, AXL, BBOX1, BCHE, CAMK2D, CAV1, CFC1/CFC1B, CFH, CFI, CGNL1, CNN1, CP, CPE, CRABP2, DAB2, DDR2, DMD, DPP4, EHF, ELF3, EPCAM, F2R, FAM134B, FGF9, FHL1, GAS1, GLT8D2, GPR133, HSD17B2, JUNB, KITLG, KLF12, LAMA4, LSAMP, LYPD6B, LYZ, MAOB, ME1, MECOM, METTL7A, MSRB3, NES, NPL, PAPP, PDGFB, PRKAR2B, PRSS1/PRSS3, PRSS12, PTHLH, RGS4, SFRP1, SLC26A4, SNCA, SORBS2, TBX3, TNNT3, TNXB, TSPAN8, VIM	63
neurological disorder	4.86E-05	AGTR1, AKR1C1/AKR1C2, ALDH1A1, ARHGAP18, ARHGAP6, AXL, BBOX1, BCHE, CAMK2D, CAV1, CGNL1, CNN1, CP, CPE, DMD, DPP4, F2R, FAM134B, FHL1, GAS1, GLT8D2, HSD17B2, JUNB, KITLG, KLF12, LAMA4, LSAMP, MAOB, ME1, MECOM, NES, NPL, PDGFB, PRKAR2B, PRSS1/PRSS3, PRSS12, RGS4, RNF128, SFRP1, SLC26A4, SNCA, SORBS2, TNXB, VIM	44
colony formation of cells	5.22E-05	ALDH1A1, CAV1, EHF, ELF3, EPCAM, JUNB, KITLG, MECOM, PDGFB, PTHLH, SFRP1	11
cell movement	5.36E-05	AGTR1, AXL, CAV1, CFH, CNN1, DAB2, DDR2, DPP4, ELF3, EPCAM, F2R, FGF9, FHL1, JUNB, KITLG, NES, PAPP, PDGFB, PRSS1/PRSS3, PTHLH, RGS4, SEMA6D, SFRP1, SLC16A4, VIM	25

Table 4 – IPA canonical pathways

Top Canonical Pathways		
Name	p-value	Molecules
G-Protein Coupled Receptor Signaling	2.36E-03	F2R, RGS4, RGS2, FZD7, AGTR1, PRKAR2B, CAMK2D, PTHLH, GPR133
C21-Steroid Hormone Metabolism	4.37E-03	AKR1C1/AKR1C2, HSD17B2
Androgen and Estrogen Metabolism	5.23E-03	AKR1C1/AKR1C2, HSD17B2, STS
cAMP-Mediated Signaling	7.50E-03	RGS4, RGS2, AGTR1, PRKAR2B, CAMK2D
Clathrin-mediated Endocytosis Signaling	1.33E-02	FGF9, F2R, PDGFB, DAB2
Complement System	1.51E-02	CFI,CFH
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	2.45E-02	FZD7,PRSS1/PRSS3,PDGFB,CAMK2D ,SFRP1
Tryptophan Metabolism	2.51E-02	AOX1,MAOB,ALDH1A1
Bile Acid Biosynthesis	3.09E-02	AKR1C1/AKR1C2,ALDH1A1
Histidine Metabolism	3.31E-02	MAOB,ALDH1A1
Ovarian Cancer Signaling	3.98E-02	FGF9,FZD7,PRKAR2B
Pyruvate Metabolism	4.68E-02	ME1,ALDH1A1
Valine, Leucine and Isoleucine Degradation	4.79E-02	AOX1,ALDH1A1
Lysine Degradation	4.79E-02	BBOX1,ALDH1A1

The above canonical pathways are significantly associated with the genes of our candidate list.

Table 5 – IPA molecular networks

ID	Molecules in Network	Focus Molecules	Top Functions
1	14-3-3, Ap1, AXL , Calcineurin protein(s), Calmodulin, CAMK2D , CaMKII, CAV1 , CNN1 , Collagen Alpha1, Collagen type IV, CP , DAB2 , DMD , EHF , ELF3 , EPCAM , FHL1 , GAS1 , GFPT2 , Hsp27, JUNB , LYZ , NES , Nfat (family), NFkB (complex), Pdgf (complex), PDGF BB, PDGFB , RGS2 , RNF128 , SFRP1 , Tgf beta, TSPAN8 , VIM	22	Cancer, Cellular Movement, Gene Expression
2	ADCY, AGTR1 , Alp, ARHGAP6 , CFH , CPE , Cyclin A, DPP4 , ERK1/2, F2R , FGF9 , FSH, FZD7 , G protein alpha1, Gpcr, GPR133 , GTPASE, hCG, Insulin, Lh, Mapk, ME1 , p85 (pik3r), PAPPA , Pka, PLC, PRKAR2B , PRSS1/PRSS3 , PTHLH , Ras, RGS4 , STS , TBX3 , Trypsin, Vegf	17	DNA Replication, Recombination, and Repair, Lipid Metabolism, Small Molecule Biochemistry
3	ABL1, AKR1C1/ AKR1C2 , AOX1 , ARHGAP18 , BBOX1 , beta-estradiol, CCND1, CDK2-Cyclin D1, CMAS, CPXM1, DHRS7, ECM2, ERBB2, EZH2, FAM134B , FAM65B , FOXN3, Fxyd3, KLF12 , KRAS, LYPD6B , MAPK3, MPZL2, OSBPL5, OSM, ProSAPiP1 , PRR15L, RB1, SEMA6D , SLC22A18AS, ST6GALNAC1 , STK11, SVIP , TMEM37 , TTC39A	13	Cancer, Cellular Development, Cellular Growth and Proliferation
4	Ca2+, CALCA, CALML3, CD52, CD101, CFI , CGNL1 , CHST2, CLEC10A, CLTC, CRABP2 , CRABP, DEFA4, EMR2, HS3ST1 , HS6ST1, IL13, IL17B, LAMA4 , MYB, NDST1, NPS, PCDH17 , PID1 , PRSS12 , Retinoic acid-CRABP2, Retinoic acid-CRABP2-RAR-RXR, SLC26A4 , SLC7A2, STAMBPL1 , STRA6, TFEC, TPCN1, tretinoin, VGLL3	11	Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry
5	ALDH1A1 , BCHE , C11orf82, CARD17, CKMT1A/CKMT1B, DDR2 , DEFB104A/DEFB104B, dihydrotestosterone, Egfbp2, FAM105B, FOXQ1 , GBP6, HIF1A, HOXA10, HSD17B2 , IFNG, IL1B, LYPD1 , MSRB3 , MYCN, NPL , progesterone, RAB20, RNASE4 , RNASE7, S100A3, SCLY, SCUBE1, SLC16A4 , SLC2A9, SLC5A2, TREM3, TWIST2 , UAP1, ZNF217	11	Cell-To-Cell Signaling and Interaction, Endocrine System Development and Function, Small Molecule Biochemistry
6	26s Proteasome, Akt, Akt-Calmodulin-Hsp90-Nos3, ANGPTL1, Caspase, CDC42SE1, ERK, HERP, Hsp90, Igkv1-117, IL1, IL22R1-IL10R2, Jnk, KITLG , LAG3, LSAMP , MAOB , MASI, MECOM , MTCH2, P38, MAPK, PI3K (complex), PIK3IP1, Pkc(s), PPM1L, RAGE, RETNLB, SEC14L2, SLC20A2, SNCA , SORBS2 , Tnfrsf22/Tnfrsf23, TNNT3 , TNXB , TRAF1-TRAF2-TRAF3	8	Cell Death, Lipid Metabolism, Small Molecule Biochemistry
7	DVL1 , PRICKLE2	1	Cellular Assembly and Organization, Tissue Development, Nervous System Development and Function
8	METTL7A , miR-146a/miR-146b/miR-146b-5p, miR-155 (human, mouse)	1	Cancer, Gastrointestinal Disease, Genetic Disorder

This table depicts candidate genes (described as focus molecules - **bolded**) along with related molecules, in molecular networks defined by biological functions and processes.

Table 6 - TCGA regional copy number alterations

Chromosome arm	Gene Candidates	Gain or Loss %	q-value
1q	CRABP2, NES, RGS4, DDR2, NPL, RGS2, CFH, ELF3	47% gain	2.11E-11
3q	LSAMP, AGTR1, CP, MECOM, BCHE	59% gain	0
4p	HS3ST1	65% loss	0
4q	SNCA, CFI, PRSS12, CAMK2D, CPE, SORBS2	67% loss	0
5q	F2R, GFPT2	51% loss	6.75E-12
6p	TNXB, FAM65B, FOXQ1	47% gain 42% loss	0.00937 gain 0.226 loss
6q	MEI, LAMA4, ARHGAP18,	59% loss	6.12E-13
7q	PRKAR2B, SLC26A4, CAV1, PRSS1	47% gain	0.000201
8p	SFRP1	74% loss	0
9q	ALDH1A1, GAS1, PAPPA,	60% loss	0
11p	EHF, LRG4, BBOX1, SVIP, TNNT3	49% loss	0.00196
12p	PTHLH	54% gain	0.00019
13q	FGF9, PCDH17, KLF12, FAM155A	64% loss	0
14q	RNASE4	49% loss	1.09E-07
15q	SEMA6D, CGNL1, NRG4,	55% loss	7.22E-15
16q	HSD17B2	80% loss	0
17q	ST6GALNAC1	75% loss	0
18q	LOC100128893, GATA6	68% loss	0
19p	JUNB, CNN1	60% loss	1.86E-15
19q	AXL	54% loss	0
20p	ProSAPiP1	56% gain	0.00019
22q	PDGFB	79% loss	0

Table 7 – Clinical characteristics of the patient cohort

		Number of Patients	Time (range)
Age		196	64 yr (52 - 81 yr)
Stage	I	10	
	II	21	
	III	139	
	IV	26	
Res. Disease	<1cm	60	
	1-2cm	17	
	> 2cm	66	
	milliary	5	
Survival Time	incidence of death	86	35 mo (0 - 134 mo)
Dis. Progression Time	incidence of progression	144	22 mo (0 - 134 mo)
CP staining	0	52	
	0<2	105	
	2	8	
	missing	31	

FIGO staging was used. Residual disease at surgery (Res. Disease) was evaluated by a gynecologic oncologist. Survival time is in months (mo) from the date of primary resection until the event of death due to OC or until the last contact date with the patient. Disease progression time (Dis. Progression) is from the date of primary resection. CP staining is the staining intensity observed on the TMA with an antibody against CP. Intensity is scored as follows: 0-negative, 1-weak, 2-moderate.

REFERENCES

- Abdel-Rahman WM, Kalinina J, Shoman S, Eissa S, Ollikainen M, Elomaa O, Eliseenkova AV, Butzow R, Mohammadi M, Peltomäki P. 2008. Somatic FGF9 mutations in colorectal and endometrial carcinomas associate with membranous B-catenin. *Human Mutation* 29(3):390-397.
- Agarwal A, Covic L, Sevigny LM, Kaneider NC, Lazarides K, Azabdaftari G, Sharifi S, Kuliopulos A. 2008. *Mol Cancer Ther* 7(9):2746-2757.
- Agarwal A, Tressel SL, Kaimal R, Balla M, Lam FH, Covic L, Kuliopulos A. 2010. Identification of a metalloprotease-chemokine signaling system in the ovarian cancer microenvironment: implications for antiangiogenic therapy. *Canc Res* 70(14):5580-5890.
- Apte SM, Bucana CD, Killian JJ, Gerhenson DM, Fidler IJ. 2004. Expression of platelet-derived growth factor and activated receptor in clinical specimens of epithelial ovarian cancer and ovarian carcinoma cell lines. *Gynecologic Oncology* 93:78-86.
- Ardura JA, Rayego-Mateos S, Ramila D, Ruiz-Ortega M, Esbrit P. 2010. Parathyroid hormone-related protein promotes epithelial-mesenchymal transition. *J Am Soc Nephrol* 21:237-248
- Arora S, Bisanz KM, Peralta LA, Basu GD, Choudhary A, Tibes R, Azorsa DO. 2010. RNAi screening of the kinome identifies modulators of cisplatin response in ovarian cancer cells. *Gynecologic Oncology* 118:220-227.
- Auersperg N, Wong AST, Choi K-C, Kang SK, Leung PCK. 2001. Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr. Rev.* 22:225-288.
- Baranova A, Gowder S, Naouar S, King S, Schlauch K, Jarrar M, Ding Y, Cook B, Chandhoke V, Christensen A. 2006. Expression profile of ovarian tumors: distinct signature of Sertoli-Leydig cell tumor. *Int J Gynecol Cancer* 26:1963-1972.
- Benedet JL and Pecorelli S. 2000. Staging Classifications and Clinical Practice Guidelines for Gynaecological Cancers. *International Journal of Gynecology and Obstetrics* 70:207-312.
- Berek JS, Chalas E, Edelson M, Moore DH, Burke WM, Cliby WA, Berchuck A. 2010. Prophylactic and risk-reducing bilateral salpingo-oophorectomy - Recommendations based on risk of ovarian cancer. *Obstet Gynecol* 116:733-743.

Berns EMJJ, Bowtell DD. 2012. The changing view of high-grade serous ovarian cancer. *Cancer Res* 72:2701-2704.

Bignotti E, Tassi RA, Calza S, Ravaggi A, Romani C, Rossi E, Falchetti M, Odicino FE, Pecorelli S, Santin AD. 2006. Differential gene expression profiles between tumor biopsies and short-term primary cultures of ovarian serous carcinomas: Identification of novel molecular biomarkers for early diagnosis and therapy. *Gynecologic Oncology* 103:405-416.

Birch AH, Quinn MCJ, Filali-Mouhim A, Provencher DM, Mes-Masson A-M, Tonin PN. 2008. Transcriptome analysis of serous ovarian cancers identifies differentially expressed chromosome 3 genes. *Molecular Carcinogenesis* 47:56-65.

Bohnkamp J and Schoneberg T. 2011. Cell Adhesion Receptor GPR133 Couples to Gs Protein. *The Journal of Biological Chemistry* 286(49):41912-41916.

Boirc A, Covic L, Agarwal A, Jacques S, Sherifi S, Kuliopulos A. 2005. PAR1 is a matrix metalloprotease-1 receptor that promotes invasion and tumorigenesis of breast cancer cells. *Cell* 120:303-313.

Bonome T, Lee J-Y, Park D-C, Radonovich M, Pise-Masison C, Brady J, Gardner GJ, Hao K, Wong WH, Barrett JC, Lu KH, Sood AK, Gershenson DM, Mok SC, Birrer MJ. 2005. Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. *Cancer Res* 65(22):10602-10612.

Buckhaults P, Zhang Z, Chen Y-C, Wang T-L, Croix BS, Saurabh S, Bardelli A, Morin PJ, Polyak K, Hruban RH, Velculescu VE, Shih I-M. 2003. Identifying tumor origin using a gene expression-based classification map. *Cancer Research* 63:4144-4149.

Buys SS, Patridge E, Black A, Johnson CC, Lamerato L, Isaacs C, Reding DJ, Greenlee RT, Yokochi LA, Kessel B, Crawford ED, Church TR, Andriole GL, Weissfeld JL, Fouad MN, Chia D, O'Brien B, Ragard LR, Clapp JD, Rathmell JM, Riley TL, Hartge P, Pinsky PF, Zhu CS, Izmirian G, Kramer BS, Miller AB, Xu J-L, Prorok PC, Gohagan JK, Berg CD. 2011. Effect of Screening on Ovarian Cancer Mortality. *JAMA* 305(22):2295-2303.

Calebiro D, Nikolaev VO, Persani L, Lohse MJ. 2010. Signaling by internalized G-protein-coupled receptors. *Trends in Pharmacological Sciences* 31(5):221-228.

Canadian Cancer Society's Steering Committee on Cancer Statistics. *Canadian Cancer Statistics 2011*. Toronto, ON: Canadian Cancer Society; 2011.

Chivukula M, Niemeier LA, Edwards R, Nikiforova M, Mantha G, McManus K, Carter G. 2011. Carcinomas of distal fallopian tube and their association with tubal intraepithelial carcinoma: do they share a common "precursor" lesion? *Loss*

of heterozygosity and immunohistochemical analysis using PAX 2, WT-1, and P53 markers. *ISRN Obstetrics and Gynecology* 2011:1-8.

Cody NAL, Ouellet V, Manderson EN, Quinn MCJ, Filiali-Mouhim A, Tellis P, Zietarska M, Provencher DM, Mes-Masson A-M, Chevrette M, Tonin PN. 2007. Transfer of chromosome 3 fragments suppress tumorigenicity of an ovarian cancer cell line monoallelic for chromosome 3p. *Oncogene* 26:618-632.

Cody NAL, Zietarska M, Filiali-Mouhim A, Provencher DM, Mes-Masson A-M, Tonin PN. 2008. Influence of monolayer, spheroid, and tumor growth conditions on chromosome 3 gene expression in tumorigenic epithelial ovarian cancer cell lines. *BMC Medical Genomics* 34:1-16.

Cregger M, Berger AJ, Rimm DL. 2006. Immunohistochemistry and Quantitative Analysis of Protein Expression. *Arch Pathol Lab Med* 130:1026-1030.

Crum CP, Drapkin R, Kindelberger D, Medeiros F, Miron A, Lee Y. 2007. Lessons from BRCA: The Tubal Fimbria Emerges as an Origin for Pelvic Serous Cancer. *Clinical Research and Medicine* 5(1)35-44.

Dorsam RT and Gutkind JS. 2007. G-protein-coupled receptors and cancer. *Nature Reviews* 7:79-94.

Drummond AE, Tellback M, Dyson M, Findlay JK. 2007. Fibroblast growth factor-9, a local regulator of ovarian cancer. *Endocrinology* 148(8):3711-3721.

Feeley KM and Wells M. 2001. Precursor lesions of ovarian epithelial malignancy. *Histopathology* 38:87-95.

Ferguson SSG. 2001. Evolving concepts in G protein-coupled receptor endocytosis: the role in receptor desensitization and signaling. *Pharmacol Rev* 53(1):1-24.

Farley J, Ozbun LL, Birrer MJ. 2008. Genomic analysis of epithelial ovarian cancer. *Cell Research* 18:538-548.

Fu S, Naing A, Macus CF, Kuo MT, Kurzrock R. 2012. Overcoming platinum resistance through the use of a copper-lowering agent. *Mol Cancer Ther* 11:1221-1225.

Fukunishi H, Yukimura N, Murata K, Gotoh A, Kitazawa R, Kitazawa. 1994. Immunohistochemical study on the expression of parathyroid hormone-related protein in ovarian tumors. *J Bone Miner Met* 12(suppl 1):S153-155.

Garson K, Shaw TJ, Clark KV, Yao D-S, Vanderhyden BC. 2005. Models of ovarian cancer - are we there yet? *Molecular and Cellular Endocrinology* 239:15-26.

George AJ, Thomas WG, Hannan RD. 2010. The renin-angiotensin system and cancer: old dog, new tricks. *Nature Reviews* 10:745-759.

Godwin AK, Testa JR, Hamilton TC. 1993. The biology of ovarian cancer development. *Cancer* 71:530-536.

Gortzak-Uzan L, Ignatchenko A, Evangelou A, Agochiya M, Brown KA, St. Onge P, Kireeva I, Schmitt-Ulms G, Brown TJ, Murphy J, Rosen B, Shaw P, Jurisica I, Kislinger T. 2007. A proteome resource of ovarian cancer ascites: integrated proteomic and bioinformatic analyses to identify putative biomarkers. *Journal of Proteome Research* 7:339-351.

Grisaru-Granovsky S, Salah Z, Maoz M, Russ D, Beller U, Bar-Shavit RB. 2005. Differential expression of protease activated receptor 1 (PAR1) and pY397FAK in benign and malignant human ovarian tissue samples. *Int. J. Cancer*. 113:372-378.

Gupte J, Swaminath G, Danao J, Tian H, Li Y, Wu X. 2012. Signaling property study of adhesion G-protein-coupled receptors. *FEBS Letters* 586:1214-1219.

Hall J, Brown R, Paul J. 2007. An exploration into study design for biomarker identification: issues and recommendations. *Cancer Genomics and Proteomics* 4:111-120.

Hanrahan AJ, Schultz N, Westfall ML, Sakr RA, Giri DD, Scarperi S, Janikariman M, Olvero N, Stevens EV, She Q-B, Aghajanian C, King TA, de Stanchina E, Spriggs DR, Heguy A, Taylor BS, Sander C, Rosen N, Levine DA, Solit DB. 2012. Genomic complexity and AKT dependence in serous ovarian cancer. *Cancer Discovery* 2(1):56-67.

Helleman J, Smid M, Jansen MPH, van der Burg MEL, Berns EMJJ. 2010. Pathway analysis of gene lists associated with platinum-based chemotherapy resistance in ovarian cancer: the big picture. *Gynecologic Oncology* 117:170-176.

Hendrix ND, Wu R, Kuick R, Schwartz DR, Fearon ER, Cho KR. 2006. Fibroblast growth factor 9 has oncogenic activity and is a downstream target of wnt signaling in ovarian endometrioid adenocarcinomas. *Cancer Res* 66:1354-1362.

Hennessy BT, Coleman RL, Markman M. 2009. Ovarian cancer. *Lancet* 374:1371-1382.

Hennig EE, Mikula M, Rubel T, Dadlex M, Ostrowski J. 2011. Comparative kinome analysis to identify putative colon tumor biomarkers. *J Mol Med* 90:447-456.

Henriksen R, Funa K, Wilander E, Backstrom T, Ridderheim M, Oberg K. 1993. Expression and prognostic significance of platelet-derived growth factor and its receptors in epithelial ovarian neoplasms. *Cancer Research* 53:4550-4554.

Hibbs K, Skubitz K, Pambuccian SE, Casey RC, Burleson KM, Oegema Jr. TR, Thiele KK, Grindle SM, Bliss RL, Skubitz APN. 2004. Differential gene expression in ovarian carcinoma - identification of potential biomarkers. *Am J Pathol* 165(2):397-414

Hough CD, Cho KR, Zonderman AB, Schwartz DR, Morin PJ. 2001. Coordinately up-regulated genes in ovarian cancer. *Canc Res* 61:3869-3876.

Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* 37(1):1-13.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4(1):44-57.

Hunn J and Rodriguez GC. 2012. Ovarian cancer: etiology, risk factors, and epidemiology. *Clinical Obstetrics and Gynecology* 55(1):3-23.

Hurst JH, Henkel PA, Brown AL, Hooks SB. 2008. Endogenous RGS proteins attenuate G α -mediated lysophosphatidic acid signaling pathways in ovarian cancer cells. *Cellular Signaling* 20:381-389.

Hurst JH, Mendpara N, Hooks SB. 2008. Regulator of G-protein signalling expression and function in ovarian cancer cell lines. *Cellular and Molecular Biology Letters* 14:153-174.

Ino K, Shibata K, Kajiyama H, Yamamoto E, Nagasaka T, Nawa A, Nomura S, Kikkawa F. 2006. Angiotensin II type I receptor expression in ovarian cancer and its correlation with tumor angiogenesis and patient survival. *British Journal of Cancer* 94:552-560.

Jemal A, Center MM, DeSantis C, Ward EM. 2010. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 19(8):1893-1907.

Jimenez-Marin A, Collado-Romera M, Ramirez-Boo M, Arce C, Garrido JJ. 2009. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proceedings* 3(Suppl 4):S6

Khatri P, Sirota M, Butte AJ. 2012. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 8(2):e1002375.

King TD, Zhang W, Suto MJ, Li Y. 2011. Frizzled7 as an emerging target for cancer therapy. *Cellular Signalling* 24:846-851.

Klebig C, Seitz S, Korsching E, Kristiansen G, Gustavus D, Scherneck S, Petersen I. 2005. Profile of differentially expressed genes after transfer of chromosome 17 into the breast cancer cell line CAL51. *Genes, Chromosomes and Cancer* 44:233-246.

Kobel M, Kalloger SE, Boyd N, McKinney S, Mehl E, Palmer C, Leung S, Bowen NJ, Ionescu DN, Rajput A, Prentice LM, Miller D, Santos J, Swenerton K, Gilks BC, Huntsman D. 2008. Ovarian carcinoma subtypes are different disease: implications for biomarker studies. *PLoS Med* 5(12):e232

Lappano R and Maggiolini. 2011. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature Reviews: Drug Discovery* 10:47-60.

Lappano R and Maggiolini M. 2012. GPCRs and cancer. *Acta Pharmacologica Sinica* 33:351-362.

Le Page C, Provencher D, Maugard CM, Ouellet V, Mes-Masson A-M. 2004. Signature of a silent killer: expression profiling in epithelial ovarian cancer. *Expert Review of Molecular Diagnostics* 4(2):157-167.

Le Page C, Marineau A, Bonza PK, Rahimi K, Cyr L, Labouba I, Madore J, Delvoye N, Mes-Masson A-M, Provencher DM, Cailhier J-F. 2004. BTN3A2 expression in epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. *PLoS ONE* 7(6):e38541.

Lee C and Meenakshi R. 2004. Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biology* 5:231.

Lee CM, Lo H-W, Shao R-P, Wang S-C, Xia W, Gershenson DM, Hung M-C. 2004. Selective activation of ceruloplasmin promoter in ovarian tumors: potential use for gene therapy. *Cancer Research* 64:1788-1793.

Lee Y, Miron A, Drapkin R, Nucci MR, Medeiros F, Saleemuddin A, Garber J, Birch C, Mou H, Gordon RW, Cramer DW, McKeon FD, Crum CP. 2007. A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *J Pathol* 211:26-35.

Leung EHL, Leung PCK, Auersperg N. 2001. Differentiation and growth potential of human ovarian surface epithelial cells expressing temperature-sensitive SV40 T antigen. *In Vitro Cell. Dev. Biol.-Animal* 37:515-521.

Lu KH, Patterson AP, Wang L, Marquez RT, Atkinson EN, Baggerly KA, Ramoth LR, Rosen DG, Liu J, Hellstrom I, Smith D, Hartmann L, Fishman D, Berchuck A, Schmandt R, Whitaker R, Gershenson DM, Mills GB, Bast Jr RC.

2004. Selection of potential markers for epithelial ovarian cancer with gene expression arrays and recursive descent partition analysis. *Clinical Cancer Research* 10:3291-3300.

Ma S, Yang Y, Wang C, Hui N, Gu L, Zhong H, Cai Z, Wang Q, Zhang Q, Li N, Cao X. 2009. Endogenous human CaMKII inhibitory protein suppresses tumor growth by inducing cell cycle arrest and apoptosis through down-regulation of phosphatidylinositide 3-kinase/Akt/HDM2 pathway. *The Journal of Biological Chemistry* 284(37):24773-24782.

Machida S, Saga Y, Takei Y, Mizuno I, Takayama T, Kohno T, Konna R, Ohwada M, Suzuki M. 2005. Inhibition of peritoneal dissemination of ovarian cancer by tyrosin kinase receptor inhibitor SU6668 (TSU-68). *Int. J. Cancer*. 114:224-229.

Mackenzie SL, Gillespie MT, Scurry JP, Planner RS, Martin TJ, Danks JA. 1994. Parathyroid hormone-related protein and human papillomavirus in gynecological tumors. *Int. J. Cancer* 56:324-330.

Marinissen MJ and Gutkind JS. 2001. G-protein-coupled receptors and signaling networks: emerging paradigms. *TRENDS in Pharmacological Sciences* 22(7):368-376.

Marquez R, Baggerly KA, Patterson AP, Liu J, Broaddus R, Frumovitz M, Atkinson EN, Smith DI, Hartmann L, Fishman D, Berchuck A, Whitaker R, Gershenson DM, Mills GB, Bast Jr RC, Lu KH. 2005. Patterns of gene expression in different histotypes of epithelial ovarian cancer correlate with those in the normal fallopian tube, endometrium, and colon. *Clin Cancer Res* 11:6116-6126.

Matei D, Emerson RE, Lai Y-C, Baldrige LA, Rao J, Yiannoutsos C, Donner DB. Autocrine activation of PDGFR α promotes the progression of ovarian cancer. *Oncogene* 25:2060-2069.

Matei D, Chang DD, Jeng M-H. 2004. Imatinib mesylate (gleevec) inhibits ovarian cancer cell growth through a mechanism dependent on platelet-derived growth factor receptor α and Akt inactivation. *Clinical Cancer Research* 10:681-690.

Matei D, Emerson RE, Schilder J, Menning N, Baldrige LA, Johnson CS, Breen T, McClean J, Stephens D, Whalen C, Sutton G. Imatinib mesylate in combination with docetaxel for the treatment of patients with advanced, platinum-resistance ovarian cancer and primary peritoneal carcinomatosis. *Cancer* 133(4):723-732.

McCauley LK and Martin TJ. Twenty-five years of PTHrP progress: from cancer hormone to multifunctional cytokine. *Journal of Bone and Mineral Research* 27(6):1231-1239.

Medeiros F, Muto MG, Lee Y, Elvin JA, Callahan MJ, Feltmate C, Garber JE, Cramer DW, Crum CP. 2006. The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am J Surg Pathol* 30:230-236.

Mills GB and Moolenaar WH. 2003. The emerging role of lysophosphatidic acid in cancer. *Nature Reviews-Cancer* 3:582-591.

Mok SC, Chan WY, Wong KK, Cheung KK, Lau CC, Ng SW, Baldini A, Colitti CV, Rock CO, Berkowitz RS. 1998. DOC-2, a candidate tumor suppressor gene in human epithelial ovarian cancer. *Oncogene* 16:2381-2387.

Ono K, Tanaka T, Tsunoda T, Kitahara O, Kihara C, Okamoto A, Ochiai K, Takagi T, Nakamura Y. 2000. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Research* 60:5007-5011.

Ouellet V, Guyot M-C, Le Page C, Filali-Mouhim A, Lussier C, Tonin PN, Provencher DM, Mes-Masson A-M. 2006. Tissue array analysis of expression microarray candidates identifies markers associated with tumor grade and outcome in serous epithelial ovarian cancer. *Int. J. Cancer*. 119:599-607.

Ozols RF. 2002. Update on the management of ovarian cancer. *The Cancer Journal* 8(Supple 1):S22-S30.

Permuth-Wey J, Chen YA, Tsai T-T, Chen Z, Qu X, Lancaster JM, Stockwell H, Dagne G, Iverson E, Risch H, Barnholtz-Sloan J, Cunningham JM, Vierkant RA, Fridley BL, Sutphen R, McLaughlin J, Narod SA, Goode EL, Schildkraut JM, Fenstermacher D, Phelan CM, Sellers TA. 2011. Inherited variants in mitochondrial biogenesis genes may influence epithelial ovarian cancer risk. *Epidemiol Biomarkers Prev* 20(6):1131-1145.

Pitteri SJ, JeBailey L, Faca VM, Thorpe JD, Silva MA, Ireton RC, Horton MB, Wang H, Pruitt LC, Zhang Q, Cheng KH, Urbain N, Hanash SM, Dinulescu DM. 2009. Integrated proteomic analysis of human cancer cells and plasma from tumor bearing mice for ovarian cancer biomarker discovery. *PLoS ONE* 4(11):e7916.

Plaxe SC, Deligdisch L, Dottino PR, Cohen CJ. 1990. Ovarian intraepithelial neoplasia demonstrated in patients with stage I ovarian carcinoma. *Gynecologic Oncology* 38(3):367-372

Pliarchopoulou K and Pectasides D. 2010. Epithelial ovarian cancer: focus on targeted therapy. *Critical Reviews in Oncology/Hematology* 1446:1-7.

Pode-Shakked N, Harari-Steinberg O, Haberman-Ziv Y, Rom-Gros E, Bahar S, Omer D, Metsuyanin S, Buzhor E, Jacob-Hirsch J, Goldstein RS, Mark-Danieli M, Dekel B. 2011. Resistance or sensitivity of Wilms' tumor to anti-FZD7

antibody highlights the Wnt pathway as a possible therapeutic target. *Oncogene* 30:1664-1680.

Provencher DM, Lounis H, Champoux L, Tetrault M, Manderson EN, Wang JC, Eydoux P, Savoie R, Tonin PN, Mes-Masson A-M. 2000. Characterization of four novel epithelial ovarian cancer cell lines. *In Vitro Cellular and Developmental Biology - Animal* 36(6):357-361.

Quinn MCJ, Wilson DJ, Young F, Dempsey AA, Arcand SL, Birch AH, Wojnarowicz PM, Provencher D, Mes-Masson A-M, Englert D, Tonin PN. 2009. The chemiluminescence based Zplex automated workstation focus array reproduces ovarian cancer Affymetrix GeneChip expression profiles. *Journal of Translational Medicine* 7:55

Quinn MCJ, Filali-Mouhim A, Provencher DM, Mes-Masson A-M, Tonin PN. 2009. Reprogramming of the Transcriptome in a Novel Chromosome 3 Transfer Tumor Suppressor Ovarian Cancer Cell Line Model Affected Molecular Networks That Are Characteristic of Ovarian Cancer. *Molecular Carcinogenesis* 48:648-661.

Rajeevan MS, Vernon SD, Taysavang N, Unger ER. 2001. Validation of Array-Based Gene Expression Profiles by Real-Time (Kinetic) RT-PCR. *Journal of Molecular Diagnostics* 3(1):26-31.

Richesson RL and Krischer J. 2007. Data standards in clinical research: gaps, overlaps, challenges and future directions. *J Am Med Inform Assoc.* 14:687-696.

Risch H. 1998. Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. *Journal of the National Cancer Institute* 90(23):1774-1786.

Rodriguez-Mora O, LaHair MM, Howe CJ, McCubrey JA, Franklin RA. 2005. Calcium/calmodulin-dependent protein kinases as potential targets in cancer therapy. *Expert Opin. Ther. Targets* 9(4):791-808.

Rodriguez GC, Walmer DK, Cline M, Krigman H, Lessey BA, Whitaker RS, Dodge R, Hughes CL. Effect of Progestin on the Ovarian Epithelium of Macaques: Cancer Prevention Through Apoptosis? *Journal of the Society for Gynecologic Investigation* 5:271-276.

Rodriguez GC, Nagarsheth NP, Lee KL, Bentley RC, Walmer DK, Cline M, Whitaker RS, Isner P, Berchuck A, Dodge RK, Hughes CL. 2002. Progestin-induced apoptosis in the macaque ovarian epithelium: differential regulation of transforming growth factor-beta. *J Natl Can* 94:50-60.

Roh MH, Yassin Y, Miron A, Mehra KK, Mehrad M, Monte NM, Mutter GL, Nucci MR, Ning G, Mckeon FD, Hirsch MS, Wa X, Crum CP. High-grade

fimbrial-ovarian carcinomas are unified by altered p53, PTEN and PAX2 expression. *Modern Pathology* 23:1316–1324.

Rokhlin OW, Taghiyev AF, Bayer KU, Bumcrot D, Koteliansk VE, Glover RA, Cohen MB. Calcium/calmodulin-dependent kinase II plays an important role in prostate cancer cell survival. *Cancer Biology & Therapy* 6(5):732-742.

Roland IH, Yang W-L, Yang D-H, Daly MB, Ozols RF, Hamilton TC, Lynch HT, Godwin AK, Xu XX. 2003. Loss of surface and cyst epithelial basement membranes and preneoplastic morphologic changes in prophylactic oophorectomies. *Cancer* 98(12):2607-2623.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 Microarray Software Suite. *Methods in Enzymology* 411:134-193.

Salehi F, Dunfield L, Phillips KP, Krewski D, Vanderhyden BC. 2008. Risk factors for ovarian cancer: an overview with emphasis on hormonal factors. *Journal of Toxicology and Environmental Health - Part B* 11:301-321.

Santin AD, Zhan F, Bellone S, Palmieri M, Cane S, Bignotti E, Anfossi S, Gokden M, Dunn D, Roman JJ, O'Brian TJ, Tian E, Cannon MJ, Shaughnessy Jr J, Pecorelli S. 2004. Gene expression profiles in primary ovarian serous papillary tumors and normal ovarian epithelium: identification of candidate molecular markers for ovarian cancer diagnosis and therapy. *Int. J. Cancer*: 112:14-25.

Schmid S, Bieber M, Zhang F, Zhang M, He B, Jablons D, Teng NNH. 2011. Wnt and Hedgehog gene pathway expression in serous ovarian cancer. *Int J Gynecol Cancer* 21:975-980.

Scully RE. 1995. Pathology of ovarian cancer precursors. *Journal of Cellular Biochemistry* 23(Supple):208-218.

Segal E, Friedman N, Kaminski N, Regev A, Koller D. 2005. From signatures to models: understanding cancer using microarrays. *Nature Genetics Supplement* 37:S38-S45.

Seidman JD, Horkayne-Szakaly I, Haiba M, Boice CR, Kurman RJ, Ronnett BM. 2004. The histological type and stage distribution of ovarian carcinomas of surface epithelial origin. *International Journal of Gynecological Pathology* 23:41–44."

Seitz S, Korsching E, Weime Jr, Jacobsen A, Arnold N, Meindl A, Arnold W, Gustavus D, Klebig C, Petersen I, Scherneck S. 2006. Genetic background of different cancer cell lines influences the gene set involved in chromosome 8 mediated breast tumor suppression. *Genes, Chromosomes & Cancer* 45:612–627.

Shetty V, Hafner J, Shah P, Nickens Z, Philip R. 2012 . Investigation of ovarian cancer associated sialylation changes in N-linked glycopeptides by quantitative proteomics. *Clinical Proteomics* 9(10):1-19.

Shih I-M and Kurman RJ. A proposed model based on morphological and molecular genetic analysis. *American Journal of Pathology*, 164(5):1511-1518.

Shridhar V, Lee J, Pandita A, Iturria S, Avula R, Staub J, Morrissey M, Calhoun E, Sen A, Kalli K, Keeney G, Roche P, Cliby W, Lu K, Schmandt R, Mills GB, Bast Jr. RC, James CD, Couch FJ, Hartmann JC, Lillie J, Smith DI. 2001. Genetic analysis of early- versus late-stage ovarian tumors. *Cancer Research* 61:5895–5904.

Siegel R, Naishadham D, Jemal A. 2012. Cancer Statistics. 2012. *Ca Cancer J* 62:10–29.

Starksen NF, Harsh GR, Gibbs VC, Williams LT. Regulated expression of the platelet-derived growth factor A chain gene in microvascular endothelial cells. *The Journal Of Biological Chemistry* 262(30):14381-14384.

Stirling D, Evans DGR, Pichert G, Shenton A, Kirk EN, Rimmer S, Steel CM, Lawson S, Busby-Earle RMC, Walker J, Lalloo FI, Eccles DM, Lucassen AM, Porteous ME. 2005. Screening for familial ovarian cancer: failure of current protocols to detect ovarian cancer at an early stage according to the international federation of gynecology and obstetrics system. *J Clin Oncol* 23(24):5588-5596.

Stronach EX, Sellar GC, Blenkiron C, Rabiasz GJ, Taylor KJ, Miller EP, Massie CE, Al-Nafussi A, Smyth JF, Porteous DJ, Gabra H. 2003. Identification of clinically relevant genes on chromosome 11 in a functional model of ovarian cancer tumor suppression. *Cancer Res* 63:8648-8655.

Sueblinvong T and Carney ME. 2009. Current understanding of risk factors in ovarian cancer. *Current Treatment Options in Oncology* 10:67–81.

Suganuma T, Ino K, Shibata K, Kajiyama H, Nagasaka T, Mizutani S, Kikkawa F. 2005. Functional expression of the angiotensin II type 1 receptor in human ovarian carcinoma cells and its blockade therapy resulting in suppression of tumor invasion, angiogenesis, and peritoneal dissemination. *Clin Cancer Res* 11(7):2686-2694.

The Cancer Genome Atlas Research Network. 2011. Integrated genomic analysis of ovarian carcinoma. *Nature* 474:609-615.

Tinker AV, Boussioutas A, Bowtell DDL. 2006. The challenges of gene expression microarrays for the study of human cancer. *Cancer Cell* 9:333-339.

Tonin PN, Hudson TJ, Rodier F, Bossolasco M, Lee PD, Novak J, Manderson EN, Provencher D, Mes-Masson A-M. 2001. Microarray analysis of gene

expression mirrors the biology of an ovarian cancer model. *Oncogene* 20:6617-6626.

Tortolero-Luna G and Mitchel MF. 1995. The epidemiology of ovarian cancer. *Journal of Cellular Biochemistry Supple* 23:200-207.

Ueno K, Hazama S, Mitomori S, Nishioka M, Suehiro Y, Hirata H, Oka M, Imai K, Dahiya R, Hinoda Y. 2009. Down-regulation of frizzled-7 expression decreases survival, invasion and metastatic capabilities of colon cancer cells. *British Journal of Cancer* 101:1374-1381.

Vanti WB, Nguyen T, Cheng R, Lynch KR, George SR, O'Dowd BF. 2003. Novel human G-protein-coupled receptors. *Biochemical and Biophysical Research Communications* 305:67–71.

Varela AS, Saez JJBL, Senra DQ. 1997. Serum ceruloplasmin as a diagnostic marker of cancer. *Cancer Letters* 121:139-145.

Vaughan S, Coward JI, Bast Jr RC, Berchuck A, Berek JS, Brenton JD, Coukos G, Crum CC, Drapkin R, Etemadmoghadam D, Friedlander M, Gabra H, Kaye SB, Lord CJ, Lengyel E, Levine DA, McNeish IA, Menon U, Mills GB, Nephew KP, Oza AM, Sood AK, Stronach EA, Walczak H, Bowtell DD, Balkwill FR. 2011. Rethinking ovarian cancer: recommendations for improving outcomes. *Nature Reviews Cancer* 11:719-725.

Vu T-K H, Hung DT, Wheaton VI, Coughlin SR. 1991. Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* 64:1057-1069.

Waldemarson S, Krogh M, Alaiya A, Kirik U, Schedvins K, Auer G, Hansson KM, Ossola R, Aebbersold R, Hookeun Lee H. Protein expression changes in ovarian cancer during the transition from benign to malignant. *J. Proteome Res.* 11:2876-2889.

Wang S, Yanga Q, Fung K-M, Lina H-K. 2008. AKR1C2 and AKR1C3 mediated prostaglandin D2 metabolism augments the PI3K/Akt proliferative signaling pathway in human prostate cancer cells. *Molecular and Cellular Endocrinology* 289:60–66.

Wang F-Q, Fisher J, Fishman DA. 2011. MMP-1-PAR1 axis mediates LPA-induced epithelial ovarian cancer (EOC) invasion. *Gynecologic Oncology* 120:247–255.

Watson P. 2002. The importance of tumor banking: bridging no-mans-land in cancer research. *Expert Rev. Anticancer Ther.* 2(1):1–3.

Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockart DJ, Burger RA, Hampton GM. 2001. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *PNAS* 98(3):1176-1181.

Wojnarowicz PM, Bresnahan A, Arcand SL, Filali-Mouhim A, Provencher DM, Mes-Masson A-M, Tonin PN. 2008. Construction of a chromosome 17 transcriptome in serous ovarian cancer identifies differentially expressed genes. *Int J Gynecol Cancer* 18:963–975.

Wysolmerski JJ and Broadus AE. 1994. Hypercalcemia of malignancy: the central role of parathyroid hormone-related protein. *Annu. Rev. Med.* 45:189-200.

Xie Y, Wolff DW, Wei T, Wang B, Deng C, Kirui JK, Jiang H, Qin J, Abel PW, Tu Y. 2009. Breast cancer migration and invasion depend on proteasome degradation of regulator of G-protein signaling 4. *Cancer Res* 69:5743-5751.

Xu Y, Gaudette DC, Boynton JD, Frankel A, Fang X-J, Sharma A, Hurteau J, Casey G, Goodbody A, Mellors A, Holub BJ, Mills GB. 1995. Characterization of an ovarian cancer activating factor in ascites from ovarian cancer patients. *Clinical Cancer Research* 1:1223-1232.

Yancik R. 1993. Age contrasts in incidence, histology, disease stage at diagnosis, and mortality. *Cancer* 71:517-523.

Yang D-H, Smith ER, Cohen C, Wu H, Patriotis C, Godwin AK, Hamilton TC, Xu X-X. 2002. Molecular events associated with dysplastic morphologic transformation and initiation of ovarian tumorigenicity. *Cancer* 94(9):2380-2392.

Yang L, Wu X, Wang Y, Zhang K, Wu J, Yuan Y-C, Deng X, Chen L, Kim CCH, Lau S, Somlo G, Yen Y. 2011. FZD7 has a critical role in cell proliferation in triple negative breast cancer. *Oncogene* 30:4437-4446.

Yap TA, Carden CP, Kaye SB. 2009. Beyond chemotherapy: targeted therapies in ovarian cancer. *Nature Reviews Cancer* 9:167-181.

Yasui T, Uemura H, Irahara M, Aono T. 1997. Effects of transforming growth-factor-beta on the production of parathyroid hormone-related peptide in a human ovarian cancer cell line in vitro. *J. Obstet. Gynaecol. Res.* 23(3):231-238.

Zhang J, Li N, Yu J, Zhang W, Cao X. 2001. Molecular cloning and characterization of a novel calcium/calmodulin-dependent protein kinase II inhibitor from human dendritic cells. *Biochemical and Biophysical Research Communications* 285:229–234.

Zhang H, Xu X, Gajewiak J, Tsukahara R, Fujiwara Y, Liu J, Fells JI, Perygin D, Parrill AL, Tigyi G, Prestwich GD. 2009. Dual activity lysophosphatidic acid

receptor pan-antagonist/autotaxin inhibitor reduces breast cancer cell migration in vitro and causes tumor regression in vivo. *Cancer Res* 69:5441-5449 .

Zorn KK, Jazaeri AA, Awtrey CS, Gardner GJ, Mok SC, Boyd J, Birrer MJ. 2003. Choice of normal ovarian control influences determination of differentially expressed genes in ovarian cancer expression profiling studies. *Clinical Cancer Research* 9:4811–4818. October 15, 2003

Zorn KK, Bonome T, Gangi L, Chandramouli GVR, Awtrey CS, Gardner GJ, Barrett JC, Boyd J, Birrer MJ. 2005. Gene expression profiles of serous, endometrioid, and clear cell subtypes of ovarian and endometrial cancer. *Clin Cancer Res* 11:6422-6430.

APPENDICES

Appendix I – David functional annotation clustering.

Functional Annotation Clustering

[Help and Manual](#)

Current Gene List: List_1

Current Background: Human Genome U133 Plus 2 Array

91 DAVID IDs

Options Classification Stringency Medium

Rerun using options

Create Sublist

48 Cluster(s)

[Download File](#)

Annotation Cluster 1		Enrichment Score: 3.55	G		Count	P_Value	Benjamini
☐	SP_PIR_KEYWORDS	signal	RT		33	1.4E-5	3.1E-3
☐	UP_SEQ_FEATURE	signal peptide	RT		33	1.5E-5	6.2E-3
☐	UP_SEQ_FEATURE	disulfide bond	RT		28	2.6E-5	5.5E-3
☐	SP_PIR_KEYWORDS	disulfide bond	RT		28	5.2E-5	5.8E-3
☐	UP_SEQ_FEATURE	glycosylation site: N-linked (GlcNAc...)	RT		34	2.3E-4	3.2E-2
☐	SP_PIR_KEYWORDS	glycoprotein	RT		35	2.5E-4	1.8E-2
☐	SP_PIR_KEYWORDS	Secreted	RT		18	1.7E-3	9.2E-2
☐	GOTERM_CC_FAT	extracellular space	RT		11	3.8E-3	4.6E-1
☐	GOTERM_CC_FAT	extracellular region part	RT		13	5.6E-3	3.6E-1
☐	GOTERM_CC_FAT	extracellular region	RT		20	5.8E-3	2.7E-1
Annotation Cluster 2		Enrichment Score: 2.19	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	embryonic appendage morphogenesis	RT		5	1.5E-3	3.5E-1
☐	GOTERM_BP_FAT	embryonic limb morphogenesis	RT		5	1.5E-3	3.5E-1
☐	GOTERM_BP_FAT	appendage morphogenesis	RT		5	2.2E-3	3.5E-1
☐	GOTERM_BP_FAT	limb morphogenesis	RT		5	2.2E-3	3.5E-1
☐	GOTERM_BP_FAT	chordate embryonic development	RT		8	2.5E-3	3.4E-1
☐	GOTERM_BP_FAT	limb development	RT		5	2.6E-3	3.1E-1
☐	GOTERM_BP_FAT	appendage development	RT		5	2.6E-3	3.1E-1
☐	GOTERM_BP_FAT	embryonic development ending in birth or egg hatching	RT		8	2.7E-3	2.9E-1
☐	GOTERM_BP_FAT	embryonic forelimb morphogenesis	RT		3	5.1E-3	3.7E-1
☐	GOTERM_BP_FAT	forelimb morphogenesis	RT		3	6.9E-3	4.1E-1
☐	GOTERM_BP_FAT	embryonic morphogenesis	RT		6	2.9E-2	5.1E-1
☐	GOTERM_BP_FAT	embryonic organ development	RT		3	2.5E-1	9.0E-1
☐	GOTERM_BP_FAT	skeletal system development	RT		4	2.7E-1	9.1E-1
Annotation Cluster 3		Enrichment Score: 1.79	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	mammary gland development	RT		4	5.0E-3	3.8E-1
☐	GOTERM_BP_FAT	gland development	RT		5	6.9E-3	3.9E-1
☐	GOTERM_BP_FAT	negative regulation of cell differentiation	RT		4	1.2E-1	7.5E-1
Annotation Cluster 4		Enrichment Score: 1.73	G		Count	P_Value	Benjamini
☐	SP_PIR_KEYWORDS	oxidoreductase	RT		9	6.1E-3	2.0E-1
☐	GOTERM_BP_FAT	oxidation reduction	RT		10	8.3E-3	3.5E-1

☐	GOTERM_MF_FAT	electron carrier activity	RT		4	1.3E-1	9.3E-1
	Annotation Cluster 5	Enrichment Score: 1.63	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	protein processing	RT		5	3.2E-3	3.1E-1
☐	GOTERM_BP_FAT	protein maturation	RT		5	4.4E-3	3.7E-1
☐	GOTERM_BP_FAT	protein maturation by peptide bond cleavage	RT		4	1.1E-2	4.0E-1
☐	SP_PIR_KEYWORDS	cleavage on pair of basic residues	RT		4	1.4E-1	8.3E-1
☐	GOTERM_BP_FAT	proteolysis	RT		9	2.2E-1	8.7E-1
	Annotation Cluster 6	Enrichment Score: 1.63	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	response to extracellular stimulus	RT		6	8.0E-3	3.6E-1
☐	GOTERM_BP_FAT	response to nutrient levels	RT		5	2.5E-2	5.1E-1
☐	GOTERM_BP_FAT	response to drug	RT		5	3.4E-2	5.3E-1
☐	GOTERM_BP_FAT	response to nutrient	RT		4	4.5E-2	5.8E-1
	Annotation Cluster 7	Enrichment Score: 1.41	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	inflammatory response	RT		7	9.1E-3	3.7E-1
☐	GOTERM_BP_FAT	response to wounding	RT		8	2.8E-2	5.0E-1
☐	GOTERM_BP_FAT	defense response	RT		8	4.9E-2	5.8E-1
☐	GOTERM_BP_FAT	response to bacterium	RT		4	8.1E-2	6.6E-1
☐	KEGG_PATHWAY	Complement and coagulation cascades	RT		3	9.4E-2	9.6E-1
	Annotation Cluster 8	Enrichment Score: 1.26	G		Count	P_Value	Benjamini
☐	SP_PIR_KEYWORDS	growth factor	RT		4	2.6E-2	4.1E-1
☐	GOTERM_MF_FAT	growth factor activity	RT		4	6.4E-2	9.0E-1
☐	KEGG_PATHWAY	Pathways in cancer	RT		6	9.9E-2	8.9E-1
	Annotation Cluster 9	Enrichment Score: 1.23	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	muscle organ development	RT		6	6.7E-3	4.2E-1
☐	GOTERM_BP_FAT	striated muscle tissue development	RT		4	3.0E-2	5.0E-1
☐	GOTERM_BP_FAT	muscle tissue development	RT		4	3.4E-2	5.3E-1
☐	GOTERM_BP_FAT	cellular homeostasis	RT		7	4.6E-2	5.8E-1
☐	GOTERM_BP_FAT	skeletal muscle organ development	RT		3	5.4E-2	5.8E-1
☐	GOTERM_BP_FAT	skeletal muscle tissue development	RT		3	5.4E-2	5.8E-1
☐	GOTERM_CC_FAT	membrane raft	RT		3	1.9E-1	8.5E-1
☐	GOTERM_BP_FAT	protein localization	RT		4	8.8E-1	1.0E0
	Annotation Cluster 10	Enrichment Score: 1.18	G		Count	P_Value	Benjamini
☐	GOTERM_MF_FAT	enzyme activator activity	RT		7	1.2E-2	9.7E-1
☐	GOTERM_MF_FAT	GTPase activator activity	RT		5	3.7E-2	9.7E-1
☐	GOTERM_MF_FAT	GTPase regulator activity	RT		5	2.0E-1	9.8E-1
☐	GOTERM_MF_FAT	nucleoside-triphosphatase regulator activity	RT		5	2.1E-1	9.7E-1
	Annotation Cluster 11	Enrichment Score: 1.16	G		Count	P_Value	Benjamini
☐	SP_PIR_KEYWORDS	steroid metabolism	RT		3	2.4E-2	4.9E-1
☐	KEGG_PATHWAY	Steroid hormone biosynthesis	RT		3	4.4E-2	9.4E-1
☐	GOTERM_BP_FAT	steroid metabolic process	RT		3	3.1E-1	9.3E-1
	Annotation Cluster 12	Enrichment Score: 1.12	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	positive regulation of catalytic activity	RT		8	2.6E-2	5.0E-1
☐	GOTERM_BP_FAT	regulation of hydrolase activity	RT		6	4.0E-2	5.8E-1

☐	GOTERM_BP_FAT	positive regulation of molecular function	RT		8	4.5E-2	5.8E-1
☐	GOTERM_BP_FAT	positive regulation of phospholipase C activity	RT		3	5.1E-2	5.8E-1
☐	GOTERM_BP_FAT	activation of phospholipase C activity	RT		3	5.1E-2	5.8E-1
☐	GOTERM_BP_FAT	positive regulation of phospholipase activity	RT		3	5.7E-2	5.9E-1
☐	GOTERM_BP_FAT	regulation of phospholipase activity	RT		3	6.0E-2	5.9E-1
☐	GOTERM_BP_FAT	positive regulation of lipase activity	RT		3	6.6E-2	6.1E-1
☐	GOTERM_BP_FAT	regulation of lipase activity	RT		3	8.6E-2	6.7E-1
☐	GOTERM_BP_FAT	positive regulation of hydrolase activity	RT		3	2.7E-1	9.1E-1
☐	GOTERM_BP_FAT	intracellular signaling cascade	RT		7	7.2E-1	1.0E0
Annotation Cluster 13		Enrichment Score: 1.11	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	regulation of blood coagulation	RT		4	1.1E-3	4.7E-1
☐	GOTERM_BP_FAT	regulation of cell proliferation	RT		13	1.4E-3	4.2E-1
☐	GOTERM_BP_FAT	regulation of coagulation	RT		4	1.6E-3	3.1E-1
☐	GOTERM_BP_FAT	regulation of response to external stimulus	RT		5	1.3E-2	4.1E-1
☐	GOTERM_BP_FAT	positive regulation of catalytic activity	RT		8	2.6E-2	5.0E-1
☐	GOTERM_BP_FAT	regulation of hydrolase activity	RT		6	4.0E-2	5.8E-1
☐	GOTERM_BP_FAT	regulation of MAP kinase activity	RT		4	4.2E-2	5.8E-1
☐	GOTERM_BP_FAT	regulation of protein kinase activity	RT		6	4.4E-2	5.8E-1
☐	GOTERM_BP_FAT	positive regulation of molecular function	RT		8	4.5E-2	5.8E-1
☐	GOTERM_BP_FAT	regulation of kinase activity	RT		6	4.9E-2	5.9E-1
☐	GOTERM_BP_FAT	regulation of peptidyl-tyrosine phosphorylation	RT		3	5.6E-2	5.8E-1
☐	GOTERM_BP_FAT	regulation of transferase activity	RT		6	5.6E-2	5.8E-1
☐	GOTERM_BP_FAT	regulation of protein amino acid phosphorylation	RT		4	7.2E-2	6.3E-1
☐	GOTERM_BP_FAT	regulation of protein modification process	RT		5	8.2E-2	6.6E-1
☐	GOTERM_BP_FAT	regulation of phosphorylation	RT		6	1.2E-1	7.4E-1
☐	GOTERM_BP_FAT	regulation of cellular protein metabolic process	RT		6	1.2E-1	7.5E-1
☐	GOTERM_BP_FAT	positive regulation of protein kinase activity	RT		4	1.3E-1	7.6E-1
☐	GOTERM_BP_FAT	regulation of phosphorus metabolic process	RT		6	1.3E-1	7.6E-1
☐	GOTERM_BP_FAT	regulation of phosphate metabolic process	RT		6	1.3E-1	7.6E-1
☐	GOTERM_BP_FAT	positive regulation of kinase activity	RT		4	1.4E-1	7.6E-1
☐	GOTERM_BP_FAT	positive regulation of transferase activity	RT		4	1.5E-1	7.7E-1
☐	GOTERM_BP_FAT	MAPKKK cascade	RT		3	2.7E-1	9.1E-1
☐	GOTERM_BP_FAT	positive regulation of cellular protein metabolic process	RT		3	3.7E-1	9.5E-1
☐	GOTERM_BP_FAT	second-messenger-mediated signaling	RT		3	3.8E-1	9.6E-1

☐	GOTERM_BP_FAT	positive regulation of protein metabolic process	RT		3	3.9E-1	9.6E-1
☐	GOTERM_MF_FAT	identical protein binding	RT		5	5.1E-1	1.0E0
☐	GOTERM_BP_FAT	protein kinase cascade	RT		3	6.2E-1	1.0E0
☐	GOTERM_BP_FAT	intracellular signaling cascade	RT		7	7.2E-1	1.0E0
☐	GOTERM_BP_FAT	macromolecular complex subunit organization	RT		3	9.1E-1	1.0E0
Annotation Cluster 14		Enrichment Score: 1.11	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	tube development	RT		5	3.5E-2	5.3E-1
☐	GOTERM_BP_FAT	development of primary male sexual characteristics	RT		3	5.1E-2	5.8E-1
☐	GOTERM_BP_FAT	reproductive developmental process	RT		5	5.9E-2	5.9E-1
☐	GOTERM_BP_FAT	male sex differentiation	RT		3	6.3E-2	6.0E-1
☐	GOTERM_BP_FAT	development of primary sexual characteristics	RT		3	1.6E-1	7.8E-1
☐	GOTERM_BP_FAT	sex differentiation	RT		3	2.0E-1	8.6E-1
Annotation Cluster 15		Enrichment Score: 1.08	G		Count	P_Value	Benjamini
☐	UP_SEQ_FEATURE	lipid moiety-binding region:GPI-anchor amidated serine	RT		3	2.9E-2	8.8E-1
☐	SP_PIR_KEYWORDS	lipoprotein	RT		7	9.2E-2	7.4E-1
☐	UP_SEQ_FEATURE	propeptide:Removed in mature form	RT		4	9.9E-2	9.7E-1
☐	GOTERM_CC_FAT	anchored to membrane	RT		4	1.2E-1	8.7E-1
☐	SP_PIR_KEYWORDS	gpi-anchor	RT		3	1.2E-1	8.1E-1
Annotation Cluster 16		Enrichment Score: 1.07	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	protein maturation by peptide bond cleavage	RT		4	1.1E-2	4.0E-1
☐	UP_SEQ_FEATURE	active site:Charge relay system	RT		5	1.3E-2	6.7E-1
☐	SP_PIR_KEYWORDS	Serine protease	RT		4	2.4E-2	4.2E-1
☐	SP_PIR_KEYWORDS	Protease	RT		7	2.7E-2	4.0E-1
☐	GOTERM_MF_FAT	serine-type endopeptidase activity	RT		4	5.0E-2	9.7E-1
☐	GOTERM_MF_FAT	serine-type peptidase activity	RT		4	7.2E-2	8.8E-1
☐	GOTERM_MF_FAT	serine hydrolase activity	RT		4	7.4E-2	8.6E-1
☐	UP_SEQ_FEATURE	domain:Peptidase S1	RT		3	8.2E-2	9.7E-1
☐	GOTERM_MF_FAT	peptidase activity, acting on L-amino acid peptides	RT		7	8.5E-2	8.8E-1
☐	INTERPRO	Peptidase S1/S6, chymotrypsin/Hap, active site	RT		3	9.4E-2	1.0E0
☐	GOTERM_MF_FAT	peptidase activity	RT		7	1.0E-1	9.0E-1
☐	INTERPRO	Peptidase S1 and S6, chymotrypsin/Hap	RT		3	1.1E-1	1.0E0
☐	SMART	Tryp_SPc	RT		3	1.2E-1	1.0E0
☐	SP_PIR_KEYWORDS	metalloprotease	RT		3	1.7E-1	8.7E-1
☐	SP_PIR_KEYWORDS	hydrolase	RT		11	2.1E-1	8.9E-1
☐	GOTERM_BP_FAT	proteolysis	RT		9	2.2E-1	8.7E-1
☐	SP_PIR_KEYWORDS	zymogen	RT		3	2.7E-1	9.1E-1
☐	GOTERM_MF_FAT	metallopeptidase activity	RT		3	2.9E-1	9.9E-1
☐	GOTERM_MF_FAT	endopeptidase activity	RT		4	3.5E-1	9.9E-1
Annotation Cluster 17		Enrichment Score: 1.07	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	regulation of system process	RT		7	7.7E-3	4.1E-1
☐	GOTERM_BP_FAT	regulation of muscle contraction	RT		4	8.0E-3	3.7E-1

☐	GOTERM_BP_FAT	regulation of response to external stimulus	RT		5	1.3E-2	4.1E-1
☐	GOTERM_BP_FAT	regulation of vasoconstriction	RT		3	1.6E-2	4.5E-1
☐	GOTERM_BP_FAT	regulation of smooth muscle contraction	RT		3	2.0E-2	4.8E-1
☐	GOTERM_BP_FAT	cation homeostasis	RT		6	2.2E-2	5.0E-1
☐	GOTERM_BP_FAT	homeostatic process	RT		10	2.4E-2	5.2E-1
☐	GOTERM_BP_FAT	chemical homeostasis	RT		8	2.4E-2	5.1E-1
☐	GOTERM_BP_FAT	regulation of lipid metabolic process	RT		4	2.5E-2	5.0E-1
☐	GOTERM_BP_FAT	positive regulation of catalytic activity	RT		8	2.6E-2	5.0E-1
☐	GOTERM_BP_FAT	ion homeostasis	RT		7	2.7E-2	5.1E-1
☐	GOTERM_BP_FAT	regulation of hydrolase activity	RT		6	4.0E-2	5.8E-1
☐	GOTERM_BP_FAT	calcium ion transport	RT		4	4.4E-2	5.8E-1
☐	GOTERM_BP_FAT	positive regulation of molecular function	RT		8	4.5E-2	5.8E-1
☐	GOTERM_BP_FAT	cellular homeostasis	RT		7	4.6E-2	5.8E-1
☐	GOTERM_BP_FAT	cellular cation homeostasis	RT		5	5.3E-2	5.8E-1
☐	GOTERM_BP_FAT	cellular ion homeostasis	RT		6	5.8E-2	5.9E-1
☐	GOTERM_BP_FAT	cellular chemical homeostasis	RT		6	6.2E-2	5.9E-1
☐	GOTERM_BP_FAT	di-, tri-valent inorganic cation transport	RT		4	7.3E-2	6.3E-1
☐	GOTERM_BP_FAT	regulation of peptidase activity	RT		3	8.4E-2	6.7E-1
☐	GOTERM_BP_FAT	cellular metal ion homeostasis	RT		4	9.7E-2	7.0E-1
☐	GOTERM_BP_FAT	metal ion homeostasis	RT		4	1.1E-1	7.3E-1
☐	GOTERM_BP_FAT	positive regulation of cell communication	RT		5	1.1E-1	7.3E-1
☐	GOTERM_BP_FAT	cellular di-, tri-valent inorganic cation homeostasis	RT		4	1.3E-1	7.6E-1
☐	GOTERM_BP_FAT	regulation of homeostatic process	RT		3	1.4E-1	7.6E-1
☐	GOTERM_BP_FAT	cytosolic calcium ion homeostasis	RT		3	1.4E-1	7.6E-1
☐	GOTERM_BP_FAT	di-, tri-valent inorganic cation homeostasis	RT		4	1.5E-1	7.7E-1
☐	GOTERM_BP_FAT	positive regulation of multicellular organismal process	RT		4	1.6E-1	7.8E-1
☐	GOTERM_BP_FAT	metal ion transport	RT		5	2.5E-1	9.0E-1
☐	GOTERM_BP_FAT	cellular calcium ion homeostasis	RT		3	2.7E-1	9.1E-1
☐	GOTERM_BP_FAT	circulatory system process	RT		3	2.8E-1	9.2E-1
☐	GOTERM_BP_FAT	blood circulation	RT		3	2.8E-1	9.2E-1
☐	GOTERM_BP_FAT	calcium ion homeostasis	RT		3	2.8E-1	9.2E-1
☐	GOTERM_BP_FAT	positive regulation of transport	RT		3	3.6E-1	9.5E-1
☐	GOTERM_BP_FAT	cation transport	RT		5	3.6E-1	9.5E-1
☐	KEGG_PATHWAY	Calcium signaling pathway	RT		3	3.9E-1	9.9E-1
☐	GOTERM_BP_FAT	ion transport	RT		6	4.1E-1	9.7E-1
☐	GOTERM_CC_FAT	intrinsic to plasma membrane	RT		8	5.3E-1	9.6E-1
☐	GOTERM_BP_FAT	intracellular signaling cascade	RT		7	7.2E-1	1.0E0
☐	GOTERM_CC_FAT	integral to plasma membrane	RT		6	8.1E-1	1.0E0
Annotation Cluster 18		Enrichment Score: 1.07	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	tube development	RT		5	3.5E-2	5.3E-1
☐	GOTERM_BP_FAT	lung development	RT		3	1.1E-1	7.3E-1

☐	GOTERM_BP_FAT	respiratory tube development	RT		3	1.1E-1	7.3E-1
☐	GOTERM_BP_FAT	respiratory system development	RT		3	1.2E-1	7.5E-1
Annotation Cluster 19		Enrichment Score: 1.06	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	blood vessel development	RT		5	5.2E-2	5.8E-1
☐	GOTERM_BP_FAT	vasculature development	RT		5	5.5E-2	5.9E-1
☐	GOTERM_BP_FAT	reproductive developmental process	RT		5	5.9E-2	5.9E-1
☐	GOTERM_BP_FAT	blood vessel morphogenesis	RT		3	3.4E-1	9.5E-1
Annotation Cluster 20		Enrichment Score: 1.01	G		Count	P_Value	Benjamini
☐	GOTERM_CC_FAT	plasma membrane	RT		30	5.9E-3	2.1E-1
☐	SP_PIR_KEYWORDS	cell membrane	RT		16	3.3E-2	4.4E-1
☐	SP_PIR_KEYWORDS	membrane	RT		36	8.2E-2	7.2E-1
☐	UP_SEQ_FEATURE	topological domain: Cytoplasmic	RT		21	8.5E-2	9.7E-1
☐	UP_SEQ_FEATURE	topological domain: Extracellular	RT		16	1.6E-1	9.9E-1
☐	UP_SEQ_FEATURE	transmembrane region	RT		27	1.8E-1	1.0E0
☐	SP_PIR_KEYWORDS	transmembrane	RT		27	1.9E-1	8.8E-1
☐	GOTERM_CC_FAT	intrinsic to membrane	RT		32	2.3E-1	8.5E-1
☐	GOTERM_CC_FAT	integral to membrane	RT		28	5.0E-1	9.6E-1
Annotation Cluster 21		Enrichment Score: 0.93	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	response to hormone stimulus	RT		7	1.8E-2	4.6E-1
☐	GOTERM_BP_FAT	response to steroid hormone stimulus	RT		5	2.3E-2	5.2E-1
☐	GOTERM_BP_FAT	response to endogenous stimulus	RT		7	2.7E-2	5.1E-1
☐	GOTERM_BP_FAT	response to drug	RT		5	3.4E-2	5.3E-1
☐	GOTERM_BP_FAT	response to organic substance	RT		9	5.0E-2	5.8E-1
☐	GOTERM_BP_FAT	response to glucocorticoid stimulus	RT		3	7.3E-2	6.3E-1
☐	GOTERM_BP_FAT	response to corticosteroid stimulus	RT		3	8.4E-2	6.7E-1
☐	GOTERM_BP_FAT	response to estrogen stimulus	RT		3	1.2E-1	7.4E-1
☐	GOTERM_BP_FAT	learning or memory	RT		3	1.3E-1	7.6E-1
☐	GOTERM_CC_FAT	membrane fraction	RT		8	1.6E-1	8.5E-1
☐	GOTERM_CC_FAT	insoluble fraction	RT		8	1.9E-1	8.5E-1
☐	GOTERM_CC_FAT	cell fraction	RT		9	2.6E-1	8.7E-1
☐	GOTERM_BP_FAT	behavior	RT		5	2.6E-1	9.1E-1
☐	GOTERM_BP_FAT	response to radiation	RT		3	3.1E-1	9.3E-1
☐	GOTERM_BP_FAT	response to abiotic stimulus	RT		4	3.5E-1	9.5E-1
☐	GOTERM_BP_FAT	neurological system process	RT		6	5.6E-1	9.9E-1
☐	GOTERM_BP_FAT	cognition	RT		4	6.5E-1	1.0E0
Annotation Cluster 22		Enrichment Score: 0.91	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	cytoskeleton organization	RT		6	9.4E-2	7.0E-1
☐	GOTERM_BP_FAT	actin cytoskeleton organization	RT		4	1.3E-1	7.6E-1
☐	GOTERM_BP_FAT	actin filament-based process	RT		4	1.5E-1	7.7E-1
Annotation Cluster 23		Enrichment Score: 0.91	G		Count	P_Value	Benjamini
☐	GOTERM_MF_FAT	monocarboxylic acid binding	RT		3	3.7E-2	1.0E0
☐	GOTERM_MF_FAT	carboxylic acid binding	RT		3	2.0E-1	9.7E-1
☐	GOTERM_MF_FAT	lipid binding	RT		5	2.5E-1	9.8E-1
Annotation Cluster 24		Enrichment Score: 0.84	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	negative regulation of cell communication	RT		6	1.2E-2	4.2E-1

☐	GOTERM_BP_FAT	negative regulation of signal transduction	RT		5	3.4E-2	5.3E-1
☐	GOTERM_BP_FAT	positive regulation of cell communication	RT		5	1.1E-1	7.3E-1
☐	GOTERM_BP_FAT	regulation of MAPKKK cascade	RT		3	1.3E-1	7.5E-1
☐	GOTERM_BP_FAT	positive regulation of multicellular organismal process	RT		4	1.6E-1	7.8E-1
☐	GOTERM_BP_FAT	positive regulation of signal transduction	RT		4	2.3E-1	8.8E-1
☐	GOTERM_BP_FAT	negative regulation of cell proliferation	RT		4	3.4E-1	9.4E-1
☐	GOTERM_BP_FAT	regulation of protein kinase cascade	RT		3	4.1E-1	9.7E-1
☐	GOTERM_BP_FAT	protein localization	RT		4	8.8E-1	1.0E0
Annotation Cluster 25		Enrichment Score: 0.84	G		Count	P_Value	Benjamini
☐	INTERPRO	Intermediate filament protein	RT		3	5.3E-2	1.0E0
☐	GOTERM_MF_FAT	structural molecule activity	RT		8	6.0E-2	9.2E-1
☐	SP_PIR_KEYWORDS	coiled coil	RT		7	9.4E-1	1.0E0
Annotation Cluster 26		Enrichment Score: 0.81	G		Count	P_Value	Benjamini
☐	GOTERM_MF_FAT	enzyme activator activity	RT		7	1.2E-2	9.7E-1
☐	UP_SEQ_FEATURE	lipid moiety-binding region:S-palmitoyl cysteine	RT		3	2.3E-1	1.0E0
☐	SP_PIR_KEYWORDS	palmitate	RT		3	2.8E-1	9.2E-1
☐	GOTERM_BP_FAT	intracellular signaling cascade	RT		7	7.2E-1	1.0E0
Annotation Cluster 27		Enrichment Score: 0.81	G		Count	P_Value	Benjamini
☐	GOTERM_CC_FAT	sarcomere	RT		3	1.1E-1	8.7E-1
☐	GOTERM_CC_FAT	myofibril	RT		3	1.3E-1	8.7E-1
☐	GOTERM_CC_FAT	contractile fiber part	RT		3	1.3E-1	8.5E-1
☐	GOTERM_CC_FAT	contractile fiber	RT		3	1.5E-1	8.4E-1
☐	GOTERM_MF_FAT	cytoskeletal protein binding	RT		5	3.3E-1	9.9E-1
Annotation Cluster 28		Enrichment Score: 0.8	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	tube development	RT		5	3.5E-2	5.3E-1
☐	GOTERM_BP_FAT	skeletal system development	RT		4	2.7E-1	9.1E-1
☐	GOTERM_BP_FAT	cell-cell signaling	RT		5	4.3E-1	9.7E-1
Annotation Cluster 29		Enrichment Score: 0.78	G		Count	P_Value	Benjamini
☐	GOTERM_CC_FAT	extracellular matrix part	RT		4	2.7E-2	5.9E-1
☐	GOTERM_CC_FAT	extracellular matrix	RT		4	3.0E-1	9.0E-1
☐	GOTERM_CC_FAT	proteinaceous extracellular matrix	RT		3	5.4E-1	9.6E-1
Annotation Cluster 30		Enrichment Score: 0.73	G		Count	P_Value	Benjamini
☐	UP_SEQ_FEATURE	short sequence motif:Cell attachment site	RT		3	7.2E-2	9.7E-1
☐	SP_PIR_KEYWORDS	cell adhesion	RT		5	1.4E-1	8.2E-1
☐	GOTERM_BP_FAT	cell adhesion	RT		6	3.5E-1	9.5E-1
☐	GOTERM_BP_FAT	biological adhesion	RT		6	3.5E-1	9.5E-1
Annotation Cluster 31		Enrichment Score: 0.71	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	regulation of locomotion	RT		4	9.6E-2	7.0E-1
☐	GOTERM_BP_FAT	regulation of cell migration	RT		3	2.5E-1	9.0E-1
☐	GOTERM_BP_FAT	regulation of cell motion	RT		3	3.0E-1	9.3E-1
Annotation Cluster 32		Enrichment Score: 0.7	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	defense response	RT		8	4.9E-2	5.8E-1
☐	GOTERM_BP_FAT	innate immune response	RT		3	1.8E-1	8.3E-1

☐	GOTERM_BP_FAT	immune response	RT		3	8.8E-1	1.0E0
	Annotation Cluster 33	Enrichment Score: 0.7	G		Count	P_Value	Benjamini
☐	UP_SEQ_FEATURE	region of interest:Head	RT		3	4.7E-2	9.5E-1
☐	UP_SEQ_FEATURE	region of interest:Tail	RT		3	4.9E-2	9.3E-1
☐	GOTERM_CC_FAT	cytoskeletal part	RT		5	7.4E-1	9.9E-1
☐	SP_PIR_KEYWORDS	coiled coil	RT		7	9.4E-1	1.0E0
	Annotation Cluster 34	Enrichment Score: 0.69	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		10	7.8E-3	3.9E-1
☐	GOTERM_BP_FAT	positive regulation of macromolecule metabolic process	RT		12	7.8E-3	3.8E-1
☐	GOTERM_BP_FAT	positive regulation of nitrogen compound metabolic process	RT		10	9.5E-3	3.7E-1
☐	GOTERM_BP_FAT	positive regulation of cellular biosynthetic process	RT		10	1.4E-2	4.2E-1
☐	GOTERM_BP_FAT	positive regulation of biosynthetic process	RT		10	1.5E-2	4.4E-1
☐	GOTERM_BP_FAT	in utero embryonic development	RT		5	1.7E-2	4.6E-1
☐	GOTERM_BP_FAT	positive regulation of macromolecule biosynthetic process	RT		9	2.9E-2	5.1E-1
☐	GOTERM_BP_FAT	epithelial cell differentiation	RT		4	2.9E-2	5.0E-1
☐	GOTERM_BP_FAT	epithelium development	RT		5	3.0E-2	4.9E-1
☐	GOTERM_BP_FAT	positive regulation of transcription, DNA-dependent	RT		7	5.1E-2	5.9E-1
☐	GOTERM_BP_FAT	positive regulation of RNA metabolic process	RT		7	5.3E-2	5.9E-1
☐	GOTERM_BP_FAT	positive regulation of transcription	RT		7	9.4E-2	7.0E-1
☐	GOTERM_BP_FAT	positive regulation of gene expression	RT		7	1.1E-1	7.3E-1
☐	GOTERM_MF_FAT	transcription repressor activity	RT		5	1.1E-1	9.1E-1
☐	GOTERM_BP_FAT	negative regulation of cell differentiation	RT		4	1.2E-1	7.5E-1
☐	GOTERM_MF_FAT	transcription factor activity	RT		9	1.7E-1	9.6E-1
☐	GOTERM_BP_FAT	negative regulation of cellular biosynthetic process	RT		6	2.1E-1	8.6E-1
☐	GOTERM_BP_FAT	negative regulation of biosynthetic process	RT		6	2.2E-1	8.7E-1
☐	GOTERM_MF_FAT	sequence-specific DNA binding	RT		6	2.4E-1	9.8E-1
☐	GOTERM_BP_FAT	heart development	RT		3	3.4E-1	9.5E-1
☐	GOTERM_BP_FAT	positive regulation of transcription from RNA polymerase II promoter	RT		4	3.4E-1	9.5E-1
☐	SP_PIR_KEYWORDS	repressor	RT		4	3.6E-1	9.3E-1
☐	GOTERM_BP_FAT	regulation of transcription from RNA polymerase II promoter	RT		6	3.9E-1	9.6E-1
☐	GOTERM_BP_FAT	negative regulation of macromolecule metabolic process	RT		6	3.9E-1	9.6E-1
☐	GOTERM_MF_FAT	RNA polymerase II transcription factor activity	RT		3	4.1E-1	1.0E0
☐	GOTERM_MF_FAT	transcription regulator activity	RT		10	4.8E-1	1.0E0
☐	GOTERM_BP_FAT	regulation of transcription, DNA-	RT		11	4.8E-1	1.0E0

☐		dependent	RT		11	4.9E-1	9.0E-1
☐	GOTERM_BP_FAT	regulation of RNA metabolic process	RT		11	5.2E-1	9.9E-1
☐	GOTERM_BP_FAT	negative regulation of nitrogen					
☐	GOTERM_MF_FAT	compound metabolic process	RT		4	5.6E-1	9.9E-1
☐	GOTERM_BP_FAT	transcription factor binding	RT		4	5.8E-1	1.0E0
☐	GOTERM_BP_FAT	negative regulation of transcription, DNA-dependent	RT		3	6.0E-1	1.0E0
☐	GOTERM_BP_FAT	negative regulation of macromolecule biosynthetic process	RT		4	6.0E-1	1.0E0
☐	GOTERM_BP_FAT	negative regulation of RNA metabolic process	RT		3	6.1E-1	1.0E0
☐	GOTERM_MF_FAT	transcription cofactor activity	RT		3	6.3E-1	1.0E0
☐	SP_PIR_KEYWORDS	transcription regulation	RT		10	6.4E-1	9.9E-1
☐	SP_PIR_KEYWORDS	dna-binding	RT		9	6.7E-1	9.9E-1
☐	SP_PIR_KEYWORDS	Transcription	RT		10	6.7E-1	9.9E-1
☐	GOTERM_BP_FAT	regulation of transcription	RT		14	6.7E-1	1.0E0
☐	GOTERM_MF_FAT	transcription activator activity	RT		3	7.0E-1	1.0E0
☐	GOTERM_BP_FAT	negative regulation of transcription	RT		3	7.4E-1	1.0E0
☐	GOTERM_BP_FAT	negative regulation of gene expression	RT		3	7.8E-1	1.0E0
☐	GOTERM_BP_FAT	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		3	7.9E-1	1.0E0
☐	GOTERM_BP_FAT	transcription	RT		9	9.1E-1	1.0E0
☐	GOTERM_MF_FAT	DNA binding	RT		9	9.6E-1	1.0E0
☐	GOTERM_CC_FAT	nuclear lumen	RT		5	9.7E-1	1.0E0
☐	SP_PIR_KEYWORDS	nucleus	RT		15	9.7E-1	1.0E0
Annotation Cluster 35		Enrichment Score: 0.68	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	response to hypoxia	RT		4	4.2E-2	5.8E-1
☐	GOTERM_BP_FAT	response to oxygen levels	RT		4	4.6E-2	5.7E-1
☐	GOTERM_CC_FAT	cell surface	RT		4	3.1E-1	8.9E-1
☐	GOTERM_CC_FAT	cytoplasmic vesicle	RT		5	4.9E-1	9.6E-1
☐	GOTERM_MF_FAT	identical protein binding	RT		5	5.1E-1	1.0E0
☐	GOTERM_CC_FAT	vesicle	RT		5	5.2E-1	9.7E-1
Annotation Cluster 36		Enrichment Score: 0.61	G		Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	anti-apoptosis	RT		4	1.1E-1	7.3E-1
☐	GOTERM_BP_FAT	negative regulation of apoptosis	RT		5	1.4E-1	7.6E-1
☐	GOTERM_BP_FAT	negative regulation of programmed cell death	RT		5	1.4E-1	7.6E-1
☐	GOTERM_BP_FAT	negative regulation of cell death	RT		5	1.4E-1	7.6E-1
☐	GOTERM_BP_FAT	pattern specification process	RT		4	1.9E-1	8.3E-1
☐	GOTERM_BP_FAT	regionalization	RT		3	3.0E-1	9.3E-1
☐	GOTERM_BP_FAT	negative regulation of macromolecule metabolic process	RT		6	3.9E-1	9.6E-1
☐	GOTERM_BP_FAT	regulation of apoptosis	RT		6	4.8E-1	9.8E-1
☐	GOTERM_BP_FAT	regulation of programmed cell death	RT		6	4.8E-1	9.8E-1
☐	GOTERM_BP_FAT	regulation of cell death	RT		6	4.9E-1	9.8E-1
Annotation Cluster 37		Enrichment Score: 0.66	G		Count	P_Value	Benjamini

Annotation Cluster	Enrichment Score	RT		Count	P_Value	Benjamini
INTERPRO	EGF-like region, conserved site	RT		4	2.1E-1	1.0E0
INTERPRO	EGF-like, type 3	RT		3	2.8E-1	1.0E0
INTERPRO	EGF-like	RT		3	2.9E-1	1.0E0
SMART	EGF	RT		3	3.0E-1	1.0E0
SP_PIR_KEYWORDS	egf-like domain	RT		3	3.1E-1	9.3E-1
Annotation Cluster 38	Enrichment Score: 0.49	G		Count	P_Value	Benjamini
GOTERM_CC_FAT	Golgi apparatus	RT		8	2.1E-1	8.7E-1
SP_PIR_KEYWORDS	golgi apparatus	RT		5	3.3E-1	9.2E-1
GOTERM_CC_FAT	Golgi apparatus part	RT		3	5.0E-1	9.6E-1
Annotation Cluster 39	Enrichment Score: 0.48	G		Count	P_Value	Benjamini
SP_PIR_KEYWORDS	g-protein coupled receptor	RT		5	2.0E-1	8.9E-1
GOTERM_BP_FAT	cell surface receptor linked signal transduction	RT		12	2.1E-1	8.6E-1
GOTERM_BP_FAT	G-protein coupled receptor protein signaling pathway	RT		7	2.5E-1	9.0E-1
SP_PIR_KEYWORDS	transducer	RT		5	2.6E-1	9.3E-1
SP_PIR_KEYWORDS	receptor	RT		8	3.8E-1	9.4E-1
INTERPRO	GPCR, rhodopsin-like superfamily	RT		3	6.1E-1	1.0E0
INTERPRO	7TM GPCR, rhodopsin-like	RT		3	6.2E-1	1.0E0
Annotation Cluster 40	Enrichment Score: 0.41	G		Count	P_Value	Benjamini
GOTERM_CC_FAT	actin cytoskeleton	RT		5	6.6E-2	7.9E-1
GOTERM_CC_FAT	cytoskeleton	RT		10	3.1E-1	8.9E-1
GOTERM_CC_FAT	cytoskeletal part	RT		5	7.4E-1	9.9E-1
GOTERM_CC_FAT	intracellular non-membrane-bounded organelle	RT		13	7.7E-1	9.9E-1
GOTERM_CC_FAT	non-membrane-bounded organelle	RT		13	7.7E-1	9.9E-1
Annotation Cluster 41	Enrichment Score: 0.33	G		Count	P_Value	Benjamini
GOTERM_CC_FAT	axon	RT		3	2.3E-1	8.6E-1
GOTERM_CC_FAT	neuron projection	RT		3	5.8E-1	9.7E-1
GOTERM_CC_FAT	cell projection	RT		4	7.6E-1	9.9E-1
Annotation Cluster 42	Enrichment Score: 0.29	G		Count	P_Value	Benjamini
GOTERM_BP_FAT	membrane invagination	RT		3	3.6E-1	9.5E-1
GOTERM_BP_FAT	endocytosis	RT		3	3.6E-1	9.5E-1
GOTERM_BP_FAT	membrane organization	RT		3	6.4E-1	1.0E0
GOTERM_BP_FAT	vesicle-mediated transport	RT		3	8.4E-1	1.0E0
Annotation Cluster 43	Enrichment Score: 0.28	G		Count	P_Value	Benjamini
GOTERM_CC_FAT	nuclear envelope	RT		3	3.2E-1	8.9E-1
GOTERM_CC_FAT	organelle envelope	RT		4	6.7E-1	9.8E-1
GOTERM_CC_FAT	envelope	RT		4	6.8E-1	9.8E-1
Annotation Cluster 44	Enrichment Score: 0.26	G		Count	P_Value	Benjamini
GOTERM_CC_FAT	cytoplasmic vesicle part	RT		3	2.9E-1	9.0E-1
GOTERM_CC_FAT	cytoplasmic vesicle	RT		5	4.9E-1	9.6E-1
GOTERM_CC_FAT	vesicle	RT		5	5.2E-1	9.7E-1
GOTERM_CC_FAT	cytoplasmic membrane-bounded vesicle	RT		3	8.2E-1	1.0E0
GOTERM_CC_FAT	membrane-bounded vesicle	RT		3	8.3E-1	1.0E0
Annotation Cluster 45	Enrichment Score: 0.19	G		Count	P_Value	Benjamini
SP_PIR_KEYWORDS	metal-binding	RT		16	4.7E-1	9.7E-1
GOTERM_MF_FAT	transition metal ion binding	RT		17	5.0E-1	1.0E0

☐	SP_PIR_KEYWORDS	zinc	RT		11	6.2E-1	9.9E-1
☐	GOTERM_MF_FAT	ion binding	RT		24	6.3E-1	1.0E0
☐	GOTERM_MF_FAT	zinc ion binding	RT		13	6.6E-1	1.0E0
☐	GOTERM_MF_FAT	cation binding	RT		23	6.9E-1	1.0E0
☐	GOTERM_MF_FAT	metal ion binding	RT		22	7.6E-1	1.0E0
☐	SP_PIR_KEYWORDS	zinc-finger	RT		6	9.2E-1	1.0E0
Annotation Cluster 46					Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	regulation of transcription from RNA polymerase II promoter	RT		6	3.9E-1	9.6E-1
☐	GOTERM_MF_FAT	RNA polymerase II transcription factor activity	RT		3	4.1E-1	1.0E0
☐	GOTERM_CC_FAT	organelle lumen	RT		8	9.0E-1	1.0E0
☐	GOTERM_CC_FAT	nucleolus	RT		3	9.1E-1	1.0E0
☐	GOTERM_CC_FAT	membrane-enclosed lumen	RT		8	9.1E-1	1.0E0
☐	GOTERM_CC_FAT	intracellular organelle lumen	RT		7	9.5E-1	1.0E0
☐	GOTERM_CC_FAT	nuclear lumen	RT		5	9.7E-1	1.0E0
Annotation Cluster 47					Count	P_Value	Benjamini
☐	UP_SEQ_FEATURE	active site:Proton acceptor	RT		6	2.2E-1	1.0E0
☐	SP_PIR_KEYWORDS	phosphotransferase	RT		3	2.7E-1	9.2E-1
☐	SP_PIR_KEYWORDS	transferase	RT		8	5.3E-1	9.8E-1
☐	GOTERM_BP_FAT	phosphorylation	RT		5	6.6E-1	1.0E0
☐	SP_PIR_KEYWORDS	kinase	RT		4	6.7E-1	1.0E0
☐	UP_SEQ_FEATURE	domain:Protein kinase	RT		3	6.8E-1	1.0E0
☐	GOTERM_BP_FAT	protein amino acid phosphorylation	RT		4	7.3E-1	1.0E0
☐	INTERPRO	Protein kinase_core	RT		3	7.4E-1	1.0E0
☐	UP_SEQ_FEATURE	binding site:ATP	RT		3	7.5E-1	1.0E0
☐	GOTERM_BP_FAT	phosphorus metabolic process	RT		5	8.0E-1	1.0E0
☐	GOTERM_BP_FAT	phosphate metabolic process	RT		5	8.0E-1	1.0E0
☐	GOTERM_MF_FAT	protein kinase activity	RT		3	8.7E-1	1.0E0
☐	UP_SEQ_FEATURE	nucleotide phosphate-binding region:ATP	RT		3	9.5E-1	1.0E0
☐	GOTERM_MF_FAT	adenyl nucleotide binding	RT		6	9.6E-1	1.0E0
☐	GOTERM_MF_FAT	purine nucleoside binding	RT		6	9.6E-1	1.0E0
☐	GOTERM_MF_FAT	nucleoside binding	RT		6	9.6E-1	1.0E0
☐	GOTERM_MF_FAT	adenyl ribonucleotide binding	RT		5	9.8E-1	1.0E0
☐	GOTERM_MF_FAT	purine nucleotide binding	RT		6	9.9E-1	1.0E0
☐	SP_PIR_KEYWORDS	atp-binding	RT		3	9.9E-1	1.0E0
☐	SP_PIR_KEYWORDS	nucleotide-binding	RT		4	9.9E-1	1.0E0
☐	GOTERM_MF_FAT	purine ribonucleotide binding	RT		5	1.0E0	1.0E0
☐	GOTERM_MF_FAT	ribonucleotide binding	RT		5	1.0E0	1.0E0
☐	GOTERM_MF_FAT	nucleotide binding	RT		6	1.0E0	1.0E0
☐	GOTERM_MF_FAT	ATP binding	RT		3	1.0E0	1.0E0
Annotation Cluster 48					Count	P_Value	Benjamini
☐	GOTERM_BP_FAT	cell death	RT		4	7.7E-1	1.0E0
☐	GOTERM_BP_FAT	death	RT		4	7.8E-1	1.0E0
☐	GOTERM_BP_FAT	programmed cell death	RT		3	8.6E-1	1.0E0

108 terms were not clustered.

Appendix II – IPA biological functions,

carcinoma, cancer, adenocarcinoma, tumorigenesis, reproductive system disorder, gynecological disorder, cardiovascular disorder, invasion of cells, metastatic colorectal cancer, endometrial cancer, genital tumor, invasion of ovarian cancer cell lines, uterine cancer, proliferation of cells, development of forelimb, tissue development, genetic disorder, neurological disorder, colony formation of cells, cell movement, synthesis of terpenoid, migration of cancer cells, proliferation of connective tissue cells, metastasis, development of blood vessel, developmental disorder, serous ovarian carcinoma process, organismal abnormalities of organ, migration of cells, morphogenesis of organism, clearance of D-glucose, invasion of tumor cell lines, synthesis of lipid, entrance of Ca^{2+} , formation of skeletal muscle, synthesis of steroid, migration of endothelial cells, metabolism of lipid, synthesis of polyols, limb development, hypertrophy of cardiomyocytes, atypical hemolytic uremic syndrome, invasion of prostate cancer cells, colorectal cancer, movement of tumor cell lines, morphogenesis of forelimb, ovarian tumor, gastrointestinal tract cancer, development of connective tissue, proliferation of epithelial cells, modification of lipid, hypertrophy, neoplasia of cells, morphogenesis of limb, diameter of cells, oxidation of retinaldehyde, recruitment of bone marrow precursor cells, vascularization of connective tissue, metastasis of tumor cell lines, fibrosis of organ, neuromuscular disease, disease of central nervous system, tumorigenesis of melanoma cell lines, quantity of phosphatidic acid, growth of prostate cancer cell lines, growth of tumor cell lines, astrogliosis, colony formation of connective tissue cells, transformation, cerebrovascular dysfunction, digestive organ tumor, neuropathy, development of skeletal muscle, angiogenesis, coagulation of blood, degeneration of cardiomyocytes, exocytosis of Weibel-Palade bodies, synthesis of inositol phosphate, cellularity, costimulation of cells, neurodegenerative disorder, anoikis of breast cancer cell lines, endocytosis by tumor cell lines, proliferation of osteoclast precursor cells, tubulation of epithelial tissue, ovarian cancer, migration of breast cancer cell lines, migration of tumor cell lines, proliferation of leukocyte cell lines, frequency of tumor, morphology of body region, primary pulmonary hypertension, proliferation of erythroid cells, transactivation of Ets element, size of cardiomyocytes, hypertrophy of cells, Alzheimer's disease, quantity of metal, quantity of vesicles, oxidation of organic chemical, atrophy, skeletal and muscular disorder, catalepsy, length of body region, proliferation of embryonic stem cell lines, dilation of heart ventricle, proliferation of cancer cells, tumorigenesis of tumor cell lines, cell cycle progression, morphology of cells, binding of bone marrow cells, cell cycle progression of cancer cells, formation of cataract, quantity of caveolae, formation of membrane ruffles, proliferation of kidney cells, vasculogenesis, squamous cell tumor, ischemic stroke, proliferation of embryonic cell lines, proliferation of mesenchymal cells, head and neck cancer, proliferation of B-lymphocyte derived cell lines, adhesion of breast cell lines, development of pericardium, dystrophy, pinocytosis by cells, psychological disorder, learning, proliferation of embryonic cells, prostate cancer, differentiation, fibrosis, synthesis of carbohydrate, heart rate, inflammatory response, progressive motor neuropathy, differentiation of cells, formation of filaments, quantity of

osteoclasts, colony formation of leukemia cell lines, development of mammary duct, extension of lamellipodia, fibrosis of interstitial tissue, inactivation of MAP kinase, quantity of cyclic nucleotides, function of heart, quantity of Ca^{2+} , release of lipid, differentiation of connective tissue cells, quantity of monoamines, astrocytosis of brain, cognition disorder, formation of glioma, metabolism of cells, survival of melanoma cell lines, transactivation of cAMP response element, formation of muscle, tubulation of cells, invasion of cancer cells, fatty acid metabolism, invasion of prostate cancer cell lines, Marfan's syndrome, cell spreading of kidney cell lines, degranulation of BMMC cells, migration of bone cancer cell lines, nutritional disorder, morphology of organ, size of bone, vasoconstriction of blood vessel, hematopoiesis of cells, colony formation of myeloid progenitor cells, contraction of connective tissue cells, function of tissue, quantity of glutathione, growth of breast cancer cell lines, tumorigenesis of cells, co-stimulation of T lymphocytes, desensitization of cells, erythropoiesis of cells, metastasis of melanoma cell lines, morphology of heart, proliferation of smooth muscle cells, ingestion by mice, smoothened signaling pathway, Early-onset reducing body x-linked severe myopathy, Parkinson disease 4, autosomal dominant, Lewy body, X-linked dominant scapuloperoneal myopathy, X-linked myopathy with postural muscle atrophy, Aceruloplasminemia, alignment of active zone, alignment of junctional folds, apoptosis of cd56+ natural killer cells, area of left common carotid artery, area of right common carotid artery, arrest in cell cycle progression of prostate cancer cells, arrest in differentiation of melanoma cell lines, assembly of Z line, atrophy of cerebellum, atrophy of olivary nucleus, atrophy of pontine nucleus, atrophy of renal medulla, auto-oxidation of L-cysteine, auto-oxidation of cystine, binding of U46619, binding of myosin filaments, binding of synaptic vesicles, blood pressure of corpus cavernosum penis, branching of N-glycan, branching of airway, branching of skin cell lines, cardiomyopathy of left ventricle, catabolism of 5-hydroxytryptamine, catabolism of norepinephrine, cell cycle progression of type A spermatogonia, chemotherapy resistance of lung cancer cell lines, colony formation of non-small-cell lung cancer cells, colony formation of osteoprogenitor cells, colony formation of small cell lung cancer cells, cranial chondrodystrophy, cytotoxicity of daunorubicin, deficiency of ferroxidase, deformation of skin, degeneration of yolk sac, delay in G2/M phase transition of fibroblasts, delay in tubulation of epithelial cells, density of collagen fibrils, density of pericytes, development of burst-forming erythroid cells, development of colony forming multilineage cells, development of costameres, development of cytoplasmic inclusions, development of epithelial bud, development of fat pad, development of fibrillar inclusions, development of granulocyte-macrophage progenitor cells, development of limbic system, development of mammary lesion, development of mesangium, development of neural retina, diameter of left common carotid artery, diameter of left ventricle, diameter of oocytes, diameter of right common carotid artery, differentiation of periosteal cells, differentiation of pneumocytes, disorganization of rod outer segments, disruption of fibrils, disruption of vesicle aggregates, dissemination of melanoma cell lines, distribution of melanocytes, entry into cell cycle progression of fibrosarcoma cells, entry into cell cycle progression of melanoma cells, erythroderma, exit from G0/G1 phase transition, exit from quiescence of

fibroblast cell lines, expansion of basal layer of epidermis, expansion of mesenchyme, expansion of suprabasal layer of epidermis, extravasation of breast cancer cell lines, factor h deficiency, focal nodular hyperplasia, formation of airspace, formation of calcium stone, formation of multipotential hemopoietic progenitor cells, frequency of mammary tumor, fusion of gonadal cell lines, generation of mast cells, growth of type A spermatogonia, heterotaxia, hydraulic conductivity of venule, hydrolysis of cocaine, hyperproliferation of B-lymphocyte derived cell lines, hyperproliferation of fibroblasts, hypertension of pulmonary artery, hypertrophy of islets of Langerhans, hypo-apoptosis of myeloid cells, induction of mammary gland, interaction of acute myeloblastic leukemia cells, interaction of cervical cancer cell lines, internalization of lactosylceramide, invasion of papillary thyroid carcinoma, leakage of vesicles, length of body trunk, length of proliferative zone, lipolysis of white adipose tissue, loss of fibrils, maintenance of microvessel, mass of epithelial tissue, metabolism of cocaine, metabolism of embryonic cell lines, metabolism of epithelial cell lines, metabolism of fibroblast cell lines, metabolism of kidney cell lines, metabolism of squamous cell carcinoma cell lines, metastasis of papillary thyroid carcinoma, metastasis of pheochromocytoma cells, microform holoprosencephaly, migration of mesonephric cells, mineralization of bone cancer cell lines, mitogenesis of endometrial stromal cells, morphology of malleus, morphology of nuclear matrix, morphology of parenchyma, morphology of styloid process of temporal bone, neoplasia of hematopoietic cells, neovascularization of tibia, non-erythrodermic inflammatory skin disease, oxidation of acetaldehyde, pendred's syndrome, pigmentation of epidermis, polarization of fibroblasts, positioning of endodermal cells, presence of melanocytes, production of cellular inclusion bodies, production of cytoplasmic aggregates, proliferation of Ewing's sarcoma cells, quantity of caveolar membranes, quantity of membrane blebs, quantity of mesenchyme, quantity of perikaryon, reduction of progesterone, regeneration of tibia, remodeling of saphenous vein, replenishment of synaptic vesicles, replication of mesangial cells, retraction of mesothelial cells, stimulation of high proliferative potential colony-forming cells, strength of skin, survival of acute myeloblastic leukemia cells, survival of melanoblasts, survival of tubular cells, synthesis of neointima, tenascin-X-deficiency, thickness of vascular smooth muscle cells, transition of breast cell lines, ulnar-mammary syndrome, development of bone marrow cells, adhesion of mast cells, colony formation of cancer cells, Parkinson's disease, end stage renal disease, aggregation of cells, brain cancer, contraction of smooth muscle, degeneration of tissue, mucinous ovarian cancer, oxidation of protein, metabolism of terpenoid, growth of cells, morphology of fibroblast cell lines, development of embryonic tissue, hematological process, metabolism of acylglycerol, activation of protein binding site, proliferation of bone marrow cells, quantity of phosphatidylinositol, metabolism of amine, edema of lung, formation of cytoplasmic aggregates, growth of muscle cells, transport of divalent cations, quantity of epinephrine, retraction of cells, transformation of breast cell lines, vacuolation of cells, squamous-cell carcinoma, quantity of connective tissue cells, development of bone, immune response, clear-cell ovarian carcinoma, fibrosis of kidney, hematopoiesis of bone marrow cells, release of fatty acid, encephalopathy, cell death of osteoblasts, conversion of hormone,

interphase of leukemia cell lines, metabolism of tretinoin, morphology of blood vessel, proliferation of tumor, fibrosis of heart, neurological disorder of organ, substance-related disorder, colony formation of tumor cell lines, degeneration of cells, activation of blood platelets, contraction of cells, quantity of catecholamine, growth of fibroblast cell lines, muscle contraction, metabolism of dopamine, opioid-related disorder, sprouting of endothelial cells, heart disease, hyperproliferation, biosynthesis of estrogen, bradycardia, metastasis of tumor cells, senescence of cells, synthesis of DNA, metabolism of reactive oxygen species, influx of Ca^{2+} , hydrolysis of phosphatidylinositol, quantity of carbohydrate, metabolism of carbohydrate, formation of tissue, differentiation of brown adipocytes, endometrial ovarian cancer, proliferation of stromal cells, tumorigenesis of fibroblast cell lines, Carney complex type 1, Ehlers-Danlos syndrome type III, activation of breast cell lines, aggregation of insect cell lines, angiogenesis of infarct, angiogenesis of melanoma cell lines, apoptosis of primordial germ cells, apoptosis of type A spermatogonia, arrest in G2/M phase transition of leukemia cell lines, atrophy of seminiferous tubules, attraction of hematopoietic progenitor cells, branching morphogenesis of mammary duct, cell cycle progression of erythroid cells, cell viability of mesothelioma cells, cell viability of neuroglia, cell-cell adhesion of leukemia cell lines, cellularity of oligodendroglioma, circulation of brain, concentration of retinaldehyde, concentration of tretinoin, contraction of microvascular endothelial cells, cytotoxicity of $\text{cd}56^{+}$ natural killer cells, cytotoxicity of dopaminergic neurons, deamination of dopamine, degeneration of mammary primordial, degeneration of myelin figure, delay in modification of organ, desensitization of hippocampal neurons, development of papilla, development of stapes, diameter of adipocytes, differentiation of blood-derived mast cells, differentiation of bone cancer cell lines, differentiation of melanoblasts, digestive process of rodents, dilatation of the vestibular aqueduct, disassembly of intermediate filaments, dysfunction of regulatory T lymphocytes, efflux of iodide, elongation of fibroblast cell lines, epithelial-mesenchymal transition of prostate cancer cell lines, exit from S phase of fibroblast cell lines, familial visceral amyloidosis, ostertag type, fibrosis of ductal epithelium, fibrosis of pericytes, formation of nipple, formation of pulmonary artery, formation of punctate structures, frequency of colony-forming granulocyte-macrophages, function of brown adipose tissue, fusion of leukemia cell lines, fusion of lymphoblastoid cell lines, hereditary sensory neuropathy type 2, hyperplasia of cardiomyocytes, hyperplasia of pneumocytes, hypopharynx carcinoma, hypoplasia of bone marrow, hypoplasia of premaxilla, infiltration of pulmonary alveolus, injury of neuroblastoma cell lines, intimal hyperplasia of femoral artery, invasion of perioptic mesenchyme, loss of coronal suture, loss of mesangial cells, metastasis of adenocarcinoma cells, metastasis of fibrosarcoma cells, microhemorrhage of brain, migration of melanoblasts, mineralization of cartilage matrix, mineralization of growth plate, mineralization of skeleton, mineralization of vascular smooth muscle cells, morphology of ductal epithelium, morphology of forelimb, morphology of hindlimb, necrosis of oligodendroglioma, neovascularization of growth plate, neurodegeneration of cholinergic neurons, organization of cardiomyocytes, outgrowth of mammary primordial, oxidation of 4-androstene-3,17-dione, oxidation of 5-hydroxytryptamine, oxidation of

epinephrine, oxidation of norepinephrine, oxidation of phenethylamine, patterning of rhombencephalon, phosphaturia, production of breast milk, progression of oligodendroglioma

The above list contains only terms found to be statistically significantly associated with our candidate list.

Appendix III – IPA canonical pathways

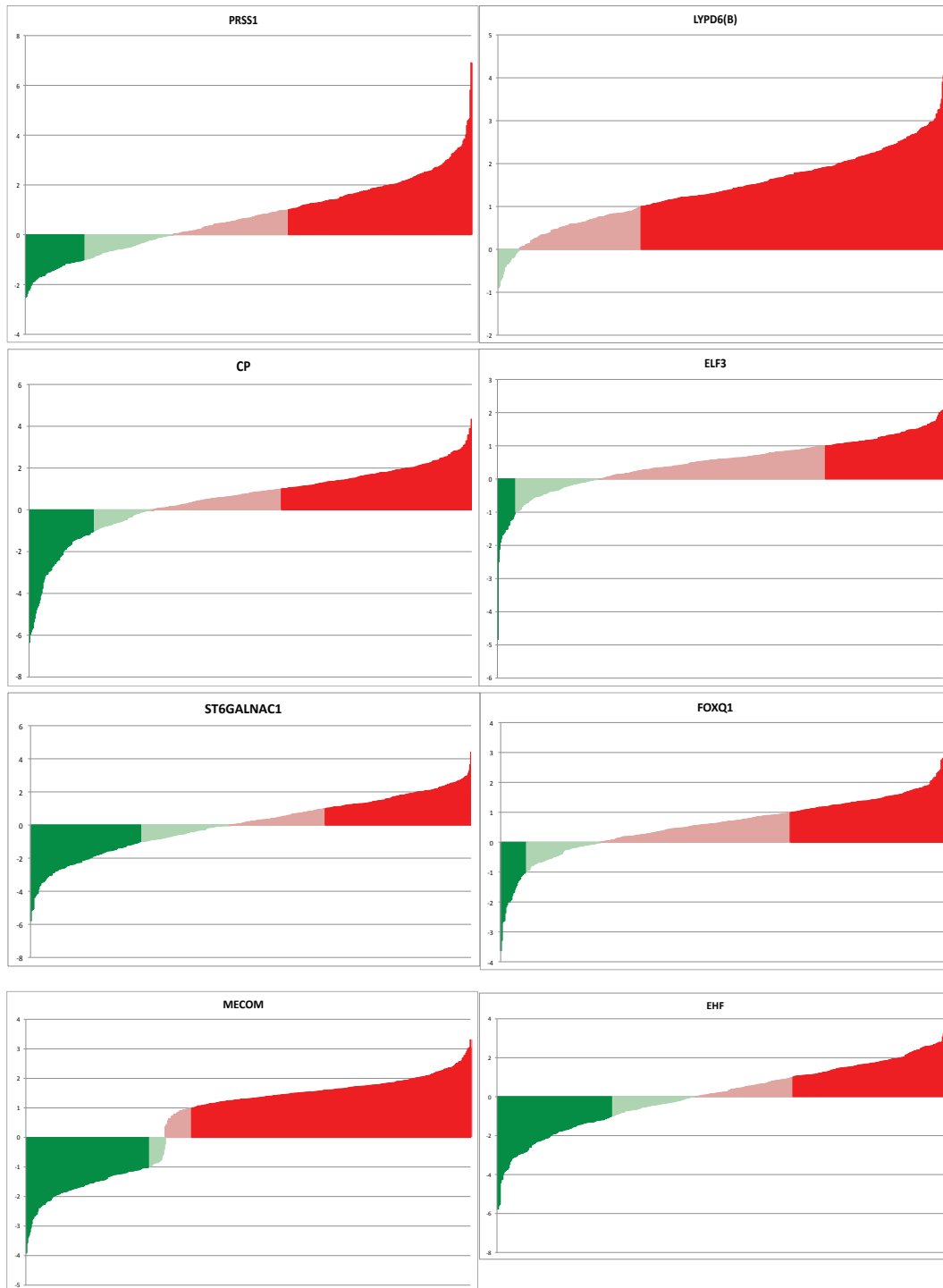
© 2000-2011 Ingenuity Systems, Inc. All rights reserved.				
Ingenuity Canonical Pathways	-log(p-value)	pvalue	Ratio	Molecules
G-Protein Coupled Receptor Signaling	2.63E00	0.0023	1.7E-02	F2R,RGS4,RGS2,FZD7,AGTR1,PRKAR2B,CAMK2D,PTHLH,GPR133
C21-Steroid Hormone Metabolism	2.36E00	0.0044	2.94E-02	AKR1C1/AKR1C2,HSD17B2
Androgen and Estrogen Metabolism	2.28E00	0.0052	2.29E-02	AKR1C1/AKR1C2,HSD17B2,STS
cAMP-mediated signaling	2.12E00	0.0076	2.28E-02	RGS4,RGS2,AGTR1,PRKAR2B,CAMK2D
Clathrin-mediated Endocytosis Signaling	1.88E00	0.0132	2.33E-02	FGF9,F2R,PDGFB,DAB2
Complement System	1.82E00	0.0151	5.71E-02	CFI,CFH
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1.61E00	0.0245	1.5E-02	FZD7,PRSS1/PRSS3,PDGFB,CAMK2D,SFRP1
Tryptophan Metabolism	1.6E00	0.0251	1.28E-02	AOX1,MAOB,ALDH1A1
Bile Acid Biosynthesis	1.51E00	0.0309	1.87E-02	AKR1C1/AKR1C2,ALDH1A1
Histidine Metabolism	1.48E00	0.0331	1.72E-02	MAOB,ALDH1A1
Ovarian Cancer Signaling	1.4E00	0.0398	2.11E-02	FGF9,FZD7,PRKAR2B
Pyruvate Metabolism	1.33E00	0.0468	1.47E-02	ME1,ALDH1A1
Valine, Leucine and Isoleucine Degradation	1.32E00	0.0479	1.85E-02	AOX1,ALDH1A1
Lysine Degradation	1.32E00	0.0479	1.44E-02	BBOX1,ALDH1A1
PXR/RXR Activation	1.3E00	0.0501	2.25E-02	PRKAR2B,ALDH1A1
Tyrosine Metabolism	1.28E00	0.0525	1.01E-02	AOX1,MAOB
Xenobiotic Metabolism Signaling	1.24E00	0.0575	1.32E-02	MAOB,HS3ST1,CAMK2D,ALDH1A1
Melatonin Signaling	1.23E00	0.0589	2.6E-02	PRKAR2B,CAMK2D
Aminosugars Metabolism	1.2E00	0.0631	1.64E-02	NPL,GFPT2
Arginine and Proline Metabolism	1.2E00	0.0631	1.13E-02	MAOB,ALDH1A1
PDGF Signaling	1.19E00	0.0646	2.53E-02	PDGFB,CAV1
Nitric Oxide Signaling in the Cardiovascular System	1.15E00	0.0708	2E-02	PRKAR2B,CAV1
Calcium Signaling	1.14E00	0.0724	1.45E-02	PRKAR2B,CAMK2D,TNNT3
Dopamine Receptor Signaling	1.13E00	0.0741	2.11E-02	MAOB,PRKAR2B
RAR Activation	1.12E00	0.0759	1.6E-02	CRABP2,PRKAR2B,ALDH1A1
TR/RXR Activation	1.05E00	0.0891	2.08E-02	AKR1C1/AKR1C2,ME1
Melanocyte Development and Pigmentation Signaling	1.05E00	0.0891	2.17E-02	KITLG,PRKAR2B
Parkinson's Signaling	1.05E00	0.0891	5.56E-02	SNCA
Ascorbate and Aldarate Metabolism	1.05E00	0.0891	1.25E-02	ALDH1A1
G Beta Gamma Signaling	1.03E00	0.0933	1.71E-02	PRKAR2B,CAV1
LPS/TL-1 Mediated Inhibition of RXR Function	1.02E00	0.0955	1.36E-02	MAOB,HS3ST1,ALDH1A1
Glioma Signaling	9.78E-01	0.1052	1.79E-02	PDGFB,CAMK2D
Neuropathic Pain Signaling In Dorsal Horn Neurons	9.4E-01	0.1148	1.85E-02	PRKAR2B,CAMK2D
Actin Cytoskeleton Signaling	9.33E-01	0.1167	1.26E-02	FGF9,F2R,PDGFB
Synaptic Long Term Potentiation	9.12E-01	0.1225	1.75E-02	PRKAR2B,CAMK2D
Renin-Angiotensin Signaling	8.72E-01	0.1343	1.59E-02	AGTR1,PRKAR2B
14-3-3-mediated Signaling	8.6E-01	0.1380	1.67E-02	VTM,SNCA
p70S6K Signaling	8.17E-01	0.1524	1.54E-02	F2R,AGTR1
O-Glycan Biosynthesis	8.04E-01	0.1570	2.27E-02	ST6GALNAC1
Sonic Hedgehog Signaling	8.04E-01	0.1570	3.03E-02	PRKAR2B
Serotonin Receptor Signaling	7.91E-01	0.1618	2.17E-02	MAOB
GNRH Signaling	7.83E-01	0.1648	1.37E-02	PRKAR2B,CAMK2D

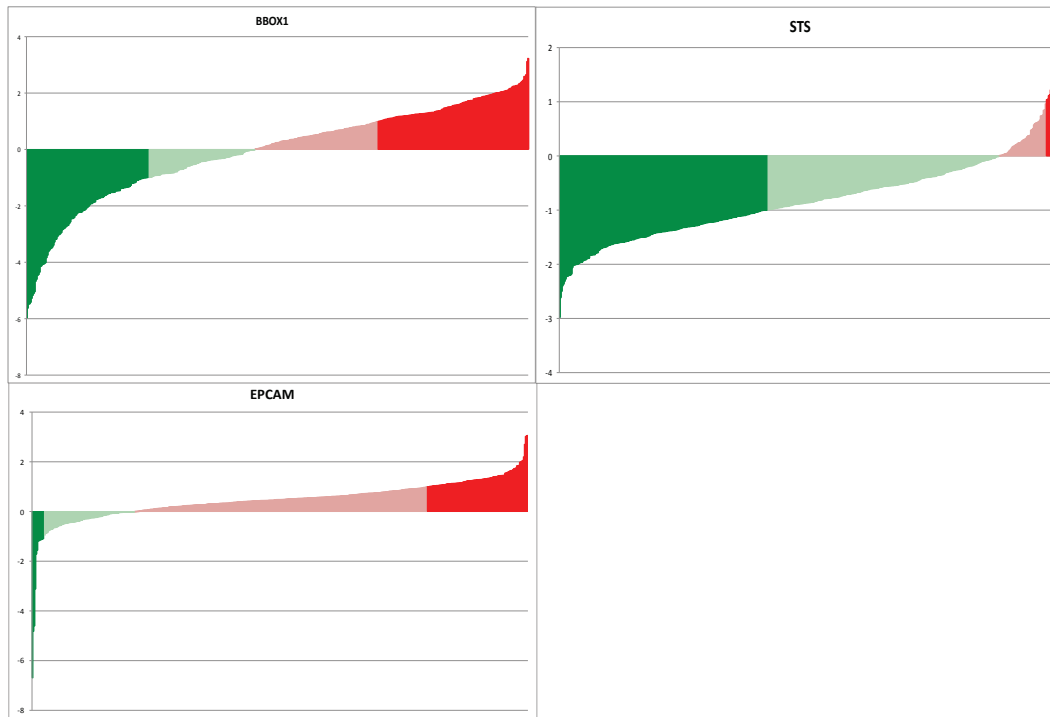
GNRH Signaling	7.83E-01	0.1648	1.37E-02	PRKAR2B,CAMK2D
Human Embryonic Stem Cell Pluripotency	7.83E-01	0.1648	1.3E-02	FZD7,PDGFB
Retinol Metabolism	7.78E-01	0.1667	1.61E-02	ALDH1A1
Mitochondrial Dysfunction	7.67E-01	0.1710	1.14E-02	MAOB,SNCA
Coagulation System	7.3E-01	0.1862	2.63E-02	F2R
Axonal Guidance Signaling	7.27E-01	0.1875	9.24E-03	FZD7,SEMA6D,PDGFB,PRKAR2B
Glutamate Metabolism	7.19E-01	0.1910	1.33E-02	GFPT2
Hepatic Fibrosis / Hepatic Stellate Cell Activation	7.17E-01	0.1919	1.36E-02	PDGFB,AGTR1
Phenylalanine Metabolism	7.08E-01	0.1959	9.01E-03	MAOB
Glioblastoma Multiforme Signaling	6.84E-01	0.2070	1.22E-02	FZD7,PDGFB
Neuroprotective Role of THOP1 in Alzheimer's Disease	6.78E-01	0.2099	1.85E-02	PRKAR2B
β -alanine Metabolism	6.33E-01	0.2328	1.08E-02	ALDH1A1
Chondroitin Sulfate Biosynthesis	6.24E-01	0.2377	1.54E-02	HS3ST1
Cysteine Metabolism	6.24E-01	0.2377	1.15E-02	HS3ST1
Keratan Sulfate Biosynthesis	6.16E-01	0.2421	1.85E-02	HS3ST1
Wnt/ β -catenin Signaling	6.06E-01	0.2477	1.15E-02	FZD7,SFRP1
Acute Phase Response Signaling	5.94E-01	0.2547	1.12E-02	CRABP2,CP
Amyloid Processing	5.93E-01	0.2553	1.79E-02	PRKAR2B
Phototransduction Pathway	5.86E-01	0.2594	1.54E-02	PRKAR2B
CREB Signaling in Neurons	5.83E-01	0.2612	9.9E-03	PRKAR2B,CAMK2D
NRF2-mediated Oxidative Stress Response	5.62E-01	0.2742	1.04E-02	AOX1,JUNB
Role of NFAT in Cardiac Hypertrophy	5.59E-01	0.2761	9.62E-03	PRKAR2B,CAMK2D
Propanoate Metabolism	5.58E-01	0.2767	8.26E-03	ALDH1A1
ERK/MAPK Signaling	5.45E-01	0.2851	9.8E-03	ELF3,PRKAR2B
Glioma Invasiveness Signaling	5.44E-01	0.2858	1.67E-02	F2R
Retinoic acid Mediated Apoptosis Signaling	5.38E-01	0.2897	1.47E-02	CRABP2
Butanoate Metabolism	5.38E-01	0.2897	7.69E-03	ALDH1A1
Estrogen-Dependent Breast Cancer Signaling	5.32E-01	0.2938	1.43E-02	HSD17B2
Thrombin Signaling	5.23E-01	0.2999	9.66E-03	F2R,CAMK2D
IL-15 Signaling	5.2E-01	0.3020	1.49E-02	AXL
GM-CSF Signaling	5.14E-01	0.3062	1.49E-02	CAMK2D
Breast Cancer Regulation by Stathmin1	5.1E-01	0.3090	9.52E-03	PRKAR2B,CAMK2D
Molecular Mechanisms of Cancer	5.06E-01	0.3119	7.92E-03	FZD7,PRKAR2B,CAMK2D
Glycine, Serine and Threonine Metabolism	4.86E-01	0.3266	6.85E-03	MAOB
Renal Cell Carcinoma Signaling	4.81E-01	0.3304	1.35E-02	PDGFB
Chemokine Signaling	4.81E-01	0.3304	1.37E-02	CAMK2D
Macropinocytosis Signaling	4.81E-01	0.3304	1.32E-02	PDGFB
Basal Cell Carcinoma Signaling	4.76E-01	0.3342	1.37E-02	FZD7
Caveolar-mediated Endocytosis Signaling	4.66E-01	0.3420	1.18E-02	CAV1
Role of Wnt/GSK-3 β Signaling in the Pathogenesis of Influenza	4.61E-01	0.3459	1.23E-02	FZD7
BMP signaling pathway	4.61E-01	0.3459	1.25E-02	PRKAR2B
Leptin Signaling in Obesity	4.51E-01	0.3540	1.19E-02	PRKAR2B
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis	4.46E-01	0.3581	8.33E-03	FZD7,SFRP1
Acute Myeloid Leukemia Signaling	4.42E-01	0.3614	1.22E-02	KITLG
Metabolism of Xenobiotics by Cytochrome P450	4.38E-01	0.3648	5.08E-03	AKR1C1/AKR1C2
VDR/RXR Activation	4.38E-01	0.3648	1.23E-02	HSD17B2
Glycolysis/Gluconeogenesis	4.29E-01	0.3724	7.46E-03	ALDH1A1
FGF Signaling	4.12E-01	0.3873	1.11E-02	FGF9

Pathway	1.0E-01	0.0001	1.0E-06	Pathway
Crosstalk between Dendritic Cells and Natural Killer Cells	4.08E-01	0.3908	1.04E-02	CAMK2D
Colorectal Cancer Metastasis Signaling	4.02E-01	0.3963	7.75E-03	FZD7,PRKAR2B
Bladder Cancer Signaling	4E-01	0.3981	1.09E-02	FGF9
α -Adrenergic Signaling	3.96E-01	0.4018	9.43E-03	PRKAR2B
Factors Promoting Cardiogenesis in Vertebrates	3.96E-01	0.4018	1.05E-02	FZD7
CDK5 Signaling	3.96E-01	0.4018	1.06E-02	PRKAR2B
PAK Signaling	3.92E-01	0.4055	9.35E-03	PDGFB
Glycerolipid Metabolism	3.92E-01	0.4055	6.45E-03	ALDH1A1
Virus Entry via Endocytic Pathways	3.89E-01	0.4083	1E-02	CAV1
Atherosclerosis Signaling	3.89E-01	0.4083	9.26E-03	PDGFB
PPAR Signaling	3.85E-01	0.4121	9.35E-03	PDGFB
IL-1 Signaling	3.81E-01	0.4159	9.35E-03	PRKAR2B
Chronic Myeloid Leukemia Signaling	3.78E-01	0.4188	9.52E-03	MECOM
IGF-1 Signaling	3.64E-01	0.4325	9.35E-03	PRKAR2B
iCOS-iCOSL Signaling in T Helper Cells	3.6E-01	0.4365	8.2E-03	CAMK2D
Nicotinate and Nicotinamide Metabolism	3.6E-01	0.4365	7.35E-03	AOX1
Fatty Acid Metabolism	3.57E-01	0.4395	5.38E-03	ALDH1A1
HGF Signaling	3.54E-01	0.4426	9.52E-03	ELF3
Type I Diabetes Mellitus Signaling	3.44E-01	0.4529	8.26E-03	CPE
Sphingosine-1-phosphate Signaling	3.32E-01	0.4656	8.26E-03	PDGFB
Corticotropin Releasing Hormone Signaling	3.29E-01	0.4688	7.35E-03	PRKAR2B
PKC δ Signaling in T Lymphocytes	3.29E-01	0.4688	7.04E-03	CAMK2D
Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency	3.26E-01	0.4721	8.77E-03	FZD7
RhoA Signaling	3.24E-01	0.4742	8.77E-03	ARHGAP6
Androgen Signaling	3.21E-01	0.4775	6.94E-03	PRKAR2B
Glycerophospholipid Metabolism	3.1E-01	0.4898	5.43E-03	BCHE
G α 12/13 Signaling	3.04E-01	0.4966	7.81E-03	F2R
P2Y Purigenic Receptor Signaling Pathway	3.02E-01	0.4989	7.3E-03	PRKAR2B
Insulin Receptor Signaling	2.92E-01	0.5105	7.14E-03	PRKAR2B
PI3K Signaling in B Lymphocytes	2.87E-01	0.5164	6.99E-03	CAMK2D
Protein Kinase A Signaling	2.83E-01	0.5212	6.1E-03	PRKAR2B,CAMK2D
Cellular Effects of Sildenafil (Viagra)	2.8E-01	0.5248	6.62E-03	PRKAR2B
Aryl Hydrocarbon Receptor Signaling	2.68E-01	0.5395	6.29E-03	ALDH1A1
Relaxin Signaling	2.68E-01	0.5395	6.33E-03	PRKAR2B
AMPK Signaling	2.68E-01	0.5395	5.95E-03	PRKAR2B
Cardiac β -adrenergic Signaling	2.66E-01	0.5420	6.49E-03	PRKAR2B
Hepatic Cholestasis	2.58E-01	0.5521	5.68E-03	PRKAR2B
B Cell Receptor Signaling	2.47E-01	0.5662	6.41E-03	CAMK2D
Tight Junction Signaling	2.3E-01	0.5888	6.1E-03	PRKAR2B
PPAR α /RXR α Activation	2.16E-01	0.6081	5.38E-03	PRKAR2B
Ephrin Receptor Signaling	1.98E-01	0.6339	5E-03	PDGFB

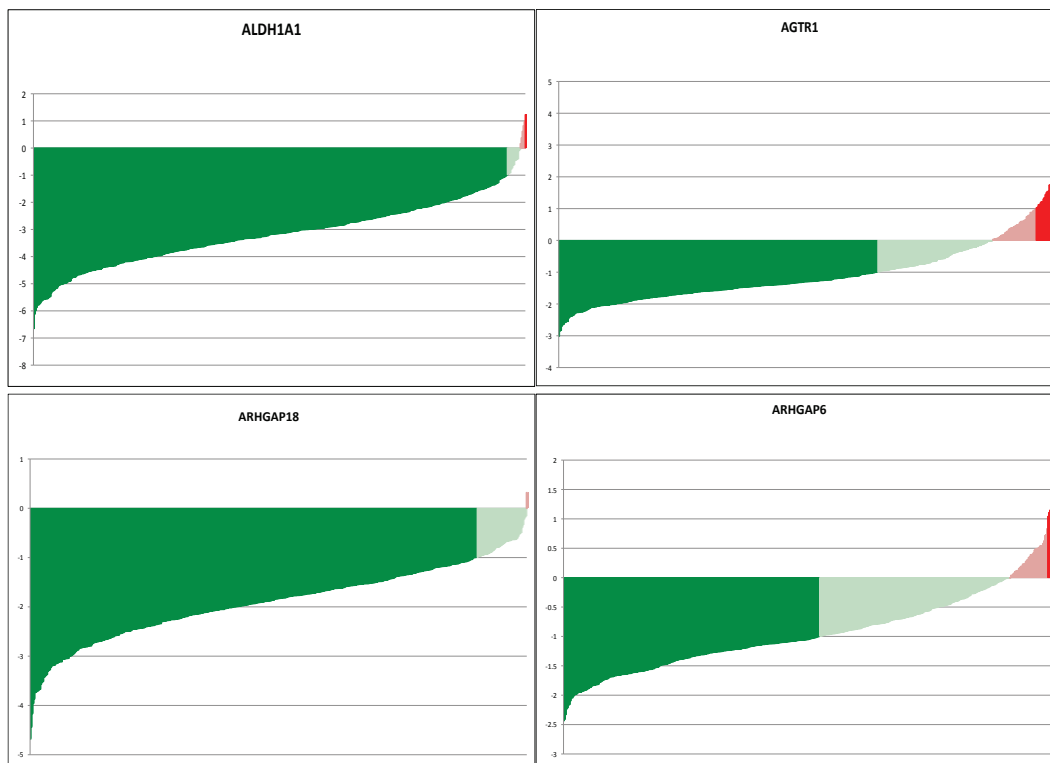
Appendix IV – TCGA expression profiles

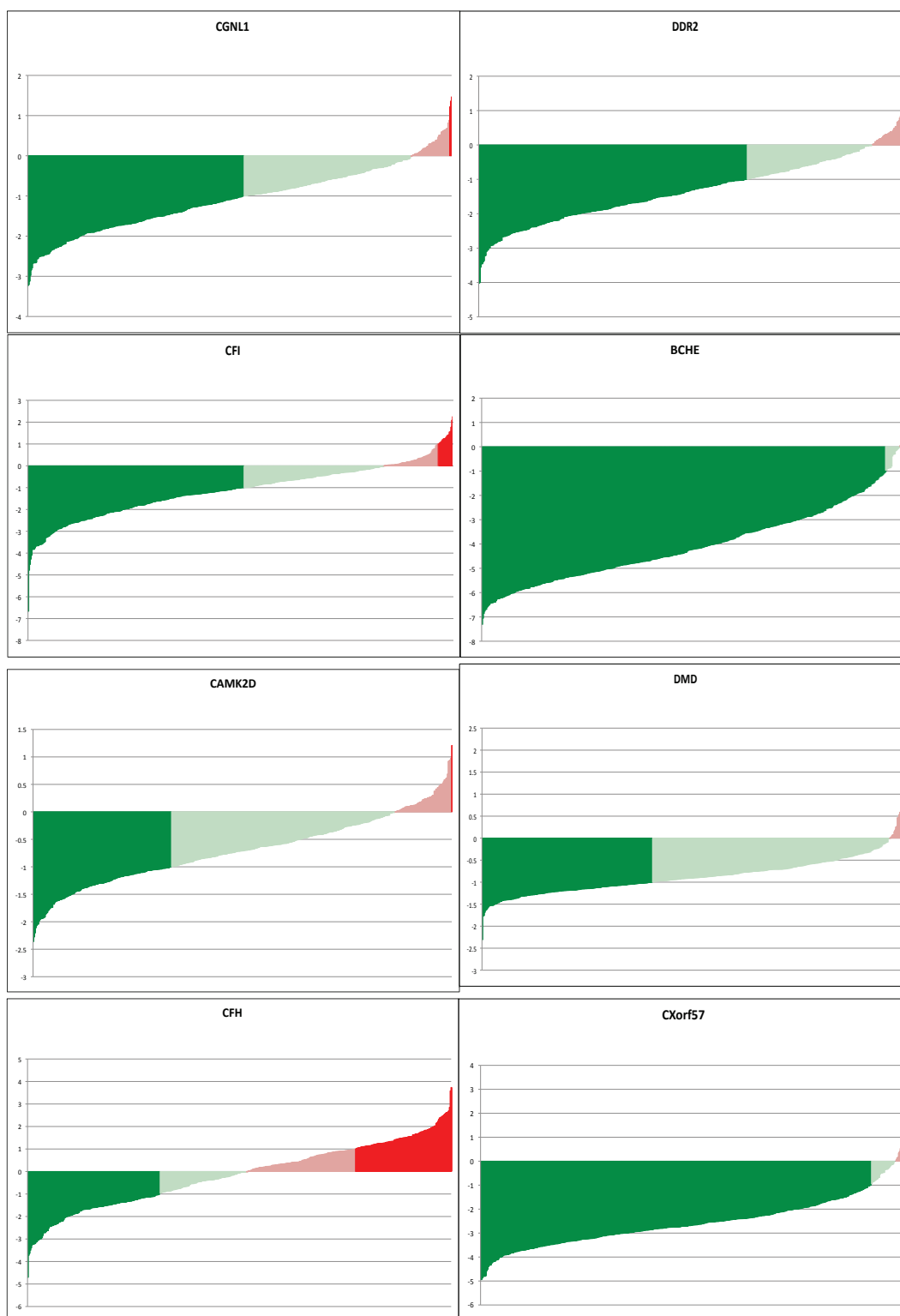
Category 1 Genes

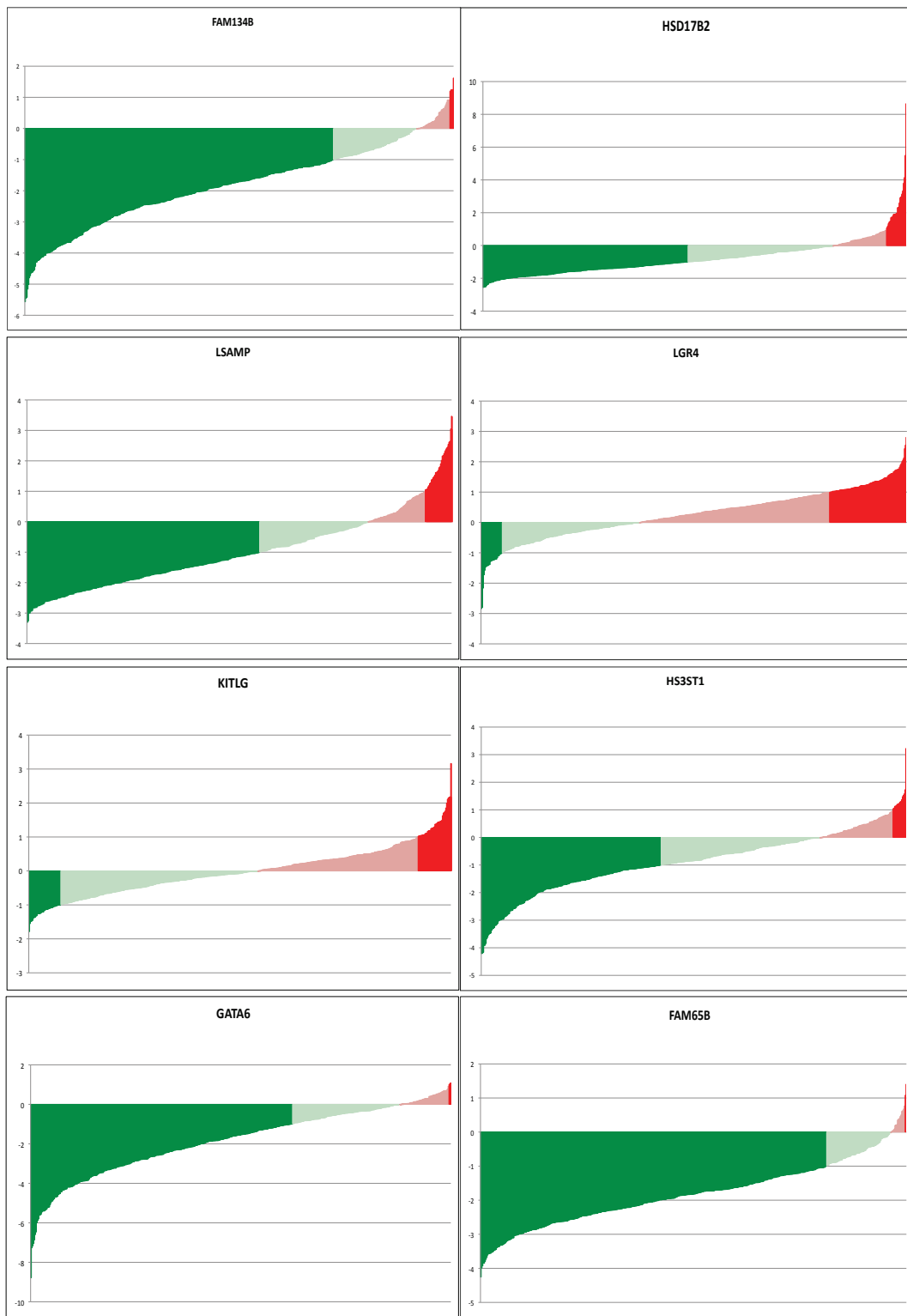


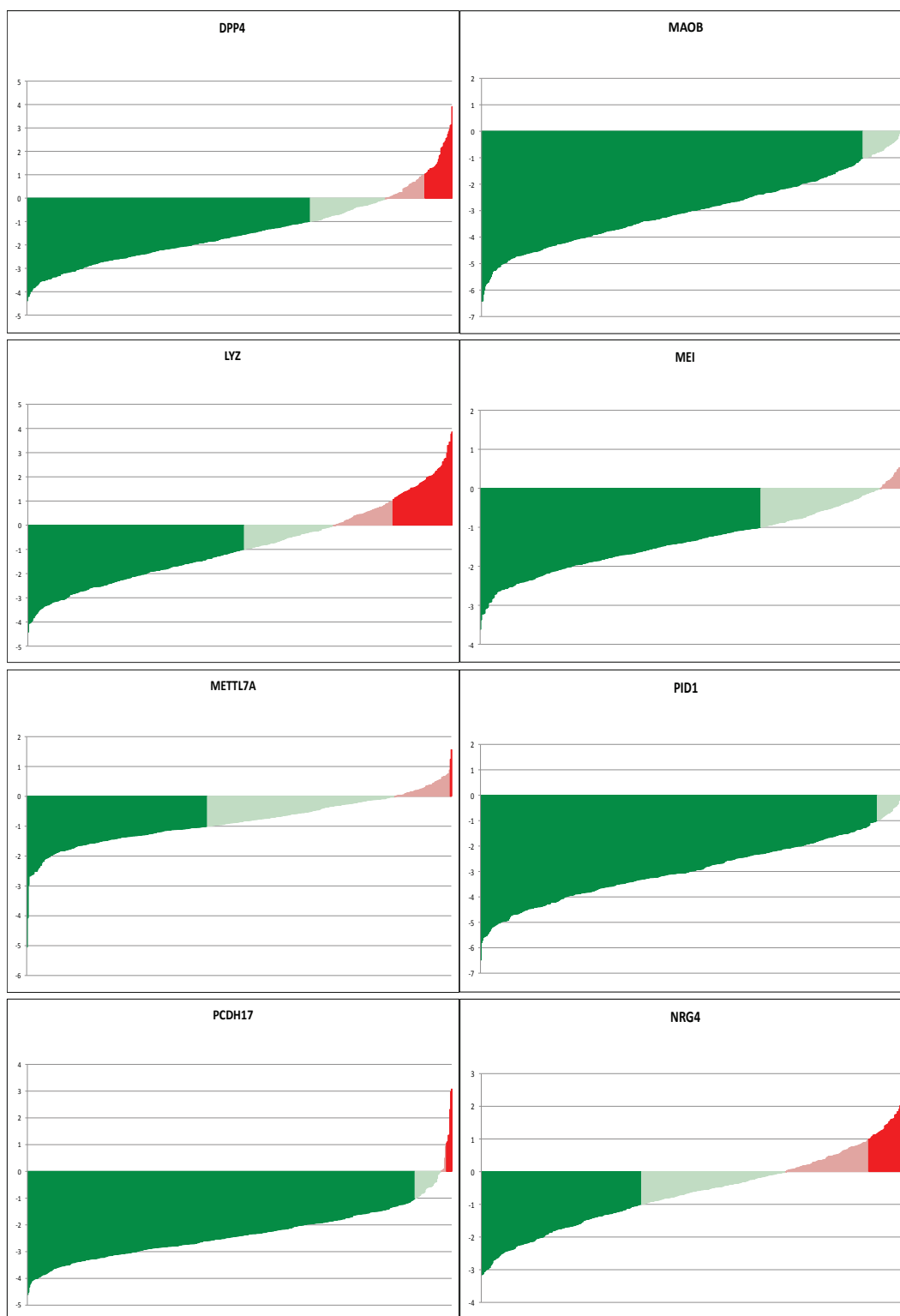


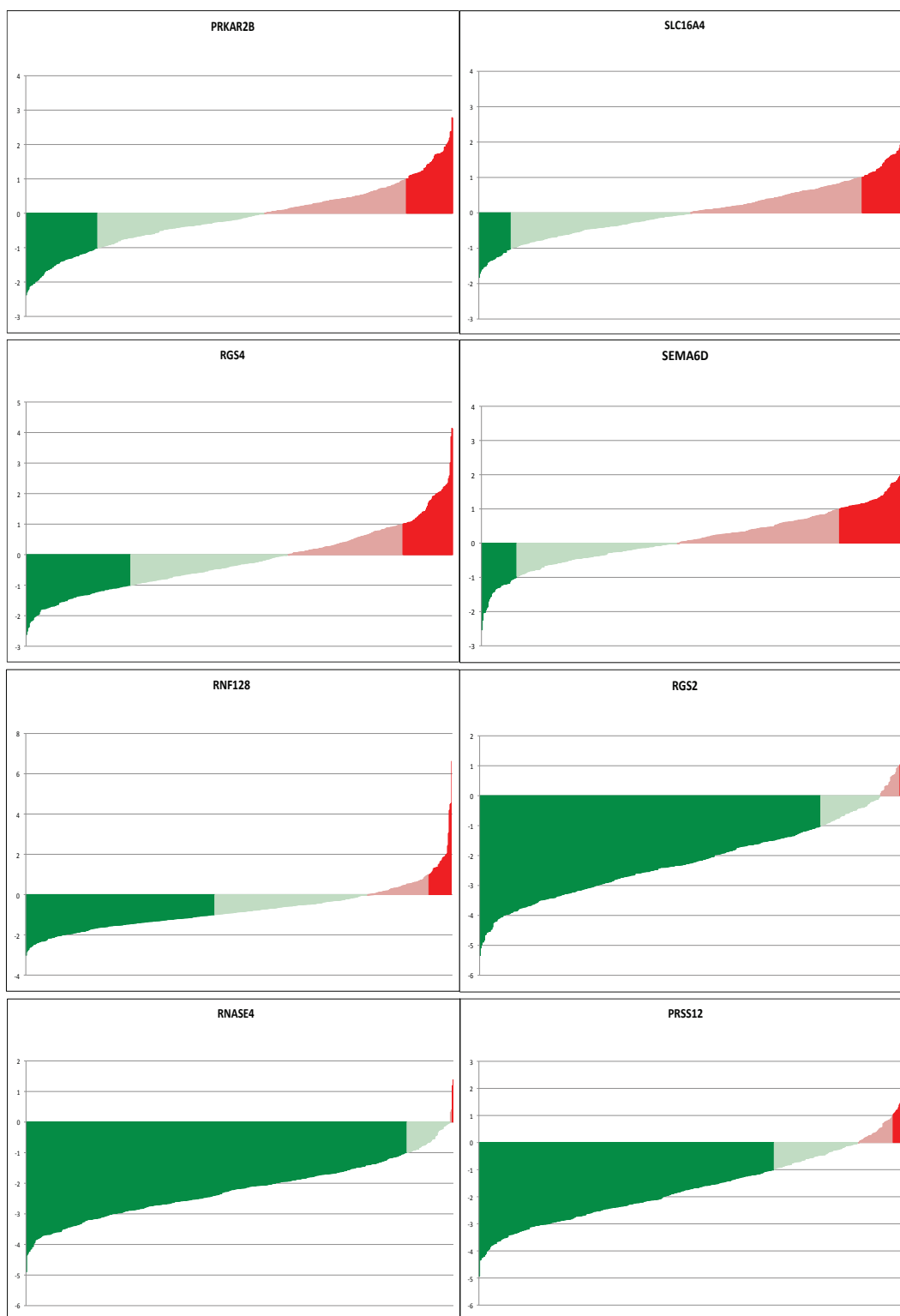
Category 2 Genes

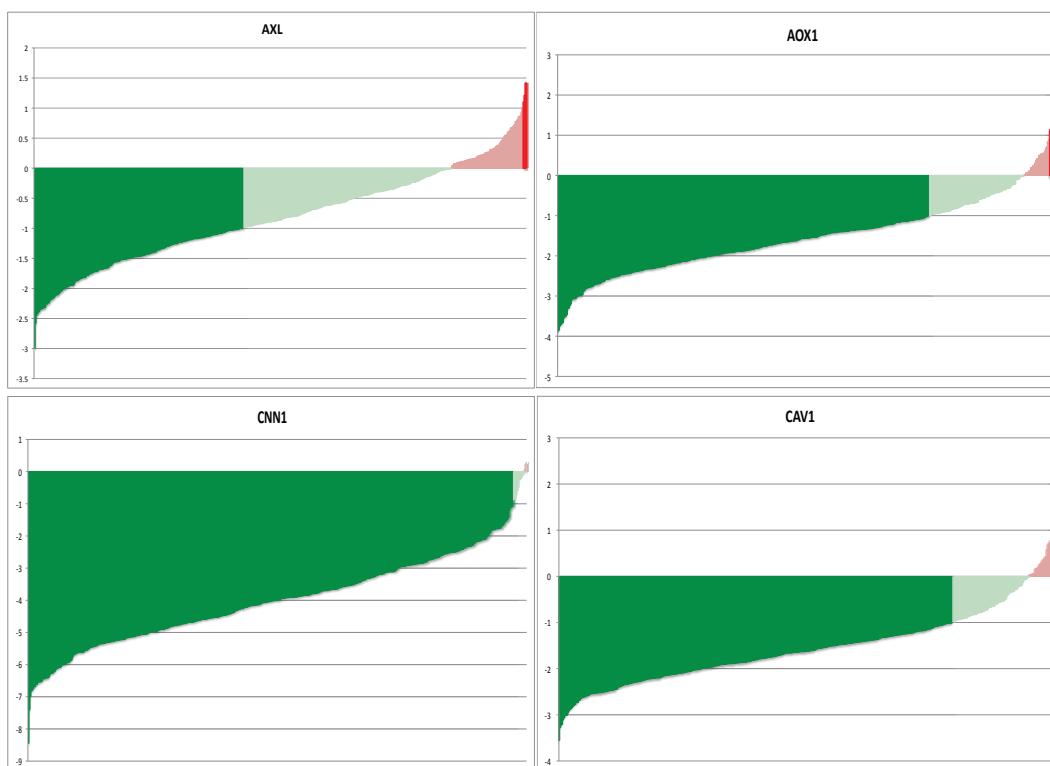
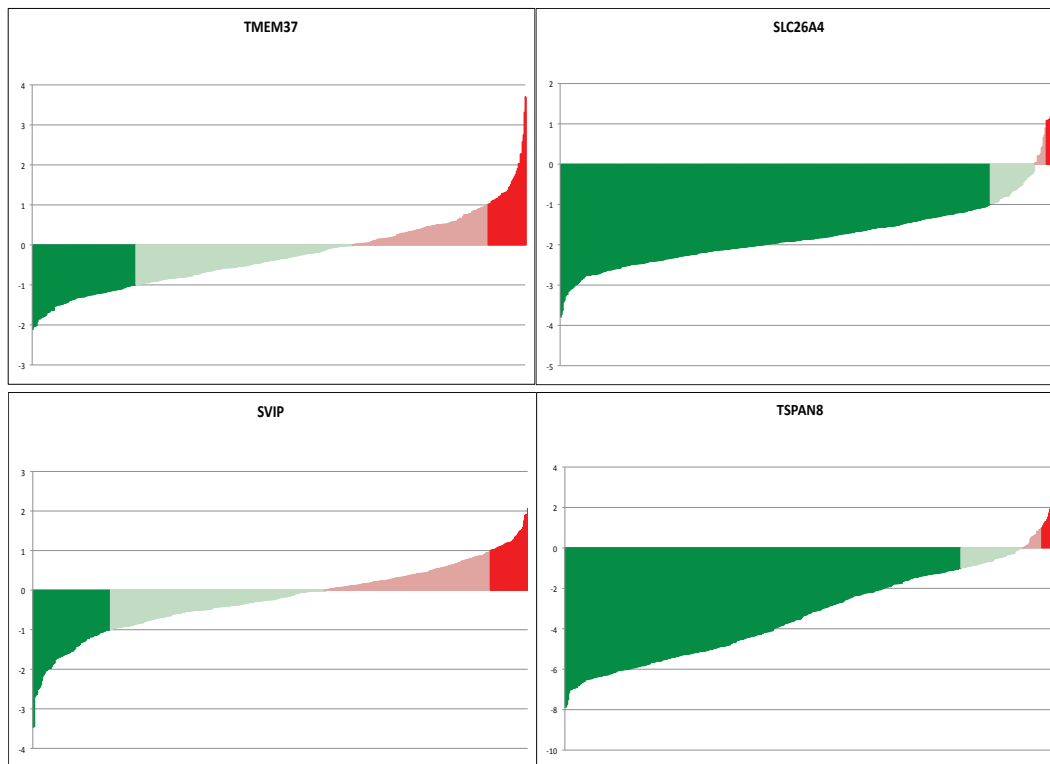


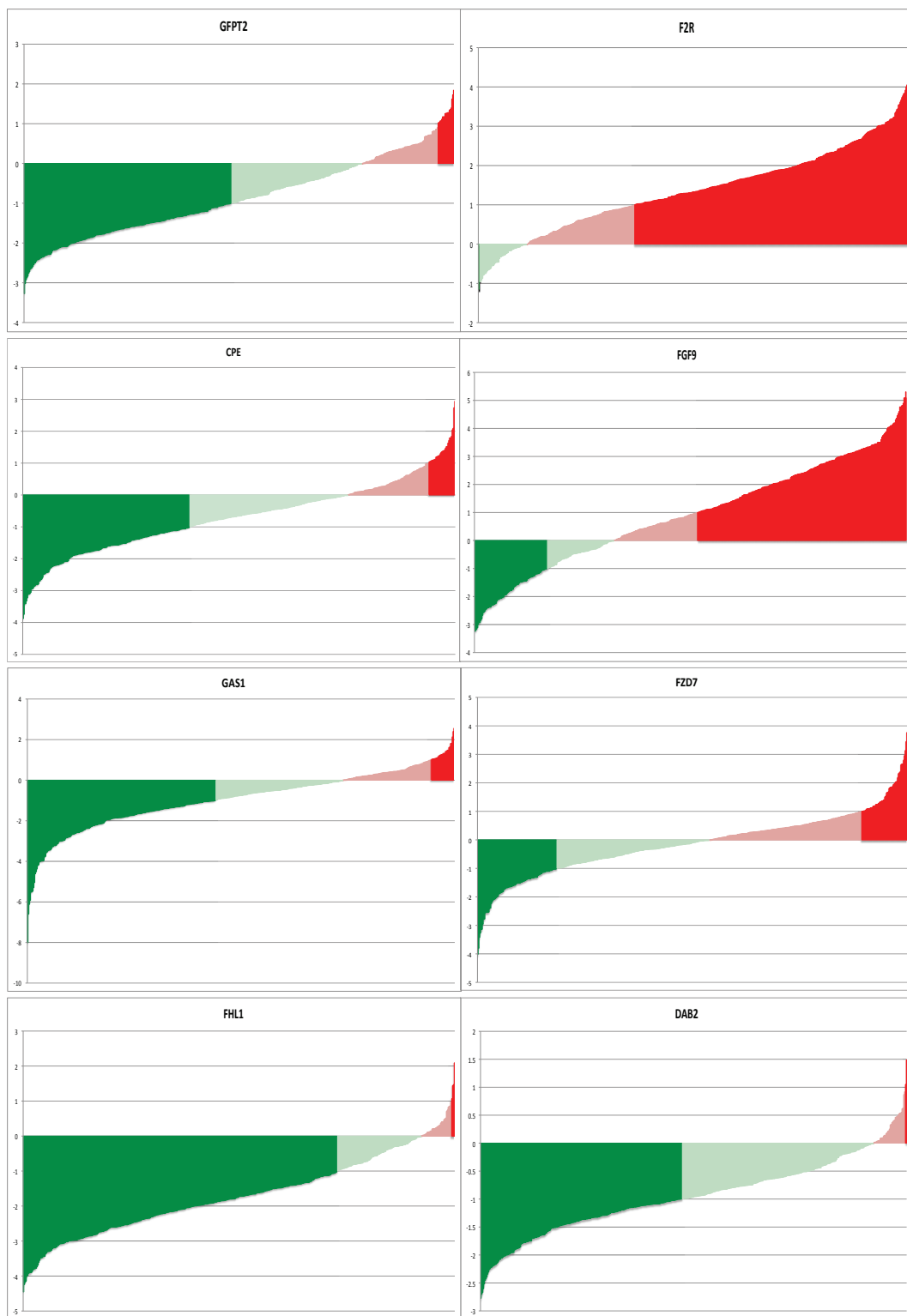


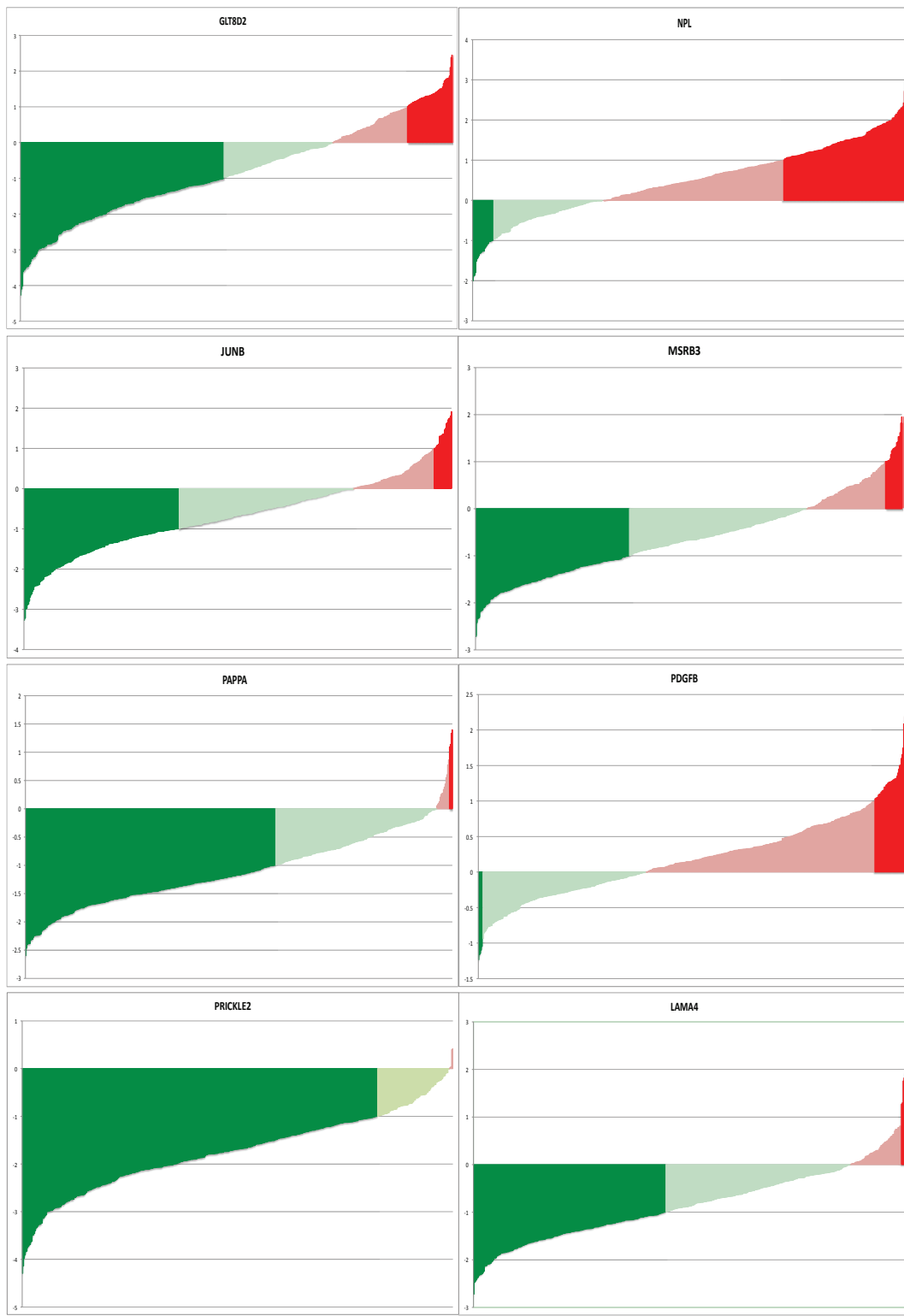


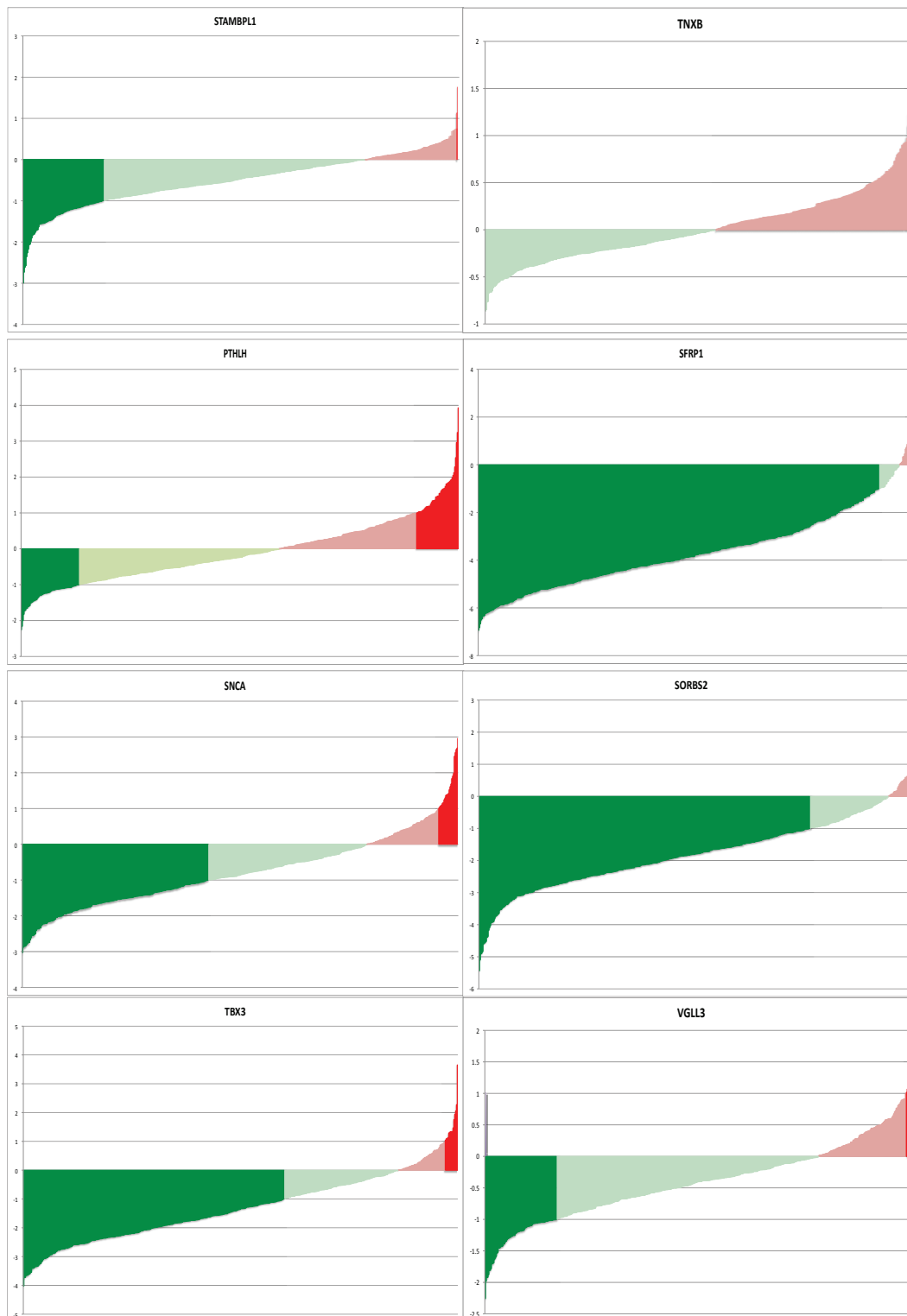


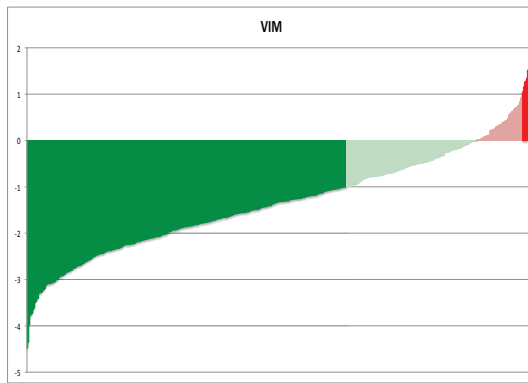












Category 4 Genes

