

DATA MINING THE SEROUS OVARIAN TUMOR TRANSCRIPTOME

by

Timothy Lee Tickle

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Bioinformatics and Computational Biology

Charlotte

2011

Approved by:

Dr. M. Taghi Mostafavi

Dr. Christine Richardson

Dr. Susan Sell

Dr. David L. Tait

Dr. Jennifer Weller

© 2011
Timothy Lee Tickle
ALL RIGHTS RESERVED

ABSTRACT

TIMOTHY LEE TICKLE. Data mining the serous ovarian tumor transcriptome.
(Under the direction of DR. M. TAGHI MOSTAFAVI AND DR. JENNIFER
WALSH WELLER)

Ovarian cancer is the most lethal gynecologic cancer in the United States. If caught in early stages, patient survival rate is 94%, late stage survival rates drop to 28%. It is because most cases are caught in late stages that high mortality is seen. Correct diagnosis is dependent on the presence of symptoms: ~90% of diagnosed ovarian cancers are symptomatic. These symptoms tend to be unfocused and not acute. The goal of this project is to develop a transcript-level data set measuring ovarian tumor expression and associated paracrine signaling for later biomarker research. To this end, laser capture microdissection was used with exon based oligonucleotide arrays to measure the transcriptome of benign and malignant (Type II) serous ovarian surface epithelial-stromal tumors. In addition to profiling tumor, surrounding stromal tissue expression was measured to examine potential paracrine signaling. In total, ~270 million measurements were performed using 50 microarrays. An initial analysis was performed to measure quality, and to compare our measurements against known ovarian cancer properties as established in the molecular genetics literature. Using ontological annotation and *de novo* pathway generation methods, major trends were defined in the data set including the following: apical surface and tight junction activity, mitotic activity, tumor suppression in benign tumors, epithelial-mesenchymal transitioning, known ovarian tumor oncogene activity, and evidence of paracrine signaling. A list of differentially expressed transcripts was defined which may be explored as biomarkers. The potential for meaningful future analysis is diverse. This data set will contribute to the capacity of the cancer genetics community to perform high resolution exploration of serous ovarian epithelial-stromal surface tumors, aiding in developing better diagnostics and therapeutics.

DEDICATION

This is dedicated to all the lives touched by cancer and those who fight for their cures.

ACKNOWLEDGEMENTS

I would like to thank the Blumenthal Cancer Center, the Department of Gynecologic Oncology, the Sonya Hanko Wyatt Ovarian Cancer Research Endowment at Carolinas Medical Center, the Hemby Cancer Research Endowment, and the Health Services Foundation (Carolinas HealthCare System) for their continued support and passion for ovarian cancer research. Thank you to the Microarray Core Facility at the Cannon Research Center (Carolinas Medical Center) for hybridizing the exon samples and for access to Partek[®] Genomics Suite[™]. Thank you to the Bioinformatics Research Center at the North Carolina Research Campus for access to software and expertise. Thank you to the GAANN fellowship program, for their support and interest in my academic experience. Thank you to the College of Computing and Informatics and the Department of Bioinformatics and Genomics for their guidance and untold support in navigating my PhD candidacy.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xii
CHAPTER 1: INTRODUCTION	1
1.1 Ovarian Tumor Clinicopathologic Features	1
1.2 Ovarian Tumor Histology	3
1.3 Serous Ovarian Surface Epithelial-Stromal Tumorigenesis	5
1.4 Tumors and the Surrounding Stroma	8
1.5 Ovarian Tumor Epidemiology	10
1.6 Ovarian Cancer Biomarkers	12
1.7 Alternative Splicing and Tumorigenesis	14
1.8 Microarray Technology	15
1.9 Prior Studies	18
1.10 Experimental Design	21
1.11 Summary	22
CHAPTER 2: MATERIALS AND METHODS	24
2.1 Rationale for the Key Features of the Wet Lab Methods	24
2.1.1 Laser Capture Microdissection	24
2.1.2 Microdissection Schema	25
2.2 Wet Lab Methods	25
2.2.1 Patient Sample Harvesting	27
2.2.2 Hemotoxylin and Eosin staining (Histopathology Confirmation)	28
2.2.3 HistoGene Staining	29
2.2.4 Laser Capture Microdissection	29
2.2.5 RNA Extraction	29
2.2.6 Tissue Scraped RNA Extraction	30
2.2.7 RNA Purification	30

2.2.8	RNA Integrity Measurements	31
2.2.9	Complementary DNA Generation	31
2.2.10	SPIA Amplification	31
2.2.11	Nanodrop Amplified Complementary DNA Quantification	32
2.2.12	ST-Complementary DNA Conversion	32
2.2.13	Fragmentation and Labeling	33
2.2.14	Exon Microarray Hybridization Recipe	33
2.3	Analysis Pipeline	33
2.3.1	<i>In Silico</i> Quality Control	33
2.3.2	Initial Data Generation	34
2.3.3	Inferential Statistics: Gene-level	35
2.3.4	Inferential Statistics: Transcript-level	37
2.3.5	Biological Interpretation	38
CHAPTER 3: RESULTS		42
3.1	Study Sample Details	42
3.2	Laser Capture Microdissection	45
3.3	<i>In Silico</i> Quality Control	46
3.3.1	Rudimentary Microarray Processing Quality Control	46
3.3.2	73 Samples	49
3.3.3	Outlier Experimental Samples	53
3.3.4	Sample Factors: Random and Experimental	53
3.4	Gene-level Analysis	59
3.5	Transcript-level Analysis	62
3.6	Biological Contextualization	65
3.6.1	Chromosomal Location	65
3.6.2	Gene Ontology Analysis	68
3.6.3	Pathway Analysis	68

	viii
3.6.4 Gene Interaction Analysis	69
CHAPTER 4: DISCUSSION	74
4.1 Notable Comparisons	74
4.2 Control Sample Evaluations	74
4.3 Ontology Enrichment	75
4.3.1 Cell Junctions	77
4.3.2 Spindle and Condensed Chromosomes	87
4.3.3 Terms Related to Cell Projection	88
4.3.4 Terms Related to Plasma Membranes	91
4.3.5 Cytoskeleton	101
4.3.6 Extracellular Region	101
4.4 Major Trends in Ontologies	105
4.5 NrCAM Expression	108
4.6 Chromosomal Location	109
4.7 Deriving a Gene Network	109
4.8 Future Directions	113
BIBLIOGRAPHY	115
APPENDIX A: LCM IMAGES	136
APPENDIX B: TRANSCRIPT / TRANSCRIPT CLUSTER LISTS	147
APPENDIX C: BIOLOGICAL CONTEXT RESULTS DETAIL	168

LIST OF TABLES

TABLE 2.1: Planned <i>a posteriori</i> contrasts performed after the ANOVA.	37
TABLE 3.1: Sampling model design.	43
TABLE 3.2: 75 samples counts detail.	44
TABLE 3.3: Basic patient demographic and pathology details.	44
TABLE 3.4: Experiment sample count after all levels of quality control.	54
TABLE 3.5: <i>A posteriori</i> contrasts (transcript cluster level).	61
TABLE 3.6: Transcript level <i>a posteriori</i> comparisons and contrasts.	64
TABLE 3.7: Summary of benign biological annotation.	66
TABLE 3.8: Summary of malignant biological annotation.	67
TABLE 4.1: Summary of unique Cellular Component term counts.	76
TABLE 4.2: Summary of transcripts related to cell junctions (1 of 3).	84
TABLE 4.3: (continued).	85
TABLE 4.4: (continued).	86
TABLE 4.5: Summary of transcripts related to spindles.	89
TABLE 4.6: Summary of transcripts related to cell projections.	92
TABLE 4.7: Summary of transcripts related to plasma membrane (1 of 6).	95
TABLE 4.8: (continued).	96
TABLE 4.9: (continued).	97
TABLE 4.10: (continued).	98
TABLE 4.11: (continued).	99
TABLE 4.12: (continued).	100
TABLE 4.13: Summary of transcripts related to the cytoskeleton (1 of 2).	102
TABLE 4.14: (continued).	103
TABLE 4.15: Summary of transcripts related to the extracellular region.	106
TABLE B.1: Complete benign tumor transcript list (1 of 4).	147
TABLE B.2: (continued).	148

TABLE B.3: (continued).	149
TABLE B.4: (continued).	150
TABLE B.5: Complete benign tumor transcript-cluster list (1 of 3).	151
TABLE B.6: (continued).	152
TABLE B.7: (continued).	153
TABLE B.8: Complete malignant tumor transcript list (1 of 3).	154
TABLE B.9: (continued).	155
TABLE B.10:(continued).	156
TABLE B.11:Complete malignant tumor transcript-cluster list (1 of 3).	157
TABLE B.12:(continued).	158
TABLE B.13:(continued).	159
TABLE B.14:RefSeq numbers and fold-changes for benign T transcripts (1 of 3).	160
TABLE B.15:(continued).	161
TABLE B.16:(continued).	162
TABLE B.17:(continued).	163
TABLE B.18:RefSeq numbers and fold-changes (malignant, tumor) (1 of 3).	164
TABLE B.19:(continued).	165
TABLE B.20:(continued).	166
TABLE B.21:RefSeq numbers and fold-changes for the benign TS transcript.	167
TABLE C.1: DAVID results for the benign transcript lists (1 of 4).	168
TABLE C.2: (continued).	169
TABLE C.3: (continued).	170
TABLE C.4: (continued).	171
TABLE C.5: DAVID results for the malignant transcript lists (1 of 6).	172
TABLE C.6: (continued).	173
TABLE C.7: (continued).	174
TABLE C.8: (continued).	175

TABLE C.9: (continued).	176
TABLE C.10:(continued).	177
TABLE C.11:DAVID results for the benign transcript cluster lists (1 of 2).	178
TABLE C.12:(continued).	179
TABLE C.13:DAVID results for the malignant transcript cluster lists (1 of 6).	180
TABLE C.14:(continued).	181
TABLE C.15:(continued).	182
TABLE C.16:(continued).	183
TABLE C.17:(continued).	184
TABLE C.18:(continued).	185

LIST OF FIGURES

FIGURE 1.1:	Human Exon 1.0 ST microarray chip synthesis.	19
FIGURE 2.1:	Microdissection schema of benign sample.	26
FIGURE 2.2:	Microdissection schema of malignant sample.	27
FIGURE 2.3:	Flow chart of wet lab protocols used in sample preparation.	40
FIGURE 2.4:	Sample preparation protocol quality checks.	41
FIGURE 3.1:	Percentages of sample pairing schemes.	43
FIGURE 3.2:	An image progression, capturing normal epithelium.	46
FIGURE 3.3:	Hemotoxylin and Eosin staining of sample B05.	47
FIGURE 3.4:	An image progression, capturing tumor and adjacent stroma.	47
FIGURE 3.5:	Box plots of the intensities of each microarray.	48
FIGURE 3.6:	The mean signal beyond noise measurements.	49
FIGURE 3.7:	All samples, visualized with PCA.	50
FIGURE 3.8:	Principle components analysis of resampled failed samples.	51
FIGURE 3.9:	PCA of samples with their representative technical replicates.	52
FIGURE 3.10:	PCA of failed and passed replicates.	52
FIGURE 3.11:	PCA of experimental samples.	54
FIGURE 3.12:	PCA of experimental samples by tissue type and cohort.	55
FIGURE 3.13:	PCA of experimental samples visualized by clinician.	57
FIGURE 3.14:	PCA of experimental samples visualized by sample preparation.	57
FIGURE 3.15:	PCA of experimental samples visualized by tissue of origin.	58
FIGURE 3.16:	Comparison of transcript and transcript cluster list elements.	60
FIGURE 3.17:	Comparison of increasing and decreasing transcript cluster lists.	60
FIGURE 3.18:	Comparison of increasing and decreasing transcript list elements.	63
FIGURE 3.19:	Comparison of transcript Vs transcript cluster list element.	63
FIGURE 3.20:	Benign transcript list interaction network (STRING).	70
FIGURE 3.21:	Malignant transcript list interaction network (STRING).	71

FIGURE 3.22: Benign transcript list interaction network (Gene2Network).	72
FIGURE 3.23: Malignant transcript list interaction network (Gene2Network).	73
FIGURE 4.1: GO terms associated with the benign transcript lists.	79
FIGURE 4.2: GO terms associated with the malignant transcript lists.	80
FIGURE 4.3: Predicted benign protein interaction network.	111
FIGURE 4.4: Predicted malignant protein interaction network.	112
FIGURE A.1: Image progression, capturing normal epithelium (sample M01).	136
FIGURE A.2: Image progression, capturing stromal tissue (sample M08).	137
FIGURE A.3: Image progression, capturing stromal tissue (sample B07).	137
FIGURE A.4: Image progression, capturing tumor (sample B16).	137
FIGURE A.5: Images of H&E staining for tissue samples B16 and B22.	138
FIGURE A.6: Image progression, capturing stromal tissue (sample M09).	138
FIGURE A.7: Image progression, capturing tumor (sample M16).	138
FIGURE A.8: Image progression, capturing normal epithelium (sample M03).	139
FIGURE A.9: Image progression, capturing stromal tissue (sample M10).	139
FIGURE A.10: Image progression, capturing sample M17 (T) and M25 (S).	140
FIGURE A.11: Image progression, capturing stromal tissue (sample M11).	140
FIGURE A.12: Image progression, capturing tumor tissue (sample M18).	140
FIGURE A.13: Image progression, capturing stromal tissue (M26).	141
FIGURE A.14: Image progression, capturing normal epithelium (sample B03).	141
FIGURE A.15: Image progression, capturing stromal tissue (sample M12).	141
FIGURE A.16: Image progression, capturing stromal tissue (sample B09).	142
FIGURE A.17: Image progression, capturing normal epithelium (sample B04).	142
FIGURE A.18: Image progression, capturing stromal tissue (sample B10).	143
FIGURE A.19: Image progression, capturing samples B17 (T) and B23 (S).	143
FIGURE A.20: Image progression, capturing normal epithelium (sample M05)	144
FIGURE A.21: Image progression, capturing stromal tissue (sample B11).	144

FIGURE A.22: Image progression, capturing normal epithelium (sample B05).	144
FIGURE A.23: Image progression, capturing stromal tissue (sample B12).	144
FIGURE A.24: Image progression, capturing normal epithelium (sample M06).	145
FIGURE A.25: Image progression, capturing stromal tissue (sample M13).	145
FIGURE A.26: Image progression, capturing samples M20 (T) and M28 (S).	145
FIGURE A.27: Image progression, capturing stromal tissue (sample B13).	146
FIGURE A.28: Image progression, capturing tumor tissue (sample M21).	146

CHAPTER 1: INTRODUCTION

One may expect a cancer derived from a small organ not essential to living to be easily managed. Although removal of the ovary requires invasive surgery and maintaining physiological balance necessitates hormone therapy, both oophorectomies (specifically in conjunction with abdominal hysterectomies) and hormone therapy are common-place medical therapies, reinforcing the notion that this is a non-essential organ. Indeed, when diagnosed as a stage 1 tumor, 5 year survival rates for patients are at 90% [1]. Yet, 20,000 cases were estimated to occur in 2008 with 15,000 associated deaths [1]. Why is ovarian cancer the most lethal gynecologic cancer in the United States?

1.1 Ovarian Tumor Clinicopathologic Features

Ovarian cancer is elusive because the organ is inaccessible and early symptoms are mild, if distinguishable at all. The reality of any cancer is grim; ovarian cancers subtle symptoms and complex nature magnify its lethal potential. Empirical medicine is based first on diagnosis through family history and patient symptoms and only secondarily on diagnostic tests. For this disease, direct observation is not possible, only symptoms remain to suggest further investigation, which is currently comprised of exploratory surgery, carrying its own risks.

Women with ovarian cancer do not necessarily present symptoms, with perhaps 20% of diagnosed ovarian cancers being asymptomatic [2] [3]. Despite this, most asymptomatic ovarian cancers are discovered at an early stage, with the exception of serous cancers. Asymptomatic serous cancers are found to be in late stages in 55% of diagnoses [4]. There is evidence that asymptomatic cancers are found as late stage

not because of lack of detection but because of a separate trend, where late-stage serous carcinomas have a shorter symptom duration than early-stage cancers [4].

Although deemed a “silent killer”, ovarian cancer usually does present symptoms, but these are often so generic that they go unrecognized [5] [6] [7]. One retrospective study of 200 ovarian cancer patients indicated that 90% of all cases present symptoms, and 100% of advanced stages present one or more symptoms [8]. The National Cancer Institute (NCI) booklet on epithelial ovarian cancer defines common symptoms as “pressure or pain in the abdomen, pelvis, back, or legs”, “a swollen or bloated abdomen”, “nausea, indigestion, gas, constipation, or diarrhea”, and “feeling very tired” [9]. Although intended to increase awareness of ovarian cancer and prevention, these symptoms are vague, and so often associated with other common illnesses, including bladder infection and irritable bowel syndrome [6] that their impact is doubtful. Patients frequently self-diagnose symptoms as arising from age, menopause, or a benign condition [3], and do not seek any clinical advice at all.

Evaluation of the above symptoms as predictors of diagnosis have shown some to be marginally predictive [10]. The most predictive individual symptoms were found to be unusual bleeding, masses, and “distended and hard abdomens”. When combining these symptoms with the occurrence of pain, the index had a 74% sensitivity and 77% specificity in prediction. More specifically, if stratified by histology, each epithelial tumor histology has specific symptoms [4]. Patients with mucinous tumors most commonly presented swollen abdomen (60%); patients with endometrioid tumors presented bleeding (19%); patients with serous tumors presented abnormal bowel symptoms (47%). All of these differences were significant statistically [4]. Still, it is clear that ovarian cancer symptoms are vague, unspecific, and vary with histological type, lending little help to diagnosis.

A peripheral molecular signature detectable in circulating fluid would be advantageous, and considerable resources have been invested in discovering such markers.

For a small cohort of patients, there appears to be an inherited genetic component; BRCA-linked inheritance has been the most studied. This group is distinguished by the mean age of diagnosis, 54 years rather than 62 years for sporadic cancer. Pathological features of the tumors in those with the inherited predisposition were not found to differ from sporadic tumors [11], but their advanced stage tumors had a lower recurrence (50% less) and a better prognosis ($p=0.004$). This has sometimes been ascribed to the inability of BRCA mutants to repair DNA damage, making post-surgical remnants of tumors more susceptible to chemotherapy [11].

1.2 Ovarian Tumor Histology

According to the World Health Organization, primary ovarian tumors are separated into the following histological subtypes: surface epithelial-stromal tumors, sex cord-stromal tumors, germ cell tumors, miscellaneous tumors, and lymphoid and haematopoietic tumors [12]. The most common histological subtype is surface epithelial-stromal tumors (occurring 90% of the time) [13]. Within this category several subcategories exist, including serous, mucinous, endometrioid, clear cell, and several less common types of tumors. A fascinating characteristic of these tumors is their disparate histological features. Many of these tumors more closely resemble tissues from other parts of the peritoneal cavity than they do normal ovarian surface epithelium. For example, serous carcinomas most closely resemble the fallopian tube epithelium, while endometrioid carcinomas resemble the uterine corpus; mucinous tumor cells exhibit intracellular mucin and resemble endocervix cells, as well as the gastric pylorus and intestine [14].

The differences in histopathological subtypes are mirrored by differences in clinicopathological symptoms, and these extend from the phenotype to the genotype. A study of the expression of 21 protein biomarkers in 500 ovarian carcinomas of a range of subtypes (high-grade serous, clear cell, endometrioid, mucinous, and low-grade serous) showed characteristic signatures for each subtype. The expression level dif-

ferences were more effective in survival analysis predictions when treated as discrete histological subtypes [15]. This heterogeneity raises serious questions about whether the origin and progression of diseases classified as ovarian cancers truly belong in the same disease category [15].

The study described here focused on a single subtype. Serous surface epithelial-stromal tumors were selected, since they are the most commonly observed form (occurring approximately 32% of the time) and are the most deadly (5 year relative survival rate of 42%) [13]. In addition to normal tissue the study included samples from benign (neoplastic) serous tumors, which typically occur as cystadenomas and cyadenofibromas; these may be large unilocular cysts or multilocular cysts surrounding pockets of serous fluid. Malignant tumors occur as adenocarcinomas, adenocarcinofibromas, cystadenocarcinomas, and cystadenocarcinofibromas. Most predominantly, as high grade carcinomas, these tumors have papillary structures, but a minority exist as glandular structures or solid sheets of cells [14]. Malignant tumors may also be composed of poorly differentiated cells [14]. An intermediate group exists, borderline tumors, which resemble benign neoplasms but are a multilayer tissue structure with infrequent evidence of proliferation [16]; they may be a step in progression between neoplastic and malignant tumors since they sometimes exhibit stromal invasion, but they do not necessarily progress to malignancy [14].

Staging is an additional method of classifying serous ovarian tumors, reflecting the location of the tumor as noted during surgery. Stages range from I to IV, with stage I indicating that the tumor was restricted to the ovary, II indicating pelvic extension, III indicating metastasis to the peritoneal cavity beyond the pelvis, and stage IV indicating metastasis outside of the peritoneal cavity [14]. Of the major ovarian surface epithelial-stromal tumors, serous tumors are the only type in which staging is least likely to be reported as stage I, with only 13% of cases reported as belonging to that category, while mucinous, endometrioid, and clear cell tumors are

reported as stage I in 71% , 48% , and 63% of cases, respectively [16].

1.3 Serous Ovarian Surface Epithelial-Stromal Tumorigenesis

The events that set the stage for the initiation of tumorigenesis in ovarian surface epithelial-stromal tumors are currently not known. Several mechanisms have been proposed and are described here. While appearing contradictory at first, additional information has revealed an unexpected common root, and there is reason to believe that the initiating events will soon be understood.

Early evidence suggested that ovarian surface epithelial-stromal tumors originated from the ovarian surface epithelium (hence the name). This came from studies in the aged domestic hen (*Gallus gallus*, female) [17]. Domestic hens are known to suffer naturally from a high incidence of ovarian cancer [18], leading to the addition of restricted ovulator chickens (and mutant types) to the short list of approved model animals for ovarian tumor research [19].

A plausible contributing factor for the susceptibility to ovarian cancers of aged domestic hens is their state of near constant ovulation. Ovulation is the production of a cell that forms the basis of a new life, but ironically, the ovum itself arises in an act of destruction. Follicles develop in the ovarian stroma, which lies directly underneath the ovarian surface epithelium. Yet, unlike many other organs that develop biological products, there are no ducts or vessels for the secretion of the ovules. Each productive cycle of ovulation ruptures the ovary, releasing the ovule to travel into the fallopian tubes and on to the uterus. This breaking free damages the ovarian surface causing a wound that must heal. The damage from ovulation is dramatically clear when comparing the ovaries of younger and older individuals: young females exhibit a relatively smooth ovarian surface, while ovaries of mature females have a heavily scarred surface. This irregular surface lends itself to the development of invaginations, which may either be consumed by the ovary or persist as cysts. In this “incessant ovulation” model, the cysts are required to initiate the tumorigene-

sis process. The incessant ovulation model does not explain the similarity between some serous ovarian surface epithelial-stromal tumors with the fallopian tube epithelium. Some explanation is required here: while the ovarian surface epithelium is continuous with the mesothelial lining of the peritoneal surfaces, and shares many histological features with those cells, ovarian epithelial carcinomas themselves are not mesothelioma-like. Instead they are Müllerian-like, meaning that they resemble the organs of the Müllerian system, which includes the fallopian tubes, uterus, cervix and the upper portion of the vagina, but not the ovary. The simplest explanation for this observation is that serous ovarian surface epithelium-stromal tumors are secondary site tumors, originating from epithelial cells in the Müllerian system [20]. This is currently a minority view, with most researchers preferring a model in which the ovarian surface epithelium undergoes a metaplastic transformation during the initial steps of tumorigenesis, causing it to resemble Müllerian-derived cells [20]. If a metaplastic transformation occurs one would expect to find precursor epithelial lesions, but no unambiguous examples have been found [21] [22]; thus a *de novo* event is sought as the source of this etiology [23].

With respect to the minority view that these tumors are secondary lesions whose source is the fallopian tubes, several arguments are advanced. One is that early studies did not consider the fallopian tubes, so the absence of evidence is not evidence of absence [23]. With a rising interest in the secondary-Müllerian system, fallopian tubes were inspected for associated precursor lesions and tumors [24] [25] [26] [27]. These studies found tubal carcinomas and similarities with ovarian carcinomas were noted, although a limitation of the studies is that they focused on individuals with BRCA mutations [28]. Specifically, when microscopically examining serial dissections of fallopian tubes, 60%-70% of ovarian and peritoneal high grade serous carcinomas showed evidence of tubal involvement [29] [30]. A study examining gene expression using microarrays showed a greater similarity between high-grade serous carcinomas

and fallopian tubes than to the ovarian surface epithelium [31].

Given this dichotomy in models for etiology, some investigators now divide ovarian epithelial serous carcinomas according to molecular profile and speed of progression [32], rather than stage or histopathology type. This avoids any presumptions about cause that might preclude considering associations of potential merit. For example, in one study, carcinomas were assigned, across histological subtypes, to one of two classes. Type I carcinomas were defined as low-grade serous carcinomas, mucinous carcinomas, endometrioid carcinomas, clear cell carcinomas, and malignant Brenner tumors (which show urothelial differentiation). Within this class, serous carcinomas predominantly exhibit KRAS and BRAF mutations and follow the classical tumor progression from benign adenomas/adenofibromas through proliferative tumors and noninvasive carcinomas to invasive low-grade carcinomas [32]. Type II tumors were defined to contain high-grade serous carcinomas, undifferentiated carcinomas, and malignant mixed mesodermal tumors or carcinosarcomas. Within this second class, serous tumors exhibit p53 mutations and HLA-G expression [32]. Tumors in this second class did not, however, have a known initiating event, and so were classified as “*de novo*” [33] [34]. There is considerable effort being invested in understanding the ramifications of previous studies of high-grade serous carcinomas if they are now cast as secondary tumors [35]. One proposal is that tubal intraepithelial carcinomas are shed from the fimbrial epithelium of the oviduct and then implant on the ovary surface, where molecular changes that are affected by the new environment continue an adapted tumorigenesis process (much like retrograde menstruation can initiate endometriosis) [23]. It is even suggested that this could be the initiation mechanism for type II serous tumors, with the outcome depending on the first molecular changes [23]. However there is as little hard evidence for this, as for the initial steps in cortical cyst transformation, so experts remain uncommitted to either at this time [23].

Two schools of thought now surround the pathogenesis of serous ovarian surface

epithelial-stromal tumors. The initiation and tumorigenesis of type I serous tumors is less controversial, particularly as molecular events and intermediate forms can be observed. However, the initiation and progression of type II serous tumors is subject to much debate; Nelly Auersperg recently suggested a theory that may resolve some of the disagreement [36]. Her argument is that, given that the fimbrial and the ovarian surface epithelium both develop from the embryonic coelomic epithelium, the epithelia of the ovaries and fallopian tube ampullae may not be fully developmentally committed, and thus are susceptible to neoplasia. Auersperg has used epithelial and mesenchymal cell molecular markers to show the existence of an overlapping region of ovarian surface and fimbrial epithelium. This region is classified as a “transitional epithelium of common origin”, giving advocates of both the ovarian surface epithelial-derived genesis and the secondary-Müllerian fimbrial-derived genesis common grounds for their evidence for their respective tumorigenesis models [36].

1.4 Tumors and the Surrounding Stroma

Paracrine signaling (signaling between cells) is an important part of cellular functionality that has been, thus far, largely ignored in the tumor research community [37]. However, cells communicate with and affect one another, performing as units in a complex network; evidence exists that this occurs in tumors [38]. Many studies begin by dissecting cells of interest from their environment, in order to enhance signals specific to those cell types. Other studies macrodissect cells of interest and stroma elements together, confounding signals in down-stream analysis. We were interested in measuring not only expression in tumorigenic cells but surrounding stroma cells which may exhibit associated paracrine signaling. Thus, we collected cells from the surrounding stromal environment of ovarian serous carcinomas.

While a relative recent focus, several studies have looked at characteristics of stroma surrounding endothelial cells, pericytes, leukocytes, and fibroblasts. Carcinoma-associated fibroblasts (CAFs) are the most frequently encountered element of tumor

exposed stroma [39] [40], and may contribute to what are considered the “hallmarks of cancer” [38] based on their: self-sufficiency in growth signals, evasion of apoptosis, sustained angiogenesis, and capacity for tissue invasion/metastasis [37].

In normal environments, cells influence others with mitogenic signaling, sending growth signals to surrounding cells. CAFs express a number of specific mitogenic factors, including hepatocyte growth factor (HGF), fibroblast growth factor (FGF) and Wnt pathway signaling molecules [41]. Several studies in mice show that aberrant signaling in just one factor can induce a tumor, including notch signaling in epidermal keratinocytes and TGF β functionality in fibroblasts [42] [43].

A requirement in all cancers is the suppression of the apoptotic pathway, by which cells suicide. Stroma cells produce survival signals that surrounding cells use to counter the initiation of apoptosis. More specifically, CAFs express insulin-like growth factor-1 and -2 (IGF1/2) [44] [45]. Additionally, increases in collagen in the extracellular matrix (produced by CAFs), have been shown to increase integrin signaling [46].

Angiogenesis, or the growth of a blood supply to nourish tumors, is important to sustain tumor growth. In one study covering a panel of known angiogenic factors, all were more highly expressed in stroma cells than the actual tumor [47], with the exception of VEGF. It was shown in mice that the VEGF promoter was more active in CAFs at the onset of primary site and metastatic tumors [48]. CAFs have been shown to have a role in recruitment of cells for MCF-7-ras human breast carcinoma vascularization [49].

Lastly, with respect to the role of CAFs in metastasis, tumors must invade surrounding tissue: CAFs secrete TGF β and facilitate the epithelial-mesenchymal transition needed for tumor invasion of surrounding cells [41]. CAFs secrete proteases, required to degrade fibers tethering tumor cells in their current environment, permitting mobility [37].

1.5 Ovarian Tumor Epidemiology

Without a well-defined pathogenesis model, uncovering epidemiological factors contributing to steps in the initiation and progression of ovarian tumors is difficult. Some risk factors have been defined but their biological bases are not understood. Protective and risk factors span biological, medical, environmental, and genetic sources.

Few reproductive factors appear to reduce the risk of ovarian surface epithelial-stromal tumors. Parity and pregnancy, being biological functions directly associated with ovaries, have been shown to have some association, with higher parity reducing the risk of ovarian cancer. In one study, one birth gave patients a 40% risk reduction with an additional 14% per additional birth [50]. This protective effect seems to hold for all ovarian surface epithelial-stromal tumors excluding mucinous tumors [51]. Incomplete pregnancies provide a slightly lower protective rate [52] [53]. An associated process, lactation, is marginally preventive, seen most clearly between women who have never breastfeed and those who have breastfed for 18 months or more (2% risk decrease per month) [54].

Oral contraceptives have been found to reduce the risk factor of ovarian cancer onset in almost all studies [55]. Although reduction in risk has been associated with less than 1 year of use, it appears that greater risk reduction occurs with more extensive use. One study reported a 50% reduction in risk after 5 years of use [56], even if use was 15 years previously [57].

If incessant ovulation is a primary risk, it would seem that ovulation-stimulation treatments would carry an increased risk. Indeed, two studies claimed fertility drugs produced an increased risk of developing ovarian tumors [58] [59], but to date only weak, or nonexistent, associations have been found between ovarian cancer and drugs causing superovulation [60] [61] [62]. The need for larger studies with longer follow-up has been expressed to better clarify this potential risk factor [63].

A bilateral oophorectomy is the most extreme preventive factor known, but the

procedure carries serious consequences, including sterility, changes in hormone production and psychological effects [55]. Interestingly, as much as an 80% reduction of risk of ovarian cancer has been shown following tubal ligation alone, carrying forward as much as 20 years post-surgery [64] [65]. A hysterectomy also is preventive although less so than tubal ligation, with a decrease of 33% [66].

Other reproductive system events increase ovarian cancer risks.

Endometriosis, in which uterine endometrial tissue implants outside the uterus (typically through retrograde menstruation), is one such risk factor for ovarian cancer onset. Endometriosis has most specifically been linked to clear cell and endometrioid tumors, with evidence showing significant co-occurrence with ovarian cancers [67] and increased ovarian tumor risk correlating with a history of endometriosis [68]. Research is currently being conducted to understand at what level endometriosis may be a causative factor for ovarian tumors [55].

Several environmental factors have been shown to affect the chance of ovarian tumor occurrence. Use of talcum powder in the genital region can increase risk by 30% [69]. Serous histopathology may specifically have a stronger percent influence [70] [71]. The structural similarity of talcum powder to asbestos may play some role in this risk [72]. Additionally, drinking green and black teas have shown a dose-responsive protective capacity [73]. Obesity in early adulthood correlated to a general increase in ovarian tumor occurrence [74]. Other environmental factors are under study including smoking, alcohol use, and caffeine, but have conflicting bodies of evidence on their relative associated risk [55].

The most significant known risk factor for ovarian tumors is genetic. In particular, BRCA1/2 mutations increase the risk of ovarian tumors, specifically serous tumors. Of the 10% of ovarian tumors which for which a genetic component is implicated, 90% map to BRCA1/2 mutations [55]. An individual with a BRCA1/2 mutation has a 15% to 66% increase in risk of developing an ovarian tumor in her lifetime [75].

1.6 Ovarian Cancer Biomarkers

An ideal diagnostic biomarker has several characteristics: it is stable in biological fluids, easy to purify and detect, always discrete between the conditions to be tested and only between those conditions, occurs in a medium that is safely and readily available, it is produced from the very beginning of the condition and continues to be produced by all offspring cells. Ideally, ovarian cancer cells from the very first transformation would produce a unique and stable protein or nucleic acid that was shed into the blood or urine, and this molecule would never be produced by healthy cells or those with other pathologies. The reality is that cancer seems to be due to dysregulation of several large neighborhoods of genes that are used in healthy cells, often with increasingly chaotic signals over time [38]. Thus an effective test will likely be based on one or more panels of transcripts (or their products) with readout comprising characteristic changes but not on/off phenotypes.

Currently, some cancers have biomarkers which are used in a clinical environment. Several clinical tests, including *BRCAAnalysis*[®], measure for genes with a known association with inherited ovarian and breast cancer (BRCA1/2 mutations). Agendia[®] provides several gene expression products that are FDA-cleared for use in clinical settings. MammaPrint[®] is a 70 gene expression panel which measures risk of metastasis and recurrence in early stage breast cancer [76]. BluePrint[™] is an 80-gene expression product used to classify breast tumors into subtypes of clinical importance [77]. Additionally Pathwork[®] Diagnostics has created a “Tissue of Origin Test” based on a 1,550 gene expression signature that is focused on indicating the primary site of poorly-differentiated and undifferentiated primary tumors, and metastatic tumors in general[20].

Given the successful translation of gene expression panels to diagnostics in breast cancer research, it is reasonable to hope for a similar outcome for correctly identified ovarian tumor biomarkers. Genes with favorable characteristics for inclusion in an

ovarian cancer diagnostic biomarker panel include: CA-125, BRCA1/2, HER2/NEU, AKT2, c-fms, BCL-2, FGF3, MET, p53, SPARC, nm23, and KRAS, some of which are described in more detail below.

Historically, CA-125 (cancer antigen 125 or MUC16) was the first notable ovarian tumor biomarker [78]. CA-125 expression is consistently exhibited in serous carcinoma. When immunohistochemistry was performed on 43 serous papillary ovarian carcinomas 37 samples indicated CA-125 expression [79]. CA-125 levels are currently monitored in clinical settings, but only after onset of ovarian tumors is confirmed. CA-125 changes are not specific enough to indicate tumorigenesis and are also triggered by other changes in the peritoneal cavity, including endometriosis. It has been suggested that screening based on CA-125 levels, ultrasound, or serum protein tests alone do not produce the sensitivity or specificity needed for screening. An approach combining modalities should be investigated [80].

Although made famous for associations with inherited breast cancer, BRCA1 and BRCA2 (Breast Cancer Gene 1 and 2) are also biomarkers for inherited ovarian cancer. BRCA1/2 mutation is also a strong factor in tubal serous epithelial tumors originating in the fallopian tubes [81]. Depending on how type II serous ovarian tumorigenesis is defined to be derived (from the fallopian tube or ovarian surface epithelium), inclusion of BRCA1/2 mutations may be complicated, since BRCA1/2 dysregulation is not specific to ovarian tumors (since it is inherited in breast cancer) [82] [83]. If tubal serous carcinoma is defined as a different tumor type from ovarian serous tumors, two BRCA1/2-expressing tumors may be difficult to distinguish.

When immortalized (non-tumorous) ovarian epithelial cell lines (T29 and T80) were transduced with HER2/NEU and exposed to subcutaneous and intraperitoneal environments, undifferentiated and papillary carcinomas arose. Papillary carcinomas from the T29 line injected intraperitoneally matched human papillary serous carcinoma in immunophenotype and histologically [84]. Although this is not evidence that

singly mutating HER2/NEU would produce carcinomas (particularly given the immortalizing of the cell lines and the apparent need of the peritoneal environment), this is strong evidence of potential activity in human papillary serous ovarian carcinoma tumorigenesis.

Protein 53 (p53) is an important part of a cell's response to damage arising in the cell-cycle, which can affect apoptosis. P53 is expressed in approximately 50% of human tumors [85]. Proliferic expression in cancer has made p53 a well studied tumor suppressor. Specific to the peritoneal cavity, p53 is known to express in tumors of the ovary, the fallopian tube, and the peritoneum [86] [87] [88]. Such ubiquitous expression in tumors is useful in developing markers for therapeutic success[89] but not for specific diagnostic markers. That is, if p53 products can be detected peripherally, p53 will not be a marker of a specific tumor type but, coupled with ovarian specific biomarkers, may increase sensitivity to a tumor positive condition.

Ras, a small GTPase superfamily member, exhibits mutations in many tumors. In relation to serous ovarian tumors, K-Ras mutations are found in >70% of low malignant potential serous ovarian tumors but only in 12.5% of serous ovarian carcinomas in general [90]. When serous carcinomas are defined as type I and type II, KRAS mutations are expressed in type I but not type II carcinoma [91]. Patients with type I carcinoma have a better prognosis than type II, KRAS and genes with similar expression could be candidates on a biomarker panel targeted at defining (better) prognosis.

1.7 Alternative Splicing and Tumorigenesis

Alternative splicing is the process where multiple mature mRNA transcripts can be derived from one pre-mRNA transcript. Alternative splicing is estimated to occur in 70% of genes (referred to as transcript-clusters) [92]. Alternative splicing has been associated with several diseases including: Frasier syndrome, frontotemporal dementia, Parkinsonism, atypical cystic fibrosis, retinitis pigmentosa, and spinal muscular

atrophy [93]. Given the increased potential for functional transcription, cancer has been documented as associated with such splicing. Many transcript-clusters recognized to be associated with cancer exhibit alternative splicing including MUC1 [94], VEGF [95], p53 [96], and BRCA1 [97]. Alternative splicing was measured in a study of mixed ovarian surface epithelial-stromal tumors using a high-throughput RT-PCR system (LISA) on 600 transcript clusters [98]. Carcinoma and normal samples could be separated by the expression of 46 transcript isoforms under unsupervised clustering. Affymetrix Human Exon 1.0 ST microarrays were also shown to be effective in measuring transcript level expression and alternative splicing in cancer (colon) [99]. The success of the colon cancer study, depending on the density of oligonucleotide microarrays that can measure differential expression, our ovarian tumors, were measured with these Affymetrix Human Exon 1.0 ST microarrays. If successful, this could document exact transcript association with tumors without the confounding effects of transcript-cluster level analysis.

1.8 Microarray Technology

The seminal manuscript “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid” by Watson and Crick concluded by alluding to the complementary nature of DNA strands [100]. As our understanding of nucleotide polymers has matured, modern molecular biology has defined a multitude of techniques exploiting the complimentary nature of double-strand nucleic acid bonding. Hybridization assays go back to the “blotting” techniques of Southern [101], but have been refined by miniaturization and advances in DNA synthesis technology to produce highly parallel platforms called microarrays or chips [102]. Oligonucleotide microarray technology combines sequences known to occur in target genomes with circuit fabrication techniques to make a device that allows one to interrogate, or probe, nucleic acids characteristic of a genome. The target pool that comes from the sample may either be the genomic DNA, which is amplified and labeled for hybridization for a genetic

screen, or the mRNA, which is first transformed to cDNA and then amplified and labeled for hybridization in an expression screen [103].

Commercial arrays vary by oligonucleotide density and length. Those used in most biomedical studies are from the manufacturing company Affymetrix, which lays down 25-mer oligonucleotides that are complementary to targeted regions of the model genome. Our interest was in the class of chips that reports on all defined exons in the human genome.

Specifically, Human Exon 1.0 ST microarrays (Affymetrix, Inc.) were used to measure the expression of genes in samples from patients with benign and malignant serous ovarian surface epithelial-stromal tumors. These microarrays are designed to have multiple (ideally 4) probes for every exon (probes are aggregated to give a ProbeSet defining the exon) belonging to a transcript cluster associated with a gene. Approximately 5.4 million probes are employed in exon microarrays to interrogate approximately 1 million exons [104]. These exons define 300,000 different transcript clusters (90,000 express multiple transcripts). Almost 800,000 probes sets of the 1.4 million ProbeSets are supported by evidence from EST and GENSCAN sources [104]. However, for a large number of computationally predicted genes, no evidence for expression has ever been found. It may be that the correct conditions for eliciting a response has never been imposed, or these may be pseudogenes, whose promoters and other regulatory elements no longer function. Similarly, detailed work on the gene models that describe intron and exon boundaries and alternative splice forms is only slowly increasing, so some of the probes may respond in unexpected ways. The manufacture process is described briefly below, with emphasis on factors that affect analysis of the signal.

The oligonucleotide probe design process is key to all subsequent steps. The confidence that the Affymetrix array designers consider that an analyst can place in a probe response is provided in the annotation files. Each ProbeSet is given an

“evidence level”: ProbeSets with the strongest support are considered “Core”, the biological contextualization and design of these ProbeSets are supported by Refseq evidence and full-length mRNA GenBank entries (with full CDS data) [105]. The next level of evidence is defined as “Extended”: these are supported by other biological proof such as GenBank mRNA entries not associated with full-length CDS data, ESTs, microRNA registry entries, vegaGene and vegaPseudoGene entries, as well as others. Next, “Full” ProbeSets are supported by *in silico* gene prediction (algorithms encompassed in GeneScan, GenScanSubOptimal, exoniphy, RNAGene, sgpGene, and TwinScan). Finally, “Free” ProbeSets are composed of probes which, when combined into ProbeSets, do not have annotations that comment on the full ProbeSet, and “Ambiguous” ProbeSets can be assigned to multiple transcript clusters [105] and thus signal from these ProbeSets live up to their label.

Once probes have been designed, the physical molecules are created using photolithography. An inert surface, typically a glass or silicon wafer, is coated with an ultraviolet light reactive chemical. A series of masks directs the building of the probes, one base with UV-sensitive reactive groups at a time. Masks have small holes allowing the UV light to reach regions to be activated so that the next base can be added. When complete, the inert surface has chains of nucleotides of 25 base in a length which have an explicit design based on the targets of the chip (Figure 1.1 left). The chemistry used is not highly efficient, so a large number of incomplete probes surround each full-length probe. Hybridization conditions are recommended that limit non-specific binding to these shorter probes, but false signals are always a risk.

Once chips have been created, patient samples can be prepared for hybridization on GeneChips. Sample preparation methodologies are tuned to the molecule type of interest and aim of the project. Although many parts of the sample preparation have been commercialized as kits, this step is known to introduce a great deal of variation

into the final measurements. In gene expression experiments the common threads are: isolation of RNA from the biological material, cDNA conversion, reamplification of the strand of interest, fragmentation of the amplification product (50 - 100 bases in this study), and attachment of a biotin-based label [103]. Many quality control measures are associated with the process.

Fragmented and labeled cDNA is combined with a hybridization solution that is added to GeneChips so that probes are exposed to the targets in the solution. Under correct conditions, only target that is fully complementary to a probe will bind stably, an association persisting through all subsequent processing steps (Figure 1.1 middle). After 18 hours of hybridization, all sample cDNA not hybridized to GeneChip probes are washed out of the chip, and a detection dye is added to biotinylated target that remains. At this point GeneChips are scanned by a laser to induce fluorescence from the detection dyes (Figure 1.1 right) [103]. Software integrated with the scanner assigns the photon flux detected per pixel to a specified spot (reported as the spot intensity). Subsequent quality control steps are used to make sure the scanner and chip were correctly aligned, that the signal was above a noise threshold, that the conformation of the spot met expected standards, etc. Those spots that pass the criteria for acceptability are averaged (most probes are present at multiple locations) and then aggregated according to some rule (generally weighted, not simply added before averaging) to yield a ProbeSet value that is the starting point for all further bioinformatics analysis. The specific details of our protocol are given in the Wet Lab Methods section of Materials and Methods chapter, to follow.

1.9 Prior Studies

The combined experimental model and transcript-level resolution of this experiment will be unique in ovarian surface epithelial-stromal tumor research. The first principles which will generate this data set do, however, apply to other oligonucleotide based arrays. Although the analysis reported in literature is based on transcript-

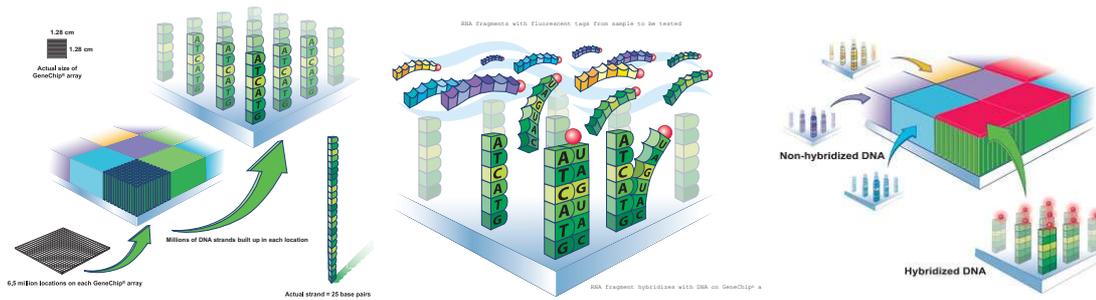


Figure 1.1: Human Exon 1.0 ST microarray chip synthesis. Photolithography is used to create oligonucleotide probes (left). Single-stranded oligonucleotide probes hybridize to biotin-labeled, complementary sample cDNA (middle). After washing, lasers scan chips to excite biotinylated dye, measuring hybridized cDNA (right). (Source Affymetrix.com)

cluster level data, this does not mean some similarity in results between platforms will not be found. At a minimum, given proper design, transcript-clusters that do not exhibit alternative splicing should have similarity. As well, in future work, it may be possible to reconstruct transcript-level array results and attempt to combine similar published data sets with this data set. With this potential, some prior studies are of interest to consider here.

The potential for biomarker identification was immediately realized by early ovarian tumor oligonucleotide microarray studies (starting in approximately 2000). Standardized and specialized microarrays were developed for analysis of transcript-cluster expression. Researchers struggled with producing adequate input sample, especially normal epithelium. Two approaches were used to amplify starting material [106]. One method, attempting to transform samples (control and test) and then allowing the cell culture to grow before RNA extraction could increase RNA yield but required non-tumorous control tissue and led to modifications in tumor expression through cell line transformation [107] [108]. Immunomagnetic-enrichment of normal epithelium nucleic-acid was also attempted but exposed RNA to the risk of degradation [109] [110]. Other limitations were already recognized at this time, including: the lack of standards, the sparse use of microdissection, the difficulty in obtaining

required starting material quantities, and the practical limitation of the features of the microarray; if the transcript-cluster does not have an associated probe on the microarray, it is not measured [106].

Learning from initial experimentation, the microarray community developed standards. Minimum Information About a Microarray Experiment (MIAME) [111], Microarray Gene Expression Markup Language (MAGE-ML) [112], and the Gene Expression Omnibus (GEO) [113] were defined to standardize experiment annotation and information, data storage, data sharing, and centralize expression data storage. The capacity of measurement for microarrays increased. Considering Affymetrix GeneChips, the original chip capacity used 18 μ features grouped in >8,700 ProbeSets to measure \sim 8,500 transcripts; (Human Genome Focus Array), currently the Human Genome U133 Plus 2.0 microarrays use 11 μ features grouped into >54,000 ProbeSets to measure \sim 47,400 transcripts [114].

Unfortunately, the experimental model of earlier OSE studies did not control for histopathological subtype. Typically histopathology of all carcinomas were viewed as one highly variable pathology, and the group was treated as one cohort. Efforts to define panels of genes associated with prognostics, response to chemotherapy, early detection, and tumor characterization were performed. Studies often did not perform tumor microdissection [115] [116] [117] [118] [119] [120]. Due to the frequency of onset, serous carcinoma often comprised the majority of samples in these studies, but other major subtypes were included, leading to mixed expression signals. Although laser cutting/capture microdissection (LCM) was not performed, microscopic inspection of tissues was performed and percentage tumor counts were given as a measure of tissue heterogeneity and expected levels of signal mixing. This is only useful if baseline signals are known. Some studies quoted as low as a median of 64% percent tumor cells in samples. Not surprisingly, these studies produced non-overlapping or minimal common content[121].

With the emerging focus on stratification by histopathology, recent studies have become more specific. The research community continues to lack commitment to standardizing LCM as a microarray experiment process, perhaps due to cost or expertise.

Large consortiums have been funded to profile ovarian tumors. One such project, the Cancer Genome Atlas project, includes a repository of 588 ovarian cystadenocarcinomas (tumor and associated controls) that have been measured by various high-throughput methods including Copy Number Variation (genes and miRNAs), gene expression, miRNA expression, DNA methylation, and somatic mutation. Although not microdissected, samples were carefully controlled for sample tumor ratio and RNA quality. These large, public projects provide data freely, allowing smaller experiments to combine data with theirs and gain additional power in analysis [122].

Certainly, next efforts in ovarian cancer biomarker detection will need to combine and make sense of the varied studies already performed. The combination of prior studies with new transcript-level studies and high through-put sequencing data will require highly integrated approaches in order to leverage this data.

1.10 Experimental Design

The recent understanding that ovarian surface epithelial-stromal tumor histopathological subtypes are separate diseases puts prior studies into question. By stratifying the design to account for subtype, analysis power will increase and more accurate measurements of expression may be performed. Changing ideas about tumorigenesis lead to questions about basic experimental models. This study takes the opportunity to apply these changing concepts and include higher resolution of measurement offered by exon-based oligonucleotide chips.

Serous ovarian surface epithelial-stromal tumors are the most common and most deadly histopathological subtype and were the focus of this experiment. Both traditionally accepted benign and malignant tumors may offer clues to biomarker selection.

In an attempt to select categories that fit both tumorigenesis theories, the malignant tumor cohort was Type II serous tumors. Benign tumors are not accepted to be the sole progenitor of malignant tumors. The tumorigenesis theories offer alternative narratives (*in situ* genesis and fallopian tube metastasis). A common control will need to be used, to provide an unambiguous reference. At the same time we are restricted to the material the surgeons can collect and that our IRB approval permits. Traditionally, tumorigenesis theories assume ovarian surface epithelial-stromal tumors to be derived from normal ovarian surface epithelium (NOSE). Therefore, NOSE was used as a control in this study. The alternative to NOSE, fimbrial epithelium, is related to NOSE through the embryonic coelomic epithelium. This relatedness and the hypothesis of the ‘transitional epithelium of common origin’ gives support for using the traditional NOSE control. Paracrine signaling will also be explored as an additional facet to this study. Stroma in proximity of tumors and distant from tumors will be compared in samples with both benign and malignant tumors.

1.11 Summary

Unfortunately, the answer to the question “Why is ovarian cancer the most lethal gynecologic cancer in the United States?” is that we dont know...yet. We are unsure of the epidemiology and etiology of the tumors, we are unclear in our understanding of the tumorigenesis model. Ovarian tumor mechanisms are as vague as the symptoms.

Progress is occurring. While few epidemiological factors are agreed upon (including BRCA1/2 genes, a history of parity, lactation, and oral contraceptive use), molecular mechanisms have begun to define the heterogeneity of ovarian tumors as a diversity of separate diseases.

Candidate biomarkers have been identified by a number of groups but do not exhibit the specificity and sensitivity needed for tumor detection. The discovery of biomarkers suitable for early diagnostics, using noninvasive techniques for sample acquisition (ideally a peripheral blood draw) was the idealized long term application

for this data set. This requires that products of early events in a few cells be secreted and stable, features not addressed in this study.

CHAPTER 2: MATERIALS AND METHODS

2.1 Rationale for the Key Features of the Wet Lab Methods

Several concepts were essential to the development of the series of wet lab protocols used in this study. Before protocols are presented, the rationale is given for the choices made, which are grounded in our experimental design and best practices as established in the literature. Firstly, we used LCM because macrodissected samples exhibit several levels of heterogeneity; there is already natural heterogeneity in the ovary and there are many histological subtypes as well as the possibility of polyclonal tumors. Next, since the mechanism of tumorigenesis of ovarian tumors is under debate, normal ovarian cancer epithelium was included as a common baseline. Since signaling may explain some of the history as well as the current activity of the tumor, stromal cells in the surrounding environment (both proximal and distal to the tumor) were profiled.

2.1.1 Laser Capture Microdissection

One of sources of complexity associated with ovarian tumors is the heterogeneity of samples. Diversity in samples can be due to any number of factors including but not limited to natural elements occurring in the organ (for example stroma, follicles, and tunica albuginea), distinct histopathological subtypes, polyclonal tumors, necrosis, immune response, and bleeding. Laser Capture Microdissection (LCM) allows specific subpopulations of tissues to be identified, separated and collected for further analysis. Microscopic identification of target cell populations is followed by placement of an LCM cap with thermoplastic film over the cells, after which a low-power infrared laser beam is focused on the film over the cells. When the film physically contacts

the cells, it attaches to them. When the cap is removed, the cells attached to the film are removed at the same time while non-attached cells are left on the slide [123]. The infrared-laser, although it activates the film and sticks it to the cells is low enough energy that it does not compromise the integrity of the biological material. Material on the cap can then be removed with extraction reagents.

2.1.2 Microdissection Schema

The experimental plan was to LCM-purify four categories of cells from every sample in one of the two class types, benign or malignant serous ovarian surface epithelial-stromal tumors. The four categories were normal epithelium, tumor cells, proximal stroma cells and distal stroma cells. Benign tumors represent early, possibly non-progressive events for Type I tumors, and the malignant tumors represent some stage of the Type II tumors, since they were further selected to match the high grade classification. The following labels are used for the samples: NE is normal ovarian surface epithelium, T is tumor, TS is proximal tumor-stroma microenvironment, and S is distal stroma. Figures 2.1 and 2.2 illustrate the microdissection schema, in a benign sample and malignant sample respectively. Not all samples yielded high-quality material in all 4 subsets of cells; in particular NE was difficult to obtain in sufficient quantity.

2.2 Wet Lab Methods

The following summarizes the main steps for preparing samples for microarray hybridization. It follows the recommendations of LCM experts and the microarray suppliers; deviations are noted where they occurred. The time line from tissue harvesting to hybridization typically required 2-3 weeks. This complex protocol used a diversity of techniques with multiple points of failure and thus quality control was essential (Figure 2.3).

Several check-points were incorporated into the wet-lab protocol to insure a level



Figure 2.1: Microdissection of benign sample. The green arrow indicates normal surface epithelium, red indicates tumor, purple indicates stroma adjacent to tumor, and blue indicates stroma distant from tumor.

of quality and consistency Figure 2.4. The initial two checks performed on samples were based on controlling for the histopathology of the tumor sample. Certified pathologists evaluated biopsies from patients to diagnose tumor histology. Serous tumors were flagged for this study. Cryosectioned, hemotoxylin and eosin stained samples were evaluated microscopically. The histopathology of the lab sample was checked for consistency with the pathology report (to remove issues involving inconsistent macrodissection of tissue). Additionally, the presence of normal epithelium, tumor, stroma and stroma adjacent to tumors was indicated.

When samples passed microscopic histopathology checks, several checks were periodically performed enforcing quality of product. Initially, the quality of RNA from samples was measured from most samples. An Agilent 2100 Bioanalyzer was used to measure RNA integrity numbers (RIN). Although limited samples were available, the samples offering the best RIN numbers were used. This check focused the study

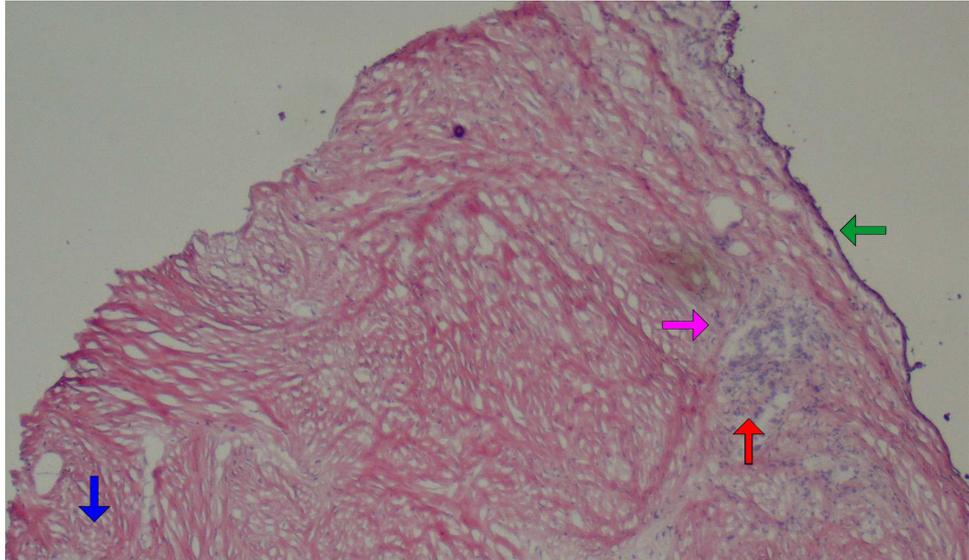


Figure 2.2: Microdissection of malignant sample. The green arrow indicates normal surface epithelium, red indicates tumor, purple indicates stroma adjacent to tumor, and blue indicates stroma distant from tumor.

on samples that would, by report, be most efficient to use in the study and give the best quality products. A Nanodrop spectrophotometer was used to measure cDNA, and ST-cDNA quality as per Nugen recommendations (having a 260/280 >1.8 and a respective quantity of 3 μ g and 5 μ g).

2.2.1 Patient Sample Harvesting

Ovaries were harvested by a clinician and placed in gamma radiation-sterilized 60 ml tubes with approximately 7.5 ml of Hanks Balanced Salt solution (Invitrogen Co., Carlsbad, CA). Ascites was collected in an empty gamma radiation-sterilized 60 ml tube (Invitrogen Co., Carlsbad, CA). Blood was collected in sterile vials. Upon receipt of sample, collected blood and ascites fluids were spun at 2,500 rotations per minute (rpm) or 1,430 relative centrifugal fields (rcf) for 15 minutes in a Beckman GS-6 centrifuge (Beckman Coulter, Inc., Fullerton, CA). During this time, ovary samples were cut into representative halves. One half was placed in a 1.5 ml cryovial (Fisher Scientific, Pittsburgh, PA), immediately placed in liquid nitrogen, and

later stored in a Taylor-Wharton K Series Cryostorage System (Taylor-Wharton-Cryogenics, Theodore, AL) with a Rees Scientific environmental monitoring system (Rees Scientific, Trenton, NJ). The other half was placed in a standard sized cryomold (Sakura Finetek USA, Inc., Torrance, CA), covered with Optimal Cutting Temperature (OCT) compound (Sakura Finetek USA, Inc., Torrance, CA), frozen on dry ice, and later stored at -80°C . Serum and buffy coat from centrifuged blood were each separately stored in 1.5 ml cryovials at -80°C . The ascities cell pellet formed after centrifuging was placed in a 1.5 ml cryovial and stored in the K Series Cryostorage System. Ascities was stored in 1.5 ml eppendorf tubes at -80°C .

2.2.2 Hemotoxylin and Eosin staining (Histopathology Confirmation)

Cryomold samples were serially sectioned into one $5\mu\text{m}$ and multiple $8\mu\text{m}$ sections using Extremus[™] Microtome Knives (C. L. Surkey, Inc, Lebanon, PA) in a Leica CM 1850 UV Cryostat (Leica Microsystems Inc., Bannockburn, IL). Sections were placed on precleaned Fisherbrand Superfrost^{®*}/Plus Microscope slides (Fisher Scientific, Pittsburgh, PA), cooled on dry ice and later stored at -80°C . $8\mu\text{m}$ sections were stored for Laser Capture Microdissection. $5\mu\text{m}$ sections were used to guide capturing and were exposed to Hemotoxylin and Eosin staining. More specifically, $5\mu\text{m}$ sectioned tissue samples were exposed to the following reagents: 95% ethanol for 1 minute, Harris' Hemotoxylin (Hemotoxylin 7211, Richard-Allan Scientific, Kalamazoo, MI) for 1 minute, distilled H_2O for 20 seconds, weak ammonium H_2O for 1 second, distilled H_2O for 20 seconds, 95% ethanol for 10 seconds, 50% Eosin (Sigma-Aldrich, St. Louis, MO) and 50% 95% ethanol for 1 second, 95% ethanol for 10 seconds 95% ethanol for 10 seconds, 100% ethanol for 10 seconds, 100% ethanol for 10 seconds, and xylene (Sciencelab.com, Inc, Houston, TX) for 10 seconds. After air-drying, VWR Micro cover glasses (VWR Scientific Inc, Media, PA) were affixed to the slides using permanent mounting media (Vector Laboratories, Inc., Burlingame, CA). Pathology was performed using a VistaVision microscope (VWR Scientific Inc, Media, PA).

Microscopic images were captured by an Olympus BX60, DP71 using the Olympus DPController and Manager (Olympus America, Inc., Center Valley, PA). Light was set to default for camera (or '9'). One touch auto-whitening was used which, on average, adjusted color composition as follows: red 0.6, green 1.0, and blue 1.2.

2.2.3 HistoGene Staining

The HistoGene LCM Frozen Section Staining kit (MDS Analytical Technologies, Toronto, Canada) was used according to the suppliers instructions to prepare fresh frozen, cryosectioned samples for LCM. Sample slides cut at 8µm were briefly thawed and subjected to 75% ethanol for 30 seconds, nuclease-free distilled H₂O for 30 seconds, HistoGene Staining solution for 20 seconds, nuclease-free distilled H₂O for 30 seconds, 75% ethanol for 30 seconds, 95% ethanol for 30 seconds, 100% ethanol for 30 seconds, and Xylene for 5 minutes.

2.2.4 Laser Capture Microdissection

After staining, samples were immediately microdissected: the first samples with a PixCell Iie instrument and the final samples with an ArcturusXT™ Laser Microdissection instrument (Applied Biosystems, Life Technologies, Co., Carlsbad, CA). Microdissection was performed on precleaned Fisherbrand Superfrost®*/Plus Microscope slides (Fisher Scientific, Pittsburgh, PA) using CapSure® HS LCM caps. When available, normal epithelium, tumor, stroma, and stroma adjacent to tumor were separately collected. The capturing period was limited to 30-45 minutes (for all cell types combined from a single slide) after the Xylene HistoGene staining kit step, with a small number of samples requiring one hour.

2.2.5 RNA Extraction

After the microdissection of each cell type ExtracSure™ devices (C. L. Surkey, Inc., Lebanon, PA) were attached to the caps and 10µl of PicoPure® RNA Isolation

kit extraction buffer (Applied Biosystems, Life Technologies, Co., Carlsbad, CA) was applied to the dissected tissues. ExtracSure™ devices were each capped with a 0.5 ml microcentrifuge tube (USA Scientific, Inc., Ocala, FL) and incubated in a mini-incubation oven (Bio-Rad Laboratories, Inc., Hercules, CA) at 42°C for 30 minutes. After incubation, ExtracSure™ devices with attached microcentrifuge tubes were centrifuged at 800xg for 2 minutes. ExtracSure™ devices were then removed, 10µl of 70% ethanol was added to each tube and tubes were stored at -20°C.

2.2.6 Tissue Scraped RNA Extraction

After the microdissection of each tissue type but before incubation, tissue scrapes from residual slide material were also performed for each patient sample, to provide additional material for inferring the RNA quality. Approximately 50µl of PicoPure® RNA Isolation kit extraction buffer (Applied Biosystems, Life Technologies, Co., Carlsbad, CA) was applied to slides after microdissection was complete to remove residual tissue. Extraction buffer with suspended tissue scrape material was pipetted into a cap with attached ExtracSure® device and capped with a 0.5 ml microcentrifuge tube (VWR Scientific Inc, Media, PA), and co-incubated with the other samples from that slide. After centrifugation 50µl of 70% ethanol was added to the tissue scrape solution before being stored at -20°C.

2.2.7 RNA Purification

The PicoPure RNA Isolation kit (Applied Biosystems, Life Technologies, Co., Carlsbad, CA) was used for the steps outlined below, according to the supplier's instructions. After column purification, 5µl of RNase-free DNase I stock solution (Qiagen, Inc., Valencia, CA) was mixed with 35µl Buffer RDD per column and added to each purification column, then allowed to incubate for 15 minutes at room temperature. After washing away the enzyme and any nucleotides resulting from its activity, samples were eluted in 11µl of Elution Buffer.

2.2.8 RNA Integrity Measurements

For assessing RNA integrity we used an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA), and the standard RNA 6000 Nano Assay protocol, as described by the manufacturer. The RNA 6000 Nano Marker was used in each well and ladders were used in each chip as positive technical controls. 1.5 μ l of sample plus RNA ladder was heated at 70°C, then 1 μ l was added to each well of an RNA Nano chip which was loaded into the Bioanalyzer. Subsequent to the run the Eukaryote Total RNA Nano electrophoresis assay was performed, yielding a quality score and mass per sample.

2.2.9 Complementary DNA Generation

Due to the nature of the samples the amount of starting material was extremely limited, so amplification was necessary to produce sufficient mass to drive the hybridization. Complimentary DNA (cDNA) generation and amplification was performed using the Whole Transcriptome WT-Ovation™ Pico RNA Amplification System Kit (NuGEN Technologies Inc, San Carlos, CA) recommended by Affymetrix. The kit was used according to the supplier's instructions. Before cDNA amplification, cDNA was purified using Agencourt® RNAClean® Beads, according to the suppliers instructions.

2.2.10 SPIA Amplification

Following the instructions supplied with the kit, the SPIA Amplification Master mix was created; 160 μ l was added to each tube of dried beads and sample, subdivided and amplified using the recommended thermoprofile for an ABI 9600 thermocycler. After removal of the beads, the amplified cDNA was further purified either with a Zymo DNA concentration and purification kit (Zymo Research Corporation, Irvine, CA) or the QIAquick™ PCR Purification Kit (Qiagen, Inc., Valencia, CA), according to each suppliers instructions. In each case the column elute was subjected to

Nanodrop quantification prior to ST-cDNA generation.

2.2.11 Nanodrop Amplified Complementary DNA Quantification

2 μ l of purified SPIA amplified cDNA was aliquotted for Nanodrop quality and quantification analysis. The Nanodrop 1000 (Thermo Fisher Scientific, Inc., Wilmington, DE) stage was initialized and blanked with nuclease-free H₂O. After blanking, a sample of nuclease-free H₂O (negative technical control) was used to verify that readings were within the prescribed ± 0.6 ng/ml range of error of the instrument. After samples were analyzed, quality and quantification scores were recorded. Due to material limitations replicates of the samples were not measured. To continue to ST-cDNA conversion a quantity of 3 μ g of SPIA amplified cDNA with an absorbence ratio of $1.8 \leq x \leq 2.0$ was required. (Nanodrop measurements were also performed for ST-cDNA protocol products requiring 5 μ g of ST-cDNA with an absorbency ratio of >1.8 before proceeding to fragmentation and labeling steps).

2.2.12 ST-Complementary DNA Conversion

The ST-cDNA is a stable intermediate of the mRNA, but does not provide the single-stranded complement to probes in sufficient quantity to drive the hybridization reaction. The asymmetric amplification was performed using the WT-Ovation[™] Exon Module (NuGEN Technologies Inc, San Carlos, CA), according to the suppliers instructions, using x μ g of qualified ST-cDNA from the above procedure. The single-stranded cDNA so generated was then purified, with either a Zymo DNA concentration and purification kit or a QIAquick[™] PCR Purification Kit, according to the suppliers instructions. The column eluates were again qualified using the Nanodrop 1000 spectrophotometer, as described above (or in the Supplementary Materials).

2.2.13 Fragmentation and Labeling

The single-stranded DNA target material is fragmented to normalize the diffusion of all targets in the hybridization reaction, and labeled to allow detection of the target. The FL-Ovation™ cDNA Biotin Module V2 (NuGEN Technologies Inc, San Carlos, CA) was used according to the suppliers instructions to perform these steps, starting with 5µg of qualified material. Samples were then stored at -20°C until hybridization could be performed.

2.2.14 Exon Microarray Hybridization Recipe

The hybridization cocktail recommended by Affymetrix for use in their stations and on these chips was modified by Nugen, specifically with respect to the use of Human Exon 1.0 ST arrays (Affymetrix, Inc., Santa Clara, CA) and Nugen's Ovation products (FL-Ovation™ cDNA Biotin Module V2 Users Guide, Appendix A.1) [124]. This modified cocktail was used for all samples. The Affymetrix hybridization, staining, washing and scanning were performed at the Microarray Core facility at Carolina's Medical Center's Cannon Research Center by certified technicians. With the exception of modifications to the hybridization cocktail noted, all reagents for these steps are included in the Affymetrix's Hybridization, Wash, and Stain (HWS) kit (Affymetrix, Inc., Santa Clara, CA). After scanning, data was burned to DVDs for later analysis. The hybridization solution was reserved and stored as a precaution against Genechip failure.

2.3 Analysis Pipeline

2.3.1 *In Silico* Quality Control

Two primary-level quality control features are built into exon chips: one is based on a handful of non-human transcripts (spike-in controls) and the second uses the distribution of all intensities on a chip [125] [126] and compares that to what is typi-

cally seen. These rudimentary quality control steps were performed immediately upon scanning of the arrays, using the Expression Console™ software (EC) from Affymetrix provided for the purpose. Briefly, the quality control report included data from the core probes, which were compared for expected intensity; the distributions of core transcript intensities of exon chips was plotted as a series of boxplots (one per chip) to identify outlier intensity distributions. Additionally, the mean intensity of Core probes (using very high-confidence probe design) and their background probes was independently assessed. Finally, Pearson and Spearman correlations were performed on the Core and Extended ProbeSet (transcript) measurements in order to identify unusual chip correlations (primarily to indicate batch effects). These measures were used to indicate whether a sample or chip had failed and needed to be replaced. Once acceptable arrays were identified more sophisticated measures were used for validating assumptions in our experimental model or more subtle batch effects using algorithms available in the Partek® Genomics Suite™ (Partek, Inc., St. Louis, Missouri).

2.3.2 Initial Data Generation

Exon-level arrays create a very large body of data that can be recombined in a number of meaningful ways. This is a benefit as it expands the number of models that can be tested, but it can easily lead to a sensation of drowning in data. As a first pass, microarray analysis was performed at the transcript cluster (gene) and transcript levels. As a first step the probe signal must be estimated (relative to noise), and we used the widely-accepted algorithm GC-RMA[127], as implemented in the Partek software environment. We focused on transcripts defined as “Core”, since the genes of interest, described in the Introduction, are all included in this category. The probes’ values are combined into ProbeSets for each exon, and exons are combined according to gene models for each transcript, and the set of transcripts is combined for each transcript cluster or gene. When performing transcript cluster level analysis, probes were combined using a Winsorized mean (truncating the top

and lowest 10 percent intensities).

2.3.3 Inferential Statistics: Gene-level

Our initial goal was to identify transcript-cluster, or gene-level inferential statistics was be performed with a 2-factor Model I Analysis of Variance (ANOVA) with multiple *a posteriori* contrasts in order to provide a list of significant differentially expressed transcript clusters.

An ANOVA was selected (with contrasts) instead of multiple t-tests for several reasons. Firstly, performing multiple t-tests as opposed to an ANOVA immediately exposes the inferential tests to a lowered confidence parameter, alpha[128]. The lowered alpha increases the chance of type I errors (false positives). Given an experimental model with 4 factors over 2 levels, it is possible to perform 6 separate t-tests. If each t-test is given an $\alpha = 0.05$ then the true alpha for the multiple t-tests is 0.26 or $1 - ((1 - 0.05)^6)$ [128]. An advantage to performing multifactorial ANOVA models is that both factors and interactions are measured. To use an ANOVAs model we must meet several assumptions about the underlying data. The theory behind parametric ANOVAs assumes a normal distribution, independent data, and homogeneity of population variances (homoscedasticity). With respect to a normal distribution of signal intensity, most microarray algorithms (D-Chip [129], RMA [130], and GC-RMA [127]) transform the signal to \log_2 scale and further scale and normalize with methods like locally weighted scatter plot smoothing (LOWESS) or quantile-quantile normalization in order to produce a more nearly normal data space for tests of differential expression. Despite these procedures, data are frequently not normal and two sample classes often do not have the same distribution; this is particularly a problem when sample numbers are low. An advantage of parametric 2-way ANOVAs is that they are quite robust to departures from normality [131]. Homocedasticity also has potential to be violated. If the expression of disease state is being measured in a microarray experiment, homoscedasticity would manifest as similar variance of gene expression. A

number of studies have indicated that in some disease states it is the variance rather than the mean of many genes that is most changed [132] so this characteristic cannot be assumed. That is, although similar expression variance may occur in a disease with a specific genetic change (for example cystic fibrosis), when comparing normal tissues with cancer tissue with large amounts of dysregulation, it would be very unusual for all genes in both tissues to have equal variance. In an experimental model with similar sizes of levels of factors, 2-way ANOVAs are robust with respect to variance on both the main factors and their interactions [131]. The assumption of data independence is more difficult to resolve. Genes are related and interact on many levels, and indeed much of bioinformatics research is currently engaged in examining regulatory networks that are interesting because the genes involved are not independent. This is a chicken and egg problem that is generally handled by first assuming independence and then iterating if significant results appear to be found in order to take this into account. Departures from independence tend to reduce the power of F-tests [131]. When working with a 1-way ANOVA, alternative, nonparametric approaches are available (for example Kruskal-Wallis [133]) but an effective 2-way nonparametric ANOVA model has not been developed [128].

The total ANOVA formula is expressed as follows.

$$Y_{ijk} = \mu + PatientTumor_i + TissueType_j + PatientTumor * TissueType_{ij} + \epsilon_{ijk} \quad (2.1)$$

Where Y is signal, μ is a common effect for the experiment, Patient Tumor represents the patient tumor experimental factor, Tissue Type is the tissue type experimental factor, the Patient Tumor*Tissue Type is each experimental cohort, and ϵ is the residual error. Iterators i, j, and k indicate each patient tumor type (benign or malignant)(i), tissue type (normal epithelium, tumor, distant stroma, or tumor-stroma

Table 2.1: Planned *a posteriori* contrasts performed after the ANOVA.

Control Cohort	Test Cohort	Contextualization
Normal Epithelium (Benign Sample)	Tumor (Benign Sample)	Benign tumor gene list
Normal Epithelium (Malignant Sample)	Tumor (Malignant Sample)	Malignant tumor gene list
Normal Epithelium (Malignant Sample)	Normal Epithelium (Benign Sample)	Differences in epithelial controls
Stroma (Benign)	Tumor-Stroma Microenvironment (Benign)	Effects to stroma of exposure to benign tumors
Stroma (Malignant)	Tumor-Stroma Microenvironment (Malignant)	Effects to stroma of exposure to malignant tumors
Stroma (Benign)	Stroma (Malignant)	Differences in stromal controls

environment)(j), and tissue sample (k).

A posteriori comparisons were performed as shown in Table 2.1.

2.3.4 Inferential Statistics: Transcript-level

Transcript-level analysis will focus on using known gene models to discover specific transcript expression. Similar to gene-level analysis, an ANOVA will be used. The ANOVA performed was also a 2-factor Model 1 ANOVA, formulated to evaluate signal at the exon level; this allows more accurate analysis, differentiating transcript clusters into individual transcript activities. The total ANOVA is expressed by:

$$\begin{aligned}
 Y_{ijklm} = & \mu + PatientTumor_i + TissueType_j + ExonID_k + \\
 & Sample(PatientTumor * TissueType)_{ijl} + PatientTumor * TissueType_{ij} + \\
 & PatientTumor * TissueType * ExonID_{ijk} + \epsilon_{ijklm} \quad (2.2)
 \end{aligned}$$

Where Y is signal, μ is a common effect for the experiment, Patient Tumor represents the patient tumor experimental factor, Tissue Type represents the tissue type

experimental factor, Exon ID represents the exon-to-exon effect independent of experimental factor, Sample(Patient Tumor*Tissue Type) represents sample-to-sample effect, Patient Tumor*TissueType represents the experimental model's cohorts, Patient Tumor*Tissue Type*Exon ID represents exon effects within the cohorts, and ϵ is the residual error. The same *a posteriori* contrasts performed in gene-level analysis were performed in the transcript-level analysis (Table 2.1). Iterators i, j, k, l, and m indicate each patient tumor type (benign or malignant)(i), tissue type (normal epithelium, tumor, distal stroma, or tumor-stroma environment)(j), exon through out samples(k), specific tissue sample (l), and exon per sample (m).

2.3.5 Biological Interpretation

Once gene lists were generated, each was assessed for significant representation in biological annotation hierarchies including gene ontology (GO) groups, biochemical and protein-protein interaction pathways, and chromosomal location, using the Database for Annotation, Visualization, and Integrated Discovery (DAVID), version 6.7, from the National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health (NIH) [134] [135]. DAVID itself accesses several external databases including: Biocarta, KEGG, Panther, Reactome, BIND, DIP, MINT, NCICB, and UCSC TFBS. These databases are used to evaluate annotative terms relating to the following categories: disease associations, molecular functionality, biological process, cellular localization, genomic position, homology, literature citations, molecular pathway, protein domain, and tissue expression. Single enrichment analysis (SEA) was performed for benign and malignant DE transcript lists (All, Up, and Down) separately, resulting in 6 result sets. All SEA was performed using Fischer's exact test, compensating for multiple comparisons with the Benjamini and Hochberg (BH) false discovery rate modification [136].

To visualize relationships between terms the tool AmiGO was used to generate a graph of relationships, and Photoshop (version 5.0) was used to better render the

AMIGO graph (without changing details or context) [137].

Two tools were used to explore *de novo* pathway generation potential, STRING 8 [138] and Genes2Network [139]. STRING 8 is a comprehensive data warehouse of protein-interactions and protein interaction prediction algorithms. STRING pulls data from IntAct [140], BioGRID [141], the Molecular Interaction Database (MINT) [142], Database of Interacting Proteins (DIP) [143], BIND [144], KEGG [145], Reactome [146], HPRD [147], EcoCyc, the NCI-Nature Pathway Interaction Database, and Gene Ontology databases [148]. Additionally STRING performs text mining of annotation in SGD, OMIM, the Interactive Fly, and PubMed abstracts. STRING is a comprehensive source of protein-interaction and so was the base resource for uncovering a gene network. Genes2Network reconstructs networks by drawing on a number of additional databases, including: BIND [144], DIP [143], IntAct [140], MINT [142], pdzbase [149], SAVI, Stelzl [150], vidal [151], ncbi hprd [147], and KEGG (mammalian) [145]. The output includes a visualization network that codes the known relationships and potential interactions reporting where z-scores are ≥ 2.5 . Gene2Network interactions were added to the STRING generated network if interactions were not represented by the STRING network, or if the Genes2Network presented an interaction between two genes with fewer intermediate connecting genes.

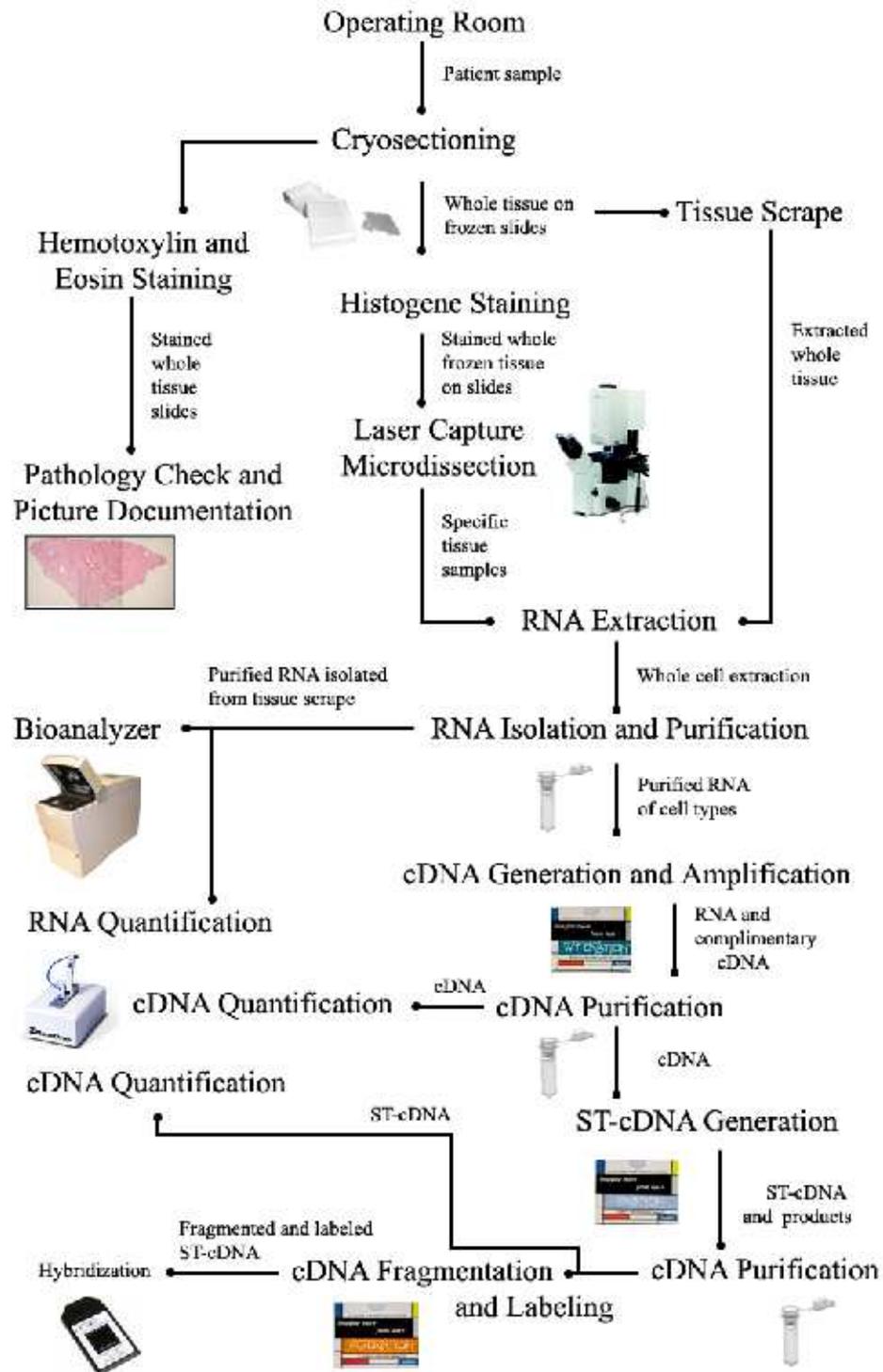


Figure 2.3: Flow chart of wet lab protocols used in microarray sample preparation.

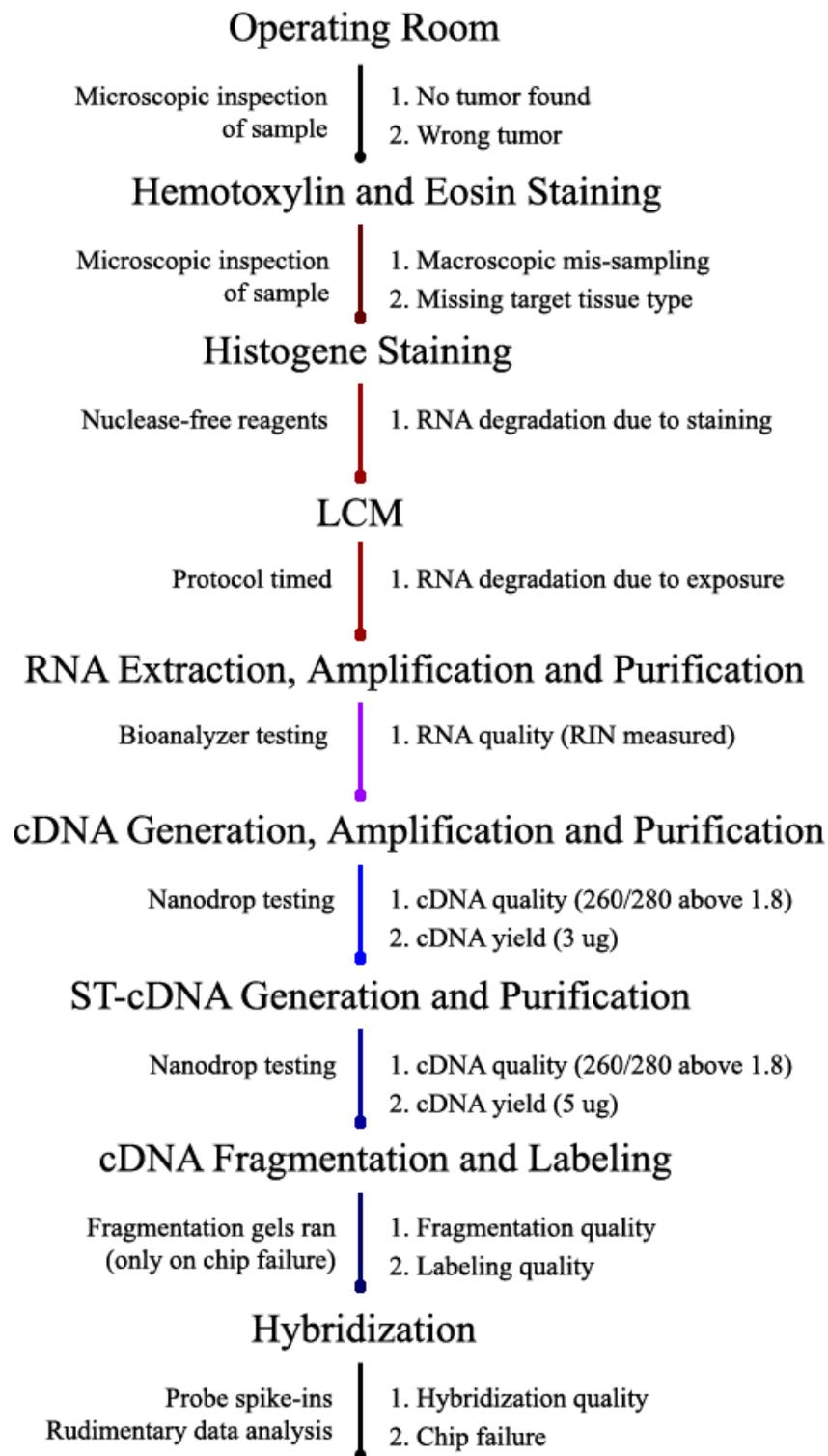


Figure 2.4: Quality checks incorporated throughout the sample preparation protocols.

CHAPTER 3: RESULTS

3.1 Study Sample Details

Ovarian tumors are histopathologically a heterogeneous group, and, as described in the first chapter; there is disagreement as to whether they derive from a common progenitor type or pathway. Under such circumstances a meaningful experimental design must focus narrowly if the study is to answer any meaningful question. In addition, human polygenic diseases require a large number of samples to achieve real statistical power, and the ability of any one clinical center to acquire that number is limited. Thus, a public ovarian cancer consortium has been established [113] to allow multiple centers to co-contribute. While CMC does not belong formally to the consortium it may contribute data, through the Gene Expression Omnibus. The bioinformatics challenge to such an endeavor is that the data are not all collected on the same platform or prepared with the same reagents, so effort is required to standardize and normalize the data before it can be combined in what are termed meta-studies [152]. Our contribution focuses on serous tumors, the most frequent ovarian tumors, both benign and malignant (type II). Malignant tumors were all type II serous tumors exhibiting a high-grade pathology and/or with co-occurrence of tumor in the fallopian tube. Tumor sample pathologies observed included simple serous cysts, serous cysts, serous cystadenomas, serous cystadenofibroma, and serous carcinomas (papillary and cuboid).

The design initially proposed had 10 samples per tissue category, as illustrated in Table 3.1. However, patients arrive according the logic of their disease and time, not ours. This explains the distribution and numbers of samples that were actually used, as shown in Table 3.2, which details the count of collected and hybridized samples

Table 3.1: Sampling model design.

	Normal Epithelium	Tumor	Stroma	Tumor-Stroma Environment
Benign	10	10	10	10
Malignant	10	10	10	10

before quality control filtering (including replicates). Paired samples are preferred, to control for age, genotype and environment, but when adequate amounts of high-quality cells of all 4 types could not be found, a minimal requirement for pairing of control and test tissues was defined (that is, normal ovarian surface epithelium with tumor and stroma with tumor-stroma environment). Lastly, since normal epithelial cells are least frequent in number and most easily removed by basic handling, for samples having well defined cells of the other types but no available normal epithelial cells, samples from age matched patients were considered acceptable. Percentages of sample pairings are shown for benign and malignant samples in Figure 3.1.

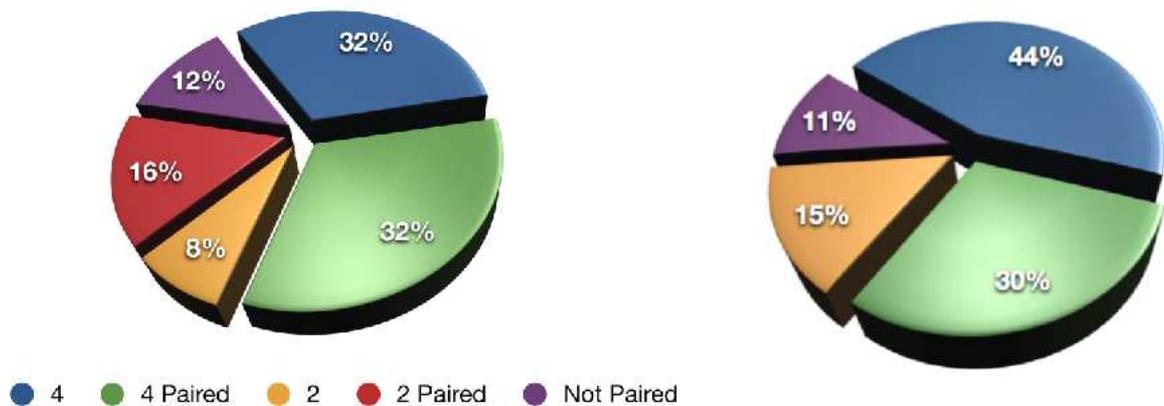


Figure 3.1: Percentages of pairing schemes for benign (left) and malignant (right) samples. “4” or “2” indicate the number of tissue samples derived from age matched patients. “4 Paired” or “2 Paired” indicate the number of tissue samples derived from the same patient. Samples counts are reported after all quality control has been performed (total count 50 samples).

Table 3.2: Total of 75 samples, details include replicates, before quality control.

	Normal Epithelium	Tumor	Stroma	Tumor-Stroma Environment
Benign	8	8	10	9
Malignant	8	10	11	11

Table 3.3: Basic patient demographic and pathology details.

Patient	Tumor Cohort	Pathology	Age of Patient
PN1	Non-tumorous	Non-tumorous	47
PN2	Non-tumorous	Non-tumorous	66
PN3	Non-tumorous	Non-tumorous	79
PN4	Non-tumorous	Non-tumorous	65
PB1	Benign	Simple Serous Cysts	69
PB2	Benign	Serous Cystadenoma	39
PB3	Benign	Serous Cystadenofibroma	63
PB4	Benign	Serous Cysts	66
PB5	Benign	Serous Cystadenofibroma	43
PB6	Benign	Serous Cysts	27
PB7	Benign	Papillary Serous Cystadenofibroma	65
PB8	Benign	Simple Cyst	65
PB9	Benign	Serous Cystadenofibroma	69
PM1	Malignant	Papillary Serous Carcinoma	75
PM2	Malignant	Serous Carcinoma	54
PM3	Malignant	Serous Carcinoma	74
PM4	Malignant	Serous Carcinoma	64
PM5	Malignant	Serous Carcinoma	80
PM6	Malignant	Serous Carcinoma	54
PM7	Malignant	Serous Carcinoma	49
PM8	Malignant	Papillary Serous Carcinoma	70

3.2 Laser Capture Microdissection

Laser capture microdissection (LCM) was performed on all samples used in this study. Although enabling the study to be more specific to tissue type expression, LCM is a technically challenging assay that requires considerable preparatory work as well as the time to perform the cell collection, neither of which preserves RNA in an intact form. Staining is necessary to allow visual confirmation of target tissues. The provided kit (Histogene Staining Kit) is an RNase-free variant of hemotoxylin staining. Pre-LCM staining, however, did not provide the same degree of visual discrimination as standard tissue staining techniques. There is a trade-off between using sufficient dye to make definitive calls and minimal enough amounts to prevent interference in subsequent manipulations. In our hands the optimal trade-off used OCT embedded samples serially cryosectioned, with some of the sections stained with Hemotoxylin and Eosin (H&E). The H&E stains were fixed to slides and used to guide LCM performed on the remaining serial sections (which were minimally exposed to staining). While it meant that not all cells in a sample were used, a time-limit was imposed during which a section was put in the LCM. This was to control for degradation of RNA as it is exposed to ambient temperatures and atmosphere. All LCM manipulations of a section were performed within 1 hour (from staining to RNA extraction buffer).

For the majority of samples, each step of LCM capture was documented with images (Appendix A). This is partly to verify that the expected cells were captured. To illustrate the process, sample B05 is shown during capturing of normal epithelium in Figure 3.2. (A reference H&E stained section of the captured area is shown in Figure 3.3.) The first panel of Figure 3.2 shows a single-layer lining of normal epithelium on the edge of a tissue section. The second panel shows capturing sites, periodically placed along the normal epithelium. The third panel was taken after capturing; areas of captured (and removed) normal epithelium and unaffected normal epithelium

are shown, to illustrate the specificity and success of normal epithelium capturing. Figure 3.4 shows the specificity of tumor and adjacent stroma collection. The first panel shows the initial capture site, the second shows just tumor cells captured (on the cap), and the third shows the section after both tumor and peripheral stroma were separately captured. Other examples are shown in Appendix A.



Figure 3.2: An image progression, capturing normal epithelium (sample B05). Before capture (left), intermittent capture (middle), after capture (right). Images adjusted 100% increased brightness, 75% increased contrast for image clarity.

3.3 *In Silico* Quality Control

3.3.1 Rudimentary Microarray Processing Quality Control

Quality control of samples was performed on .cel files (raw data) to determine if microarrays successfully hybridized. This step determines if a GeneChip in combination with the sample performs to specifications. Although every effort was made to perform quality control on the samples, uncontrolled variables are the hybridization cocktail and the GeneChip lot and age (and performance of the hybridization station and scanner). It is noted here that there currently is no validated standardization method for quality control of exon microarrays. Affymetrix, Inc. does make available the software tool “Expression Console” which enables researchers to perform rudimentary quality control. Several metrics can be performed to compare samples to each other and determine if the total microarray signal is significantly different from others in the study [125]. All measures of signal in this rudimentary quality control

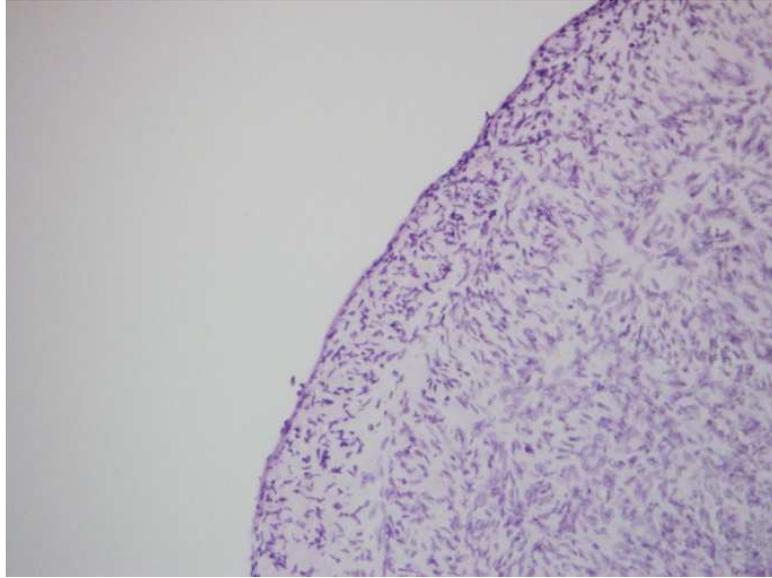


Figure 3.3: Hemotoxylin and Eosin staining of sample B05.



Figure 3.4: An image progression, capturing tumor and adjacent stroma (paired benign tumor and tumor-stroma environment samples B16 and B22).

are based on performing RMA Sketch, using only “Core” annotation type probes; however, measurements with “Extended” annotation type probes produced similar trends (data not shown).

Extracted intensity values are first displayed as box plots (Figure 3.5). Technical replicates are paired, sharing a primary label and using the letter ‘R’ to distinguish them (for example M06 and M06R). Inspection of the box plots shows considerable variation in the sample signals. The distribution of signal is one of the criteria for a quality array: failed chips have lower mean intensities and a smaller total signal range. We did not attempt to retrieve partial data from these arrays since we could not determine single probe quality criteria.

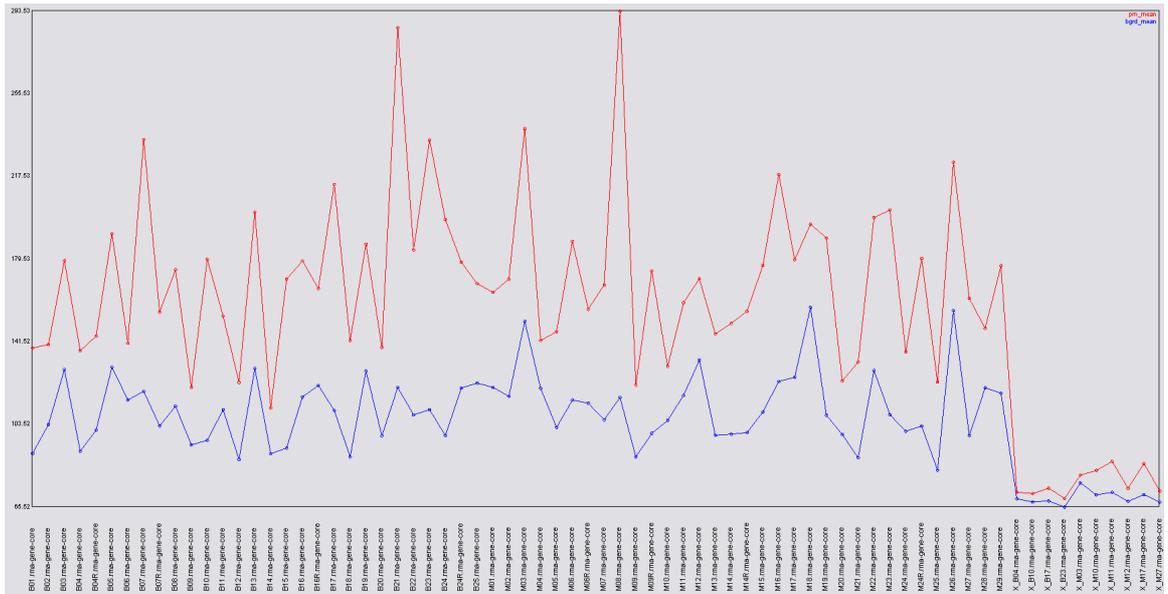


Figure 3.6: Line graph plotting the mean signal of experiment probes (red) and mean signal of background probes (blue).

3.3.2 73 Samples

Of the 75 samples prepared for this study, 73 hybridizations generated .cel files without primary errors and were subjected to the quality control stages. Samples prepared with the same lots of reagents often show correlation based on this variable rather than the clinical variable, usually called a ‘batch effect’. There are a number of simple visual analytic approaches to test for such association, one of the best-known being Principle Components Analysis (PCA). Examination of the first three components across all 73 samples, including those that were called ‘failed’, was performed on GCRMA calculated ProbeSet values. We note here that the samples from failed arrays were retrieved, rehybridized to fresh GeneChips with newly prepared sample and that those follow-up arrays were judged to have passed. In Figure 3.7, passed samples are green, the technical replicates are in blue, those in red were first rated as failed. The passed samples (study and technical replicates) clearly cluster together. Rotation of the components shows that the failed samples occupy a broad area below the others (Figure 3.7).

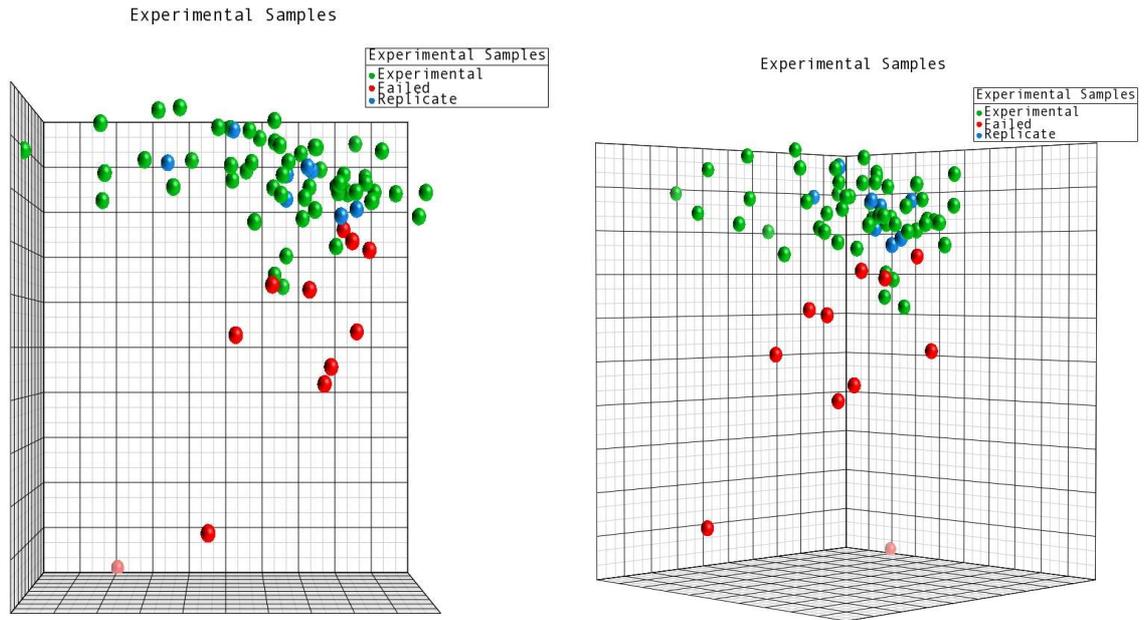


Figure 3.7: All samples, visualized with principal components analysis. Green spheres indicate samples that have passed initial quality control (QC), blue spheres represent samples that passed QC and were technical replicates, and red spheres were samples that failed QC.

Samples that initially failed and were resampled and subjected to a second round of hybridization that lead to qualification are highlighted in Figure 3.8. This figure demonstrates that the passed replicate samples now cluster with the other passed samples and not their previous arrays in the failed group. The pass/fail pairs are colored and labeled based on the specific patient number and tissue type from which they are derived. One can also examine how the samples cluster with respect to the clinical class. The closest distance is between samples M25 and M27, which came from patient 42 (tumor-stroma environment sampling).

Technical replicates should cluster most tightly of all samples with LCM purified samples. Figure 3.9 displays the clustering results, with replicates shown in the same color. All pairs are in close proximity, and since they all qualified they all group in the ‘pass’ cluster. One would expect that the variance between replicates would be smaller than the mean for all sample pairs, and this is indeed the case, shown in

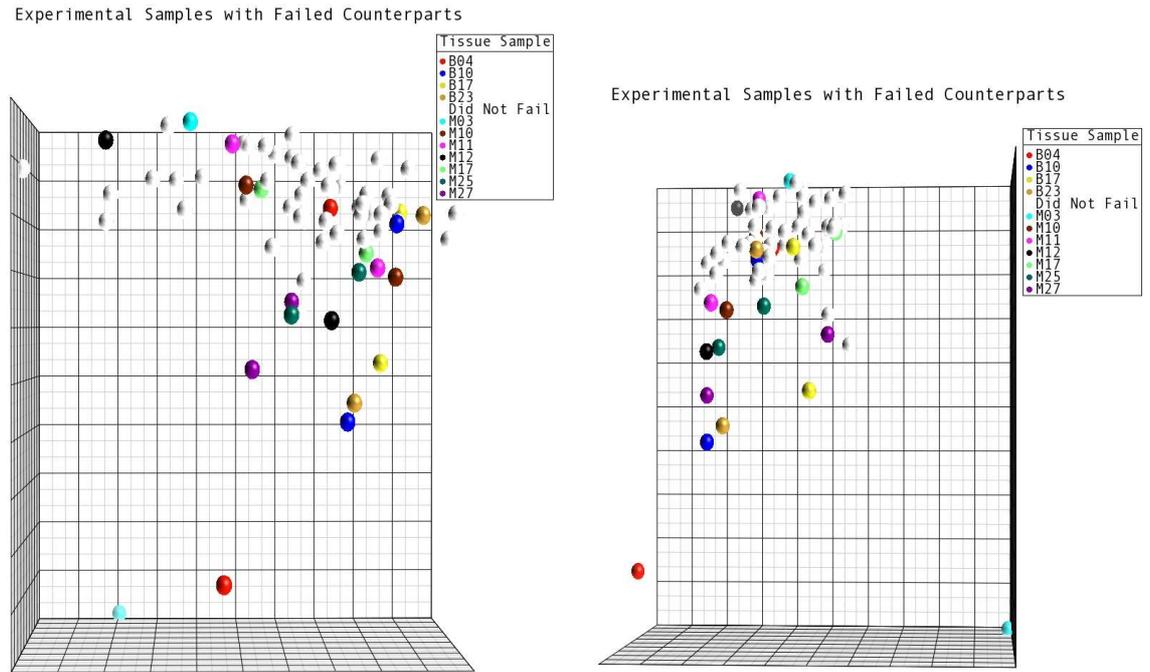


Figure 3.8: Principle components analysis of samples that initially failed but were re-sampled and analyzed. Related failed and passed samples are indicated with the same color. White samples are not relevant but could not be removed without changing the covariance structure of the PCA plot.

Figure 3.8. Since a rehybridized sample might be depleted in some components, we examined the variance between pass/fail pairs are less than the general variance seen between the passed sample and failed sample pairing in Figure 3.9. This illustrates that technical replicates are more similar by this measure than pass/fail pairs, shown in Figure 3.10 using a trio (Pass/Fail/Replicate) of samples from B04.

When more sophisticated quality controls were instituted the same general grouping of samples was observed, although specific pair-wise distances were not maintained. Thus all measures gave agreement on the removal of the samples listed, specifically in the case of one batch.

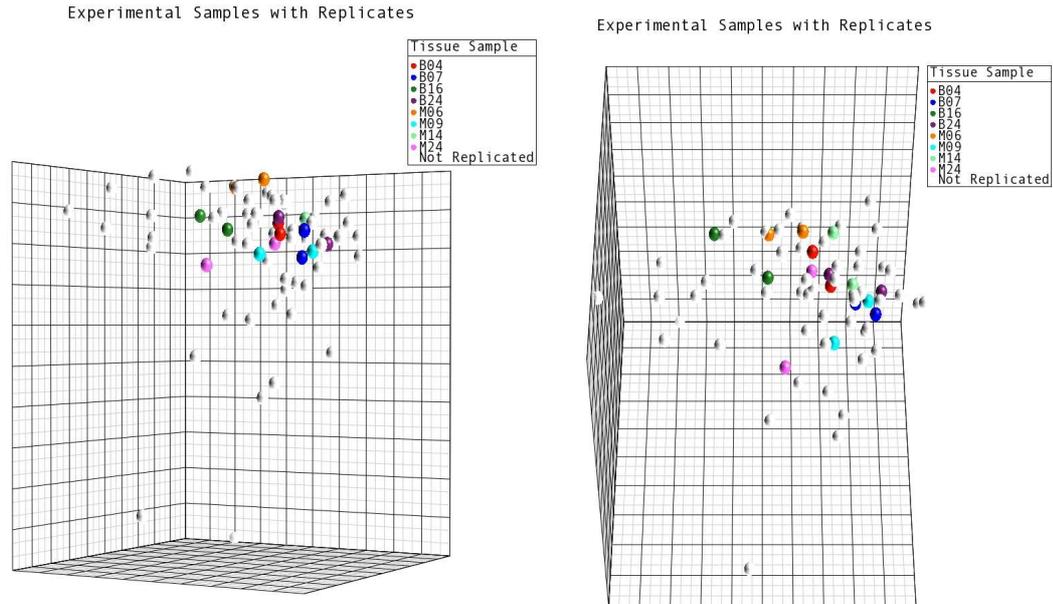


Figure 3.9: Principle components analysis of samples with their representative technical replicates. Color indicates passed samples and associated technical replicates. White samples are not relevant but could not be removed without changing the covariance structure of the PCA plot.

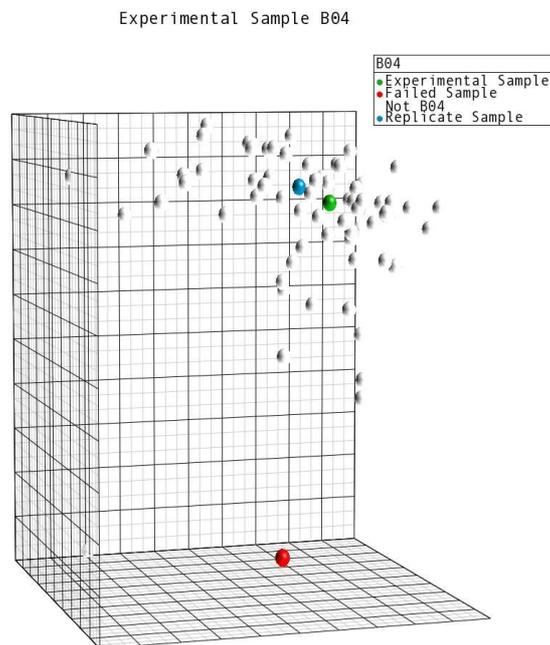


Figure 3.10: Principle components analysis of sample B04 with its technical replicate, and a failed sample (green, blue, and red respectively). White samples are not relevant but could not be removed without changing the covariance structure of the PCA plot.

3.3.3 Outlier Experimental Samples

Of the 75 samples eventually prepared, 54 samples were unique and had acceptable quality for subsequent analysis. Basic PCA of this subset of samples is shown in Figure 3.11. Color according to several classifiers was used to look for indications of substructure in the clusters. Figure 3.11 (left) compares malignant and benign samples without showing a distinct class separation. However four samples cluster together away from the main group, 2 from each class. The samples are M04, M27, B01, and B15, two are from normal epithelium, one is from tumor, and one is a tumor-stroma microenvironment sample. Interestingly, these come from only two patients: PM6 (M04 [NE] and M27 [TS]) and PB2 (B01 [NE] and B15 [T]). All but the tumor sample were outliers on the right edge of the 3D-PCA plot. Checking several of the batch factors for possible correlation with the structure did not uncover anything significant. This observation is data-cleansing method dependent; when RMA was used to calculate sample signal the structure disappeared (data not shown). The structure is not genotype-dependent as additional samples from these patients, using the same preparative reagents, did not co-cluster with the outliers. These samples were removed from the pool used in subsequent analyses since there is a clear but undefined factor exerting considerable influence on the data. After the preceding quality control steps were carried out, 50 samples remained for the analyses described below, detailed in Table 3.4.

3.3.4 Sample Factors: Random and Experimental

Many factors affect processes this complex that might influence the sample stratification, including: clinician involved in tissue harvesting, patient's other known conditions and demographics, tumor stage and origin (epithelial or mesenchymal), experimental cohort, sample preparation conditions, and hybridization processing batch. Examination of these factors for correlation with general data features showed

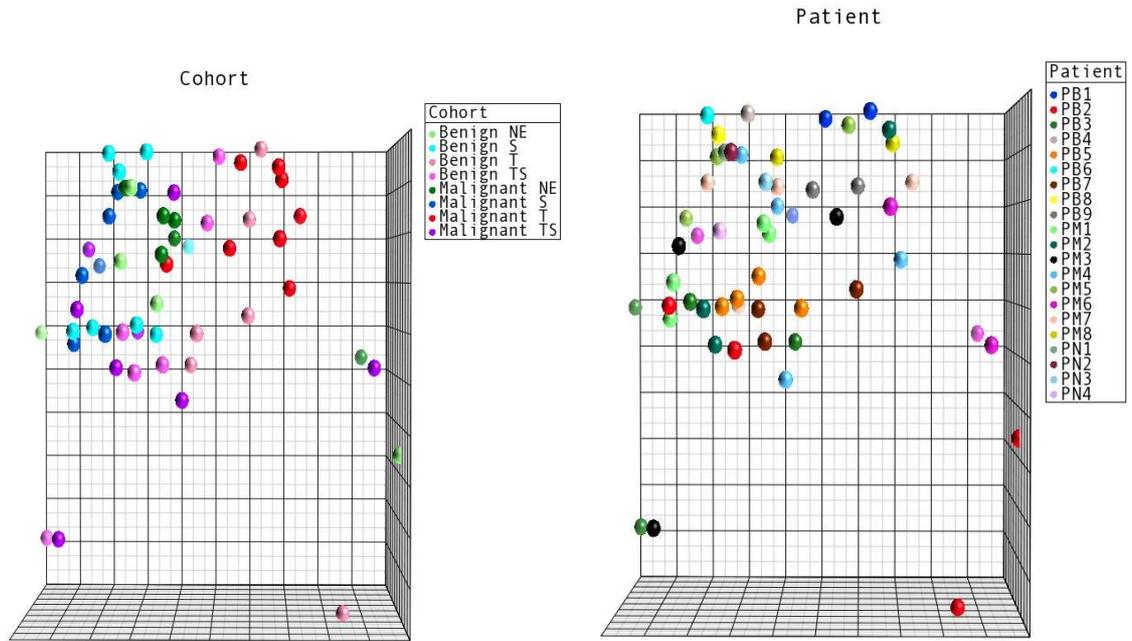


Figure 3.11: Principle components analysis of experimental samples. Coloring is by cohort (left) and coloring by patient (right) (with 4 outliers).

Table 3.4: Experiment sample count after completing all levels of quality control.

	Normal Epithelium	Tumor	Stroma	Tumor-Stroma Environment
Benign	4	5	8	6
Malignant	5	8	7	7

positive effects of varying degree for: the clinician, cohort, tissue type, tumor origin, sample preparation conditions, and hybridization processing batch. The patient tumor, patient, and age factors were not significantly correlated with general data features.

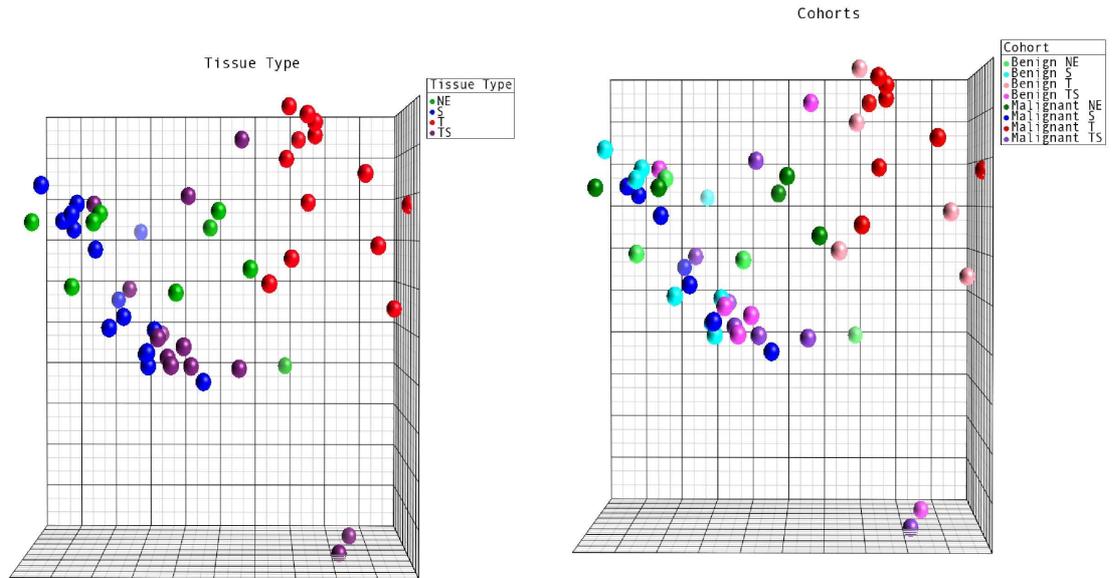


Figure 3.12: Principal components analysis of the final 50 samples, colored by tissue type (left) and cohort (right). Tissue type is indicated as: tumor (red), normal epithelium (green), stroma (blue) and tumor-stroma environment (purple). Cohorts are indicated as: tumor (shades of red), normal epithelium (shades of green), stroma (shades of blue), and tumor-stroma environment (shades of purple). The lighter shades are benign and the darker shades are malignant.

Uncontrolled factors for which some substructure was revealed included clinician (Figure 3.13), tissue origin (Figure 3.15), sample preparation (Figure 3.14), and hybridization processing batches (Figure 3.14). Most of the samples were harvested by clinicians C1 and C2 (clinician encoded as C_n, where n is 1-4); for C1 and C2 no clustering is apparent while C3 and C4 collected 4 samples which do appear to show substructure, but given the sample size this may not be significant. Since they remain within the overall cluster and do not redefine the boundary they do not need to be removed from the pool. Since the hybridization processing occurs after the sample processing, it encompasses the cumulative variation (Figure 3.14). In general, both images show a similar pattern. If samples are coded by image processing date (Figure 3.14, right) a trend is observed, in which samples appear higher in the cluster the later the scan date. This trend is typical of practice effects.

The controlled factors of interest in this study were the tissue type, patient tumor and the interaction of these two factors, designated as “cohorts”. Tissue type specified if the tissue was normal epithelium, stroma, tumor, or tumor-stroma environment. Patient tumor is defined as the tumor type of the patient the tissue was removed from (benign or malignant), no matter the malignancy of the tissue. The interaction of the factors, or cohort, included the following distinctions: normal epithelium from a patient exhibiting a benign tumor, tumor from a patient exhibiting a benign tumor, and so on for all combinations of tissue type and patient tumor. Using PCA on GCRMA intensities, the samples did not fall into clearly defined groups according to a “patient tumor” variable (data not shown), however, using either tissue type (Figure 3.12, left) or cohort (Figure 3.12, right), similar groups emerged. In Figure 3.12, left, the tumor sample group is distinct from the normal epithelium sample group, and from the stroma and tumor-stroma environment sample groups. The most dispersed sample groups are the normal epithelium and the tumor-stroma environment. The stroma group falls into a distinct cluster, completely separate from the tumor samples but overlapping the NE and TS groups. In Figure 3.12 right, we examine the four interaction variables to determine whether local structure exists within the larger tissue type clusters; no such structure can be observed. On the other hand, based on tissue origin (epithelial Vs. mesenchymal) substructure becomes apparent (Figure 3.15), although some overlap occurs.

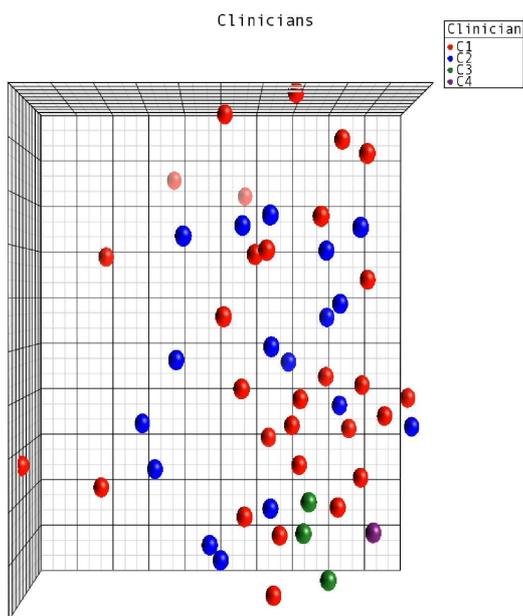


Figure 3.13: Principal components analysis of the final 50 samples, colored by clinician.

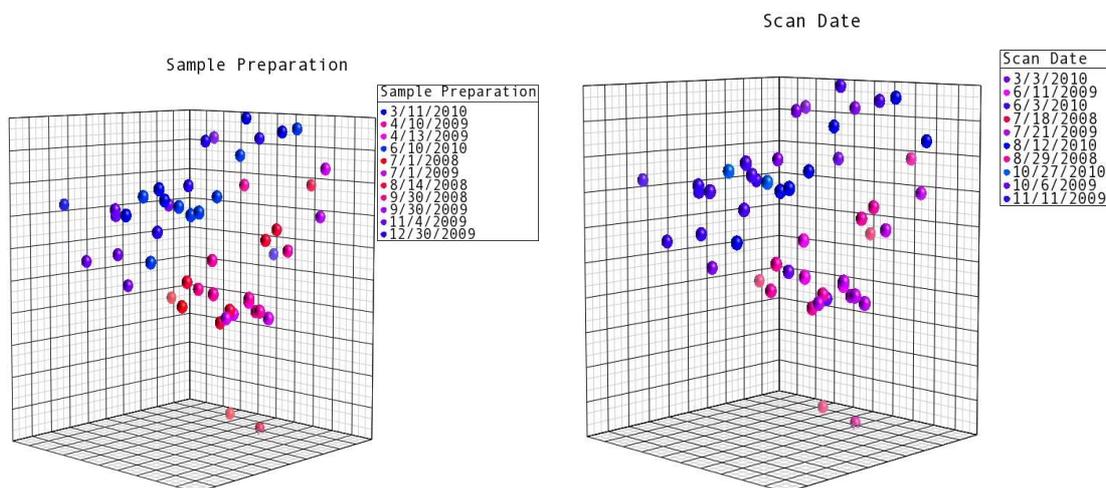


Figure 3.14: Principal components analysis of the final 50 samples, colored by sample preparation batch (left) and chip scan date (right).

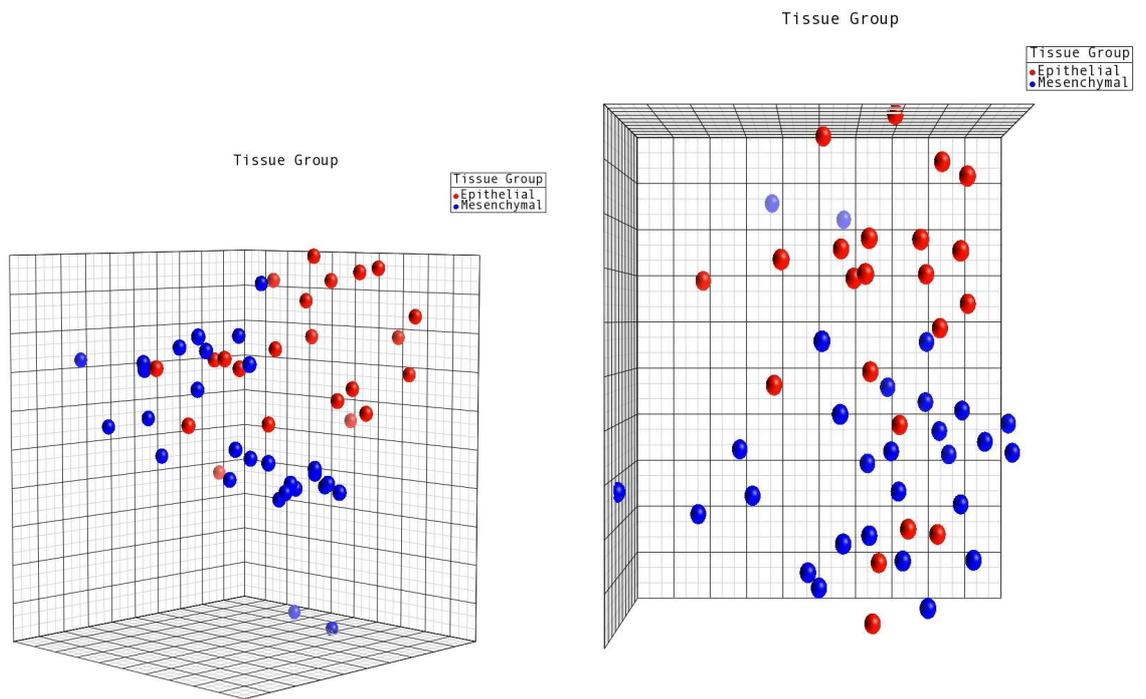


Figure 3.15: Principal components analysis of the final 50 samples colored by tissue of origin: epithelial (red) or mesenchymal (blue). Images show orthogonal views.

3.4 Gene-level Analysis

In order to perform gene-level analysis, transcripts (232,496 core ProbeSets) were collapsed into 22,027 transcript clusters (effectively genes) using a winsorized mean [153]. A 2-way ANOVA was performed, for which factor 1 was patient tumor and factor 2 the tissue type, so the interaction was patient tumor x tissue type. After performing the ANOVA, a total of 6 contrasts were performed, in order to define lists of significantly differentially expressed (DE) genes (Table 3.5). All such transcript clusters were significant with a FDR of less than or equal to 0.05 and an absolute fold change of 2 or greater.

For malignant tumors, the S x TS contrast detected no DE genes. Contrasting NE and S in malignant tumor Vs benign tumor samples detected no DE genes. In benign tumors the NE x T yielded 92 DE genes (the benign tumor cluster list) and, in malignant tumors the NE x T contrast yielded 87 DE genes (the malignant tumor cluster list) (Figure 3.16). In benign tumors the S x TS contrast revealed a single DE gene, (NRCAM, a neuronal adhesion molecule in the immunoglobulin superfamily that is involved in cell-cell adhesion and directional signaling, characterized by having many isoforms) [154] (Figure 3.16). Further inspection of the benign tumor cluster list showed the 78 of the 92 DE genes showed increased expression, while in the malignant tumor cluster list 73 of the 87 DE genes showed increased expression. Of the DE genes with increased expression, 21 were common to both lists. Both lists had a total of 14 DE genes that decreased in expression, of which 2 were common to both lists (Figure 3.17). Detailed transcript cluster lists can be found in Appendix B.

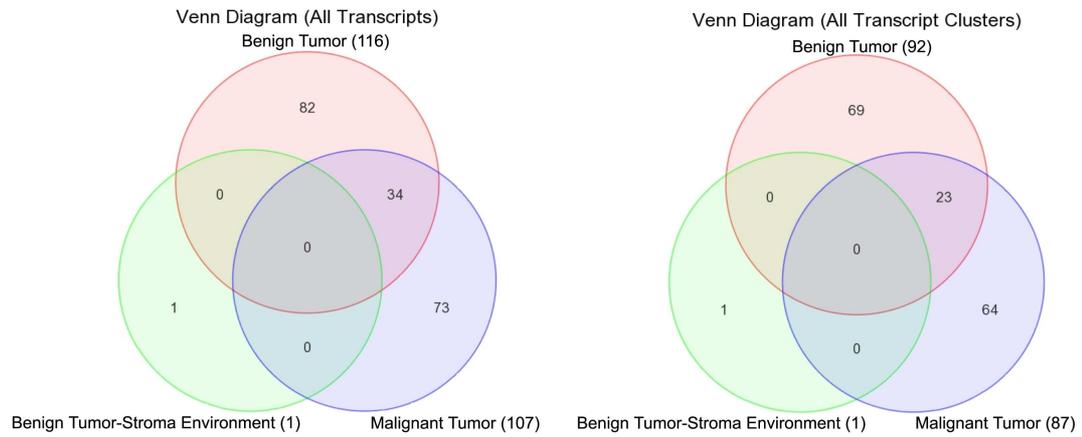


Figure 3.16: Difference and intersection counts of transcript (left) and transcript cluster (right) lists. NrCAM was found significant to both transcript and transcript cluster benign tumor-stroma environment lists.

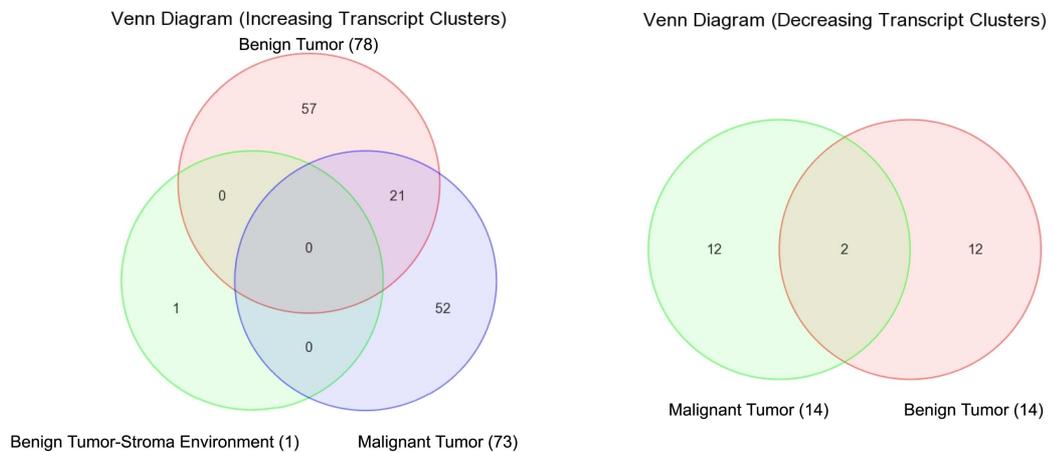


Figure 3.17: Difference and intersection counts of increasing (left) and decreasing (right) transcript cluster lists produced from the gene-level analysis. NrCAM was found significant in the increasing transcript cluster list commenting on the benign tumor-stroma environment. GNG11 and END2 were found in common in both malignant and benign tumor decreasing transcript cluster lists.

Table 3.5: *A posteriori* contrasts, performed at the transcript cluster (gene) level.

Control Cohort	Test Cohort	Gene Count	Contextualization
Normal Epithelium (Benign Sample)	Tumor (Benign Sample)	92	Benign tumor gene list
Normal Epithelium (Malignant Sample)	Tumor (Malignant Sample)	87	Malignant tumor gene list
Normal Epithelium (Malignant Sample)	Normal Epithelium (Benign Sample)	0	Differences in epithelial controls
Stroma (Benign)	Tumor-Stroma Microenvironment (Benign)	1	Effects to stroma of exposure to benign tumors
Stroma (Malignant)	Tumor-Stroma Microenvironment (Malignant)	0	Effects to stroma of exposure to malignant tumors
Stroma (Benign)	Stroma (Malignant)	0	Differences in stromal controls

3.5 Transcript-level Analysis

We selected exon arrays as the platform because it allows us to further dissect DE genes into component isoforms. Thus, after determining the DE genes, the ANOVA step was extended to the transcript level, with the same FDR and significance cut-offs. This expands the input to 221,985 ProbeSets. The same experimental factors were used as in section 3.4, and some terms were added to allow us to model exons as distinct factors (described in detail in the Inferential Statistics section of the Materials and Methods). In the benign tumor samples, the NE x T contrast revealed 116 DE transcripts (Figure 3.16, left) of which 95 were increased (Figure 3.18, left) and 21 decreased in expression (Figure 3.18, right); in the malignant tumors the NE x T contrast revealed 107 DE transcripts (Figure 3.16) of which 85 increased (Figure 3.18, left) and 22 decreased in expression (Figure 3.18, right). Across the two lists, 31 increased (Figure 3.18, left) and 3 decreased transcripts were in common (Figure 3.18, right). In the benign tumor S x TS contrast the same single DE transcript (NRCAM; NM_001037132) was identified. No DE transcripts were found for the other contrasts. All transcript DE list counts are shown in Table 3.6. Detailed transcript lists can be found in Appendix B.

Within a transcript cluster some transcripts may vary when others do not, so a second comparison was performed. The single DE gene identified in the benign tumor S x TS contrast, NRCAM, maps to the same transcript form. Of the 92 DE genes in the NE x T benign tumor list, 83 map to transcripts identified in the list of 116 DE transcripts (Figure 3.19, left). Of the 87 DE genes in the NE x T malignant tumor list, 86 mapped to the same transcript (Figure 3.19, right). The genes, transcripts and differences are discussed in detail below.

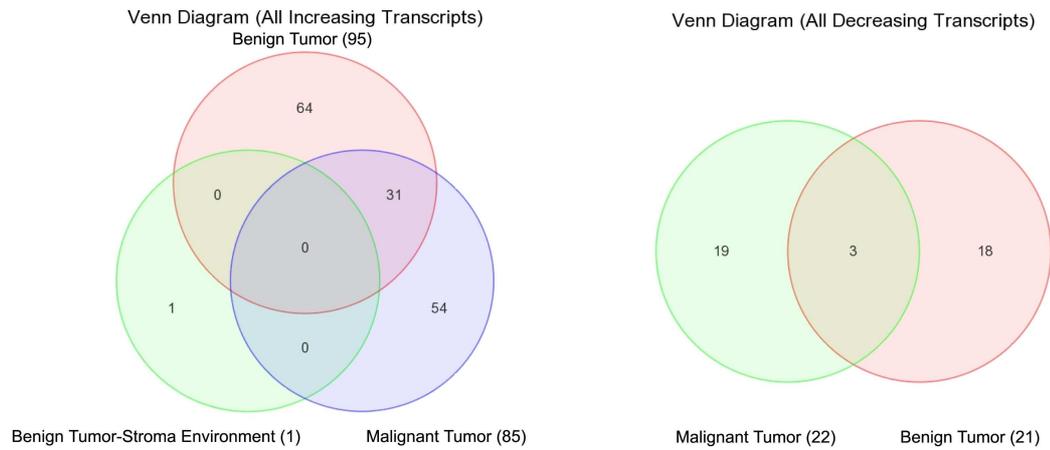


Figure 3.18: Difference and intersection counts of increasing (left) and decreasing (right) transcript lists produced from the transcript-level analysis. NrCAM was the increasing transcript found significant to the benign tumor-stroma environment. END2, GNG11, and MGP were the decreasing transcripts found significant to the benign and malignant tumor environments.

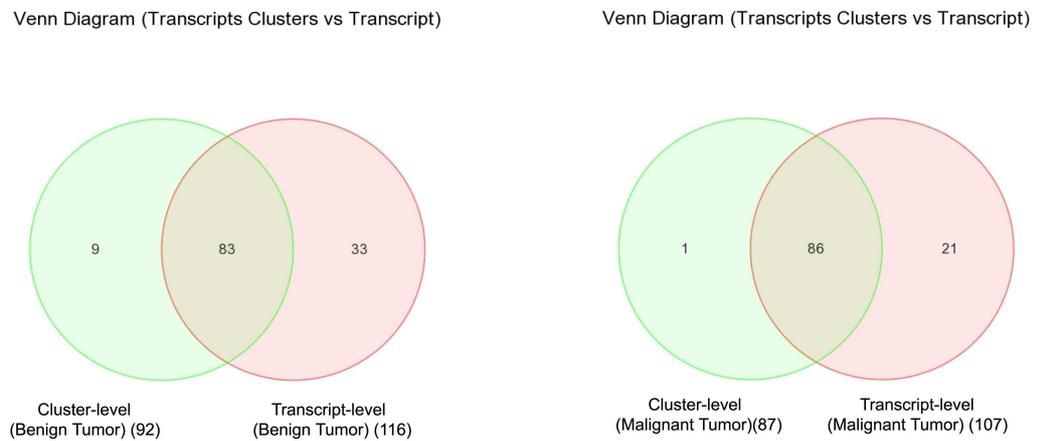


Figure 3.19: Difference and intersection counts of transcript lists Vs transcript cluster lists derived from benign tumors (left) and malignant tumors (right). KCTD1 was the transcript cluster not found when modeling malignant tumor as transcripts.

Table 3.6: *A posteriori* comparisons and contrasts, performed at the transcript level.

Control Cohort	Test Cohort	Gene Count	Contextualization
Normal Epithelium (Benign Sample)	Tumor (Benign Sample)	116	Benign tumor gene list
Normal Epithelium (Malignant Sample)	Tumor (Malignant Sample)	107	Malignant tumor gene list
Normal Epithelium (Malignant Sample)	Normal Epithelium (Benign Sample)	0	Differences in epithelial controls
Stroma (Benign)	Tumor-Stroma Microenvironment (Benign)	1	Effects to stroma of exposure to benign tumors
Stroma (Malignant)	Tumor-Stroma Microenvironment (Malignant)	0	Effects to stroma of exposure to malignant tumors
Stroma (Benign)	Stroma (Malignant)	0	Differences in stromal controls

3.6 Biological Contextualization

Significant results were found for the NE x T contrasts in benign and malignant tumors. These can be considered with respect to individual elements, or mapped in sets to pathways and systems. For the latter approach there are a number of public resources that facilitate different levels of interaction, from the genetic manipulation of metabolic pathways to the yeast two-hybrid and cross-linking methods used in protein-protein interaction studies. The results presented here used DAVID [134] [135], which has the advantage that multiple databases are accessed and the associations encompass a number of types of connections. Summary counts for results from DAVID are recorded in Table 3.7 and Table 3.8 (detailed tables are given in Appendix C). Databases are listed if significant terms resulted from them. Significance was determined by an FDR of at least 0.05. The column label “exon” indicates transcript-level results and the label “gene” indicates transcript-cluster based results. A label ‘All’ means a complete list while a label “Up” indicates the subset whose expression was increased in the direction of the contrast, while “Down” indicates the opposite. The interpretation of these results are presented in Chapter 4.

3.6.1 Chromosomal Location

It is well known that there are gross chromosomal aberrations associated with many cancers, and that there are long-range epigenetic effects in evidence. Thus co-location can be an informative feature for genes that are mutually affected even when functional roles do not overlap. DAVID reports on chromosomal location at two scales: chromosome number and cytoband, and the genes themselves have start positions relative to the reference genome build. Examination of the gene lists indicates one co-location cluster occurs for the malignant tumor increased expression at the transcript level. Chromosome 8 is significant (FDR = 0.024) with the gene set (RAD54B, SOX17, AZIN1, CHMP4, C8orf62, ESRP1, FAM49B, GRHL2, MAL2, PRKDC, and

Table 3.7: Summary of biological annotation for benign transcript and transcript-cluster lists. Abbreviations include the following usage: SP for Superfamily (Protein), PIR for Protein Information Resource, GOTerm CC for Gene Ontology Term Cellular Component followed by a number indicating specificity of terms included (the larger the more specific), FAT for Filtered Algorithm Tree, Panther MF for Protein Analysis through evolutionary relationships molecular function, UCSC TFBS for University California Santa Cruz Transcript Factor Binding Site, CGAP Sage for Cancer Genome Anatomy Project Serial Analysis of Gene Expression, GNF U133A for Genomics Institute of the Novartis Research Foundation U133A chip expression, and Est for expressed sequence tag

Database	Exon All	Exon Up	Exon Down	Gene All	Gene Up	Gene Down
SP Comment Type	1	0	0	1	0	0
SP PIR Keywords	2	3	0	0	0	0
GOTerm CC 2	3	3	0	0	0	0
GOTerm CC 3	5	6	0	0	0	0
GOTerm CC 4	5	3	0	0	0	0
GOTerm CC 5	5	7	0	0	0	0
GOTerm CC All	0	12	0	0	0	0
GOTerm CC FAT	0	12	0	0	0	0
Panther MF All	0	1	0	0	0	0
PIR Summary	2	0	0	0	0	0
PubMed	4	0	0	2	3	0
Biocarta	1	0	0	0	0	0
UCSC TFBS	75	64	0	48	57	0
CGAP Sage	4	2	0	2	1	0
GNF U133A	7	2	1	4	5	0
Unigene Est	12	7	0	9	12	0
Total	126	122	1	66	78	0

Table 3.8: Summary of biological annotation for malignant transcript and transcript-cluster lists. Abbreviations as stated in Table 3.7 with the addition of BP for biological process and EC number for enzyme commission number.

Database	Exon All	Exon Up	Exon Down	Gene All	Gene Up	Gene Down
SP Comment Type	6	10	0	12	10	0
SP PIR Keywords	0	5	0	2	4	0
GOTerm BP 2	1	0	0	0	0	0
GOTerm CC 2	0	1	1	0	0	0
GOTerm CC 3	1	0	3	2	1	1
GOTerm CC 4	0	1	1	2	2	1
GOTerm CC 5	5	3	0	4	8	1
GOTerm CC All	7	5	1	7	10	1
GOTerm CC FAT	7	15	1	7	14	1
Panther MF All	0	0	1	0	1	0
PubMed	3	7	1	3	9	0
EC Number	1	1	0	1	1	0
Reactome Pathway	0	0	0	1	1	0
UCSC TFBS	25	36	0	35	39	0
CGAP Sage	3	6	14	3	5	0
GNF U133A	8	10	7	8	8	1
Unigene Est	5	18	6	11	16	0
Total	72	119	36	98	129	6

SNTB1). Relevance to this finding to other reports is given in Chapter 4.

3.6.2 Gene Ontology Analysis

Functional roles are assigned to genes according to controlled vocabularies and the curation of domain experts. The most widely used controlled vocabulary is the Gene Ontology, or GO [148]. The hierarchy contains three independent branches, biological process (BP), cellular component (CC), and molecular function (MF), providing themes that link individual gene products. The significantly enriched CC terms in the lists for both benign and malignant tumors include a number of expected categories: plasma membrane, apical junctions, cell junctions, occluding junction, tight junctions, cytoskeleton, extracellular matrix, as well as condensed chromosome centromeric region. Only one significant BP annotation was reported for malignant tumor, cellular component organization, which was associated with a transcript list using all transcripts (up and down). Significant enrichment of MF terms was reported for two of the lists (increasing benign tumor transcripts ‘Up’ and malignant tumor transcripts ‘Down’); the terms were “Cell junction protein” and “Extracellular matrix”, respectively. Gene and transcript mappings, their concordance with current models, and other interpretations are given in Chapter 4.

3.6.3 Pathway Analysis

DAVID accesses several databases containing pathway and protein interaction information, including: Biocarta, KEGG, Panther, Reactome, BIND, DIP, MINT, NCICB, and UCSC TFBS. From these, significant interactions were reported based on data in Biocarta, Reactome, and UCSC TFBS. In the benign tumor transcript list (All) Biocarta reported significance for the SUMOylation pathway. Small Ubiquitin-Like Modifiers (SUMOs) are proteins that post-translationally modify other proteins over many processes like apoptosis, cell cycle progression, response to stress, transcriptional stability and protein stabilization; the most common outcome is to modify

localization or binding partners, and generally only a fraction of the total protein is modified. For the malignant tumor transcript list (Up), Reactome reported significant over-representation of cell division genes. Another form of interaction is the gene regulatory network, which is connected by transcription factors that co-regulate spatially distributed genes whose products are needed in a spatially and/or temporally connected function. When reducing scope from known pathways to potential interactions between elements in the lists, UCSC TFBS found many potential interactions from all lists except the decreasing ‘transcript’ and ‘transcript cluster’ lists.

3.6.4 Gene Interaction Analysis

Differential gene lists from benign and malignant tumor samples were used to derive networks of interaction. Two tools were used to search gene product interaction databases for known interactions and build a network. Initially, STRING8 was used to derive pathways for both the benign (Figure 3.20) and malignant (Figure 3.21) transcript lists. Only experimentally verified interactions were shown. Genes2Network was then used to augment the STRING8-based network with inferred relationships (Figure 3.22, benign, and Figure 3.23, malignant).

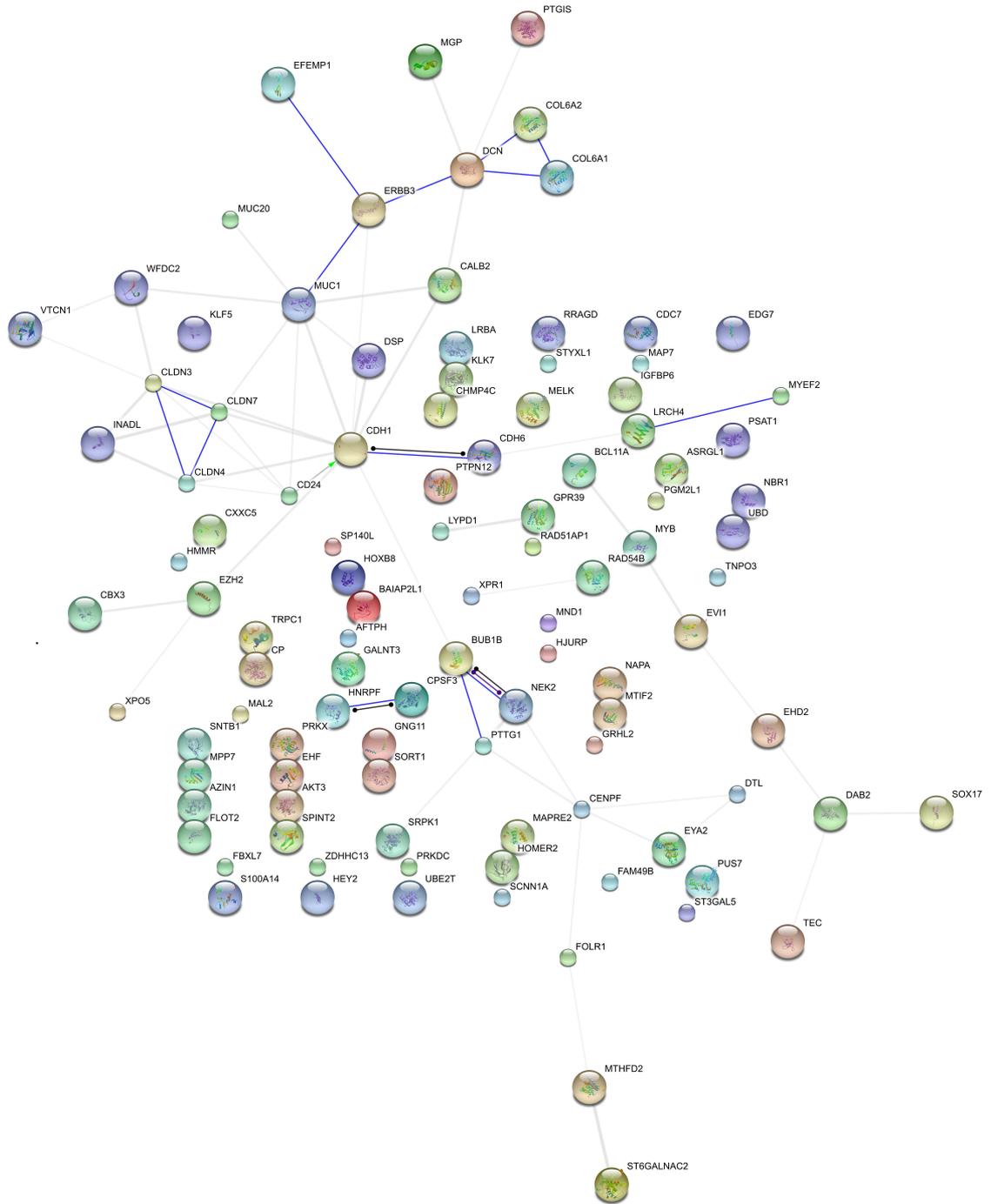


Figure 3.21: Network of interactions between transcript products from the malignant transcript list (derived from STRING8).

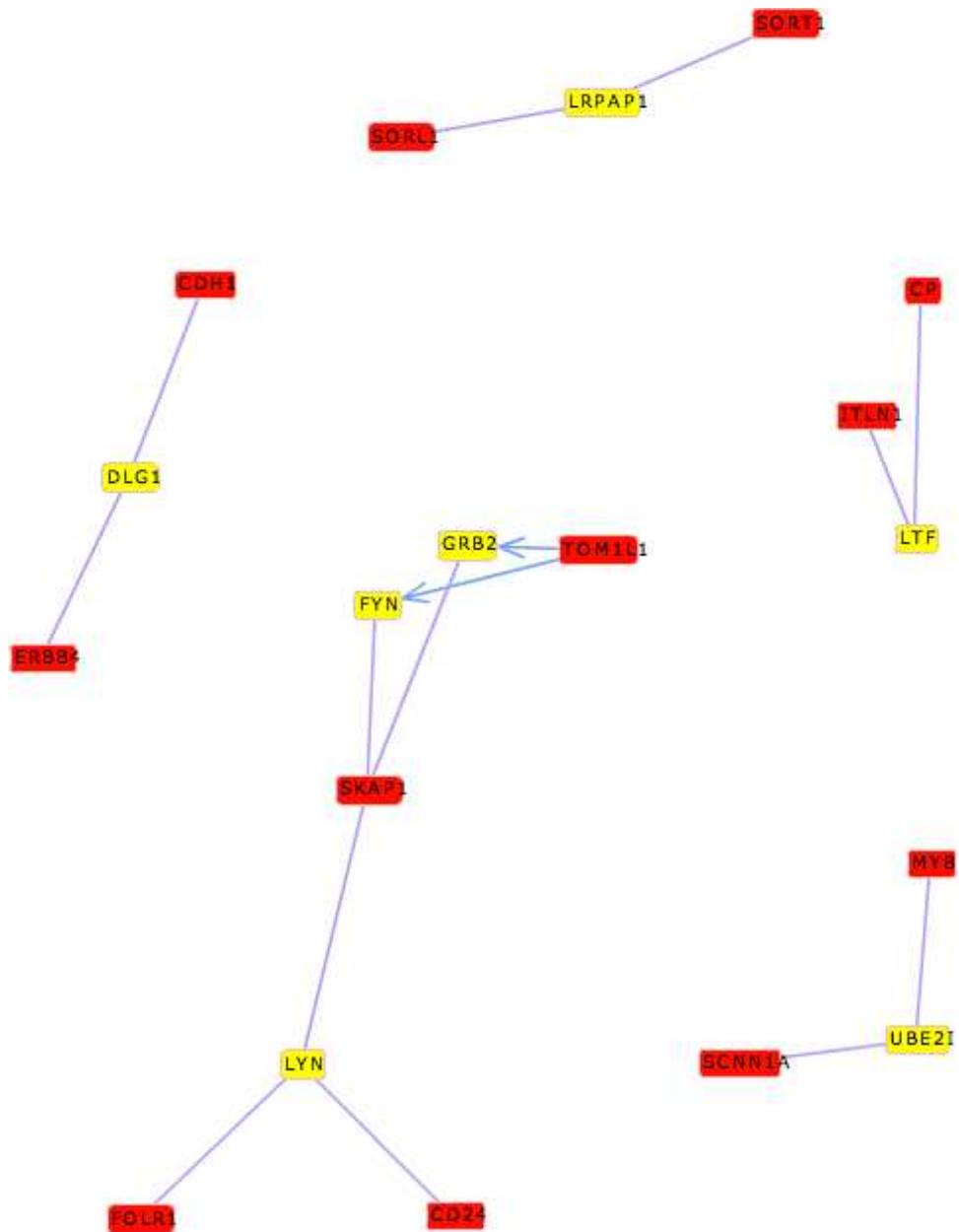


Figure 3.22: Network of interactions between transcript products from the benign transcript list (derived from Gene2Network). Red indicates a seed node of the network, yellow indicates significant nodes (z-scores > 2.5).

CHAPTER 4: DISCUSSION

4.1 Notable Comparisons

Both transcript and gene (transcript cluster) lists were derived using ANOVA models. Respectively, benign tumor, malignant tumor, and benign tumor-stroma environment had 92, 87, and 1 significant DE genes and 116, 107, and 1 transcripts. Global properties were evaluated using 3D-PCA of the 50 samples, in which the most spatially distinct clusters were the tumors (benign and malignant). The other categories clustered, but showed more overlap. Intuitively this separation makes sense given the subsequent finding that a majority of DE genes and transcripts were regulated in one direction (Up), rather than having equal numbers up and down. The overlap in the benign and malignant samples at the global level is reasonable since a number of DE genes were found in common.

The gene lists were somewhat shorter ($\sim 10\%$ fewer members) than the transcript lists. This may be due to combining transcripts into one cluster. Since this data set allows discrimination of exons, the primary focus in the following interpretation of the results is on transcript-level analysis. That is, annotation was evaluated for significance at both gene and transcript levels for gene networks, but pathway analyses were performed and evaluated only with the transcript lists, except in the rare cases where the network analysis only yielded results with gene input lists, in which case that information was carried forward to the pathway analysis.

4.2 Control Sample Evaluations

In our opinion it is very significant that samples identified as normal epithelium were very similar, regardless of the patient category. This observation is only possible

for LCM isolated cells since this cell type occurs in such a small percent of the total; thus the purification procedure was effective. When microdissected, care was taken to identify epithelium that appeared histologically normal; this visual confirmation is documented in Appendix A. In addition to confirming that the technical procedures were correctly deployed, there is biological significance in this transcriptome profile consistency: despite being in the neighborhood of abnormal cells, the epithelial cells do not exhibit signs of significant perturbation. We infer from this stable baseline that high confidence can be placed in those changes that are observed in other samples.

We identified no DE genes or transcripts in the profiles of distal stromal cells, regardless of disease classification. While not biologically interesting, it again validates the precision of the identification and sampling method, and increases our confidence in the conclusions drawn from comparisons of tumor cells and their immediate neighbors. Specifically, previous reports have shown that there is a change in stromal cells adjacent to tumors, but the reports lacked specific distance information. This study clarifies the relationship, showing that stromal cells immediately adjacent to tumor cells are affected and more distant stromal cells are not detectably influenced by the tumor with respect to transcripts. In fact, although we state that adjacent stromal cells are affected by the presence of tumor, only one transcript was found to be significantly different; this may reflect the need for more optimization in the cell selection technique, or limits in the sample size of this study, or limits in the current data cleansing methods, which are not very sensitive.

4.3 Ontology Enrichment

The complete results of the bioinformatics analyses that used Gene Ontology labels as the core unit are given in Appendix C; the following paragraphs are focused on those categories showing the greatest enrichment. In general the transcript and gene lists yielded very similar terms and significance scores, as expected given the overlap in the lists. When unique occurrences of a term were counted, the longer transcript

Table 4.1: Summary of unique Cellular Component (Gene Ontology) term counts for transcript and transcript-cluster lists associated with benign and malignant tumors.

	Benign	Malignant
Transcript (All)	8	8
Transcript (Increasing)	13	15
Transcript (Decreasing)	0	3
Transcript Cluster (All)	0	7
Transcript Cluster (Increasing)	0	13
Transcript Cluster (Decreasing)	0	3

lists yielded more terms. Not all lists yielded significantly enriched terms, including benign tumor transcripts in the Cellular Component category, as shown in Table 4.1. The relationship between these terms is visualized in Figure 4.1 (benign) and Figure 4.2 (malignant).

In Figure 4.1, there are two areas of primary enrichment, organelles and plasma membrane, the first with more cytoskeletal terms and the second with more membrane and junction terms. In Figure 4.2 a similar projection for the malignant tumor set is shown. As in the benign tumor example, there is enrichment for organelles and plasma membrane, and many of the subterms are shared. Interestingly, an additional category of terms referring to the extracellular region is found in the malignant ontologies. The “Extracellular Region” term contains two significant terms “Extracellular Matrix” and “Proteinaceous Extracellular Matrix”, both of which are associated with decreased levels of transcripts. The only other Gene Ontology term associated with transcripts present in lowered amounts is “Extrinsic to Membrane” which is only found in the malignant ontology term set. Plasma membrane-linked terms are most common between benign and malignant terms and differ only by the addition of “Apical Plasma Membrane” to the malignant set. With respect to organelle-linked terms, while the parent node is similar the leaves differ, with “Spindle” and “Condensed chromosome, centromeric region” dominating the malignant set rather than

“Cell projection” and “Cilium”. Subsequent sections examine these categories and the contributing transcripts in detail. The relationship between these terms is visualized in Figure 4.1 (benign) and Figure 4.2 (malignant).

4.3.1 Cell Junctions

Cell junctions are a major theme in both benign and malignant transcript lists, and many terms are shared: cell, cell-cell, occluding, tight, and apical junction (complex).

In the benign ontology set, it is the increased concentration of nine transcripts that leads to the enrichment of these ‘junction’ terms: ARHGAP26, CDH1, CLDN4, CLDN7, DSP, INADL, MPP7, PARD6B, PERP. Of these genes, ARHGAP26 is a rho GTPase-activating kinase that binds to focal adhesion kinase, involved in organization of the actin cytoskeleton. Defects in this gene are a cause of juvenile myelomonocytic leukemia, and this gene was found to be down-regulated in other myeloid malignancies [155]. It is regulated by Myb-b among other transcription factors. CDH1 (epithelial- or E-cadherin) is a calcium dependent adhesion glycoprotein of the cell membrane, and decreases have been correlated to the presence of gastric, breast, colorectal, thyroid and ovarian cancer. In contrast to our observations here, in which it has a higher relative level in both benign and malignant tumors, a common observation is that expression of CDH1 decreases proportionally to the severity of cancer stage [156]. It has a proposed role in the mechanism of the important epithelial-mesenchymal transition [157]. It is the central element to the benign tumor network (shown graphically in Figure 4.3), having confirmed interactions with 11 elements. It is regulated by P53 and FOXD1 among other factors. CLDN4 and CLDN7 are both members of the claudin family, integral membrane proteins that are part of the tight junction complex, and regulate cell polarity and signal transduction through the paracellular space. Claudin 4 and claudin 7 are commonly found to be over-expressed in pancreatic, prostate, breast and ovarian cancer (CLDN4) and in addition in hepatocellular, urinary, lung, neck and head and thyroid cancers (CLDN7), and, in ovarian carci-

noma, are among the most highly expressed genes [158][159]. CLDN7 has additionally been identified as an independent predictor of poor survival in ovarian carcinoma [159]. They are regulated by c-Jun, c-Fos, JunD (CLD4) and p53, MEF-2A and NF-1 (CLDN7) among other transcription factors. Other transcripts are specifically important to tight junction formation, including INADL [160], a protein important to protein-protein interactions found in tight junctions and the apical membrane of epithelial cells and with receptors that is regulated by AP-1, c-Fos, c-Jun and JunD among others. Higher expression has been associated with progression of cervical cancers associated with HPV infections [161]. MPP7 [162] is a membrane-associated guanylate kinase -palmitoylated 7 adaptor protein that participates in assembling protein complexes at the sites of cell-cell contact, such as tight junctions. It maintains epithelial cell polarity, and is regulated by FOXO3a,b and HNF-1A. PARD6B [163], is a cytoplasmic protein involved in cell polarization processes as part of complexes, serving as an adaptor, specifically in epithelial tight junctions; it is regulated by the SRC-3 coactivator oncogene, and by Egr2,3, FOXL1 and E2F-3a,4,5 as well as others. Desmosomes are also represented with DSP, desmoplakin, whose expression increases. Desmoplakin is an obligate component of desmosomes, anchoring intermediate filaments to plaques. They have been shown to gradually reduce in count with tumorigenesis in both squamous epithelium and cervix uteri [164] [165]. More specific to DSP expression, decreased expression in oral and pharyngeal squamous cell carcinomas is associated with metastases [166] [167]. DSP is regulated by c-Jun, c-Fos, NFkB, p53, c-Myc and GATA-1 among others. PERP, the TP53 apoptosis effector gene, is a component of desmosome junctions, as well as the TP53-dependent apoptotic pathway. The protein is localized to the plasma membrane, and is a tumor suppressor gene down-regulated in metastatic melanoma and mammary carcinoma [168] [169] [170].

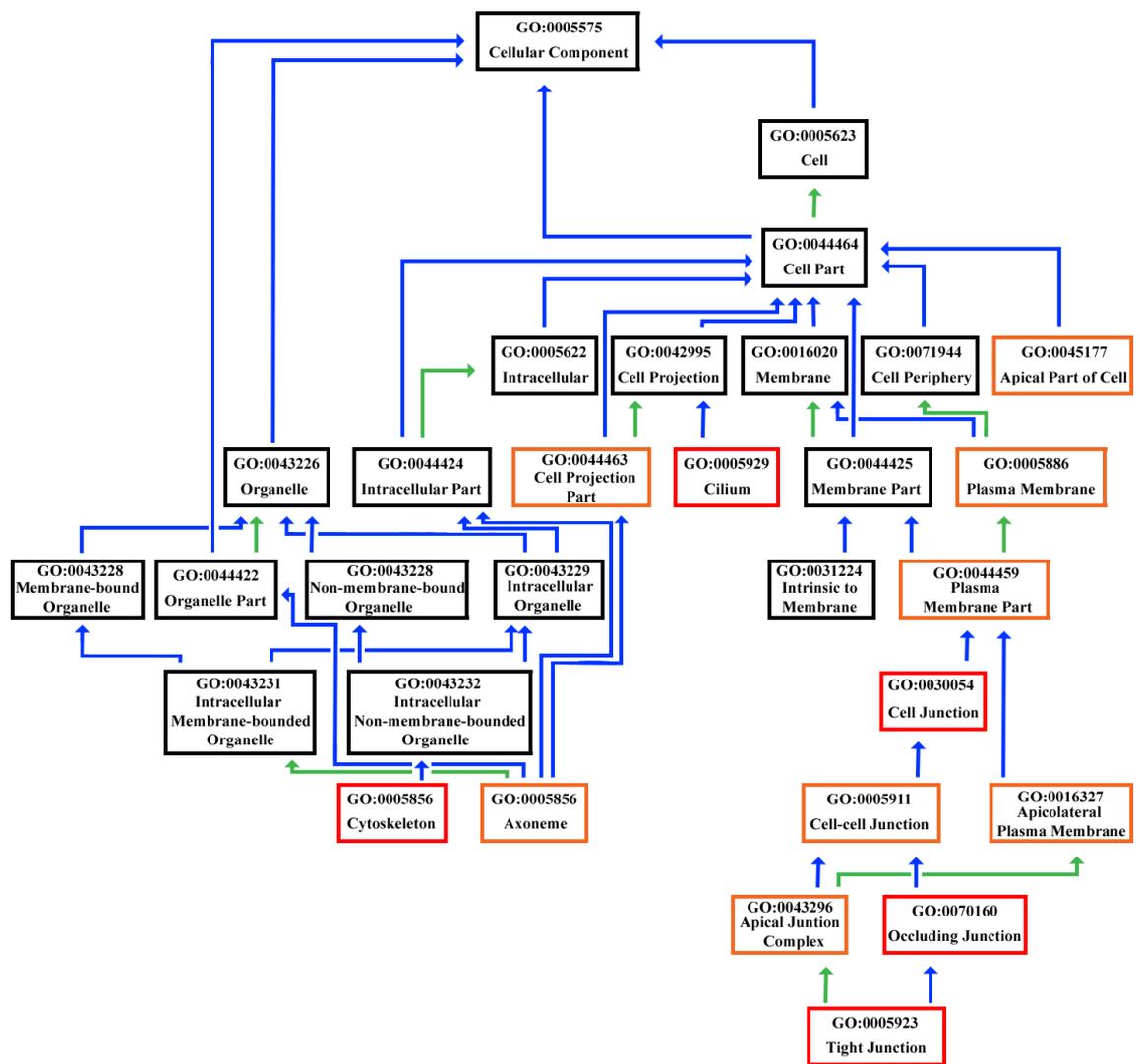


Figure 4.1: Graph of GO terms associated with the benign transcript lists. Terms are placed in similar locations between the two ontology figures so relationships between benign and malignant ontologies are clear. Border colors on ontological terms indicate which transcript list yielded the term. Orange indicates the term arose from increasing transcripts, blue indicates derivation from decreasing transcripts, and red indicates significance emerged only when all transcripts were used. A black border indicates no significance but give context in GO hierarchy. Green lines indicate an “is a part of” relationship between terms while blue indicates an “is a” relationships between terms. Originally, AmiGO was used to visualize the relationship between terms, the resulting graph was redesigned in Photoshop, for clarity, without changing details or context.

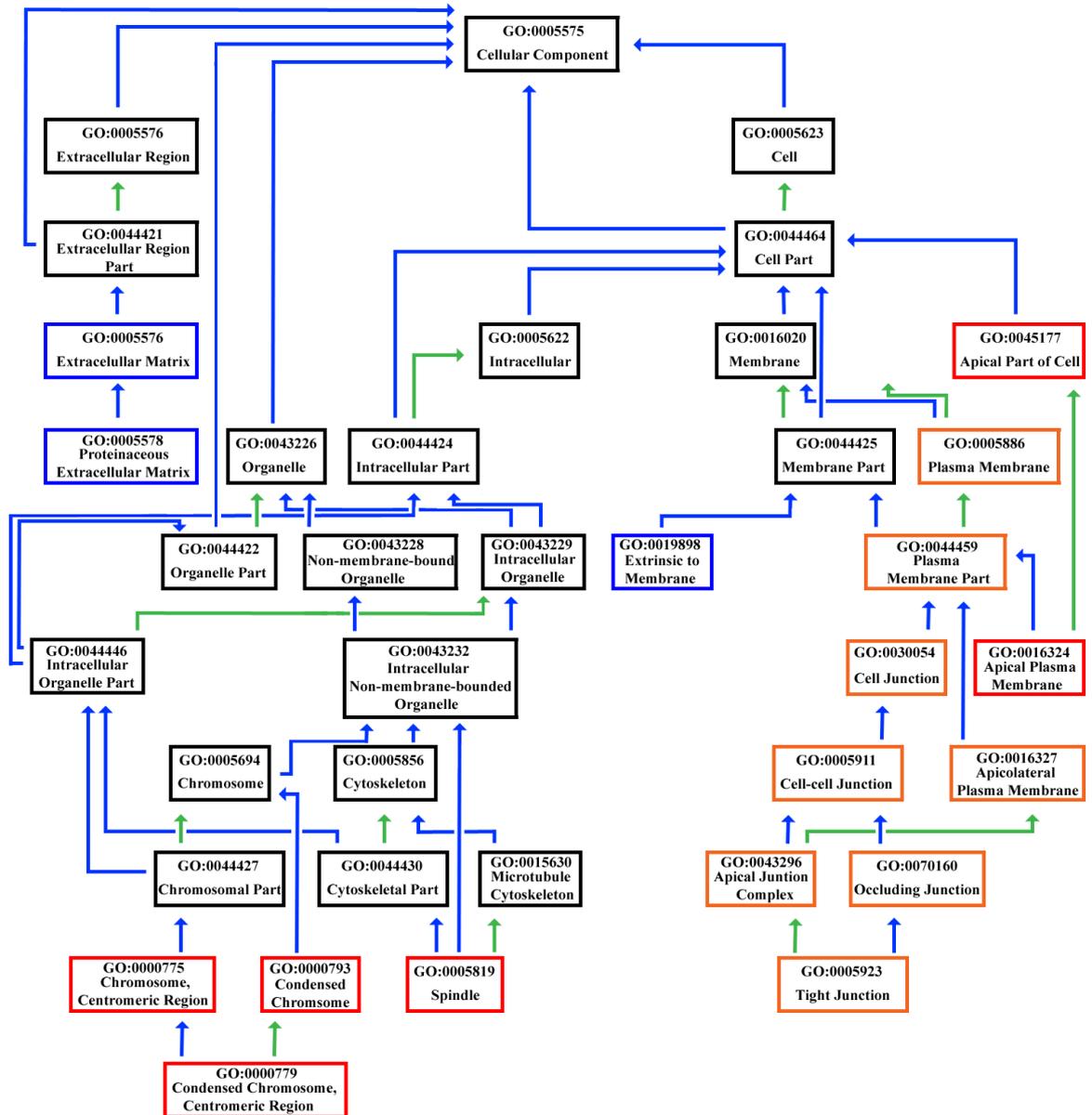


Figure 4.2: Graph of GO terms associated with the malignant transcript lists. Figure generated as stated in Figure 4.1.

Of the above genes, three (ARHGAP26, PARD6B, and PERP) are not found in the malignant transcript list. Both ARHGAP26 and PERP showed expression changes consistent with a tumor which was not malignant (PARD6B has not been tied to tumors). ARHGAP26 is expected to be decreasing in myeloid malignancies but is increasing in this study; PERP is a tumor suppressor and is known to increase in benign tumors. The expression levels in these samples may indicate that their presence signals a block to the transition to malignant tumor.

The junction-related terms in malignant tumors were driven by transcripts of the following genes: CALB2, CDH1, CLDN3, CLDN4, CLDN7, DSP, GABBR1, HOMER1, INADL, MPP7, and SNTB1, of which six are common between the benign and malignant junction terms (CDH1, CLDN4, CLDN7, DSP, INADL and MPP7). The 5 uniquely associated with malignant samples are: CALB2, CLDN3, GABBR1, HOMER1, SNTB1, of which CALB2, calbindin 2, and GABBR1, gamma-aminobutyric acid B receptor 1, exhibit decreased levels. CALB2 is an intracellular calcium-binding protein in the troponin C superfamily, and has been identified as a marker in some cancers. It is regulated by *Cart-1* and CREB among other factors. Because of its other functions it also appears with the extracellular region terms and is additionally discussed in that section. The additional claudin gene, CLDN4, is again an integral membrane protein, with receptor capabilities and similarity to an apoptosis gene in rats. Of the other genes, SNTB1 (syntrophin, beta 1) is a cytoskeletal protein of the peripheral membrane, responsible for organizing the subcellular localization of many membrane proteins, linking receptors to the actin cytoskeleton. It is regulated by p53, POU3F2 and STAT5A among factors. GABBR1 is a signal receptor for the main inhibitory neurotransmitter of the mammalian CNS, in this case for the state of prolonged inhibition, perhaps related to suppression of pain symptoms. It is considered a therapeutic target for pancreatic cancer [171]. Activity is mediated by many G-proteins that act by down-regulating activated calcium channels (of CALB2

above). It is regulated by NFkB,1, Egr-1, CREB, FOXC1 and MEF-2A among other factors. The homer homolog 1 (HOMER1) gene encodes a scaffolding protein that binds and crosslinks the cytoplasmic regions of a number of proteins, and regulates group 1 metabotropic glutamate receptor function, which inhibits apoptosis [172], coupling surface receptors to intracellular calcium release. There is evidence that isoform 3 is involved in structural changes at synapses during neuronal development. It is regulated by MEF-2A, POU3F2, CREB and other factors.

The tight junction figures prominently in the set above; tight junctions are one type of cell-to-cell adhesion in epithelial or endothelial cell sheets, forming continuous seals around the cells as barriers that prevent solutes and water from passing freely through the paracellular space. They occlude movement in the plasma membrane, segregating the apical (i.e. edge exposed to lumen) and basolateral (adjacent to cells and basement membrane) regions of the plasma membrane. Claudins are integral to tight junctions. Although a decrease in tight junctions is most frequently observed in tumorigenesis, ovarian tumors do not fit this standard, and our results are in agreement with those reports. Of the three claudins found to have significant expression increase in this study, CLDN3 is the only transcript cluster containing a transcript exclusively significant to malignant tumors. Given claudins hallmark over-expression in malignant ovarian tumors, not benign tumors, this was initially contrary to literature. One study found CLDN3/4 overexpression in ovarian carcinomas but not cystadenomas [173]. This would support our CLDN3 results, but does not explain CLDN4 or, by inference, CLDN7. The assay in the study cited used immunohistochemistry, and documented CLDN4 staining in up to 14% of serous cystadenomas. When evaluated with RT-PCR, CLDN4 transcript was present at 10-fold the level of CLDN3 transcript. A troubling aspect of the study is the absence of a positive control in the adenoma staining (negative result) group, although there was a negative control for the carcinoma staining (positive result) group [173]. Thus the disparate

CLDN4 results in benign tumors may reflect the more sensitive, quantitative nature of oligonucleotide arrays over immunohistochemistry.

Table 4.2: Summary of transcripts related to Cell Junctions, specifically “cell”, “cell-cell”, “occluding”, “tight”, and “apical junction (complex)” terms. Listed are the gene symbol associated with the transcript, the state of the tumor (specifically benign, malignant, or both), the expression relative to the control tissue (given as increasing or decreasing), and the associated ontological term/s. Terms in parentheses are related to the transcripts but are not the direct subject matter of the table (1 of 3).

Gene Symbol	Relative Expression	Tumor State	Term
ARHGAP26	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cell junction (Cytoskeleton) (Plasma membrane) (Plasma membrane part)
CALB2	Decrease Decrease (Decrease) (Decrease)	Malignant Malignant (Malignant) (Malignant)	Cell junction Cell-cell junction (Plasma membrane) (Plasma membrane part)
CDH1	Increase Increase (Increase) (Increase) (Increase) (Increase)	Both Both (Both) (Benign) (Both) (Both)	Cell junction Cell-cell junction (Apical part of cell) (Cytoskeleton) (Plasma membrane) (Plasma membrane part)
CLDN3	Increase Increase Increase Increase (Increase) (Increase) (Increase)	Malignant Malignant Malignant Malignant (Malignant) (Malignant) (Malignant)	Apical junction complex Cell junction Cell-cell junction Occluding junction Tight junction (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)
CLDN4	Increase Increase Increase Increase Increase (Increase) (Increase) (Increase)	Both Both Both Both Both (Both) (Both) (Both)	Apical junction complex Cell junction Cell-cell junction Occluding junction Tight junction (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)

Table 4.3: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
CLDN7	Increase Increase Increase Increase Increase (Increase) (Increase) (Increase)	Both Both Both Both Both (Both) (Both) (Both)	Apical junction complex Cell junction Cell-cell junction Occluding junction Tight junction (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)
DSP	Increase Increase Increase (Increase) (Increase) (Increase) (Increase)	Both Both Both (Malignant) (Benign) (Malignant) (Malignant)	Apical junction complex Cell junction Cell-cell junction (Apicolateral) (plasma membrane) (Cytoskeleton) (Plasma membrane) (Plasma membrane part)
GABBR1	Decrease (Decrease) (Decrease)	Malignant (Malignant) (Malignant)	Cell junction (Plasma membrane) (Plasma membrane part)
HOMER1	Increase (Increase) (Increase)	Malignant (Malignant) (Malignant)	Cell junction (Plasma membrane) (Plasma membrane part)
INADL	Increase Increase Increase Increase Increase (Increase) (Increase) (Increase) (Increase)	Both Both Both Both Both (Both) (Both) (Both) (Both)	Apical junction complex Cell junction Cell-cell junction Occluding junction Tight junction (Apical part of cell) (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)

Table 4.4: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
MPP7	Increase Increase Increase Increase Increase (Increase) (Increase) (Increase)	Both Both Both Both Both (Both) (Both) (Both)	Apical junction complex Cell junction Cell-cell junction Occluding junction Tight junction (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)
PAR6B	Increase Increase Increase Increase Increase (Increase) (Increase) (Increase) (Increase)	Benign Benign Benign Benign Benign (Benign) (Benign) (Benign) (Benign)	Apical junction complex Cell junction Cell-cell junction Occluding junction Tight junction (Apical part of cell) (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)
PERP	Increase Increase Increase (Increase) (Increase) (Increase)	Benign Benign Benign (Benign) (Benign) (Benign)	Apical junction complex Cell junction Cell-cell junction (Apicolateral) (plasma membrane) (Plasma membrane) (Plasma membrane part)
SNTB1	Increase (Increase) (Increase)	Malignant (Malignant) (Malignant)	Cell junction (Plasma membrane) (Plasma membrane part)

4.3.2 Spindle and Condensed Chromosomes

A unique set of terms was significantly enriched in the malignant tumor results, relating to spindle functions, which are involved with cell cycle and mitosis. “Spindle” term enrichment was due to the increased expression of 5 transcripts: NEK2, BUB1B, CENPF, RCC2, and CBX3. The Never in Mitosis Gene a-related Kinase 2, NEK2, functions in mitotic regulation as an integral component of the mitotic spindle-assembly checkpoint. It phosphorylates SGOL1, and isoforms have distinct roles related to the cell-cycle. Increased expression has been confirmed in ovarian cancer cell lines (SKOV3 and OVCAR5) as well as cervical, breast, prostate, and leukemic cell lines. In breast tissue cell lines, increases of NEK2 cause cells to become multinucleated with supernumerary centrosomes [174]. It is regulated by MIF-1, PPAR1,2, Egr-3 and other factors. Budding uninhibited by benzimidazoles 1 homolog beat (BUB1B) is a kinase involved in spindle checkpoint function, localized to the kinetochore, that helps ensure proper chromosome segregation. It is dysregulated in many tumors: for example, in bladder carcinoma an increase in expression of BUB1B was shown to be associated with variation in centromere count and, by extrapolation to chromosomal instability (CIN) [175]. It has been implicated as an apoptosis trigger in polyploid cells and may play a role in tumor suppression. It is SUMOylated by SUMO2, and is regulated by PPAR-1,2, MEF-2A, GATA-2 factors, among others. Centromere Protein F (CENPF) is known to be associated with CIN in breast cancer and to be correlated with poor patient prognosis [176]. The gene product associates with the centromere-kinetochore complex at the onset of mitosis [177] and is part of the nuclear matrix in G2; it likely has a role in chromosome segregation in mitosis. Autoantibodies to this protein are found in patients with cancer. It regulates membrane recycling. It is regulated by GATA-1, RP58, δ CREB, Pax-6 and other factors. Regulator of chromosome condensation 2 (RCC2) is required for completion of mitosis and cytokinesis and may be a guanine nucleotide exchange fac-

tor for RAC1. It has been shown to be involved in histone methylation [178], and it interacts with microtubules. It is phosphorylated upon DNA damage. Additionally, RCC2 was shown to be necessary for mitotic spindle formation in a siRNA experiment where silencing RCC2 caused spindle assembly suppression, and activation of the spindle cell-cycle checkpoint[179]. It is regulated by c-Fos, JunD, Junb and other factors. Lastly, chromobox homolog 3 (CBX3) binds DNA and is a component of heterochromatin, but also binds lamin B receptor (a nuclear membrane protein), so it may explain the association of heterochromatin with the inner nuclear membrane [180].

A subset of terms, “Condensed chromosome, centromeric region”, was significantly enriched by 4 of the above transcripts (BUB1B, CENPF, RCC2, and CBX3) and one that is unique, the Holliday junction recognition protein (HJURP). This protein plays a primary role in the incorporation and maintenance of the Histone H3 variant called CENPA. It is required for the incorporation of CENPA into nucleosomes at replicated centromeres, perhaps by stabilizing its protein during late telophase/early G1 [181]. We note that this interaction was not placed in the gene network given that CENPF, not CENPA, was found significant in this study.

Both increases in mitotic activity and associations with CIN would be expected of malignant tumors so the association of the malignant tumor samples with these terms is consistent with reports in the literature. Since the enrichment was absent in the benign tumor samples the processes described are likely required in the steps by which tumors progress to the malignant state.

4.3.3 Terms Related to Cell Projection

Benign tumor enriched terms are related to cell projection, cilia, and axonemes, due to the increased presence of five transcripts for “Cilium” (CD24, DNAH11, DNAH7, LRRC50, and SPAG17), 6 for “Cell projection part” (CD24, DNAH11, DNAH7, LRRC50, SLC34A2, SPAG17) of which 5 are common to the previous

Table 4.5: Summary of transcripts related to “Spindle” and “Condensed Chromosome Centromeric Region” terms. Listed are the gene symbol associated with the transcript, the state of the tumor (specifically benign, malignant, or both), the expression relative to the control tissue (given as increasing or decreasing), and the associated ontological term/s. Terms in parentheses are related to the transcripts but are not the direct subject matter of the table.

Gene Symbol	Relative Expression	Tumor State	Term
BUB1B	Increase Increase	Malignant Malignant	Spindle Condensed chromosome centromeric region
CBX3	Increase Increase	Malignant Malignant	Spindle Condensed chromosome centromeric region
CENPF	Increase Increase	Malignant Malignant	Spindle Condensed chromosome centromeric region
HJURP	Increase	Malignant	Condensed chromosome centromeric region
NEK2	Increase	Malignant	Spindle
RCC2	Increase Increase (Increase)	Malignant Malignant (Benign)	Spindle Condensed chromosome centromeric region (Cytoskeleton)

set, and 4 for “Axonemes” (DNAH11, DNAH7, ITLN1, LRRC50, and SPAG17), all shared with the previous sets. ITLN1, Intelectin 1, is an adipokine normally expressed by visceral adipose tissue. Distinguishing itself as the only cell projection associated transcript with decreased expression, reduction in ITLN1 expression is known in polycystic ovary syndrome, obesity, and diabetics [182]. DNAH7 [183] is the heavy chain of of axonemal dynein, the force-generating protein of respiratory cilia, as is DNAH11 [184], and the leucine rich repeat cilium-stabilizing protein LRRC50 [185], is associated with cilium structure and stability. None of these three genes have known associations with carcinoma. DNAH7 is regulated by FOXC1, MIF-1, POU3F2,1 and FOXD1, among others. DNAH11 is regulated by p53, Sp1, SRF, POU2F1 and Cart-1

among other factors. LRRC50 is regulated by NFB1, Egr-1, CREB and SREBP-1a,b among other factors. The sperm associated antigen 17, SPAG17, is involved in the structural integrity of the central apparatus of the sperm axoneme, and is regulated by POU2F1, FOXD3, Pax-6, Cdc5 among other factors. The solute carrier family 34 (Na phosphate) member 2 protein, SLC34A2, may be the main phosphate transport protein for brush border membranes. It is regulated by c-Myb, NFB,1 among other factors. CD24 antigen is a sialoglycoprotein expressed on mature granulocytes and in many B cells, linked by a glycosyl phosphatidylinositol to the cell surface that may be part of the signal transduction process. It promotes B cell proliferation but prevents their terminal differentiation, and binds protein factors important in several cancer cell lines and specifically with cell-type specific promoter activity in small cell lung cancer. It is regulated by FOXO1a, POU3F2,1, FOXF2 among other factors. CD24 increases in expression have been associated with ovarian serous borderline tumor microinvasion [186]. Perhaps this increase corresponds to early stromal invasion, given CD24s role association with borderline serous tumor microinvasion.

Cilia are a known ultrastructural feature of cells in serous cysts, but are not a feature of ovarian surface epithelium or a consistent feature of high-grade serous carcinomas [187]. There are occasional reports associating cilia to tumorigenesis [188], including a recent manuscript that cataloged a search for potential mechanisms common to primary cilium and tumorigenesis [189]. No specific mechanisms were found, but there is evidence for a complex interaction of cilium and tumor development, based on cell morphological and internal organization changes. In particular, centrioles, fundamental to separation in mitosis, are involved in the creation of cilia. In actively dividing cells, centrioles flow between two states, centrosome and cilium, according to the cell cycle stage [190]. As cells move to the S phase, cilia are reabsorbed before progression to mitosis [191]. Ciliary membranes have been shown to contain growth factors, morphogens, and hormone receptors [189]. Although mechanisms by

which cilia use the receptors and signaling components are unclear [189], the potential for ciliated membranes to affect tissue architecture is considerable as connections between the signaling and morphogens pathways (Wnt and Hedgehog signaling) are better understood. Other aspects of cilia also seem related to cancer, such as the occurrence of human inherited mutations in cilia that result in cell proliferation and cysts (for example polycystic kidney disease). Loss of ciliated membranes leads to the redistribution of receptors and consequent aberrant signaling [192] [193].

While cilium terms are enriched only for the benign tumors, spindle terms are enriched only for the malignant samples. A possible explanation is the changing role of centrioles, shuttling between the spindle and cilium as the time spent in a particular part of the cell cycle changes. That is, given that cilium-related genes show increased in expression in benign samples and spindle-related gene expression is increased in malignant samples, this shift could be due to monopolization of centrioles by the cell-cycle functions. Thus, this set of genes may be an indicator of malignancy. A lack of cilia may reduce a cells sensitivity to external signals working to suppress the tumor.

4.3.4 Terms Related to Plasma Membranes

Plasma membrane related terms were enriched and many were shared by both benign and malignant transcript lists.

Excluding cell junction terms (discussed above in section 4.3.1, as a separate group), benign plasma membrane terms are most specific for apical and apicolateral plasma membranes. The term “Apicolateral plasma membrane” is related to increased expression of the same transcripts as “Apical part of cell”, and has already been discussed above. The benign related transcripts are: INADL, CDH1, EPCAM, PARD6B, SCNN1A, and SLC34A2. Of these, INADL, CDH1, and PARD6B were discussed above, under tight junction terms and SLC34A2 under the cell projection ontology. SCNN1A is a sodium channel, non-voltage gated 1 alpha, and is not unique

Table 4.6: Summary of transcripts related to “Axoneme”, “Cell Projection Part”, and “Cilium” terms. Listed are the gene symbol associated with the transcript, the state of the tumor (specifically benign, malignant, or both), the expression relative to the control tissue (given as increasing or decreasing), and the associated ontological term/s. Terms in parentheses are related to the transcripts but are not the direct subject matter of the table.

Gene Symbol	Relative Expression	Tumor State	Term
CD24	Increase Increase (Increase) (Increase)	Benign Benign (Both) (Both)	Cell projection part Cilium (Plasma membrane) (Plasma membrane part)
DNAH7	Increase Increase Increase (Increase)	Benign Benign Benign (Benign)	Axoneme Cell projection part Cilium (Cytoskeleton)
DNAH11	Increase Increase Increase (Increase)	Benign Benign Benign (Benign)	Axoneme Cell projection part Cilium (Cytoskeleton)
ITLN1	Decrease (Decrease) (Decrease)	Benign (Benign) (Benign)	Cell projection part (Plasma membrane) (Plasma membrane part)
LRRC50	Increase Increase Increase (Increase)	Benign Benign Benign (Benign)	Axoneme Cell projection part Cilium (Cytoskeleton)
SLC34A2	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cell projection part (Apical part of cell) (Plasma membrane) (Plasma membrane part)
SPAG17	Increase Increase Increase (Increase)	Benign Benign Benign (Benign)	Axoneme Cell projection part Cilium (Cytoskeleton)

to plasma membrane as it is also associated with “Cytoskeleton”. It was included in this group since it is epithelial membrane-associated but has no direct associations to carcinoma [194]. It controls the re-adsorption of sodium in many of the gut organs, and has a role in taste perception. It is regulated by c-Fos, c-Jun, NF-B,1, p53,

PPAR-1,2 and other factors. EPCAM is the epithelial cell adhesion molecule, a membrane protein found in most normal epithelial cells and gastrointestinal carcinomas. It is calcium-independent. Although it appears only in this list because it fell below the significance limit for the malignant samples, there was increased expression in both benign and tumor samples, relative to normal epithelium, and it is central to a number of interaction networks. EPCAM is considered to be one of the most widely expressed tumor transcripts and there are multiple derived therapeutics that target this gene product [195]. Several studies have evaluated EPCAM in ovarian surface epithelial-stromal tumors, and while there is disagreement as to its effectiveness as a marker for progression, all studies found increased EPCAM expression in tumors [196] [197] [198] [199]. It is regulated by HNF-41,2 among other factors.

The ontological terms enriched in malignant samples are nested in a similar fashion to the benign terms. Six transcripts lead to the appearance of this term “Apical Part of Cell” including INADL, CDH1, EPCAM, MUC1, SCNN1A, and ERBB3. Of these, 2 are unique (MUC1 and ERBB3) to this sample class. ERBB3/HER3/Erb1 is the v-erb erythroblastic leukemia oncogene homolog 3, an epidermal growth factor receptor (EGFR, tyrosine kinase family). It is membrane bound, and can bind ligand, but requires heterodimerization with another EGFR member before signal transduction occurs; the activated pathways lead to cell proliferation and differentiation. Overexpression of this gene is reported in many cancers. One of its isoforms is secreted and functions to modulate the activity of the membrane-bound form. It is regulated by c-Jun, PPAR-1,2, c-Myb and MRF-2 among others. It belongs to a family of 4 human growth factor receptors (HER), all which are under different levels of investigation for therapeutics in multiple carcinomas including ovarian types. Mucin 1, cell surface associated, MUC1, is a membrane-bound glycosylated phosphoprotein. It binds to pathogens and also functions in cell signaling, including modulation of signaling in the ERK, SRC, and NK-B pathways, while in activated T-

cells it influences the Ras/MAPK pathway. Changes in expression, modification, and aberrant localization are all associated with carcinomas. It is regulated by NF- κ B, c-Jun, c-Fos, GR-, GATA-3 among other factors. Mucins as a group are commonly expressed in carcinomas, including reports of association specifically with ovarian surface epithelial-stromal tumors, including the serous subtype [200] [201].

Given the cross-over between plasma membrane terms and more specific terms (for instance “Apical Junction Complex”), these plasma membrane terms are driven by transcripts more specifically associated with other ontologies. Transcripts associated with EPCAM, ERBB3, and MUC1 are uniquely associated with these terms; all have well-documented associations with ovarian tumors.

Table 4.7: Summary of transcripts related to “Plasma Membrane” and “Plasma Membrane Part”. Listed are the gene symbol associated with the transcript, the state of the tumor (specifically benign, malignant, or both), the expression relative to the control tissue (given as increasing or decreasing), and the associated ontological term/s. Terms in parentheses are related to the transcripts but are not the direct subject matter of the table (1 of 7).

Gene Symbol	Relative Expression	Tumor State	Term
AMHR2	Decrease Decrease	Benign Benign	Plasma membrane Plasma membrane part
ARHGAP26	Increase Increase (Increase) (Increase)	Benign Benign (Benign) (Benign)	Plasma membrane Plasma membrane part (Cell junction) (Cytoskeleton)
CALB2	Decrease Decrease (Decrease) (Decrease)	Malignant Malignant (Malignant) (Malignant)	Plasma membrane Plasma membrane part (Cell junction) (Cell-cell junction)
CD24	Increase Increase (Increase) (Increase)	Both Both (Benign) (Benign)	Plasma membrane Plasma membrane part (Cell projection part) (Cilium)
CDH1	Increase Increase Increase (Increase) (Increase) (Increase)	Both Both Both (Both) (Both) (Benign)	Apical part of cell Plasma membrane Plasma membrane part (Cell junction) (Cell-cell junction) (Cytoskeleton)
CDH6	Increase	Both	Plasma membrane
CLDN3	Increase Increase Increase Increase Increase Increase Increase	Malignant Malignant Malignant Malignant Malignant Malignant Malignant	Apicolateral plasma membrane Plasma membrane Plasma membrane part (Apical junction complex) (Cell junction) (Cell-cell junction) (Occluding junction) (Tight junction)

Table 4.8: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
CLDN4	Increase	Both	Apicolateral plasma membrane
	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part
	(Increase)	(Both)	(Apical junction complex)
	(Increase)	(Both)	(Cell junction)
	(Increase)	(Both)	(Cell-cell junction)
	(Increase)	(Both)	(Occluding junction)
CLDN7	Increase	Both	Apicolateral plasma membrane
	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part
	(Increase)	(Both)	(Apical junction complex)
	(Increase)	(Both)	(Cell junction)
	(Increase)	(Both)	(Cell-cell junction)
	(Increase)	(Both)	(Occluding junction)
CLEC4M	Decrease	Benign	Plasma membrane
	Decrease	Benign	Plasma membrane part
COL6A1	Decrease	Malignant	Plasma membrane
	(Decrease)	(Malignant)	(Extracellular matrix)
	(Decrease)	(Malignant)	(Proteinaceous extracellular matrix)
COL6A2	Decrease	Malignant	Plasma membrane
	(Decrease)	(Malignant)	(Extracellular matrix)
	(Decrease)	(Malignant)	(Extrinsic to membrane)
	(Decrease)	(Malignant)	(Proteinaceous extracellular matrix)
CP	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part

Table 4.9: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
DSP	Increase	Both	Apicolateral plasma membrane
	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part
	(Increase)	(Both)	(Apical junction complex)
	(Increase)	(Malignant)	(Cell junction)
	(Increase)	(Malignant)	(Cell-cell junction)
	(Increase)	(Benign)	(Cytoskeleton)
DYTN	Increase	Benign	Plasma membrane
EHD2	Decrease	Both	Plasma membrane
	Decrease	Both	Plasma membrane part
	(Decrease)	(Malignant)	(Extrinsic to membrane)
EPCAM	Increase	Both	Apical part of cell
	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part
ERBB3	Increase	Malignant	Apical part of cell
	Increase	Malignant	Plasma membrane
	Increase	Malignant	Plasma membrane part
FOLR1	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part
GABBR1	Decrease	Malignant	Plasma membrane
	Decrease	Malignant	Plasma membrane part
	(Decrease)	(Malignant)	(Cell junction)
GNG11	Decrease	Both	Plasma membrane
	Decrease	Both	Plasma membrane part
	(Decrease)	(Malignant)	(Extrinsic to membrane)

Table 4.10: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
GPR39	Increase Increase	Both Both	Plasma membrane Plasma membrane part
HOMER2	Increase Increase (Increase)	Malignant Malignant (Malignant)	Plasma membrane Plasma membrane part (Cell junction)
INADL	Increase Increase Increase Increase (Increase) (Increase) (Increase) (Increase) (Increase)	Both Both Both Both (Both) (Both) (Both) (Both) (Both)	Apical part of cell Apicolateral plasma membrane Plasma membrane Plasma membrane part (Apical junction complex) (Cell junction) (Cell-cell junction) (Occluding junction) (Tight junction)
ITLN1	Decrease Decrease (Decrease)	Benign Benign (Benign)	Plasma membrane Plasma membrane part (Cell projection part)
LAMC2	Increase Increase	Benign Benign	Plasma membrane Plasma membrane part
LPAR3	Increase Increase	Both Both	Plasma membrane Plasma membrane part
LYPD1	Increase	Both	Plasma membrane
MAL2	Increase	Both	Plasma membrane
MAP7	Increase Increase	Malignant Malignant	Plasma membrane Plasma membrane part

Table 4.11: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
MPP7	Increase	Both	Apicolateral plasma membrane
	Increase	Both	Plasma membrane
	Increase	Both	Plasma membrane part
	(Increase)	(Both)	(Apical junction complex)
	(Increase)	(Malignant)	(Cell junction)
	(Increase)	(Malignant)	(Cell-cell junction)
MUC1	(Increase)	(Malignant)	(Occluding junction)
	(Increase)	(Malignant)	(Tight junction)
	Increase	Malignant	Apical part of cell
	Increase	Malignant	Plasma membrane
MYOF	Increase	Malignant	Plasma membrane part
	(Increase)	(Benign)	Plasma membrane
	(Increase)	(Benign)	(Cytoskeleton)
PARD6B	Increase	Benign	Apical part of cell
	Increase	Benign	Apicolateral plasma membrane
	Increase	Benign	Plasma membrane
	Increase	Benign	Plasma membrane part
	(Increase)	(Benign)	(Apical junction complex)
	(Increase)	(Benign)	(Cell junction)
	(Increase)	(Benign)	(Cell-cell junction)
	(Increase)	(Benign)	(Occluding junction)
PERP	(Increase)	(Benign)	(Tight junction)
	Increase	Benign	Apicolateral plasma membrane
	Increase	Benign	Plasma membrane
	(Increase)	(Benign)	Plasma membrane part
	(Increase)	(Benign)	(Apical junction complex)
PTPN3	(Increase)	(Benign)	(Cell junction)
	(Increase)	(Benign)	(Cell-cell junction)
	Increase	Benign	Plasma membrane
	(Increase)	(Benign)	Plasma membrane part
	(Increase)	(Benign)	(Cytoskeleton)

Table 4.12: (continued).

Gene Symbol	Relative Expression	Tumor State	Term
SCNN1A	Increase Increase Increase (Increase)	Both Both Both (Benign)	Apical part of cell Plasma membrane Plasma membrane part (Cytoskeleton)
SLC34A2	Increase Increase Increase (Increase)	Benign Benign Benign (Benign)	Apical part of cell Plasma membrane Plasma membrane part (Cell projection part)
SKAP1	Increase	Benign	Plasma membrane
SLC1A1	Increase Increase	Benign Benign	Plasma membrane Plasma membrane part
SLC24A6	Decrease	Benign	Plasma membrane
	Increase	Benign	Plasma membrane part
SNTB1	Increase Increase (Increase)	Malignant Malignant (Malignant)	Plasma membrane Plasma membrane part (Cell junction)
SORL1	Increase Increase	Benign Benign	Plasma membrane Plasma membrane part
SORT1	Increase Increase	Both Both	Plasma membrane Plasma membrane part
SPINT2	Increase	Malignant	Plasma membrane
ST3GAL5	Decrease Decrease	Malignant Malignant	Plasma membrane Plasma membrane part
TACSTD2	Increase Increase	Benign Benign	Plasma membrane Plasma membrane part
TRPC1	Decrease Decrease	Malignant Malignant	Plasma membrane Plasma membrane part
VTCN1	Increase	Malignant	Plasma membrane
XPR1	Increase Increase	Malignant Malignant	Plasma membrane Plasma membrane part

4.3.5 Cytoskeleton

Enrichment for the “Cytoskeleton” associated terms occurred only with benign tumors, although it is a node to the “Spindle” term that was enriched in the malignant samples. The enrichment was due to increased expression of the following transcripts: *AGBL2*, *ARHGAP26*, *CDH1*, *DSP*, *DNAH11*, *DNAH7*, *LRRC50*, *MNS1*, *MPZL2*, *MYOF*, *MYO5C*, *PTPN3*, *RCC2*, *SCNN1A*, and *SPAG17*. Several of these terms have been discussed under other headings, which is reasonable since this is a combination of cell junction and plasma membrane terms. It includes 4 transcripts not previously covered: *AGBL2*, *MNS1*, *MPZL2*, and *MYO5C*.

The ATP/GTP binding-protein like, *AGPL2*, may play a role in processing tubulin but little is known about its functions. It is regulated by *Nkx3-1(v1,2,3,4)* and NF- κ B among other factors. *MNS1* is a meiosis-specific nuclear structural 1 gene, which may function as a nuclear skeleton protein that regulates morphology during meiosis. It is regulated by *POU3F1,2*, *FOXL1* among other factors. The myelin protein zero-like 2, *MPZL2*, is strongly regulated in thymus development and in several epithelial structures in embryogenesis. It is regulated by *c-Myc*, *PPAR-1,2*, *POU2F1,2* among other factors. Myosin VC, *MYO5C*, is a myosin form involved in transferrin trafficking and secretory granule transport [202], and alternative splicing determines its associations with *RAB10*, but little else is known about its specific functions. It is regulated by *MEF-2A*, *HNF-1A*, *POU2F1,2,2B,2C*, among other factors. Compared to many of the other ontological terms, many transcripts exclusively associated with “Cytoskeleton” are weakly characterized representing need for further investigation.

4.3.6 Extracellular Region

The extracellular region terms were only enriched with the malignant transcript list and were also the only group whose phenotype was decreased expression instead of increased expression. Specific terms included: “Extracellular Matrix”, “Proteina-

Table 4.13: Summary of transcripts related to Cytoskeleton. Listed are the gene symbol associated with the transcript, the state of the tumor (specifically benign, malignant, or both), the expression relative to the control tissue (given as increasing or decreasing), and the associated ontological term/s. Terms in parentheses are related to the transcripts but are not the direct subject matter of the table (1 of 2).

Gene Symbol	Relative Expression	Tumor State	Term
AGBL2	Increase	Benign	Cytoskeleton
ARHGAP26	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cytoskeleton (Plasma membrane) (Plasma membrane part) (Cell junction)
CDH1	Increase (Increase) (Increase) (Increase) (Increase) (Increase)	Benign (Both) (Both) (Both) (Both) (Both)	Cytoskeleton (Apical part of cell) (Cell junction) (Cell-cell junction) (Plasma membrane) (Plasma membrane part)
DSP	Increase (Increase) (Increase) (Increase) (Increase) (Increase) (Increase)	Benign (Both) (Both) (Malignant) (Malignant) (Both) (Both)	Cytoskeleton (Apical junction complex) (Apicolateral) (plasma membrane) (Cell junction) (Cell-cell junction) (Plasma membrane) (Plasma membrane part)
DNAH11	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cytoskeleton (Axoneme) (Cell projection part) (Cilium)
DNAH7	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cytoskeleton (Axoneme) (Cell projection part) (Cilium)

ceous Extracellular Matrix”, and “Extrinsic to Membrane”. For the first two terms the same transcript group was responsible, including EFEMP1, COL6A1, COL6A2, DCN, and MGP. These genes are a closely connected cluster of deregulation in the malignant gene network (Figure 4.4). Many of these genes are reported to have tumor

Table 4.14: (continued).

LRRC50	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cytoskeleton (Axoneme) (Cell projection part) (Cilium)
MNS1	Increase	Benign	Cytoskeleton
MPZL2	Increase	Benign	Cytoskeleton
MYOF	Increase (Increase)	Benign (Benign)	Cytoskeleton (Plasma membrane)
MYO5C	Increase	Benign	Cytoskeleton
PTPN3	Increase (Increase) (Increase)	Benign (Benign) (Benign)	Cytoskeleton (Plasma membrane) (Plasma membrane part)
RCC2	Increase (Increase) (Increase)	Benign (Malignant) (Malignant)	Cytoskeleton (Condensed chromosome) (centromeric region) (Spindle)
SCNN1A	Increase (Increase) (Increase) (Increase)	Benign (Both) (Both) (Both)	Cytoskeleton (Apical part of cell) (Plasma membrane) (Plasma membrane part)
SPAG17	Increase (Increase) (Increase) (Increase)	Benign (Benign) (Benign) (Benign)	Cytoskeleton (Axoneme) (Cell projection part) (Cilium)

suppression functions. COL6A1 and A2 are collagen type 6 variants, beaded filament collagens found in most connective tissue that maintain structural integrity. They are extracellular matrix proteins, this family is a component of microfibrils. COL6A1 is regulated by c-Jun, c-Fos, Pax-2, Nkx2-2, RP58 and CREB, among other factors and COL6A2 is regulated by SREBP-1, Elk-1, TBP and AP-1 amongst others. Decorin, DCN, is a proteoglycan, a small cellular or pericellular matrix material, a component of connective tissue that binds to collagen fibrils and contributes to matrix assembly. DCN, is a known growth inhibitor to ovarian cancer cell lines SKOV and 2774 *in vitro* [203]. When placed in an matrigel assay, DCN did not suppress tumor cells; the study authors attributed this to collagen binding to DCN [203]. Interestingly, it is repressed

by COL6A1, and COL6A2. It is regulated by c-Jun, c-Fos, HNF-1A, POU2F1m GR- α,β among other factors. The matrix Gla protein, MGP, is found in the organic matrix of bone and cartilage. It has been shown to bind, interact and be repressed by DCN. MGP has been shown to be an important biological regulator through controlling calcification [204]. It is regulated by AP-1, Sox9, c-Fos, c-Jun, PPAR-1,2 and FOX11, among other factors. The EGF-containing fibulin-like extracellular matrix protein 1, EFEMP1, is an extracellular matrix glycoprotein, with tandemly repeated EGF repeats. It is up-regulated in malignant gliomas, as well as down-regulated in hepatocellular carcinoma and sporadic breast cancer due to promoter hypermethylation [205] [206]. It is regulated by Cart-1, HNF-1A, PPAR-1,2, FOXJ2, POU2F1 and other factors. Calbindin 2, CALB2 (or calretinin) is an intracellular calcium-binding protein in the troponin C superfamily (discussed in section 4.3.1). It was not expressed in invasive serous tumor nests in the peritoneum as well the destruction of surrounding calretinin+ mesothelial cells surrounding all metastases studied [207]. It is regulated by Cart-1, CREB and other factors.

Presence of the more specific “Extrinsic to Membrane” term was due to 5 transcripts with decreased expression: EHD2, NAPA, COL6A2, GNG11, and AKT3. COL6A2 is part of the cluster described above. The N-ethylmaleimide-sensitive factor attachment protein alpha, NADA, is involved in the process of vesicle docking and fusion processes of cells. It is part of the multi-subunit fusion apparatus, required for vesicular transport between the endoplasmic reticulum and the Golgi apparatus. It is implicated in the accurate transport of cadherins to adheren junctions of the plasma membrane [208]. It is regulated by NF-B, TGIF, SREBP-1a,b,c, and HNF-1A among other factors. This study results indicated increased expression of CDH1 (E-cadherin) and CDH6 (H-cadherin) in both benign and malignant tumors. GNG11 is a guanine nucleotide binding protein, gamma 11, that encodes a lipid-anchored cell membrane protein that participates in transmembrane signaling. Decreased expression is asso-

ciated with splenic marginal zone lymphomas. Several multimeric forms activate ion channels. It is regulated by Sox9, FOXO1a, FOXO4, HSF1, POU3F2 among other factors. EHD2, EH-domain containing 2, has been associated with membrane protein trafficking. The v-akt murine thymoma viral oncogene homolog 3, (or protein kinase B, gamma) protein, AKT3, belongs to the serine/threonine protein kinase family, regulators of cell signaling in response to insulin and growth factors. They are involved in many processes from proliferation to differentiation, apoptosis, tumorigenesis, glucose uptake, etc. It is a downstream mediator of the PI 3-K pathway, whereby Akt is recruited to the plasma membrane, where it then phosphorylates many substrates, including transcription factors (FOXO1), kinases (Raf-1, Chk1) and signaling pathway factors (Bad, MDM2). It has been shown to be protective against apoptosis-based therapeutics [209]; this may be associated with the G-M phase transition of the cell cycle which is slowed by the repression of AKT3 in ovarian carcinoma [210]. It is regulated by Sp1, FOXO1a, FOXC1, STAT3, c-Jun, CREB, FOXO3b among others.

4.4 Major Trends in Ontologies

On the whole, the enriched ontology terms shown here were consistent with literature reports. The strong enrichment of cilia terms in benign versus spindle terms in malignant tumors is interesting, and reasonable. Benign serous tumors do indeed show the same ultrastructural features of cilia that are found in serous cysts. The significantly increased cell-cycle activity in rapidly growing malignant tumors should lead to more spindle formation. Transcripts associated with centrioles, used in both spindles and cilia, may be a barometer for the transition to a malignant state, as increased mitotic activity would tend to monopolize available centrioles. We found evidence for apoptotic signaling in benign tumors, through the enhanced presence of PERP, a tumor-suppressor gene. Not only was PERP not increased in the malignant cells, HOMER1 was increased, which is an apoptosis suppressor. In the malignant tumor cells there was increased expression of cadherin, claudins, and other cell junc-

Table 4.15: Summary of transcripts related to Extracellular Region. Listed are the gene symbol associated with the transcript, the state of the tumor (specifically benign, malignant, or both), the expression relative to the control tissue (given as increasing or decreasing), and the associated ontological term/s. Terms in parentheses are related to the transcripts but are not the direct subject matter of the table.

Gene Symbol	Relative Expression	Tumor State	Term
AKT3	Decrease	Malignant	Extrinsic to membrane
COL6A1	Decrease Decrease (Decrease)	Malignant Malignant (Malignant)	Extracellular matrix Proteinaceous extracellular matrix (Plasma Membrane)
COL6A2	Decrease Decrease Decrease (Decrease)	Malignant Malignant Malignant (Malignant)	Extracellular matrix Extrinsic to membrane Proteinaceous extracellular matrix (Plasma Membrane)
DCN	Decrease Decrease	Malignant Malignant	Extracellular matrix Proteinaceous extracellular matrix
EFEMP1	Decrease Decrease	Malignant Malignant	Extracellular matrix Proteinaceous extracellular matrix
EHD2	Decrease (Decrease) (Decrease)	Malignant (Malignant) (Malignant)	Extrinsic to membrane (Plasma Membrane) (Plasma Membrane Part)
GNG11	Decrease (Decrease) (Decrease)	Malignant (Malignant) (Malignant)	Extrinsic to membrane (Plasma Membrane) (Plasma Membrane Part)
MGP	Decrease Decrease	Malignant Malignant	Extracellular matrix Proteinaceous extracellular matrix
NAPA	Decrease	Malignant	Extrinsic to membrane

tion gene products, and these form the hub of several of the *in silico* gene networks developed in these cells. While increased expression of claudins is typical in malignancies, expression was higher than corresponding normal cells in both tumor types for E-cadherin (CDH1), a well studied gene that is usually lowered in expression in malignancies. This led us to look more closely at the transcript cluster for CDH1 gene. With further evaluation, malignant CDH1 expression is measured at a 2.3 fold increase in expression, which, while increased from normal expression, is only half the signal increase measured in benign tissues (4.5 fold increase). For this transcript,

then, we do see a trend of decreased expression from benign to malignant, as expected. A decreased expression of CDH1 is an indicator of epithelial-mesenchymal transition-essential to metastasis but not of malignancy. Since the target was malignant, not metastatic, serous tumors, this observation is not a contradiction.

Other clues were found to position tumor cohorts in the tumorigenesis process. Benign tumors in our data set showed elevated levels of genes associated with apoptosis, generally repressed in malignant cells, and with a known ultrastructural feature, cilia. Cilia have been observed but are rare in serous malignancy, so they are not a definite indicator of tumorigenesis either. The most precise evidence of progression in the benign samples is the elevated expression of CD24, cited as an indicator of borderline serous tumors. If this is a robust indicator, this would place our benign tumor samples closer to the borderline malignant stage of progression.

Malignant tumors also fall within a progression line. Evidence of CIN, mitotic activity, and loss of apoptotic tumor-suppressors all invoke known aspects of the malignant state. The epithelial-to-mesenchymal transition is considered the watershed event for metastasis. Several observations indicate that this set of malignant tumors has not completed this transition, and likely are not metastatic. This includes the increased expression in CDH1 mentioned above, and also CD24, DSP and the set of tight junction-related transcripts. Although negative data is not a reliable foundation for conclusions, there are known up-regulated genes in the EMT that we looked at; for example MPP7 is increased in malignant cells but was not found significantly increased in our samples

In this context other trends from the ontologies make sense: the transition involves restructuring of the apical junctions and membrane, and these terms are enriched. The reduction in extracellular elements, particularly collagen, is typical of the destruction of the basement membrane, a necessary step before metastasis. Malignancy often involves induction of fetal genes, and the observed increase in one of the isoforms

of the plasma membrane gene, ERBB3, which is active in embryonic development and differentiation, is a commonly seen feature of malignant cells [211].

4.5 NRCAM Expression

When performing comparisons of the tumor-stroma environment, only one transcript was significant. In the benign tumor-stroma environment, NRCAM (Neuron Cell Adhesion Molecule: NM_001037132) showed a 6.77 increase in benign adjacent stromal expression over distal stroma, in contrast to a 1.31 increase in malignant samples. NRCAM is a neuronal cell adhesion molecule that binds ankyrin and promotes directional signaling during axonal cone growth. It likely has a general role in cell-cell communication, signaling from its intracellular domain to the actin cytoskeleton during directional cell migration. It has a large number of allelic variants. It is regulated by HNF-42, POU2F1, PPAR1,2, FOXD1,3 and other factors. Although mostly studied in the context of the nervous system it has been implicated in tumor biochemistry including thyroid and lung carcinomas [212] [213]. NRCAM has also been shown to indoctrinate surrounding cells into tumorigenesis (survival and proliferation) when expressed in NIH3T3 fibroblasts [214]. In that study the extracellular region of NRCAM was shown to induce proliferation by stimulating the PI3K/AKT pathway, indicating a paracrine function for NRCAM through integrin interaction [215].

In our study slight NRCAM increases were observed in malignant cells, although they did not meet the criteria for significance. It is not clear whether this is due to a real drop in expression or because of the significant reduction in the presence of material in the extracellular region overall; NRCAM as an extracellular agent may have been destroyed with other proteins. This transcript, even though it may have affected tumor tissue, was not expressed in tumor tissue. Indeed, we were only able to observe it all because of the purification method used (LCM); since NRCAM can act in both autocrine and paracrine signaling, a macrodissected sample would have

given no reason to suspect paracrine signaling over autocrine signaling.

4.6 Chromosomal Location

By comparing the distribution of significant DE genes compared to a randomly selected set one can determine whether a likely epigenetic or other large-scale aberration has occurred. Only one chromosome had significance: chromosome 8 contained 11 of the malignancy-associated increased-expression transcripts, although they are not clustered on a single cytoband. The transcripts in chromosome 8 include from PRKDC (8q11), C8orf62 (8q11.21), SOX17 (8q11.23), CHMP4C (8q21.13), RAD54B (8q22.1), ESRP1 (8q22.1), AZIN1 (8q22.3), GRHL2 (8q22.3), MAL2 (8q23), SNTB1 (8q23-q24), and FAM49B (8q24.21). Of this set, two have previously known associations with ovarian cancer, including MAL2 and RAD54B [216] [217].

4.7 Deriving a Gene Network

There are many levels at which cellular networks can be considered, from regulatory circuits with transcription factors to protein localization and interaction complexes. There are several tools that will accept a ‘significant gene list’ and search relevant databases for known networks at each of these levels. Here we used STRING8 to derive pathways for both the benign (Figure 4.3) and malignant (Figure 4.4) transcript lists. The tool allows inclusion of connecting elements at varying levels of evidence; in our case we restricted connections to experimentally verified interactions. Single time-point sampling is likely to miss important nodes in any network, and not all nodes need change for network output to change, so that an analysis based on differential expression will not pick up all nodes; biological networks are particularly fragile in this respect. Thus, after significant interacting genes were placed, a second tool, Genes2Networks, was used to infer transcripts not in the list which may be important.

Many of the trends reported in the ontological enrichment results are also apparent

in the networks, including especially the decreased expression of extracellular region genes, seen as a cluster of interconnected blue transcripts in the malignant network (Figure 4.4). As well, the 2 core nodes in both networks, CDH1, and CLDN4 are associated with tight junctions and apical membranes, highlighted in the ontological enrichment outcomes.

The networks provide a better-integrated visualization of the connectivity of ERBB3/4 modulation with other observations in the malignant data set. Also, even though the benign transcript list was longer, the benign network is less connected, with only 50 derived genes and 53 interactions compared to the malignant network, with 73 and 97 respectively. This is mirrored by the GO enrichment lists, where the benign list has 30 nodes (terms) and 44 edges compared to the malignant GO list, with 37 and 54 respectively. This is expected if more and more dysregulated genes are expressed through the accumulation of mutations proceeding through the steps of tumorigenesis.

Although it is not clear if this is a general feature, the benign gene network contains ERBB4 while the malignant network contains ERBB3. As a reminder, these are EGFR protein kinases; ERBB3 is the only HER/NEU protein that has an inactive intracellular domain, relying entirely on trans-extracellular signalling [218]. Both ERBB3 and ERBB4 bind to PI3-K and are associated with the PI3K/AKT pathway, associated with apoptosis suppression and angiogenesis, but only ERBB3 is immune to EGFR tyrosine kinase inhibitors [219]. This may be why AKT3 significantly decreased in malignant tumors. Further study in this shift between two HER proteins may uncover important therapeutic mechanisms.

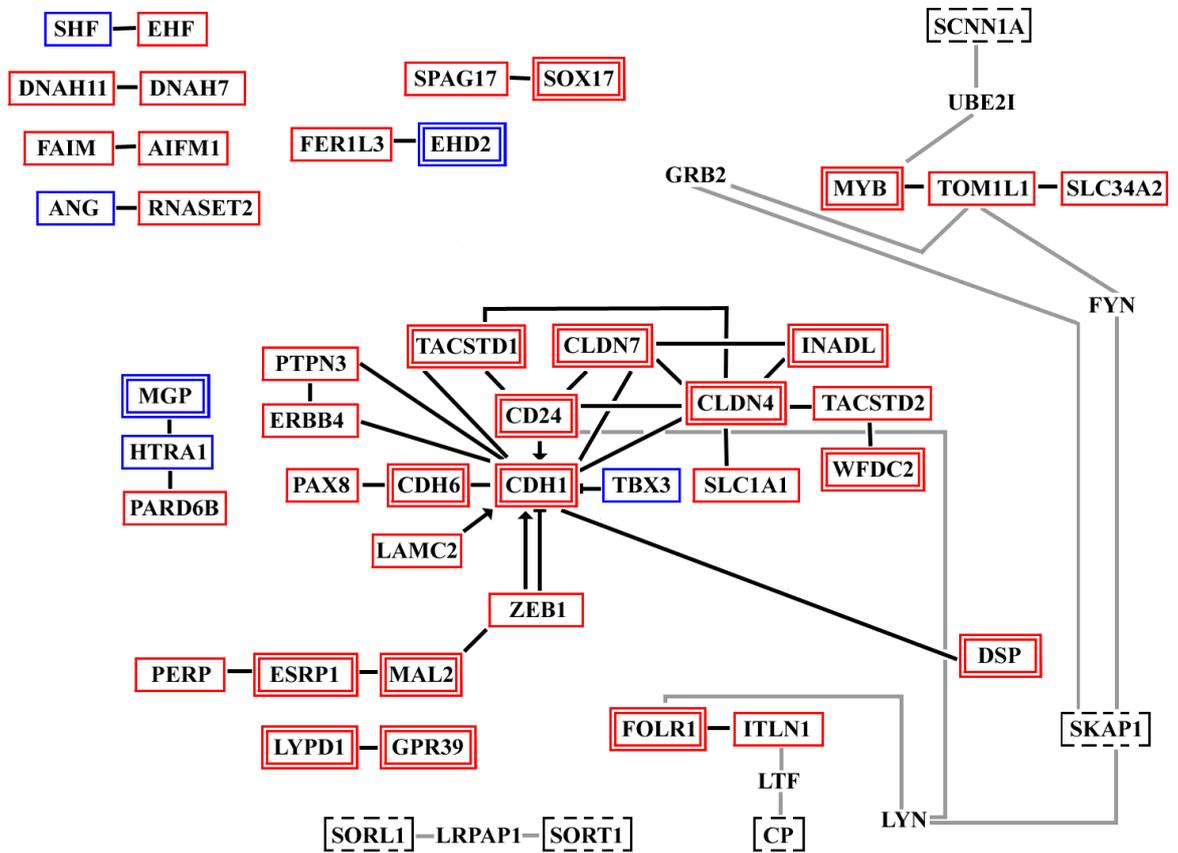


Figure 4.3: Graph of a predicted protein network derived from the benign transcript list. Gene symbols are surrounded by red borders if they were increased relative to normal, and blue if they were decreased. Double outlined boxes indicate those genes common to both benign and malignant networks (see Figure 4.4). Gene symbols with black dashed borders indicate DE transcripts with only inferred relationships. Gene symbols with no border are Genes2Network-inferred relationships but were not found to be a DE transcripts. Black lines indicate known interactions, grey lines indicate inferred relationships. Known inhibiting and promoting relationships are respectively given bars or arrows (for example CD24 inhibits CDH1).

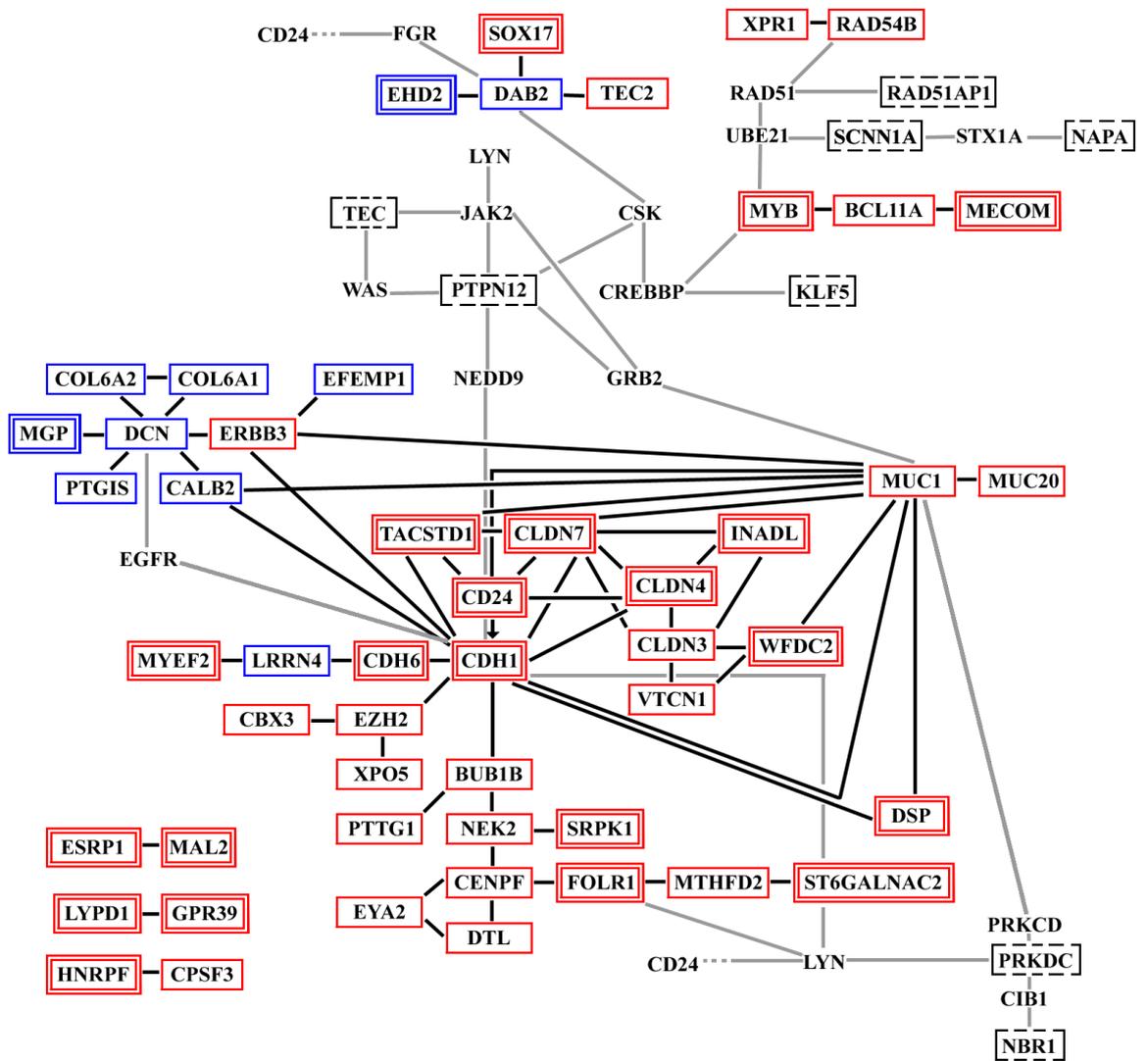


Figure 4.4: Graph of a predicted protein interaction network derived from the malignant transcript list. Figure elements are represented as in Figure 4.3.

4.8 Future Directions

The production of a high-quality genomics data set is cause for pride and also for frustration. The experimental design must accommodate some sampling problems and in some cases the ability to actually collect samples simply does not meet expectations. The technical requirements of the assays are such that extended consistency is extremely difficult to achieve. The choice of platform is subject to the vagaries of fashion in both availability of similar data sets and the algorithms and methods produced for their analysis. Frustration arises because the data set, once produced, has such rich options for subsequent data mining that knowing where to start first and acknowledging when additional efforts must be left to others is difficult. The bioinformatics level achieved in this study was rudimentary, and primarily involved sample cleansing and preliminary screening for the largest signals, to validate global consistency with published studies. This establishes the suitability of the samples for inclusion in in-depth data mining projects. While the sample numbers are quite small, high quality and sound platform choice ensure that merging the data with other studies, or extending this study, are both feasible approaches.

If study extension is considered several factors of interest immediately come to mind. Given the heterogeneity of ovarian surface epithelial-stromal tumor histopathology, the addition of other subtype cohorts would certainly be revealing if the LCM method was used in cell collection. Currently ovarian serous tumorigenesis theory is hotly debated, as there is a proposal that type II serous ovarian tumors may not rise *in situ* but result from metastases of fallopian tube serous tumors. Thus, adding samples from fallopian tube fimbriae epithelium to this study would allow rapid evaluation of the merits of the proposal, as well as other types of primary sites. A reasonable common baseline would be the progenitor tissue, fetal coelomic epithelium, as well as a comparison of that tissue with the OSE profiled here. Also, since the experimental model focused on serous tumors, the stroma (although evaluated) was not extensively

monitored (distance from tumor was either immediately proximal or distal but the interval was not carefully measured).

With respect to data mining, the tools for exon signal analysis, rather than gene signal analysis, and for transcript isoform enrichment discovery are both obvious points for bioinformatics tool development, with the advantage that predictions can be tested using wet-lab assays on the original samples. Meta-analyses, achieved by merging external data sets, and integration, based on informatics tools that bring in RNA-seq and other types of high-throughput assay data, are also clearly areas of immediate interest.

BIBLIOGRAPHY

- [1] A. Jemal, R. Siegel, J. Xu, and E. Ward, "Cancer statistics, 2010," *CA Cancer J Clin*, vol. 60, no. 5, pp. 277–300, 2010.
- [2] P. M. Webb, D. M. Purdie, S. Grover, S. Jordan, M.-L. Dick, and A. C. Green, "Symptoms and diagnosis of borderline, early and advanced epithelial ovarian cancer," *Gynecologic Oncology*, vol. 92, no. 1, pp. 232 – 239, 2004.
- [3] C. M. Olsen, J. Cnossen, A. C. Green, and P. M. Webb, "Comparison of symptoms and presentation of women with benign, low malignant potential and invasive ovarian tumors," *European Journal of Gynaecological Oncology*, vol. 28, pp. 376–80, 2007.
- [4] G. Lurie, L. R. Wilkens, P. J. Thompson, R. K. Matsuno, M. E. Carney, and M. T. Goodman, "Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: A case analysis," *Gynecologic Oncology*, vol. 119, no. 2, pp. 278 – 284, 2010.
- [5] C. Bankhead, C. Collins, H. Stokes-Lampard, P. Rose, S. Wilson, A. Clements, D. Mant, S. Kehoe, and J. Austoker, "Identifying symptoms of ovarian cancer: a qualitative and quantitative study," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 115, no. 8, pp. 1008–1014, 2008.
- [6] R. Twombly, "Cancer killer may be "silent" no more," *Journal of the National Cancer Institute*, vol. 99, no. 18, pp. 1359–1361, 2007.
- [7] C. Bankhead, C. Collins, H. Stokes-Lampard, P. Rose, S. Wilson, A. Clements, D. Mant, S. Kehoe, and J. Austoker, "Identifying symptoms of ovarian cancer: a qualitative and quantitative study," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 115, no. 8, pp. 1008–1014, 2008.
- [8] I. Lataifeh, D. E. Marsden, G. Robertson, V. Gebiski, and N. F. Hacker, "Presenting symptoms of epithelial ovarian cancer," *Australian and New Zealand Journal of Obstetrics and Gynaecology*, vol. 45, no. 3, pp. 211–214, 2005.
- [9] "What you need to know about™ ovarian cancer."
- [10] G. Lurie, P. J. Thompson, K. E. McDuffie, M. E. Carney, and M. T. Goodman, "Prediagnostic symptoms of ovarian carcinoma: A case-control study," *Gynecologic Oncology*, vol. 114, no. 2, pp. 231 – 236, 2009.
- [11] J. Boyd, Y. Sonoda, M. G. Federici, F. Bogomolny, E. Rhei, D. L. Maresco, P. E. Saigo, L. A. Almadrones, R. R. Barakat, C. L. Brown, D. S. Chi, J. P. Curtin, E. A. Poynor, and W. J. Hoskins, "Clinicopathologic features of brca-linked and sporadic ovarian cancer," *JAMA: The Journal of the American Medical Association*, vol. 283, no. 17, pp. 2260–2265, 2000.

- [12] A. Fritz, C. Percy, A. Jack, K. Shanmugaratnam, L. Sobin, D. Parkin, and S. Whelan, *International Classification of Diseases for Oncology (ICD-0)*. Geneva: World Health Organization, 3rd ed., 2000.
- [13] L. A. G. Ries, J. John L Young, G. E. Keel, M. P. Eisner, Y. D. Lin, and M.-J. D. Horner, eds., *SEER survival monograph: cancer survival among adults: US Seer program, 1988-2001, Patient and Tumor Characteristics*. Bethesda, MD: National Cancer Institute, 2007.
- [14] F. A. Tavassoli and P. Devilee, eds., *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. Lyon: IARC Press, 2003.
- [15] M. Köbel, S. E. Kalloger, N. Boyd, S. McKinney, E. Mehl, C. Palmer, S. Leung, N. J. Bowen, D. N. Ionescu, A. Rajput, L. M. Prentice, D. Miller, J. Santos, K. Swenerton, C. B. Gilks, and D. Huntsman, "Ovarian carcinoma subtypes are different diseases: Implications for biomarker studies," *PLoS Med*, vol. 5, p. e232, 12 2008.
- [16] T. Kaku, S. Ogawa, Y. Kawano, Y. Ohishi, H. Kobayashi, T. Hirakawa, and H. Nakano, "Histological classification of ovarian cancer," *Medical Electron Microscopy*, vol. 36, pp. 9–17, 2003. 10.1007/s007950300002.
- [17] T. N. Fredrickson, "Ovarian tumors of the hen," *Environmental Health Perspectives*, vol. 73, pp. 35–51, 1987.
- [18] N. Auersperg, A. S. T. Wong, K.-C. Choi, S. K. Kang, and P. C. K. Leung, "Ovarian surface epithelium: Biology, endocrinology, and pathology," *Endocr Rev*, vol. 22, no. 2, pp. 255–288, 2001.
- [19] J. R. Giles, R. G. Elkin, L. S. Trevino, M. E. Urick, R. Ramachandran, and P. A. Johnson, "The restricted ovulator chicken: A unique animal model for investigating the etiology of ovarian cancer," *International Journal of Gynecological Cancer*, vol. 20, pp. 738–44, 2010.
- [20] F. A. Monzon, M. Lyons-Weiler, L. J. Buturovic, C. T. Rigl, W. D. Henner, C. Sciulli, C. I. Dumur, F. Medeiros, and G. G. Anderson, "Multicenter validation of a 1,550-gene expression profile for identification of tumor tissue of origin," *Journal of Clinical Oncology*, vol. 27, no. 15, pp. 2503–2508, 2009.
- [21] D. A. Bell, "Origins and molecular pathology of ovarian cancer," *Modern Pathology*, vol. 18, pp. S19–S32, 2005.
- [22] K. M. Feeley and M. Wells, "Precursor lesions of ovarian epithelial malignancy," *Histopathology*, vol. 38, no. 2, pp. 87–95, 2001.
- [23] R. J. Kurman and I.-M. Shih, "The origin and pathogenesis of epithelial ovarian cancer: A proposed unifying theory," *American Journal of Surgical Pathology*, vol. 34, pp. 433–443, 2010.

- [24] M. J. Callahan, C. P. Crum, F. Medeiros, D. W. Kindelberger, J. A. Elvin, J. E. Garber, C. M. Feltmate, R. S. Berkowitz, and M. G. Muto, "Primary fallopian tube malignancies in brca-positive women undergoing surgery for ovarian cancer risk reduction," *Journal of Clinical Oncology*, vol. 25, no. 25, pp. 3985–3990, 2007.
- [25] A. Finch, P. Shaw, B. Rosen, J. Murphy, S. A. Narod, and T. J. Colgan, "Clinical and pathologic findings of prophylactic salpingo-oophorectomies in 159 brca1 and brca2 carriers," *Gynecologic Oncology*, vol. 100, no. 1, pp. 58 – 64, 2006.
- [26] F. Medeiros, M. G. Muto, Y. Lee, J. A. Elvin, M. J. Callahan, C. Feltmate, J. E. Garber, D. W. Cramer, and C. P. Crum, "The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome," *American Journal of Surgical Pathology*, vol. 30, pp. 230–6, 2006.
- [27] P. A. Shaw, M. Rouzbahman, E. S. Pizer, M. Pintilie, and H. Begley, "Candidate serous cancer precursors in fallopian tube epithelium of brca1/2 mutation carriers," *Modern Pathology*, vol. 22, pp. 1133–8, 2009.
- [28] J. M. J. Piek, R. H. M. Verheijen, P. Kenemans, L. F. Massuger, H. Bulten, and P. J. van Diest, "Brca1/2-related ovarian cancers are of tubal origin: a hypothesis," *Gynecologic Oncology*, vol. 90, no. 2, pp. 491 – 491, 2003.
- [29] D. W. Kindelberger, Y. Lee, A. Miron, M. S. Hirsch, C. Fletmate, F. Medeiros, M. J. Callahan, E. O. Garner, R. W. Gordon, C. Birch, R. S. Birkowitz, M. G. Muto, and C. P. Crum, "Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship," *American Journal of Surgical Pathology*, vol. 31, pp. 161–9, 2007.
- [30] C. G. Przybycin, R. J. Kurman, B. M. Ronnett, I.-M. Shih, and R. Vang, "Are all pelvic (nonuterine) serous carcinoma of tubal origin?," *American Journal of Surgical Pathology*, vol. 34, pp. 1407–1416, 2010.
- [31] R. T. Marquez, K. A. Baggerly, A. P. Patterson, J. Liu, R. Broaddus, M. Frumovitz, E. N. Atkinson, D. I. Smith, L. Hartmann, D. Fishman, A. Berchuck, R. Whitaker, D. M. Gershenson, G. B. Mills, R. C. Bast, and K. H. Lu, "Patterns of Gene Expression in Different Histotypes of Epithelial Ovarian Cancer Correlate with Those in Normal Fallopian Tube, Endometrium, and Colon," *Clinical Cancer Research*, vol. 11, no. 17, pp. 6116–6126, 2005.
- [32] G. Singer, R. J. Kurman, H.-W. Chang, S. K. R. Cho, and I.-M. Shih, "Diverse tumorigenic pathways in ovarian serous carcinoma," *The American Journal of Pathology*, vol. 160, pp. 1223–1228, 2002.
- [33] I.-M. Shih and R. J. Kurman, "Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis," *The American Journal of Pathology*, vol. 164, pp. 1511–1518, 2004.

- [34] R. Vang, I.-M. Shih, and R. J. Kurman, “Ovarian low-grade and high-grade serous carcinoma: Pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems,” *Advances in Anatomic Pathology*, vol. 16, pp. 267–282, 2009.
- [35] D. D. L. Bowtell, “The genesis and evolution of high-grade serous ovarian cancer,” *Nature Reviews Cancer*, vol. 10, pp. 803–8, 2010.
- [36] N. Auersperg, “The origin of ovarian carcinomas: a unifying hypothesis,” *International Journal of Gynecologic Pathology*, vol. 30, pp. 12–21, 2011.
- [37] K. Pietras and A. Östman, “Hallmarks of cancer: interactions with the tumor stroma,” *Experimental Cell Research*, vol. 316, pp. 1324–1331, 2010.
- [38] D. Hanahan, “The Hallmarks of Cancer,” *Cell*, vol. 100, pp. 57–70, Jan. 2000.
- [39] R. Kalluri and M. Zeisberg, “Fibroblasts in cancer,” *Nature Reviews Cancer*, vol. 6, pp. 392–401, 2006.
- [40] A. Östman and M. Augsten, “Cancer-associated fibroblasts and tumor growth-bystanders turning into key players,” *Current Opinion in Genetics & Development*, vol. 19, pp. 67–73, 2009.
- [41] N. A. Bhowmick, E. G. Neilson, and H. L. Moses, “Stromal fibroblasts in cancer initiation and progression,” *Nature*, vol. 432, pp. 332–337, 2004.
- [42] S. Demehri and R. Kopan, “Notch signaling in bulge stem cells is not required for selection of hair follicle fate,” *Development*, vol. 136, no. 6, pp. 891–896, 2009.
- [43] N. A. Bhowmick, A. Chytil, D. Plieth, A. E. Gorska, N. Dumont, S. Shappell, M. K. Washington, E. G. Neilson, and H. L. Moses, “Tgf- β signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia,” *Science*, vol. 303, no. 5659, pp. 848–51, 2004.
- [44] H. Strand, L. Lacina, M. Kolář, Z. Čada, Čestmír Vlček, B. Dvořánková, J. Betka, J. Plzák, M. Chovanec, J. Šáchová, J. Valach, M. Urbanová, and K. S. Jr, “Head and neck squamous cancer stromal fibroblasts produce growth factors influencing phenotype of normal human keratinocytes,” *Histochemistry and Cell Biology*, vol. 133, pp. 201–211, 2010. 10.1007/s00418-009-0661-6.
- [45] C. Lebedis, K. Chen, L. Fallavollita, T. Boutros, and P. Brodt, “Peripheral lymph node stromal cells can promote growth and tumorigenicity of breast carcinoma cells through the release of igf-1 and egf,” *International Journal of Cancer*, vol. 100, pp. 2–8, 2002.
- [46] K. R. Levental, H. Yu, L. Kass, J. N. Lakins, M. Egeblad, J. T. Erler, S. F. Fong, K. Csiszar, A. Giaccia, W. Weninger, M. Yamauchi, D. L. Gasser, and V. M. Weaver, “Matrix crosslinking forces tumor progression by enhancing integrin signaling,” *Cell*, vol. 139, pp. 891–906, 2009.

- [47] V. L. Thijssen, R. J. Brandwijk, R. P. Dings, and A. W. Griffioen, "Angiogenesis gene expression profiling in xenograft models to study cellular interactions," *Experimental Cell Research*, vol. 299, pp. 286–293, 2004.
- [48] D. Fukumura, R. Xavier, T. Sugiura, Y. Chen, E.-C. Park, N. Lu, M. Selig, G. Nielsen, T. Taksir, R. K. Jain, and B. Seed, "Tumor induction of vegf promoter activity in stromal cells," *Cell*, vol. 94, pp. 715–725, 1998.
- [49] A. Orimo, P. B. Gupta, D. C. Sgroi, F. Arenzana-Seisdedos, T. Delaunay, R. Naeem, V. J. Carey, A. L. Richardson, and R. A. Weinberg, "Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated sdf1/cxcl12 secretion," *Cell*, vol. 121, pp. 335–348, 2005.
- [50] A. S. Whittemore, R. Harris, and J. Itnyre, "Characteristics relating to ovarian cancer risk: collaborative analysis of 12 us case-control studies. ii. invasive epithelial ovarian cancers in white women. collaborative ovarian cancer group," *American Journal of Epidemiology*, vol. 136, pp. 1184–203, 1992.
- [51] T. Riman, P. W. Dickman, S. Nilsson, N. Correia, H. Nordlinder, C. M. Magnusson, and I. R. Persson, "Risk factors for invasive epithelial ovarian cancer: Results from a swedish case-control study," *American Journal of Epidemiology*, vol. 156, no. 4, pp. 363–373, 2002.
- [52] R. Harris, A. S. Whittemore, and J. Itnyre, "Characteristics relating to ovarian cancer risk: Collaborative analysis of 12 us case-control studies iii. epithelial tumors of low malignant potential in white women. collaborative ovarian cancer group.," *American Journal of Epidemiology*, vol. 136, pp. 1204–11, 1992.
- [53] B. L. Harlow, N. S. Weiss, G. J. Roth, J. Chu, and J. R. Daling, "Case-control study of borderline ovarian tumors: reproductive history and exposure to exogenous female hormones," *Cancer Research*, pp. 5849–5852, 1988.
- [54] K. Danforth, S. Tworoger, J. Hecht, B. Rosner, G. Colditz, and S. Hankinson, "Breastfeeding and risk of ovarian cancer in two prospective cohorts," *Cancer Causes and Control*, vol. 18, pp. 517–523, 2007. 10.1007/s10552-007-0130-2.
- [55] T. Sueblinvong and M. Carney, "Current understanding of risk factors for ovarian cancer," *Current Treatment Options in Oncology*, vol. 10, pp. 67–81, 2009. 10.1007/s11864-009-0108-2.
- [56] J. L. Stanford, "Oral contraceptives and neoplasia of the ovary," *Contraception*, vol. 43, pp. 543–56, 1991.
- [57] S. Franceschi, F. Parazzini, E. Negri, M. Booth, C. L. Vecchia, V. Beral, A. Tzonou, and D. Trichopoulos, "Pooled analysis of 3 european case-control studies of epithelial ovarian cancer: iii. oral contraceptive use," *International Journal of Cancer*, vol. 49, pp. 61–65, 1991.

- [58] A. S. Whittemore, R. Harris, and J. Itnyre, "Characteristics relating to ovarian cancer risk: collaborative analysis of 12 us case-control studies. ii. invasive epithelial ovarian cancers in white women," *American Journal of Epidemiology*, vol. 136, pp. 1184–1203, 1992.
- [59] M. A. Rossing, J. R. Daling, N. S. Weiss, D. E. Moore, and S. G. Self, "Ovarian tumors in a cohort of infertile women," *The New England Journal of Medicine*, vol. 331, no. 12, pp. 771–776, 1994.
- [60] L. A. Brinton, E. J. Lamb, K. S. Moghissi, B. Scoccia, M. D. Althuis, J. E. Mabie, and C. L. Westhoff, "Ovarian cancer risk after the use of ovulation-stimulating drugs," *Obstetrics and Gynecology*, vol. 103, pp. 1194–1203, 2004.
- [61] L. Lerner-Geva, J. Rabinovici, and B. Lunenfeld, "Ovarian stimulation: is there a long-term risk for ovarian, breast and endometrial cancer?," *Womens Health*, vol. 6, pp. 831–9, 2010.
- [62] A. Mahdavi, T. Pejovic, and F. Nezhat, "Induction of ovulation and ovarian cancer: a critical review of the literature," *Fertility and Sterility*, vol. 85, no. 4, pp. 819 – 826, 2006.
- [63] N. F. Vlahos, K. P. Economopoulos, and G. Creatsas, "Fertility drugs and ovarian cancer risk: a critical review of the literature," *Annals of the New York Academy of Sciences*, vol. 1205, no. 1, pp. 214–219, 2010.
- [64] H. L. Miracle-McMahill, E. E. Calle, A. S. Kosinski, C. Rodriguez, P. A. Wingo, M. J. Thun, and C. W. H. Jr, "Tubal ligation and fatal ovarian cancer in a large prospective cohort study," *American Journal of Epidemiology*, vol. 145, no. 4, pp. 349–57, 1997.
- [65] A. Green, D. Purdie, C. Bain, V. Siskind, P. Russell, M. Quinn, and B. Ward, "Tubal sterilisation, hysterectomy and decreased risk of ovarian cancer," *International Journal of Cancer*, vol. 71, no. 6, pp. 948–951, 1997.
- [66] S. E. Hankinson, D. J. Hunter, G. A. Colditz, W. C. Willett, M. J. Stampfer, B. Rosner, C. H. Hennekens, and F. E. Speizer, "Tubal ligation, hysterectomy, and risk of ovarian cancer," *JAMA: The Journal of the American Medical Association*, vol. 270, no. 23, pp. 2813–2818, 1993.
- [67] K. Takahashi, H. Kurioka, M. Irikoma, T. Ozaki, H. Kanasaki, and K. Miyazaki, "Benign or malignant ovarian neoplasms and ovarian endometriomas," *The Journal of the American Association of Gynecologic Laparoscopists*, vol. 8, no. 2, pp. 278 – 284, 2001.
- [68] M. A. Rossing, K. L. Cushing-Haugen, K. G. Wicklund, J. A. Doherty, and N. S. Weiss, "Risk of epithelial ovarian cancer in relation to benign ovarian conditions and ovarian surgery," *Cancer, Causes and Control*, vol. 19, pp. 1357–64, 2008.

- [69] M. Huncharek, J. F. Geschwind, and B. Kupelnick, “Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies,” *Anticancer Research*, vol. 23, pp. 1955–60, 2003.
- [70] L. S. Cook, M. L. Kamb, and N. S. Weiss, “Peritoneal powder exposure and the risk of ovarian cancer,” *American Journal of Epidemiology*, vol. 145, no. 5, pp. 459–65, 1997.
- [71] D. M. Gertig, D. J. Hunter, D. W. Cramer, G. A. Colditz, F. E. Speizer, W. C. Willett, and S. E. Hankinson, “Prospective study of talc use and ovarian cancer,” *Journal of the National Cancer Institute*, vol. 92, no. 3, pp. 249–252, 2000.
- [72] P. K. Mills, D. G. Riordan, R. D. Cress, and H. A. Young, “Perineal talc exposure and epithelial ovarian cancer risk in the central valley of california,” *International Journal of Cancer*, vol. 112, no. 3, pp. 458–464, 2004.
- [73] S. C. Larsson and A. Wolk, “Tea consumption and ovarian cancer risk in a population-based cohort,” *Arch Intern Med*, vol. 165, no. 22, pp. 2683–2686, 2005.
- [74] C. M. Olsen, A. C. Green, D. C. Whiteman, S. Sadeghi, F. Kolahehdooz, and P. M. Webb, “Obesity and the risk of epithelial ovarian cancer: A systematic review and meta-analysis,” *European Journal of Cancer*, vol. 43, pp. 690–709, 2007.
- [75] A. C. Antoniou, S. A. Gayther, J. F. Stratton, B. A. Ponder, and D. F. Easton, “Risk models for familial ovarian and breast cancer,” *Genetic Epidemiology*, vol. 18, no. 2, pp. 173–190, 2000.
- [76] M. J. van de Vijver, Y. D. He, L. J. van ’t Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E. T. Rutgers, S. H. Friend, and R. Bernards, “A gene-expression signature as a predictor of survival in breast cancer,” *New England Journal of Medicine*, vol. 347, no. 25, pp. 1999–2009, 2002.
- [77] H. Gevensleben, U.-J. Göhring, R. Büttner, L. C. Heukamp, G. Kunz, T. Dimpfl, C. Jackisch, O. Ortmann, U.-S. Albert, R. Bender, F. D. Snoo, O. Krijgsman, A. M. Glas, Y. H. Ergönenc, C. Vogel, A. Dykgers, C. Langwieder, M. Rees, and T. Anzeneder, “Comparison of mammaprint and target-print results with clinical parameters in german patients with early stage breast cancer,” *International Journal of Molecular Medicine*, vol. 26, no. 6, pp. 837–843, 2010.

- [78] R. C. B. Jr, M. Feeney, H. Lazarus, L. M. Nadler, R. B. Colvin, and R. C. Knapp, "Reactivity of a monoclonal antibody with human ovarian carcinoma," *The Journal of Clinical Investigation*, vol. 68, pp. 1331–1337, 1981.
- [79] A. Karaferic, D. Jovanovic, and S. Jelic, "Expression of her2/neu, estrogen and progesterone receptors, ca-125 and ca19-9 on cancer cell membrane in patients with serous and mucinous carcinoma of the ovary," *Journal of B.U.ON*, vol. 14, pp. 635–639, 2009.
- [80] J. M. Cragun, "Screening for ovarian cancer," *Cancer Control*, pp. 16–21, 2001.
- [81] I. Cass, C. Holschneider, N. Datta, D. Barbuto, A. E. Walts, and B. Y. Karlan, "Brca-mutation-associated fallopian tube carcinoma: a distinct clinical phenotype?," *Obstetrics and Gynecology*, vol. 106, pp. 1327–1334, 2005.
- [82] Y. Miki, J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L. M. Bennet, W. Ding, R. Bell, J. Rosenthal, C. Hussey, T. Tran, M. McClure, C. Frye, T. Hattier, R. Phelps, A. Haugen-Strano, H. Katcher, K. Yakumo, Z. Gholami, D. Shaffer, S. Stone, S. Bayer, C. Wray, R. Bogden, P. Dayananth, J. Ward, P. Tonin, S. Narod, P. K. Bristow, F. H. Norris, L. Helvering, P. Morrison, P. Rosteck, M. Lai, J. C. Barrett, C. Lewis, S. Neuhausen, L. Cannon-Albright, D. Goldgar, R. Wiseman, A. Kamb, and M. H. Skolnick, "A strong candidate for the breast and ovarian cancer susceptibility gene brca1," *Science*, vol. 266, pp. 66–71, 1994.
- [83] R. Wooster, G. Bignell, J. Lancaster, S. Swift, S. Seal, J. Mangion, N. Collins, S. Gregory, C. Gumbs, G. Micklem, R. Barfoot, R. Hamoudi, S. Patel, C. Rices, P. Biggs, Y. Hashim, A. Smith, F. Connor, A. Arason, J. Gudmundsson, D. Ficenec, D. Kelsell, P. Fordtonin, D. T. Bishop, N. K. Spurr, B. A. J. Ponder, R. Eeles, J. Peto, P. Devilee, C. Cornelisse, H. Lynch, S. Narod, G. Lenoir, V. Egilsson, R. B. Barkadottir, D. F. Easton, D. R. Bentley, P. A. Futreal, A. Ashworth, and M. R. Stratton, "Identification of the breast cancer susceptibility gene brca2," *Nature*, vol. 378, pp. 789–792, 1995.
- [84] J. Zheng, I. Mercado-Uribe, D. G. Rosen, B. Chang, P. Liu, G. Yang, A. Malpica, H. Naora, N. Auersperg, G. B. Mills, R. C. Bast, and J. Liu, "Induction of papillary carcinoma in human ovarian surface epithelial cells using combined genetic elements and peritoneal microenvironment," *Cell Cycle*, vol. 9, no. 1, pp. 140–146, 2010.
- [85] M. Olivier, M. Hollstein, and P. Hainaut, "Tp53 mutations in human cancers: origins, consequences, and clinical use," *Cold Spring Harbor Perspectives in Biology*, vol. 2, p. a001008, 2010.
- [86] L. Havrilesky, M. Darcy, H. Hamdan, R. L. Priore, J. Leon, J. Bell, A. Berchuck, and G. O. G. Study, "Prognostic significance of p53 mutation and p53 overexpression in advanced epithelial ovarian cancer: a gynecologic oncology group study," *Journal of Clinical Oncology*, vol. 21, pp. 3814–3825, 2003.

- [87] M. Bernardini, T. Baba, P. Lee, J. Barnett, G. Sfakianos, A. Secord, S. Murphy, E. Iversen, J. Marks, and A. Berchuck, "Expression signatures of tp53 mutations in serous ovarian cancers," *BMC Cancer*, vol. 10, no. 1, p. 237, 2010.
- [88] Y. Lee, A. Miron, R. Drapkin, M. R. Nucci, F. Medeiros, A. Saleemuddin, J. Garber, C. Birch, H. Mou, R. W. Gordon, D. W. Cramer, F. D. McKeon, and C. P. Crum, "A candidate precursor to serous carcinoma that originates in the distal fallopian tube," *Journal of Pathology*, vol. 211, pp. 26–35, 2007.
- [89] M. M. Maslon and T. R, "Drug discovery and mutant p53," *Trends in Cell Biology*, vol. 20, pp. 542–555, 2010.
- [90] M. S. Anglesio, J. M. Arnold, J. George, A. V. Tinker, R. Tothill, N. Waddell, L. Simms, B. Locandro, S. Fereday, N. Traficante, P. Russell, R. Sharma, M. J. Birrer, A. S. Group, A. deFazio, G. Chenevix-Trench, and D. D. Bowtell, "Mutation of erbb2 provides a novel alternative mechanism for the ubiquitous activation of ras-mapk in ovarian serous low malignant potential tumors," *Molecular Cancer Research*, vol. 6, pp. 1678–1690, 2008.
- [91] R. Vang, M. Shih, and R. J. Kurman, "Ovarian low-grade and high-grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems," *Advances in Anatomic Pathology*, vol. 16, pp. 267–282, 2009.
- [92] J. P. Venerables, "Aberrant and alternative splicing in cancer," *Cancer Research*, vol. 64, pp. 7647–7654, 2004.
- [93] N. A. Faustino and T. A. Cooper, "Pre-mrna splicing and human disease," *Genes and Development*, vol. 17, pp. 419–437, 2003.
- [94] M. Weiss, A. Baruch, I. Keydar, and D. H. Wreschner, "Preoperative diagnosis of thyroid papillary carcinoma by reverse transcriptase polymerase chain reaction of the mucl gene," *International Journal of Cancer*, vol. 66, pp. 55–9, 1996.
- [95] N. Cheung, M. P. Wong, S. T. Yeun, S. Y. Leung, and L. P. Chung, "Tissue-specific expression pattern of vascular endothelial growth factor isoforms in the malignant transformation of lung and colon," *Human Pathology*, vol. 29, pp. 910–914, 1998.
- [96] J. C. Bourdon, K. Fernandes, F. Murray-Zmijewski, G. Liu, A. Diot, D. P. Xirodimas, and D. P. Lane, "p53 isoforms can regulate p53 transcriptional activity," *Genes Development*, vol. 19, pp. 2122–2137, 2005.
- [97] J. D. Hoffman, S. E. Hallan, V. L. Venne, E. Lyon, and K. Ward, "Implications of cryptic splice site in the brca1 gene," *American Journal of Medical Genetics*, vol. 80, pp. 140–144, 1998.

- [98] R. Klinck, A. Bramard, L. Inkel, G. Dufresne-Martin, J. Gervais-Bird, R. Madden, E. R. Paquet, C. Koh, J. P. Venables, P. Prinos, M. Jilaveanu-Pelms, R. Wellinger, C. Rancourt, B. Chabot, and S. A. Elela, "Multiple alternative splicing markers for ovarian cancer," *Cancer Research*, vol. 68, pp. 657–663, 2008.
- [99] P. J. Gardina, T. A. Clark, B. Shimada, M. K. Staples, Q. Yang, J. Veitch, A. Schweitzer, T. Awad, C. Sunet, S. Dee, C. Davies, A. Williams, and Y. Turpaz, "Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array," *BMC Genomics*, vol. 7, 2006.
- [100] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids," *JAMA: The Journal of the American Medical Association*, vol. 269, no. 15, pp. 1966–1967, 1993.
- [101] U. Masks and E. M. Southern, "Parallel analysis of oligodeoxyribonucleotide (oligonucleotide) interactions. i. analysis of factors influencing oligonucleotide duplex formation," *Nucleic Acids Research*, vol. 20, no. 7, pp. 1675–1678, 1992.
- [102] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. Fodor, "Light-generated oligonucleotide arrays for rapid dna sequence analysis," *Proceedings of the National Academy of Sciences*, vol. 91, no. 11, pp. 5022–5026, 1994.
- [103] Affymetrix Inc, *GeneChip[®] Whole Transcript (WT) Sense Target Labeling Assay Manual*, version 4 ed., 2007.
- [104] Affymetrix, Inc, CA, *Technical Note: GeneChip[®] Exon Array Design*, 2005.
- [105] Affymetrix, Inc, CA, *Exon Probeset Annotations and Transcript Cluster Groupings*, version 1.0 ed., 2005.
- [106] G. Russo, C. Zegar, and A. Giordano, "Advantages and limitation of microarray technology in human cancer," *Oncogene*, vol. 22, pp. 6497–6507, 2003.
- [107] A. D. Santin, F. Zhan, S. Bellone, M. Palmieri, S. Cane, E. Bignotti, S. Anfossi, M. Gokden, D. Dunn, J. J. Roman, T. J. O'brien, E. Tian, M. J. Cannon, J. John Shaughnessy, and S. Percorelli, "Gene expression profiles in primary ovarian serous and papillary tumors and normal ovarian epithelium: identification of candidate molecular markers for ovarian cancer diagnosis and therapy," *International Journal of Cancer*, vol. 112, pp. 14–20, 2004.
- [108] R. S. Ismail, R. L. Baldwin, J. Fang, D. Browning, B. Y. Karlan, J. C. Gasson, and D. D. Chang, "Differential gene expression between normal and tumor-derived ovarian epithelial cells," *Cancer Research*, vol. 60, pp. 6744–6749, 2000.
- [109] K. Ono, T. Tanaka, T. Tsunoda, O. Kitahara, C. Kihara, A. Okamoto, K. Ochial, T. Takagi, and Y. Nakamura, "Identification by cDNA microarray of

- genes involved in ovarian carcinogenesis,” *Cancer Research*, vol. 60, pp. 5007–5011, 2000.
- [110] J. B. Welsh, P. P. Zarrinkar, L. M. Sapinoso, S. G. Kern, C. A. Behling, B. J. Monk, D. J. Lockhart, R. A. Burger, and G. M. H. G. M, “Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, pp. 1176–81, 2001.
- [111] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron, “Minimum information about a microarray experiment (miame) - towards standards for microarray data,” *Nature Genetics*, vol. 29, pp. 365–371, 2001.
- [112] P. T. Spellmand, M. Miller, J. Stewart, C. Troup, U. Sarkans, S. Chervitz, D. Bernhart, G. Sherlock, C. Ball, M. Lepage, M. Swiatek, W. Marks, J. Goncalves, S. Markel, D. Iordan, M. Shojatalab, A. Pizarro, J. White, R. Hubley, E. Deutsch, M. Senger, B. J. Aronow, A. Robinson, D. Bassett, C. J. Stoeckert, and A. Brazma, “Design and implementation of microarray gene expression markup language,” *Genome Biology*, vol. 3, pp. research0046–research0046.9, 2002.
- [113] T. Barrett, T. O. Suzek, D. B. Troup, S. E. Wilhite, W. C. Ngau, P. Ledoux, D. Rudnev, A. E. Lash, W. Fujibuchi, and R. Edgar, “Ncbi geo: mining millions of expression profiles,” *Nucleic Acids Research*, pp. D526–566, 2005.
- [114] “Genechip[®] human genome array: the most comprehensive coverage of all well-substantiated genes in the human genome.”
- [115] L. C. Hartmann, K. H. Lu, G. P. Linette, W. A. Cliby, K. R. Kalli, D. Gershenson, R. C. Bast, J. Stec, N. Iartchouk, D. I. SMith, J. S. Ross, S. Hoersch, V. Shridhar, J. Lillie, S. H. Kaufmann, E. A. Clark, and A. I. Damokosh, “Gene expression profiles predict early relapse in ovarian cancer after platinum-paclitaxel chemotherapy,” *Clinical Cancer Research*, vol. 11, p. 2149, 2005.
- [116] D. Spentzos, D. A. Levine, M. F. Ramoni, M. Joseph, X. Gu, J. Byod, T. A. Libermann, and S. A. Cannistra, “Gene expression signature with independent prognostic significance in epithelial ovarian cancer,” *Journal of Clinical Oncology*, vol. 22, no. 23, 2004.
- [117] A. Berchuck, E. S. Iverson, J. M. Lancaster, H. K. Dressman, M. West, J. R. Nevins, and J. R. Marks, “Prediction of optimal versus suboptimal cytoreduction of advanced-stage serous ovarian cancer with the use of microarrays,” *American Journal of Obstetrics and Gynecology*, vol. 190, pp. 910–923, 2004.

- [118] H. K. Dressman, A. Berchuck, G. Chan, J. Zhai, A. Bild, R. Sayer, J. Cragun, J. Clarke, R. S. Whitaker, L. Li, J. Gray, J. Marks, G. S. Ginsburg, A. Potti, M. West, J. R. Nevins, and J. M. Lancaster, “An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer,” *Journal of Clinical Oncology*, vol. 25, no. 5, pp. 517–525, 2007.
- [119] J. Hellen, M. P. H. M. Jansen, P. N. Span, I. L. van Staveren, L. F. A. G. Massuger, M. E. M. van Gelder, F. C. G. J. Sweep, P. C. Ewing, M. E. L. van der Burg, G. Stoter, K. Nooter, and E. M. J. J. Berns, “Molecular profiling of platinum resistant ovarian cancer,” *International Journal of Cancer*, vol. 118, pp. 1963–1971, 2006.
- [120] D. Spentzos, D. A. Levine, S. Koli, H. Otu, J. Boyd, T. A. Libermann, and S. A. Cannistra, “Unique gene expression profile based on pathologic response in epithelial ovarian cancer,” *Journal of Clinical Oncology*, vol. 23, no. 31, pp. 7911–7918, 2005.
- [121] P. A. Konstantinopoulos, D. Spentzos, and S. A. Cannistra, “Gene-expression profiling in epithelial ovarian cancer,” *Nature Clinical Practice*, vol. 5, no. 10, pp. 577–587, 2008.
- [122] T. C. G. A. T. R. Network, “Comprehensive genomic characterization defines human glioblastoma genes and core pathways,” *Nature*, vol. 455, pp. 1061–1068, 2008.
- [123] Life Technologies Corporation, Carlsbad, CA, *Arcturus^{XTM} Laser Capture Microdissection System: User Guide*, november 2010 ed., 2008.
- [124] NuGEN Technologies Inc, CA, *FL-OvationTM cDNA Biotin Module V2 User Guide*, version 05.13.08 ed.
- [125] Affymetrix, Inc., CA, *Quality Assessment of Exon Arrays*, version 1.0 ed., 2005.
- [126] Affymetrix, Inc, CA, *Exon Array Background Correction*, revision version 1.0 ed., 2005.
- [127] Z. Wu, R. A. Irizarry, R. Gentleman, F. Martinez-Murillo, and F. Spencer, “A model-based background adjustment for oligonucleotide expression arrays,” *Journal of the American Statistical Association*, vol. 99, no. 468, pp. 909–917, 2004.
- [128] H. Zar, *Biostatistical Analysis*. New Jersey: Prentice-Hall, Inc, 4th ed., 1999.
- [129] C. Li and W. Hung Wong, “Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application,” *Genome Biology*, vol. 2, no. 8, pp. research0032.1–research0032.11, 2001.

- [130] R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed, "Summaries of affymetrix genechip probe level data," *Nucleic Acids Research*, vol. 31, no. 4, p. e15, 2003.
- [131] H. R. Lindman, *Analysis of Variance in Complex Experimental Designs*. San Francisco, CA: W.H.Reeman and Company, 1974.
- [132] J. W. K. Ho, M. Stefani, C. G. dos Remedios, and M. A. Charleston, "Differential variability analysis of gene expression and its application to human diseases," *Bioinformatics*, vol. 24, pp. i390–i398, 2008.
- [133] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. pp. 583–621, 1952.
- [134] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using david bioinformatics resources," *Nature Protocols*, vol. 4, pp. 44–57, 2008.
- [135] G. Dennis, B. Sherman, D. Hosack, J. Yang, W. Gao, H. C. Lane, and R. Lempicki, "David: Database for annotation, visualization, and integrated discovery," *Genome Biology*, vol. 4, no. 5, p. P3, 2003. This is the first version of this article to be made available publicly. A peer-reviewed and modified version is now available in full at <http://genomebiology.com/2003/4/9/R60>.
- [136] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal for the Royal Statistical Society*, vol. 57, no. 1, pp. 289–300, 1995.
- [137] S. Carbon, A. Ireland, C. J. Mungall, S. Shu, B. Marshall, and S. Lewis, "Amigo: online access to ontology and annotation data," *Bioinformatics*, vol. 25, pp. 288–289, 2009.
- [138] L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, and C. von Mering, "String 8—a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Research*, vol. 37, pp. D412–D416, 2009.
- [139] S. I. Berger, J. M. Posner, and A. Ma'ayan, "Genes2networks: connecting lists of gene symbols using mammalian protein interactions databases," *BMC Bioinformatics*, vol. 8, p. 372, 2007.
- [140] B. Aranda, P. Achuthan, Y. Alam-Faruque, I. Armean, A. Bridge, C. Derow, M. Feuermann, A. T. Ghanbarian, S. Kerrien, J. Khadake, J. Kerssemakers, C. Leroy, M. Menden, M. Michaut, L. Montecchi-Palazzi, N. Neuhauser, S. Orchard, V. Perreau, B. Roechert, K. van Eijk, and H. Hermjakob, "The intact molecular interaction database in 2010," *Nucleic Acids Research*, 2009.

- [141] C. Stark, B. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, "Biogrid: a general repository for interaction datasets," *Nucleic Acids Research*, vol. 34, pp. D535–D539, 2006.
- [142] A. Ceol, A. C. Aryamontri, L. Licata, D. Peluso, L. Briganti, L. Perfetto, L. Castagnoli, and G. Cesareni, "Mint, the molecular interaction database: 2009 update," *Nucleic Acids Research*, vol. 38, pp. D532–D539, 2010.
- [143] I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte, and D. Eisenberg, "Dip: the database of interacting proteins," *Nucleic Acids Research*, vol. 28, pp. 289–291, 2000.
- [144] G. D. Bader, I. Donaldson, C. Wolting, B. F. Ouellette, T. Pawson, and C. W. Hogue, "Bind-the biomolecular interaction network database," *Nucleic Acids Research*, vol. 29, pp. 242–245, 2001.
- [145] M. Kanehisa and S. Goto, "Kegg, kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [146] G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein, "Reactome: a knowledgebase of biological pathways," *Nucleic Acids Research*, vol. 33, pp. D428–D32, 2005.
- [147] T. S. K. Prasad, R. Goel, K. Kandasamy, Shivakumar, Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venigopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Raganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, "Human protein reference database - 2009 update," *Nucleic Acids Research*, vol. 37, pp. D767–772, 2009.
- [148] T. G. O. Consortium, "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, pp. 25–29, 2000.
- [149] T. Beuming, Skrabanek, M. Y. Niv, Mukherjee, and H. Weinstein, "Pdzbases: a protein-protein interaction database for pdz-domains," *Bioinformatics*, vol. 21, no. 6, pp. 827–828, 2005.
- [150] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Bremback, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlauff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Groege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker, "A human protein-protein interaction network: a resource for annotating the proteome," *Cell*, vol. 122, no. 6, pp. 957–968, 2004.

- [151] J. F. Raul, K. Venkayesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxem, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. L. S. Cevik, C. Bex, P. Lamesch, R. S. Sikprski, J. Vandenhoute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, and M. Vidal, "Towards a proteome-scale map of the human protein-protein interaction network," *Nature*, vol. 437, pp. 1173–1178, 2005.
- [152] D. Ghosh, T. R. Barette, D. Rhodes, and A. M. Chinnaiyan, "Statistical issues and methods for meta-analysis of microarray data: a case study in prostate cancer," *Functional and Integrative Genomics*, vol. 3, no. 4, pp. 180–188, 2003.
- [153] J. W. Tukey and D. H. McLaughlin, "Less vulnerable confidence and significance procedures for location based on a single sample: trimming/winsorization," *Sankhyā:the Indian Journal of Statistics, Series A*, vol. 25, no. 3, pp. 331–352, 1963.
- [154] R. P. Lane, X.-N. C. and Kazuhiro Yamakawa, J. Vielmetter, J. R. Korenberg, and W. J. Dreyer, "Characterization of a highly conserved human homolog to the chicken neural cell surface protein bravo/nr-cam that maps to chromosome band 7q31," *Genomics*, vol. 35, pp. 456–465, 1996.
- [155] Z. Qian, J. Qian, J. Lon, D. ming Yao, Q. Chen, R. bi Ji, Y. Li, G. fei Xiao, and J. yong Li, "Gtpase regulator associated with the focal adhesion kinase (graf) transcript was down-regulated in patients with myeloid malignancies," *Journal of Experimental and Clinical Cancer Research*, vol. 29, 2010.
- [156] Y. Yuecheng, L. Hongmei, and X. Xiaoyan, "Clinical evaluation of e-cadherin expression and its regulation mechanism in epithelial ovarian cancer," *Clinical & Experimental Metastasis*, vol. 23, pp. 65–74, 2006.
- [157] D. Vergara, B. Merlot, J.-P. Lucot, P. Collinet, D. Vinatier, I. Fournier, and M. Salzet, "Epithelial-mesenchymal transition in ovarian cancer," *Cancer Letters*, vol. 291, pp. 59–66, 2010.
- [158] C. D. Hough, C. A. Sherman-Baust, E. S. Pizer, F. J. Montz, D. D. Im, N. B. Rosenshein, K. R. Cho, G. J. Figgins, and P. J. Morin, "Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer," *Cancer Research*, vol. 60, pp. 6281–6287, 2000.
- [159] L. Kleinberg, A. H. C. G. Trope, R. Reich, and B. Davidson, "Claudin upregulation in ovarian carcinoma effusion is associated with poor survival," *Human Pathology*, vol. 39, pp. 747–757, 2008.
- [160] D. Michel, J.-P. Arsanto, D. Massey-Harroche, C. Béclin, J. Wijnholds, and A. L. Bivic, "Patj connects and stabilizes apical and lateral components of

- tight junctions in human intestinal cells,” *Journal of Cell Science*, vol. 118, pp. 4049–4057, 2005.
- [161] J. Doorbar, “Molecular biology of human papillomavirus infection and cervical cancer.,” *Clin. Sci.*, vol. 110, no. 5, pp. 525–541, 2006.
- [162] V. M. Stucke, E. Timmerman, J. Vandekerckhove, K. Gevaert, and A. Hall, “The maguk protein mpp7 binds to the polarity protein hdlg1 and facilitates epithelial tight junction formation,” *Molecular Biology of the Cell*, vol. 18, pp. 1744–1755, 2007.
- [163] T. W. Hurd, L. Gao, M. H. Roh, I. G. Macara, and B. Margolis, “Direct interaction of two polarity complexes implicated in epithelial tight junction assembly,” *Nature Cell Biology*, vol. 5, pp. 137–42, 2003.
- [164] A. M. Schindler, M. A. Amaudruz, O. Kocher, G. Riotton, and G. Gabbiani, “Desmosomes and gap-junctions in various epidermoid preneoplastic and neoplastic lesions of the cervix uteri,” *Acta Cytologica*, vol. 26, pp. 466–70, 1982.
- [165] O. Kocher, M. Amaudruz, A. M. Schindler, and G. Gabbiani, “Desmosomes and gap junctions in precarcinomatous and carcinomatous conditions of squamous epithelia. an electron microscopic and morphometrical study,” *Journal of Submicroscopic Cytology*, vol. 12, pp. 267–281, 1981.
- [166] J. Depondt, A. H. Shabana, S. Florescu-Zorila, P. Gehanno, and N. Forest, “Down-regulation of desmosomal molecules in oral and pharyngeal squamous cell carcinomas as a marker for tumor growth and distant metastasis,” *European Journal of Oral Sciences*, vol. 107, pp. 183–193, 1999.
- [167] S. Papagerakis, A.-H. Shabana, B. H. Polluck, P. Papagerakis, J. Depondt, and A. Berdal, “Altered desmoplakin expression at transcriptional and protein levels provides prognostic information in human oropharyngeal cancer,” *Human Pathology*, vol. 40, pp. 1320–1329, 2009.
- [168] L. D. Attardi, E. E. Reczek, C. Cosmas, E. G. Demicco, M. E. McCurrach, S. W. Lowe, and T. Jacks, “Perp, an apoptosis-associated target of p53, is a novel member of the pmp-22/gas3 family,” *Genes & Development*, vol. 14, pp. 704–718, 2000.
- [169] L. Davies, D. Gray, D. Spiller, M. R. White, B. Damato, I. Grierson, and L. Paraoan, “P53 apoptosis mediator perp: localization, function and caspase activation in uveal melanoma,” *Journal of Cellular and Molecular Medicine*, vol. 13, pp. 1995–2007, 2009.
- [170] T. Hildebrandt, J. Preiherr, N. Tarbé, S. Klostermann, G. N. V. Muijen, and U. H. Weidle, “Identification of thw, a putative new tumor suppressor gene,” *Anticancer Research*, vol. 20, pp. 2801–2809, 2000.

- [171] H. M. Schuller, H. A. N. Al-Wadei, and M. Majidi, "Gaba_B receptor is a novel drug target for pancreatic cancer," *Cancer*, vol. 112, pp. 767–778, 2008.
- [172] R. Rong, J. Y. Ahn, H. Huang, E. Nagata, D. Kalman, J. A. Kapp, J. Tu, P. F. Worley, S. H. Snyder, and K. Ye, "Pi3 kinase enhancer-homer complex couples mglur1 to pi3 kinase, preventing neuronal apoptosis," *Nature Neuroscience*, vol. 6, pp. 1153–61, 2003.
- [173] L. B. A. Rangel, R. Agarwal, T. D'Souza, E. S. Pizer, P. L. Alò, W. D. Lancaster, L. Gregoire, D. R. Schwartz, K. R. Cho, and P. J. Morin, "Tight junction proteins claudin-3 and claudin-4 are frequently overexpressed in ovarian cancer but not in ovarian cystadenomas," *Clinical Cancer Research*, vol. 9, pp. 2567–2575, 2003.
- [174] D. G. Hayward, R. B. Clarke, A. J. Faragher, M. R. Pillai, I. M. Hagan, and A. M. Fry, "The centrosomal kinase nek2 displays elevated levels of protein expression in human breast cancer," *Cancer Research*, pp. 7370–7376, 2004.
- [175] T. Yamamoto, H. Matsuyama, Y. Chochi, M. Okuda, S. Kawauchi, R. Inoue, T. Furuya, A. Oga, K. Naito, and K. Sasaki, "Overexpression of bubr1 is associated with chromosomal instability in bladder cancer," *Cancer Genetics and Cytogenetics*, pp. 42–47, 2007.
- [176] S. L. O'Brien, A. Fagan, E. J. P. Fox, R. C. Millikan, A. C. Culhane, D. J. Brennan, A. H. McCann, S. Hegarty, S. Moyna, M. J. Duffy, D. G. Higgins, K. Jirström, G. Landberg, and W. M. Gallagher, "Cenp-f expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer," *International Journal of Cancer*, vol. 120, pp. 1434–1443, 2007.
- [177] J. Feng, H. Huang, and T. J. Yen, "Cenp-f is a novel microtubule-binding protein that is essential for kinetochore attachments and affects the duration of the mitotic checkpoint delay," *Chromosoma*, vol. 115, pp. 320–329, 2006.
- [178] J. Du, T. Li, and X. Zhu, "Involvement of cenp-f in histone methylation," *Acta Biochimica et Biophysica Sinica*, vol. 42, pp. 173–176, 2010.
- [179] C. Millinari, C. Reynaud, S. Martineau-Thuillier, S. Monier, S. Kieffer, J. Garin, P. R. Andreassen, A. Boulet, B. Goud, J.-P. Kleman, and R. L. Margolis, "The mammalian passenger protein td-60 is an rcc1 family member with an essential role in prometaphase to metaphase progression," *Developmental Cell*, vol. 5, pp. 295–307, 2003.
- [180] A. L. Nielsen, M. Oulad-Abdelghani, J. A. Ortiz, E. Remboutsika, P. Chambon, and R. Losson, "Heterochromatin formation in mammalian cells: interaction between histones and hp1 proteins," *Molecular Cell*, vol. 7, pp. 729–739, 2001.

- [181] E. M. Dunleavy, D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, Y. Nakamura, Y. Daigo, Y. Nakatani, and G. Almouzni-Pettinotti, "Hjurp is a cell-cycle-dependent maintenance and deposition factor of cenp-a at centromeres," *Cell*, vol. 137, pp. 485–497, 2009.
- [182] B. K. Tan, R. Adya, S. Farhatullah, J. Chen, H. Lehnert, and H. S. Randeva, "Metformin Treatment May Increase Omentin-1 Levels in Women With Polycystic Ovary Syndrome," *Diabetes*, vol. 59, no. 12, pp. 3023–3031, 2010.
- [183] Y. J. Zhang, W. K. O'Neal, S. H. Randell, K. Blackburn, M. B. Moyer, R. C. Boucher, and L. E. Ostrowski, "Identification of dynein heavy chain 7 as an inner arm component of human cilia that is synthesized but not assembled in a case of primary ciliary dyskinesia," *The journal of biological chemistry*, vol. 277, pp. 17906–17915, 2002.
- [184] C. Chapellin, B. Duriez, F. Magnino, M. Goossens, E. Escudier, and S. Amselem, "Isolation of several human axonemal dynein heavy chain genes: genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment," *FEBS Letters*, vol. 412, pp. 325–330, 1997.
- [185] E. van Rooijen, R. H. Giles, E. E. Voest, C. van Rooijen, S. Schulte-Merker, and F. J. van Eeden, "Lrrc50, a conserved ciliary protein implicated in polycystic kidney disease," *Journal of the American Society of Nephrology*, vol. 19, pp. 1128–1138, 2008.
- [186] Y. la Choi, S.-H. Kim, Y. K. Shin, Y.-C. Hong, S.-J. Lee, S. Y. Kang, and G. Ahn, "Cytoplasmic cd24 expression in advanced ovarian serous borderline tumors," *Gynecologic Oncology*, vol. 97, pp. 379–386, 2005.
- [187] Z. Suo, E. Karbovo, C. G. Trope, K. Metodiev, and J. M. Nesland, "Papillary serous carcinoma of the ovary: an ultrastructural and immunohistochemical study," *Ultrastructural Pathology*, vol. 28, pp. 141–7, 2004.
- [188] O. V. Plotnikova, E. A. Golemis, and E. N. Pugacheva, "Cell cycle-dependent ciliogenesis and cancer," *Cancer Research*, vol. 68, 2008.
- [189] E. S. Seeley and M. V. Nachury, "Constructing and deconstructing roles for the primary cilium in tissue architecture and cancer," *Methods Cell Biology*, vol. 94, pp. 299–313, 2009.
- [190] P. Satir and S. T. Christensen, "Overview of structure and function of mammalian cilia," *Annual Review of Physiology*, vol. 69, pp. 377–400, 2007.
- [191] C. L. Rieder, C. G. Jensen, and L. C. W. Jensen, "The resorption of primary cilia during mitosis in a vertebrate (ptk₁) cell line," *Journal of ultrastructure research*, vol. 68, pp. 173–185, 1979.

- [192] F. Lin, T. Hiesberger, K. Cordes, A. M. Sinclair, L. S. Goldstein, S. Somlo, and P. Igarashi, "Kidney-specific inactivation of the kif3a subunit of kinesin-ii inhibits renal ciliogenesis and produces polycystic kidney disease," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 5286–91, 2003.
- [193] B. J. Sifoky, W. B. Ferguson, A. L. Fuson, Y. Xie, A. Fintha, P. Komlosi, B. K. Yoder, E. M. Schwiebert, L. M. Guay-Woodford, and P. D. Bell, "Loss of primary cilia results in deregulated and unabated apical calcium entry in arpkd collecting duct cells," *American journal of physiology. Renal physiology*, vol. 290, 2006.
- [194] M. H. Meisler, L. L. Barrow, C. M. Canessa, and B. C. Rossier, "Scnn1a, an epithelial cell sodium channel gene in the conserved lineage group on mouse chromosome 6 and human chromosome 12," *Genomics*, vol. 24, pp. 185–186, 1994.
- [195] P. A. Baeuerle and O. Gires, "Epcam (cd326) finding its role in cancer," *British Journal of Cancer*, vol. 96, pp. 417–423, 2007.
- [196] V. A. Heinzelmann-Schwarz, M. Gardiner-Garden, S. M. Henshall, J. Scurry, R. A. Scolyer, M. J. Davies, M. Heinzelmann, L. H. Kalish, A. Bali, J. G. Kench, L. S. Edwards, P. M. V. Bergh, N. F. Hacker, R. L. Sutherland, and P. M. O'Brien, "Overexpression of the cell adhesion molecules ddr1, claudin 3, and ep-cam in metaplastic ovarian epithelium and ovarian cancer," *Clinical Cancer Research*, vol. 10, pp. 4427–4436, 2004.
- [197] J.-H. Kim, D. Herlyn, K. kwok Wong, D.-C. Park, J. O. Schorge, K. H. Lu, S. J. Skates, D. W. Cramer, R. S. Berkowitz, and S. C. Mok, "Identification of epithelial cell adhesion molecule autoantibody in patients with ovarian cancer," *Clinical Cancer Research*, vol. 9, pp. 4782–4791, 2003.
- [198] G. Spizzo, P. Went, S. Dirnhofer, P. Obrist, H. Moch, P. A. Baeuerle, E. Mueller-Holzner, C. Marth, G. Gastl, and A. G. Zelman, "Overexpression of epithelial cell adhesion molecule (ep-cam) is an independent prognostic marker for reduced survival of patients with epithelial ovarian cancer," *Gynecologic Oncology*, vol. 103, pp. 483–488, 2006.
- [199] S. Bellone, E. R. Siegel, E. Cocco, M. Cargnelutti, D. A. Silasi, M. Azodi, P. E. Schwartz, T. J. Rutherford, S. Pecorelli, and A. D. Santin, "Overexpression of epithelial cell adhesion molecule in primary, metastatic, and recurrent/chemotherapy-resistant epithelial ovarian cancer: implications for epithelial cell adhesion molecule-specific immunotherapy," *International Journal of Gynecological Cancer*, vol. 19, pp. 860–866, 2009.
- [200] C. H. M. J. V. Elssen, P. W. H. Frings, F. J. Bot, K. K. V. de Vijver, M. B. Huls, B. Meek, P. Hupperets, W. T. V. Germeraad, and G. M. J. Bos, "Expression

- of aberrantly glycosylated mucin-1 in ovarian cancer,” *Histopathology*, vol. 57, pp. 597–606, 2010.
- [201] L. Wang, J. Ma, F. Liu, Q. Yu, G. Chu, A. C. Perkins, and Y. Li, “Expression of muc1 in primary and metastatic human epithelial ovarian cancer and its therapeutic significance,” *Gynecologic Oncology*, vol. 105, pp. 695–702, 2007.
- [202] D. T. Jacobs, R. Weigert, K. D. Grode, J. G. Donaldson, and R. E. Cheney, “Myosin vc is a molecular motor that functions in secretory granule trafficking,” *Molecular Biology of the Cell*, vol. 20, pp. 4471–4488, 2009.
- [203] M. A. Kash, A. E. Loercher, and R. S. Freedman, “In vitro growth inhibition of ovarian cancer cells by decorin: synergism of action between decorin and carboplatin,” *Cancer Research*, vol. 59, pp. 6192–6196, 1999.
- [204] Y. Zhai, L. Chen, M. Hömme, T. Hackert, M.-L. Gross, G. F. Huffmann, F. Schaefer, and C. P. Schmitt, “Expression and function of matrix gla protein in human peritoneal mesothelial cells,” *Nephrology, dialysis, transplantation*, vol. 25, pp. 3213–3221, 2010.
- [205] S. Nomoto, M. Kanda, Y. Okamura, Y. Nishikawa, L. Qiyong, T. Fuji, H. Sugimoto, S. Takeda, and A. Nakao, “Epidermal growth factor-containing fibulin-like extracellular matrix protein 1, efemp1, a novel tumor-suppressor gene detected in hepatocellular carcinoma using double combination array analysis,” *Annals of surgical oncology*, vol. 17, pp. 923–932, 2010.
- [206] A. Sadr-Nabavi, J. Ramser, J. Volkmann, J. Naehrig, F. Wiesmann, B. Betz, H. Hellbrand, S. Engert, S. Seitz, R. Kreutzfeld, T. Sasaki, N. Arnold, R. Schmutzler, M. Kiechle, D. Niederacher, N. Harbeck, E. Dahl, and A. Meindl, “Decreased expression of angiogenesis antagonist efemp1 in sporadic breast cancer is caused by aberrant promoter methylation and points to an impact of efemp1 as molecular biomarker,” *International journal of cancer*, vol. 124, pp. 1727–1735, 2009.
- [207] E. S. Lee, A. S.-Y. Leong, Y.-S. Kim, J. han Lee, I. Kim, G. H. Ahn, H. S. Kim, and Y. K. Chun, “Calretinin, cd 34, and α -smooth muscle actin in the identification of peritoneal invasive implants of serous borderline tumors of the ovary,” *Modern Pathology*, vol. 19, pp. 364–372, 2006.
- [208] A. V. Andreeva, M. A. Kutuzov, R. Vaiskunaite, J. Profirovic, T. E. Meigs, S. Predescu, A. B. Malik, and T. Voyno-yasenetskaya, “G α_{12} interaction with α induces ve-caderin localization at endothelial junctions and regulates barrier functions,” *The journal of biological chemistry*, vol. 280, no. 34, pp. 30376–30383, 2005.
- [209] C. Page, H. J. Lin, Y. Jin, V. P. Castle, G. Nunez, M. Huang, and J. Lin, “Overexpression of akt/akt can modulate chemotherapy-induced apoptosis,” *Anticancer Research*, vol. 20, pp. 407–416, 2000.

- [210] B. E. Christiano, J. C. Chan, K. M. Hannan, N. A. Lundie, N. J. Marmy-Conus, I. G. Campbell, W. A. Phillips, M. Robbie, R. D. Hannan, and R. B. Pearson, "A specific role for akt3 in the genesis of ovarian cancer through modulation of g₂-m phase transition," *Cancer Research*, vol. 66, pp. 11718–11725, 2006.
- [211] S. J. Fuller, K. Sivarajah, and P. H. Sugden, "ErbB receptors, their ligands, and the consequences of their activation and inhibition in the myocardium," *The Journal of Molecular and Cellular Cardiology*, vol. 44, pp. 831–854, 2008.
- [212] B. Górka, J. Skubis-Zegadlo, M. Mikula, K. Bardadin, E. Paliczka, and B. Czarnocka, "Nrcam, a neuronal system cell-adhesion molecule, is induced in papillary thyroid carcinomas," *BMC Cancer*, vol. 97, pp. 531–538, 2007.
- [213] M. Taniwaki, Y. Daigo, N. Ishikawa, A. Takano, T. Tsunoda, W. Yasui, and K. Inai, "Gene expression profiles of small-cell lung cancers: molecular signatures of lung cancer," *International Journal of Oncology*, vol. 29, pp. 567–575, 2006.
- [214] M. Conacci-Sorrell, J. Zhurinsky, and A. Ben-Ze'ev, "The cadherin-catenin adhesion system in signaling and cancer," *Journal of Clinical Investigation*, vol. 109, pp. 987–991, 2002.
- [215] M. Conacci-Sorrell, A. Kaplan, S. Raveh, N. Gavert, T. Sakurai, and A. Ben-Ze'ev, "The shed ectodomain of nr-cam stimulates cell proliferation and motility, and confers cell transformation," *Cancer Research*, vol. 65, pp. 11605–11612, 2005.
- [216] J. A. Byrne, S. Maleki, J. R. Hardy, B. S. Gloss, R. Murali, J. P. Scurry, S. Fanayan, C. Emmanuel, N. F. Hacker, R. L. Sutherland, A. Defazio, and P. M. O'Brien, "Mal2 and tumor protein d52 (tpd52) are frequently overexpressed in ovarian carcinoma, but differentially associated with histological subtype and patient outcome," *BMC Cancer*, vol. 10, p. 497, 2010.
- [217] E. Kodym, R. Kodym, B. P. Chen, D. J. Chen, K. Morotomi-Yano, H. Choy, and D. Saha, "Dna-pkcd-dependent modulation of cellular radiosensitivity by a selective cyclooxygenase-2 inhibitor," *International Journal of Radiation Oncology, Biology, Physics*, vol. 69, pp. 187–193, 2007.
- [218] R. R. Jr, "The erbb/her receptor protein-tyrosine kinases and cancer," *Biochemical and Biophysical Research Communications*, vol. 319, pp. 1–11, 2004.
- [219] A. K. Koutras, G. Fountzilas, K. T. Kalogeras, I. Starakis, G. Iconomou, and H. P. Kalofonos, "The upgraded role of her3 and her4 receptors in breast cancer," *Critical Reviews in Oncology/Hematology*, vol. 74, pp. 73–78, 2010.

APPENDIX A: LCM IMAGES

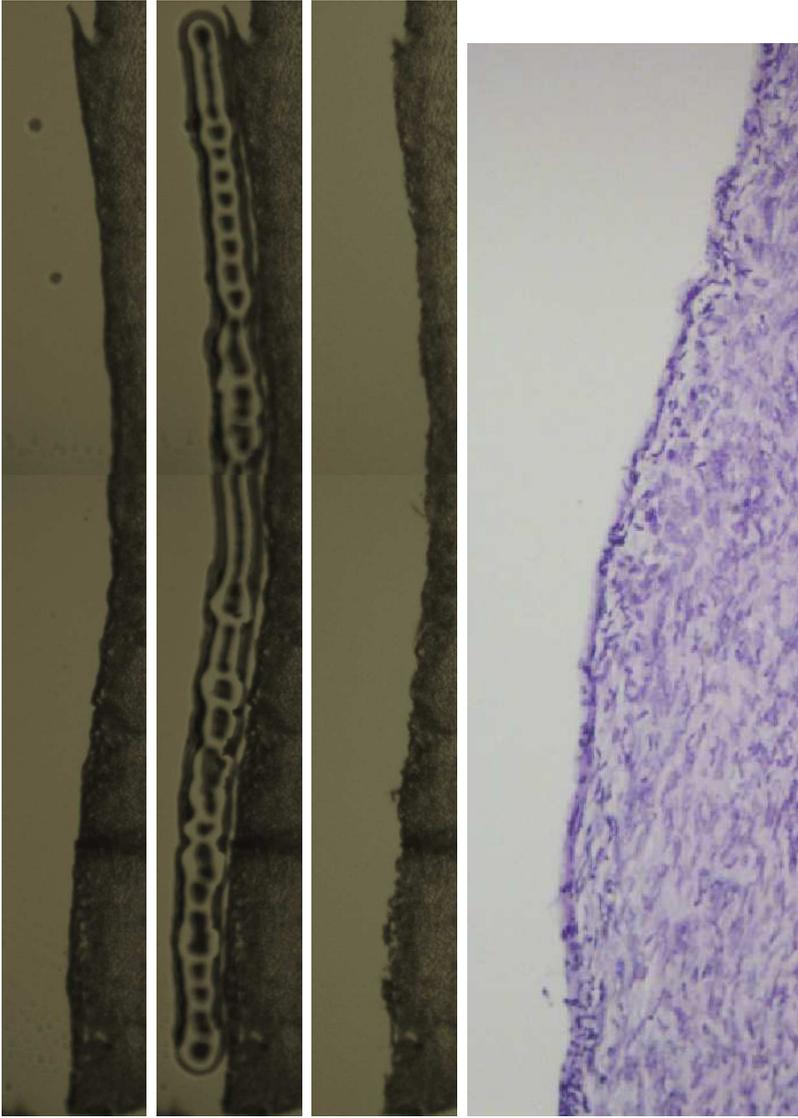


Figure A.1: Image progression, capturing normal epithelial tissue sample M01 with H&E stain. Before capture, during capture, after capture, and H&E stain (from left to right).

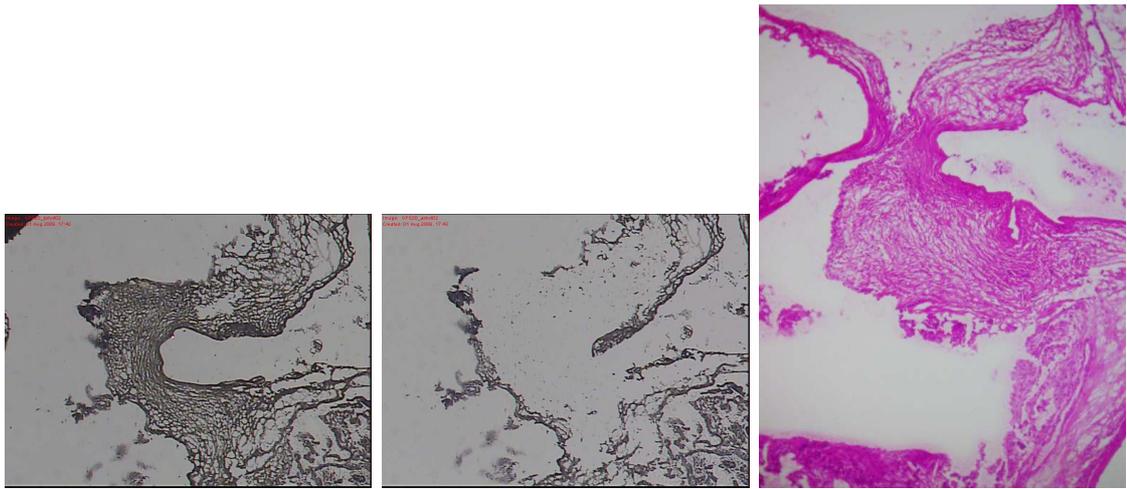


Figure A.2: Image progression, capturing stromal tissue sample M08 with H&E stain. (Left before, middle after, and right H&E staining).

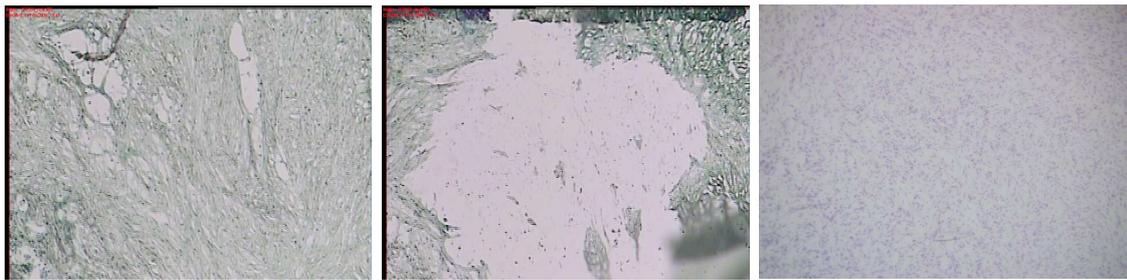


Figure A.3: Image progression, capturing stromal tissue sample B07 with H&E stain. (Left before, middle after, and right H&E staining).



Figure A.4: Image progression, capturing tumor (B16) and adjacent stromal tissue sample (B22). Images show before capturing (left), tumor captured on cap (middle), and tissue after adjacent stromal has additionally been sampled (right).

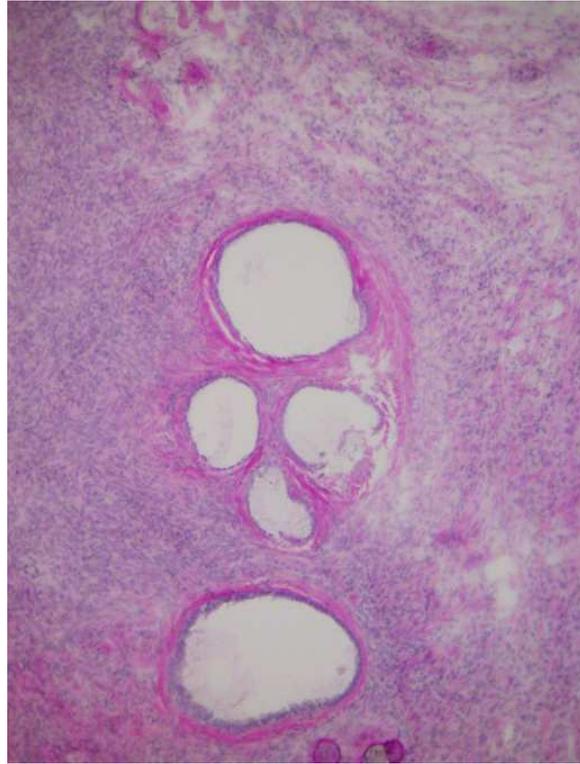


Figure A.5: Images of H&E staining for tissue samples B16 and B22.

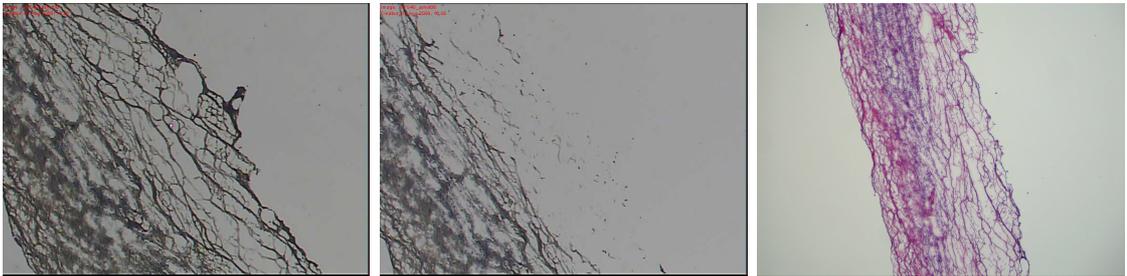


Figure A.6: Image progression, capturing stromal tissue sample M09 with H&E stain. (Left before, middle after, and right H&E staining).

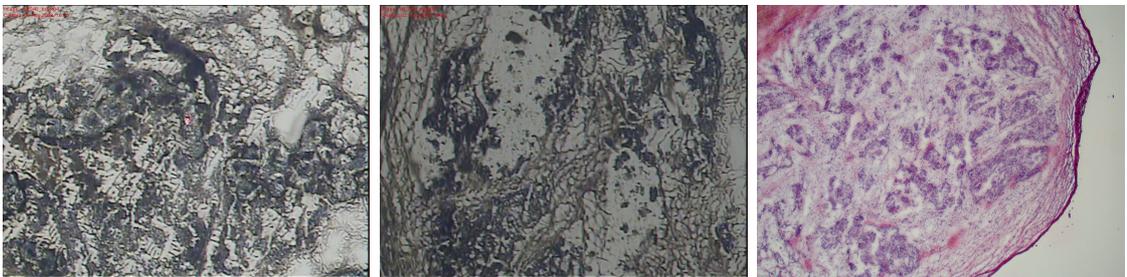


Figure A.7: Image progression, capturing tumor tissue sample M16 with H&E stain. (Left before, middle after, and right H&E staining).

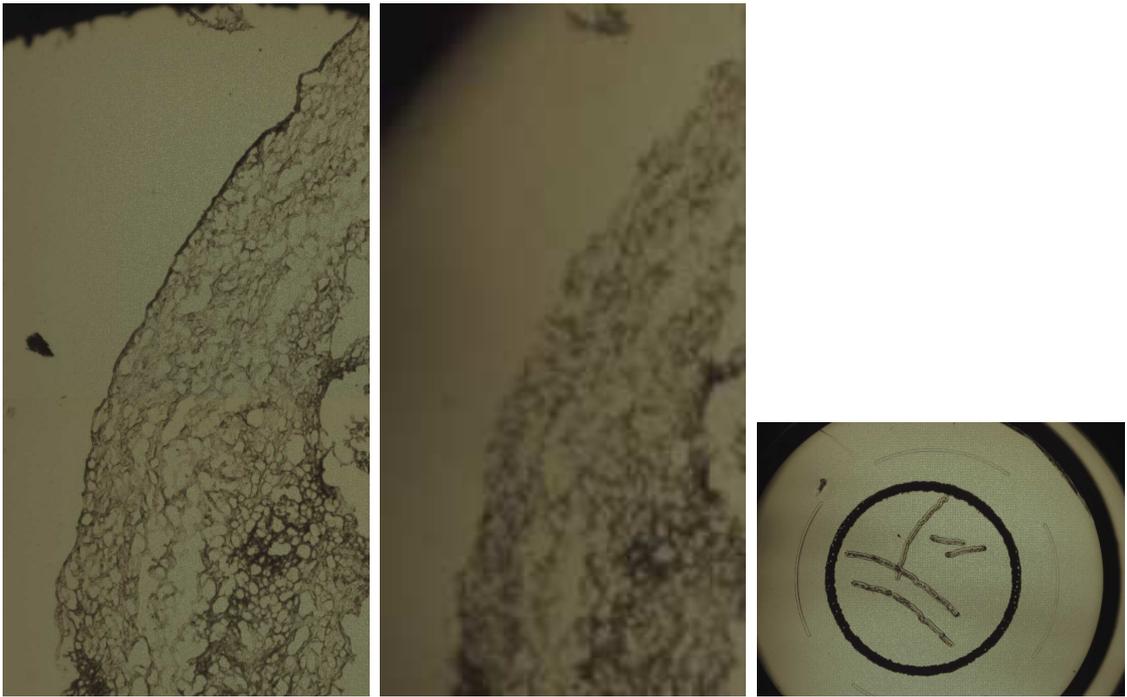


Figure A.8: Image progression, capturing normal epithelial tissue sample M03 with H&E stain. The image on the right shows the captured normal epithelium on a cap devoid of stroma

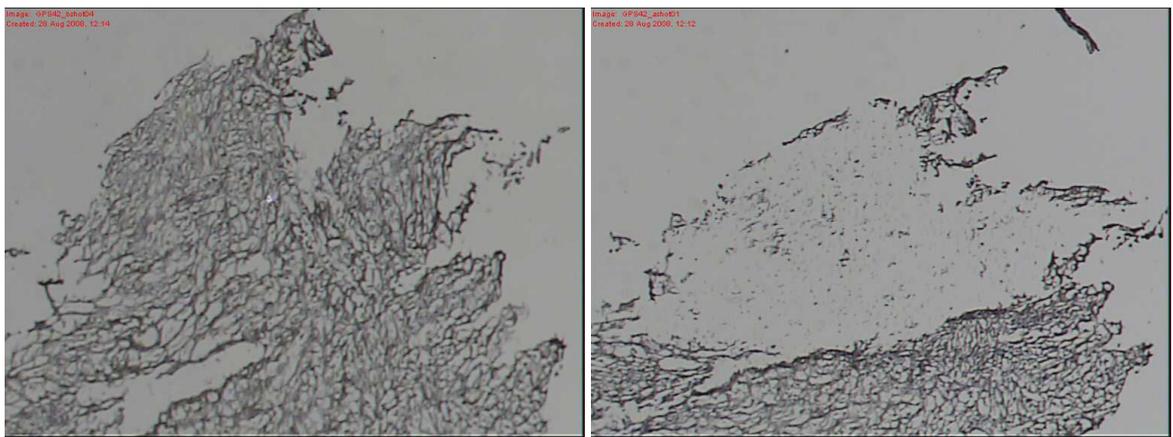


Figure A.9: Image progression, capturing stromal tissue sample M10. (Left before and right after).

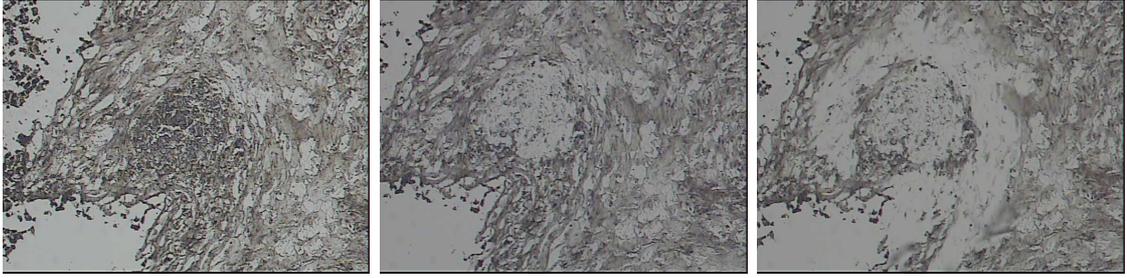


Figure A.10: Image progression, capturing tumor (M17) and adjacent stromal tissue sample (M25). Images show before capturing (left), after tumor capture (middle), and tissue after adjacent stromal has been captured (right).

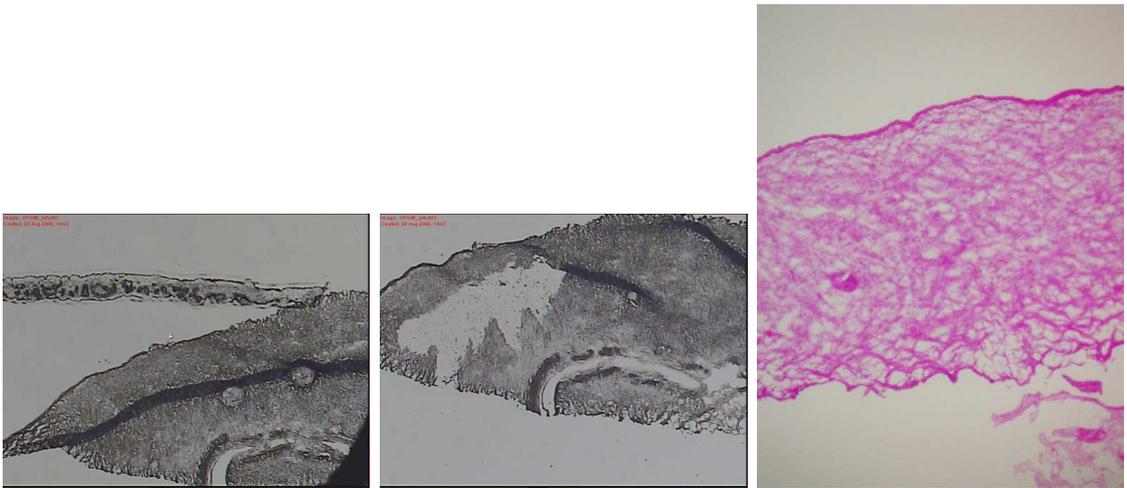


Figure A.11: Image progression, capturing stromal tissue sample M11 with H&E stain. (Left before, middle after, and right H&E staining).

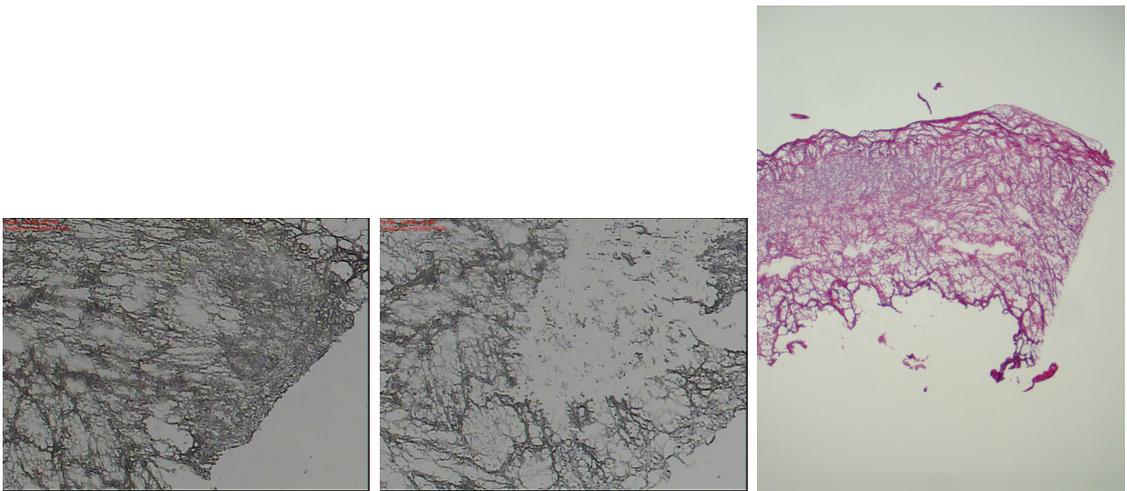


Figure A.12: Image progression, capturing tumor tissue sample M18 with H&E stain. (Left before, middle after, and right H&E staining).



Figure A.13: Image progression, capturing adjacent stromal tissue sample M26 with H&E stain. (Left before, middle after, and right H&E staining).

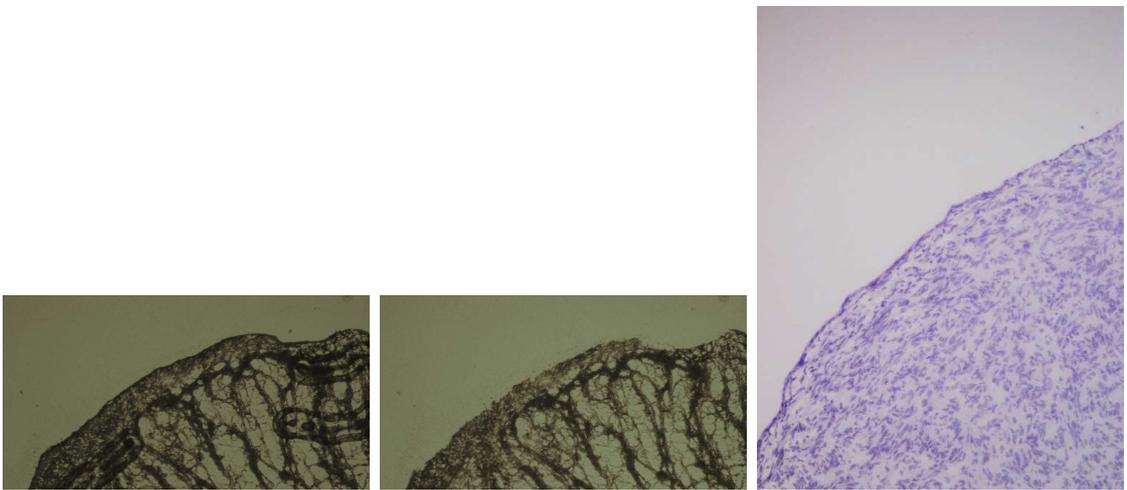


Figure A.14: Image progression, capturing normal epithelial tissue sample B03 with H&E stain. (Left before, middle after, and right H&E staining).



Figure A.15: Image progression, capturing stromal tissue sample M12 with H&E stain. (Left before, middle after, and right H&E staining).



Figure A.16: Image progression, capturing stromal tissue sample B09 with H&E stain. (Left before, middle after, and right H&E staining).

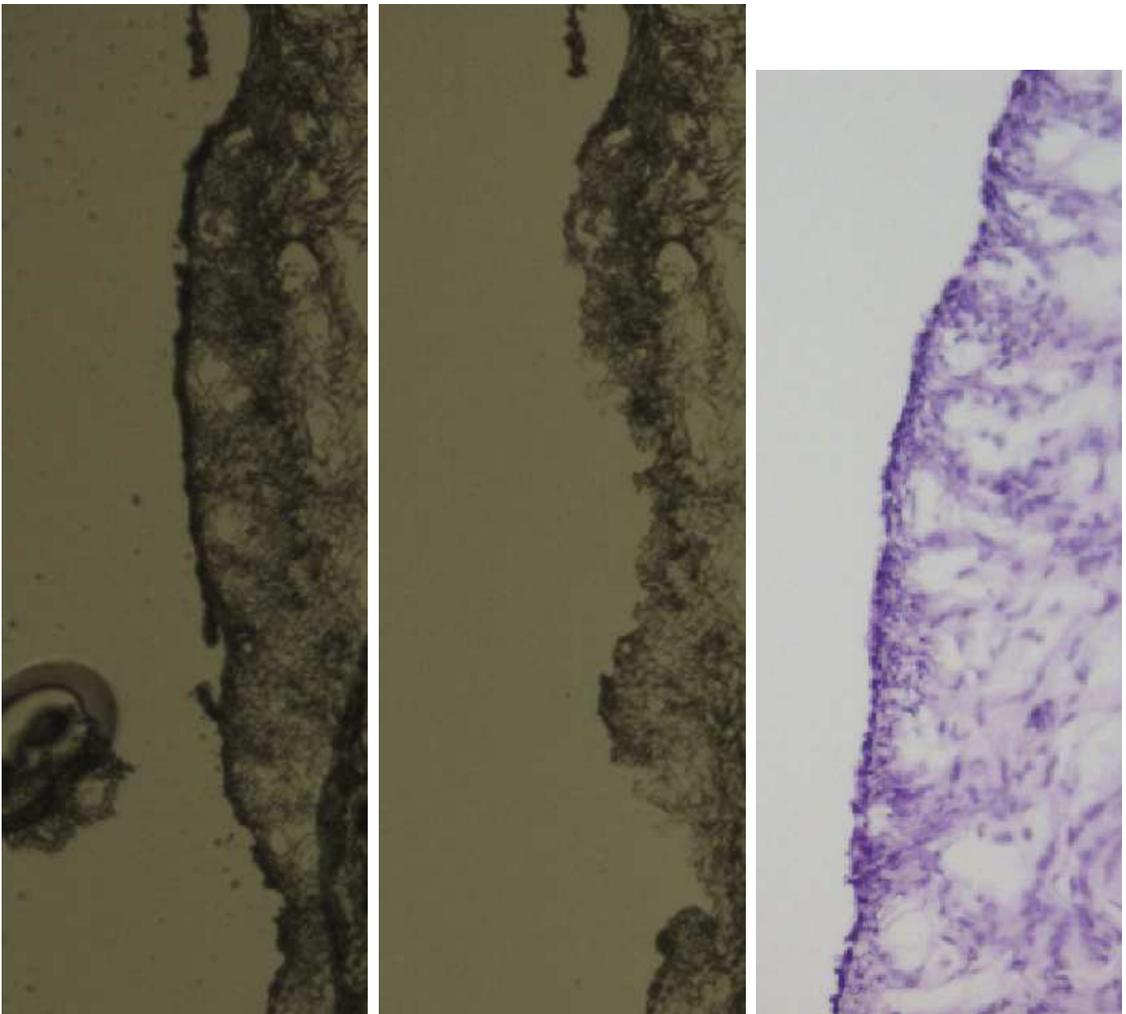


Figure A.17: Image progression, capturing normal epithelial tissue sample B04 with H&E stain. (Left before, middle after, and right H&E staining).

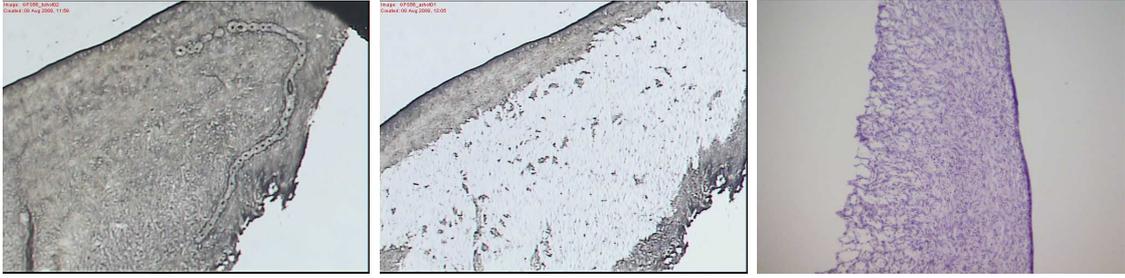


Figure A.18: Image progression, capturing stromal tissue sample B10 with H&E stain. (Left before, middle after, and right H&E staining).

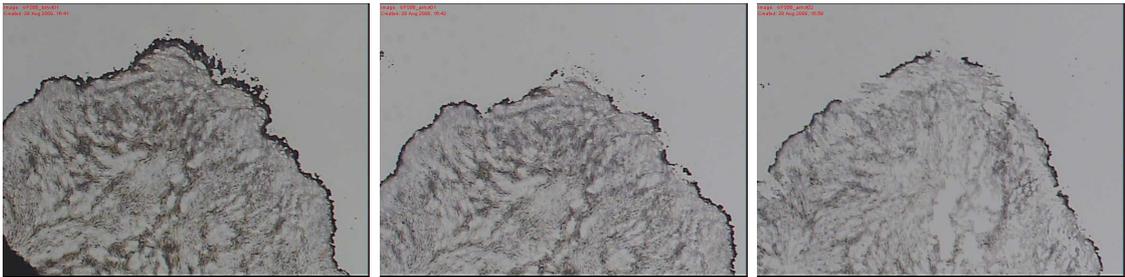


Figure A.19: Image progression, capturing tumor (B17) and adjacent stromal tissue sample (B23). Images show before capturing (left), after tumor capture (middle), and tissue after adjacent stromal has been captured (right).



Figure A.20: Image progression, capturing normal epithelial tissue sample M05 with H&E stain. (Left before, middle after, and right H&E staining).

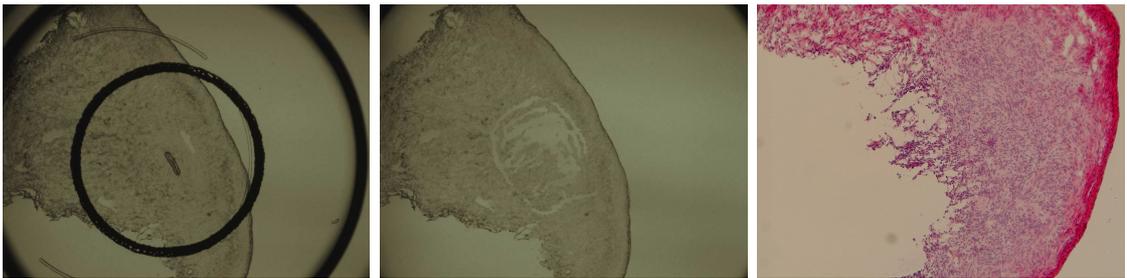


Figure A.21: Image progression, capturing stromal tissue sample B11 with H&E stain. (Left before, middle after, and right H&E staining).



Figure A.22: Image progression, capturing normal epithelial tissue sample B05. Before capturing is shown in the left image, then periodic capturing (middle image), and after periodic capturing to show specificity of capturing (right image)

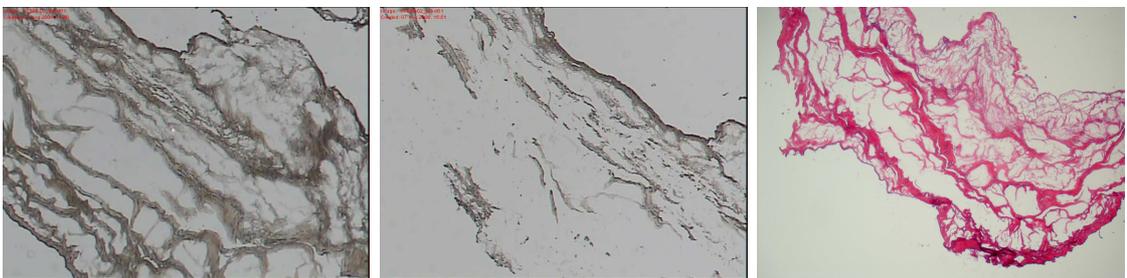


Figure A.23: Image progression, capturing stromal tissue sample B12 with H&E stain. (Left before, middle after, and right H&E staining).

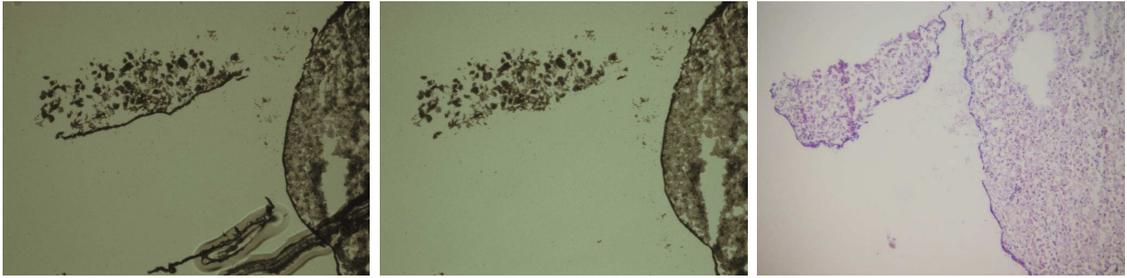


Figure A.24: Image progression, capturing normal epithelial tissue sample M06 with H&E stain. (Left before, middle after, and right H&E staining).



Figure A.25: Image progression, capturing stromal tissue sample M13 with H&E stain. (Left before, middle after, and right H&E staining).

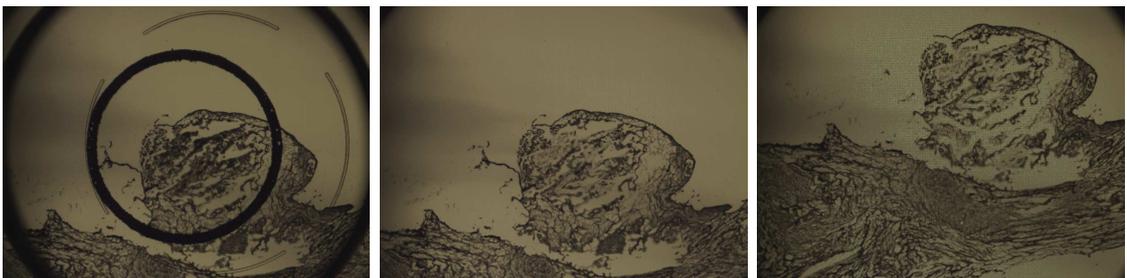


Figure A.26: Image progression, capturing tumor (M20) and adjacent stromal tissue sample (M28). Images show before capturing (left), after tumor capture (middle), and tissue after adjacent stromal has been captured (right).

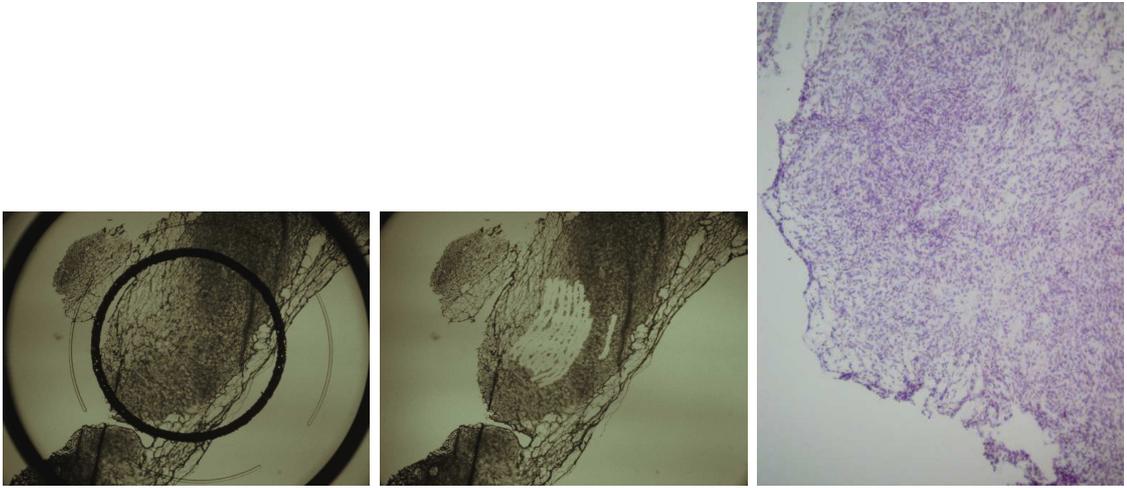


Figure A.27: Image progression, capturing stromal tissue sample B13 with H&E stain. (Left before, middle after, and right H&E staining).

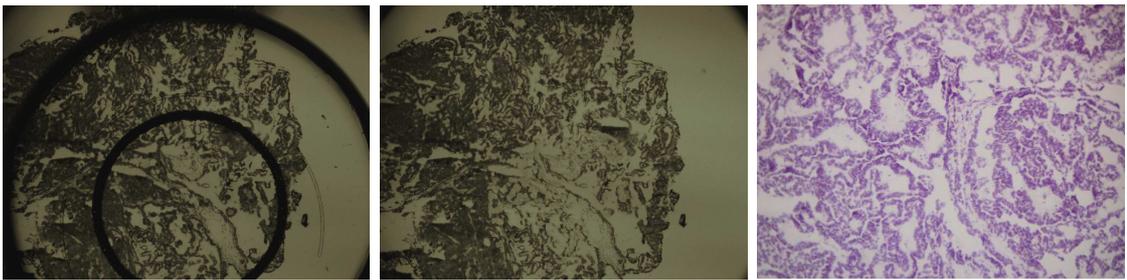


Figure A.28: Image progression, capturing tumor tissue sample M21 with H&E stain. (Left before, middle after, and right H&E staining).

APPENDIX B: TRANSCRIPT / TRANSCRIPT CLUSTER LISTS

Table B.1: Complete transcript list representing significant expression in benign tumors (1 of 4).

All Transcripts	Increasing Only	Decreasing Only
AGBL2	AGBL2	AMHR2
AHCYL1	AHCYL1	ANG
AIFM1	AIFM1	BNC1
AMHR2	ANKRD18A	C18orf34
ANG	ARHGAP26	CLEC4M
ANKRD18A	ARMC3	EHD2
ARHGAP26	ASRGL1	GNG11
ARMC3	C10orf79	GPRASP1
ASRGL1	C3orf15	HTRA1
BNC1	C6orf108	ITLN1
C10orf79	C9orf68	ITLN2
C18orf34	CAPSL	MGP
C3orf15	CASC1	MTMR11
C6orf108	CCDC11	PRG4
C9orf68	CCDC30	SHF
CAPSL	CD24	SLC24A6
CASC1	CDH1	SPOCK1
CCDC11	CDH6	TBX3
CCDC30	CLDN4	YPEL4
CD24	CLDN7	ZEB1
CDH1	CP	ZNF385B
CDH6	CXorf22	—
CLDN4	DCDC2	—
CLDN7	DNAH11	—
CLEC4M	DNAH7	—
CP	DSP	—
CXorf22	EFHC2	—
DCDC2	EHF	—

Table B.2: (continued).

All Transcripts	Increasing Only	Decreasing Only
DNAH11	EPCAM	—
DNAH7	ERBB4	—
DSP	ERP27	—
EFHC2	ESRP1	—
EHD2	FAAH2	—
EHF	FAIM	—
EPCAM	FOLR1	—
ERBB4	GALNT4	—
ERP27	GPR39	—
ESRP1	HNRNPF	—
FAAH2	HYDIN	—
FAIM	INADL	—
FOLR1	KIAA1712	—
GALNT4	LAMC2	—
GNG11	LOC100132319	—
GPR39	LPAR3	—
GPRASP1	LRRC50	—
HNRNPF	LYPD1	—
HTRA1	MAL2	—
HYDIN	MDH1B	—
INADL	MECOM	—
ITLN1	MNS1	—
ITLN2	MPP7	—
KIAA1712	MPZL2	—
LAMC2	MYB	—
LOC100132319	MYEF2	—
LPAR3	MYO5C	—
LRRC50	MYOF	—
LYPD1	OSBPL3	—
MAL2	OXR1	—
MDH1B	PARD6B	—
MECOM	PAX8	—
MGP	PERP	—
MNS1	PGM2L1	—

Table B.3: (continued).

All Transcripts	Increasing Only	Decreasing Only
MPP7	PIH1D2	—
MPZL2	PRKX	—
MTMR11	PRUNE2	—
MYB	PRUNE2	—
MYEF2	PTPN3	—
MYO5C	RBBP8	—
MYOF	RNASET2	—
OSBPL3	SCGB2A1	—
OXR1	SCNN1A	—
PARD6B	SKAP1	—
PAX8	SLC1A1	—
PERP	SLC34A2	—
PGM2L1	SORL1	—
PIH1D2	SORT1	—
PRG4	SOX17	—
PRKX	SPAG17	—
PRUNE2	SPEF2	—
PRUNE2	SRPK1	—
PTPN3	ST6GALNAC2	—
RBBP8	STX18	—
RNASET2	TACSTD2	—
SCGB2A1	TMC4	—
SCNN1A	TOM1L1	—
SHF	TPRG1	—
SKAP1	TSGA10	—
SLC1A1	VWA3A	—
SLC24A6	WDR49	—
SLC34A2	WDR52	—
SORL1	WDR77	—
SORT1	WDR78	—
SOX17	WFDC2	—
SPAG17	ZBBX	—
SPEF2	ZDHHC13	—
SPOCK1	—	—
SRPK1	—	—

Table B.4: (continued).

All Transcripts	Increasing Only	Decreasing Only
ST6GALNAC2	—	—
STX18	—	—
TACSTD2	—	—
TBX3	—	—
TMC4	—	—
TOM1L1	—	—
TPRG1	—	—
TSGA10	—	—
VWA3A	—	—
WDR49	—	—
WDR52	—	—
WDR77	—	—
WDR78	—	—
WFDC2	—	—
YPEL4	—	—
ZBBX	—	—
ZDHHC13	—	—
ZEB1	—	—
ZNF385B	—	—

Table B.5: Complete transcript-cluster list representing significant expression in benign tumors (1 of 3).

All Clusters	Increasing Only	Decreasing Only
AHCYL1	AHCYL1	AMHR2
AIFM1	AIFM1	BNC1
AMHR2	ARHGAP26	C4orf49
ARHGAP26	ARMC3	CLEC4M
ARMC3	ASRGL1	EHD2
ASRGL1	C10orf79	GNG11
BNC1	C3orf15	HTRA1
C10orf79	C9orf68	ITLN1
C3orf15	CAPSL	PRG4
C4orf49	CASC1	SALL1
C9orf68	CCDC11	SLC24A6
CAPSL	CCDC30	TBX3
CASC1	CD24	YPEL4
CCDC11	CDH1	ZEB1
CCDC30	CDH6	—
CD24	CLDN4	—
CDH1	CLGN	—
CDH6	CMTM7	—
CLDN4	CP	—
CLEC4M	DCDC2	—
CLGN	DNAH7	—
CMTM7	DSP	—
CP	EFHC2	—
DCDC2	EHF	—
DNAH7	EPCAM	—
DSP	ERBB4	—
EFHC2	ERP27	—
EHD2	ESRP1	—
EHF	FAAH2	—
EPCAM	FAIM	—
ERBB4	FOLR1	—
ERP27	GALNT4	—
ESRP1	GPR39	—
FAAH2	HNRNPF	—
FAIM	HYDIN	—
FOLR1	KIAA1712	—
GALNT4	LOC100132319	—
GNG11	LRRC50	—
GPR39	LYPD1	—

Table B.6: (continued).

All Clusters	Increasing Only	Decreasing Only
HNRNPF	MAL2	—
HTRA1	MDH1B	—
HYDIN	MECOM	—
ITLN1	MNS1	—
KIAA1712	MPP7	—
LOC100132319	MYB	—
LRRC50	MYO5C	—
LYPD1	MYOF	—
MAL2	NUP62CL	—
MDH1B	OSBPL3	—
MECOM	PAX8	—
MNS1	PERP	—
MPP7	PIH1D2	—
MYB	PRKX	—
MYO5C	PRUNE2	—
MYOF	PRUNE2	—
NUP62CL	RBBP8	—
OSBPL3	SAMD12	—
PAX8	SCGB1D2	—
PERP	SCGB2A1	—
PIH1D2	SCNN1A	—
PRG4	SKAP1	—
PRKX	SLC1A1	—
PRUNE2	SLC34A2	—
PRUNE2	SORL1	—
RBBP8	SORT1	—
SALL1	SPAG17	—
SAMD12	SPEF2	—
SCGB1D2	ST6GALNAC2	—
SCGB2A1	STX18	—
SCNN1A	STX19	—
SKAP1	TBC1D8	—
SLC1A1	TOM1L1	—
SLC24A6	TSGA10	—
SLC34A2	WDR49	—
SORL1	WDR52	—
SORT1	WDR77	—
SPAG17	WDR78	—
SPEF2	WFDC2	—

Table B.7: (continued).

All Clusters	Increasing Only	Decreasing Only
ST6GALNAC2	—	—
STX18	—	—
STX19	—	—
TBC1D8	—	—
TBX3	—	—
TOM1L1	—	—
TSGA10	—	—
WDR49	—	—
WDR52	—	—
WDR77	—	—
WDR78	—	—
WFDC2	—	—
YPEL4	—	—
ZEB1	—	—

Table B.8: Complete transcript list representing significant expression in malignant tumors (1 of 3).

All Transcripts	Increasing Only	Decreasing Only
AFTPH	AFTPH	AKT3
AKT3	ASRGL1	C16orf62
ASRGL1	AZIN1	CALB2
AZIN1	BAIAP2L1	COL6A1
BAIAP2L1	BCL11A	COL6A2
BCL11A	BUB1B	DAB2
BUB1B	CBX3	DCN
C16orf62	CD24	EFEMP1
CALB2	CDC7	EHD2
CBX3	CDH1	FBXL7
CD24	CDH6	GABBR1
CDC7	CENPF	GNG11
CDH1	CHMP4C	IGFBP6
CDH6	CLDN3	LRRN4
CENPF	CLDN4	MAPRE2
CHMP4C	CLDN7	MGP
CLDN3	CP	NAP
CLDN4	CPSF3	NBR1
CLDN7	CXXC5	PTGIS
COL6A1	DSP	SP140L
COL6A2	DTL	ST3GAL5
CP	EHF	TRPC1
CPSF3	EPCAM	—
CXXC5	ERBB3	—
DAB2	ESRP1	—
DCN	EYA2	—
DSP	EZH2	—
DTL	FAM49B	—
EFEMP1	FOLR1	—
EHD2	GALNT3	—
EHF	GPR39	—
EPCAM	GRHL2	—
ERBB3	HEY2	—
ESRP1	HJURP	—
EYA2	HMMR	—
EZH2	HNRNPF	—
FAM49B	HOMER2	—
FBXL7	HOXB8	—
FOLR1	INADL	—

Table B.9: (continued).

All Transcripts	Increasing Only	Decreasing Only
GABBR1	KLF5	—
GALNT3	KLK7	—
GNG11	LPAR3	—
GPR39	LRBA	—
GRHL2	LYPD1	—
HEY2	MAL2	—
HJURP	MAP7	—
HMMR	MECOM	—
HNRNPF	MECOM	—
HOMER2	MELK	—
HOXB8	MND1	—
IGFBP6	MPP7	—
INADL	MTHFD2	—
KLF5	MTIF2	—
KLK7	MUC1	—
LPAR3	MUC20	—
LRBA	MYB	—
LRRN4	MYEF2	—
LYPD1	NEK2	—
MAL2	PGM2L1	—
MAP7	PRKDC	—
MAPRE2	PRKX	—
MECOM	PSAT1	—
MECOM	PTPN12	—
MELK	PTTG1	—
MGP	PUS7	—
MND1	RAD51AP1	—
MPP7	RAD54B	—
MTHFD2	RRAGD	—
MTIF2	S100A14	—
MUC1	SCNN1A	—
MUC20	SNTB1	—
MYB	SORT1	—
MYEF2	SOX17	—
NAPA	SPINT2	—
NBR1	SRPK1	—
NEK2	ST6GALNAC2	—
PGM2L1	STYXL1	—

Table B.10: (continued).

All Transcripts	Increasing Only	Decreasing Only
PRKDC	TEC	—
PRKX	TNPO3	—
PSAT1	UBE2T	—
PTGIS	VTCN1	—
PTPN12	WFDC2	—
PTTG1	XPO5	—
PUS7	XPR1	—
RAD51AP1	ZDHHC13	—
RAD54B	—	—
RRAGD	—	—
S100A14	—	—
SCNN1A	—	—
SNTB1	—	—
SORT1	—	—
SOX17	—	—
SP140L	—	—
SPINT2	—	—
SRPK1	—	—
ST3GAL5	—	—
ST6GALNAC2	—	—
STYXL1	—	—
TEC	—	—
TNPO3	—	—
TRPC1	—	—
UBE2T	—	—
VTCN1	—	—
WFDC2	—	—
XPO5	—	—
XPR1	—	—
ZDHHC13	—	—

Table B.11: Complete transcript-cluster list representing significant expression in malignant tumors (1 of 3).

All Clusters	Increasing Only	Decreasing Only
AFTPH	AFTPH	AKT3
AKT3	ASRGL1	CALB2
ASRGL1	AZIN1	COL6A2
AZIN1	BAIAP2L1	EFEMP1
BAIAP2L1	BCL11A	EHD2
BCL11A	BUB1B	FBXL7
BUB1B	CBX3	GNG11
CALB2	CD24	IGFBP6
CBX3	CDC7	LRRN4
CD24	CDH6	NAPA
CDC7	CENPF	NBR1
CDH6	CHMP4C	SP140L
CENPF	CLDN3	ST3GAL5
CHMP4C	CLDN4	TRPC1
CLDN3	CLDN7	—
CLDN4	CP	—
CLDN7	CPSF3	—
COL6A2	CXXC5	—
CP	DSP	—
CPSF3	DTL	—
CXXC5	EHF	—
DSP	EPCAM	—
DTL	ERBB3	—
EFEMP1	ESRP1	—
EHD2	EZH2	—
EHF	FOLR1	—
EPCAM	GALNT3	—
ERBB3	GPR39	—
ESRP1	GRHL2	—
EZH2	HJURP	—
FBXL7	HOMER2	—
FOLR1	HOXB8	—
GALNT3	INADL	—
GNG11	KCTD1	—
GPR39	KLF5	—
GRHL2	KLK7	—

Table B.12: (continued).

All Clusters	Increasing Only	Decreasing Only
HJURP	LPAR3	—
HOMER2	LRBA	—
HOXB8	LYPD1	—
IGFBP6	MAL2	—
INADL	MAP7	—
KCTD1	MECOM	—
KLF5	MECOM	—
KLK7	MELK	—
LPAR3	MND1	—
LRBA	MPP7	—
LRRN4	MTHFD2	—
LYPD1	MTIF2	—
MAL2	MUC1	—
MAP7	MYB	—
MECOM	MYEF2	—
MECOM	NEK2	—
MELK	PGM2L1	—
MND1	PRKDC	—
MPP7	PRKX	—
MTHFD2	PSAT1	—
MTIF2	PTPN12	—
MUC1	PTTG1	—
MYB	PUS7	—
MYEF2	RAD51AP1	—
NAPA	RAD54B	—
NBR1	SCNN1A	—
NEK2	SNTB1	—
PGM2L1	SORT1	—
PRKDC	SPINT2	—
PRKX	SRPK1	—
PSAT1	ST6GALNAC2	—
PTPN12	STYXL1	—
PTTG1	UBE2T	—
PUS7	VTCN1	—
RAD51AP1	WFDC2	—
RAD54B	XPR1	—
SCNN1A	ZDHHC13	—
SNTB1	—	—

Table B.13: (continued).

All Clusters	Increasing Only	Decreasing Only
SORT1	—	—
SP140L	—	—
SPINT2	—	—
SRPK1	—	—
ST3GAL5	—	—
ST6GALNAC2	—	—
STYXL1	—	—
TRPC1	—	—
UBE2T	—	—
VTCN1	—	—
WFDC2	—	—
XPR1	—	—
ZDHHC13	—	—

Table B.14: RefSeq numbers and fold-changes for all gene symbols associated with the significant benign tumor transcript list (1 of 4).

Gene Symbol	RefSeq Number	Fold-Change
AGBL2	NM_024783	2.8
AHCYL1	NM_006621	3.34
AIFM1	NM_001130847	5.24
AMHR2	NM_020547	-2.53
ANG	NM_001145	-3.94
ANKRD18A	NM_147195	2.07
ARHGAP26	NM_015071	7.36
ARMC3	NM_173081	3.95
ASRGL1	NM_001083926	9.91
BNC1	NM_001717	-2.86
C10orf79	NM_025145	8.76
C18orf34	NM_001105528	-2.08
C3orf15	NM_033364	3.84
C6orf108	NM_199184	2.88
C9orf68	NM_001039395	3.94
CAPSL	NM_144647	3.53
CASC1	NM_001082972	4.99
CCDC11	NM_145020	9.63
CCDC30	NM_001080850	4.7
CD24	NM_013230	15.31
CDH1	NM_004360	4.51
CDH6	NM_004932	6.68
CLDN4	NM_001305	4.22
CLDN7	NM_001307	3.67
CLEC4M	NR_026707	-6.35
CP	NM_000096	37.89
CXorf22	NM_152632	3.2
DCDC2	NM_016356	8.2
DNAH11	NM_003777	3.46
DNAH7	NM_018897	3.17
DSP	NM_004415	5.55
EFHC2	NM_025184	2.31
EHD2	NM_014601	-2.36
EHF	NM_012153	7.26
EPCAM	NM_002354	8.4
ERBB4	NM_005235	3.99

Table B.15: (continued).

Gene Symbol	RefSeq Number	Fold-Change
ERP27	NM_152321	10.41
ESRP1	NM_017697	11.08
FAAH2	NM_174912	3.23
FAIM	NM_001033030	3.23
FOLR1	NM_016724	4.07
GALNT4	NM_003774	3.56
GNG11	NM_004126	-4.97
GPR39	NM_001508	30.92
GPRASP1	NM_014710	-5.87
HNRNPF	NM_004966	3.13
HTRA1	NM_002775	-3.02
HYDIN	NM_032821	3.54
INADL	NM_176877	3.03
ITLN1	NM_017625	-31.69
ITLN2	NM_080878	-2.21
KIAA1712	NM_001040157	2.21
LAMC2	NM_005562	2.71
LOC100132319	AF315716	5.38
LPAR3	NM_012152	5.92
LRRC50	NM_178452	2.83
LYPD1	NM_144586	26.23
MAL2	NM_052886	23.28
MDH1B	NM_001039845	2.69
MECOM	NM_001105077	9.35
MGP	NM_000900	-9.28
MNS1	NM_018365	6.2
MPP7	NM_173496	3.23
MPZL2	NM_144765	5.67
MTMR11	NM_001145862	-2.02
MYB	NM_001130173	3.05
MYEF2	NM_016132	4.67
MYO5C	NM_018728	2.75
MYOF	NM_013451	7.84
OSBPL3	NM_015550	3.51
OXR1	NM_018002	3.15
PARD6B	NM_032521	2.98

Table B.16: (continued).

Gene Symbol	RefSeq Number	Fold-Change
PAX8	NM_003466	4.2
PERP	NM_022121	11.04
PGM2L1	NM_173582	2.58
PIH1D2	NM_138789	4.25
PRG4	NM_005807	-10.97
PRKX	NM_005044	3.56
PRUNE2	NM_015225	3.21
PRUNE2	NM_015225	6.5
PTPN3	NM_002829	2.04
RBBP8	NM_002894	5.75
RNASET2	NM_003730	9.31
SCGB2A1	NM_002407	20.67
SCNN1A	NM_001038	5.75
SHF	NM_138356	-2.12
SKAP1	NM_003726	2.59
SLC1A1	NM_004170	2.6
SLC24A6	NM_024959	-2.05
SLC34A2	NM_001177999	19.13
SORL1	NM_003105	4.34
SORT1	NM_002959	3.85
SOX17	NM_022454	3.43
SPAG17	NM_206996	2.77
SPEF2	NM_024867	2.68
SPOCK1	NM_004598	-2.11
SRPK1	NM_003137	2.49
ST6GALNAC2	NM_006456	6.38
STX18	NM_016930	3.95
TACSTD2	NM_002353	3.76
TBX3	NM_016569	-2.19
TMC4	NM_001145303	2.5
TOM1L1	NM_005486	4.68
TPRG1	NM_198485	2.6
TSGA10	NM_182911	4.29
VWA3A	NM_173615	2.02
WDR49	NM_178824	4.36
WDR52	NM_001164496	3.21

Table B.17: (continued).

Gene Symbol	RefSeq Number	Fold-Change
WDR77	NM_024102	4.95
WDR78	NM_024763	3.42
WFDC2	NM_006103	11.45
YPEL4	NM_145008	-2.37
ZBBX	NM_024687	5.76
ZDHHC13	NM_019028	2.55
ZEB1	NM_030751	-7.04
ZNF385B	NM_152520	-2.79

Table B.18: RefSeq numbers and fold-changes for all gene symbols associated with the significant malignant tumor transcript list (1 of 3).

Gene Symbol	RefSeq Number	Fold-Change
AFTPH	NM_203437	2.28
AKT3	NM_181690	-7.3
ASRGL1	NM_001083926	6.45
AZIN1	NM_015878	3.09
BAIAP2L1	NM_018842	2.6
BCL11A	NM_022893	3.03
BUB1B	NM_001211	3.02
C16orf62	BC050464	-2.61
CALB2	NM_001740	-12.23
CBX3	NM_016587	4.74
CD24	NM_013230	7.81
CDC7	NM_001134420	2.63
CDH1	NM_004360	2.34
CDH6	NM_004932	4.5
CENPF	NM_016343	4.91
CHMP4C	NM_152284	4.51
CLDN3	NM_001306	2.27
CLDN4	NM_001305	5.49
CLDN7	NM_001307	3.82
COL6A1	NM_001848	-2.44
COL6A2	NM_001849	-2.38
CP	NM_000096	14.24
CPSF3	NM_016207	2.76
CXXC5	NM_016463	2.82
DAB2	NM_001343	-3.63
DCN	NM_001920	-15.14
DSP	NM_004415	4.37
DTL	NM_016448	3
EFEMP1	NM_004105	-16.58
EHD2	NM_014601	-2.29
EHF	NM_012153	3.8
EPCAM	NM_002354	5.67
ERBB3	NM_001982	2.63
ESRP1	NM_017697	15.83
EYA2	NM_005244	2.37
EZH2	NM_004456	2.92

Table B.19: (continued).

Gene Symbol	RefSeq Number	Fold-Change
FAM49B	BC017297	3.07
FBXL7	NM_012304	-2.01
FOLR1	NM_016724	3.06
GABBR1	NM_001470	-2.09
GALNT3	NM_004482	2.69
GNG11	NM_004126	-6.22
GPR39	NM_001508	7.09
GRHL2	NM_024915	3.16
HEY2	NM_012259	3.36
HJURP	NM_018410	2.4
HMMR	NM_001142556	2.24
HNRNPF	NM_004966	2.44
HOMER2	NM_199330	4.24
HOXB8	NM_024016	2.78
IGFBP6	NM_002178	-9.37
INADL	NM_176877	2.94
KLF5	NM_001730	2.29
KLK7	NM_139277	3.9
LPAR3	NM_012152	12.98
LRBA	NM_006726	2.9
LRRN4	NM_152611	-2.06
LYPD1	NM_144586	6.69
MAL2	NM_052886	8.73
MAP7	NM_003980	2.09
MAPRE2	NM_014268	-2.49
MECOM	NM_004991	2.91
MECOM	NM_001105077	9.92
MELK	NM_014791	2.91
MGP	NM_000900	-7.06
MND1	NM_032117	3.93
MPP7	NM_173496	2.94
MTHFD2	NR_027405	4.13
MTIF2	NM_002453	2.37
MUC1	NM_002456	4.94
MUC20	NM_001098516	2.47

Table B.20: (continued).

Gene Symbol	RefSeq Number	Fold-Change
MYB	NM_001130173	2.71
MYEF2	NM_016132	3.9
NAPA	NM_003827	-2.21
NBR1	NM_031858	-3.01
NEK2	NM_002497	3.8
PGM2L1	NM_173582	2.53
PRKDC	NM_006904	2.57
PRKX	NM_005044	3.06
PSAT1	NM_058179	2.62
PTGIS	NM_000961	-5.53
PTPN12	NM_002835	2.66
PTTG1	NM_004219	21.82
PUS7	NM_019042	2.02
RAD51AP1	NM_001130862	4.94
RAD54B	NM_012415	2.36
RRAGD	NM_021244	2.11
S100A14	NM_020672	2.34
SCNN1A	NM_001038	5.37
SNTB1	NM_021021	2.28
SORT1	NM_002959	2.97
SOX17	NM_022454	2.18
SP140L	NM_138402	-2.56
SPINT2	NM_021102	4.44
SRPK1	NM_003137	2.8
ST3GAL5	NM_003896	-2.62
ST6GALNAC2	NM_006456	3.29
STYXL1	NM_016086	2.43
TEC	NM_003215	4.39
TNPO3	NM_012470	2.03
TRPC1	NM_003304	-3.46
UBE2T	NM_014176	3.01
VTCN1	NM_024626	4.27
WFDC2	NM_006103	5.87
XPO5	NM_020750	2.17
XPR1	NM_004736	4.27
ZDHHC13	NM_019028	2.27

Table B.21: Ref seq numbers and fold-changes for all gene symbols associated with the significant benign tumor-stromal environment transcript list.

Gene Symbol	RefSeq Number	Fold-Change
NRCAM	NM_001037132	6.77

APPENDIX C: BIOLOGICAL CONTEXT RESULTS DETAIL

Table C.1: Complete results from DAVID for the benign transcript lists (1 of 4).

Annotation	Exon All	Exon Up	Exon Down
SP COMMENT TYPE	Tissue Specificity	—	—
SP PIR KEYWORDS	Polymorphism	Polymorphism	—
SP PIR KEYWORDS	Tight Junction	Tight Junction	—
SP PIR KEYWORDS	—	Cell Junction	—
GOTERM CC 2	Axoneme	Axoneme	—
GOTERM CC 2	Cell Projection Part	Cell Projection Part	—
GOTERM CC 2	Apical Part of cell	Apical Part of cell	—
GOTERM CC 3	Plasma Membrane Part	Plasma Membrane Part	—
GOTERM CC 3	Plasma Membrane	Axoneme	—
GOTERM CC 3	Axoneme	Apical Part of cell	—
GOTERM CC 3	Cell Projection Part	Plasma Membrane	—
GOTERM CC 3	Apical Part of cell	Cilium	—
GOTERM CC 3	—	Cell Projection Part	—
GOTERM CC 4	Apicolateral Plasma Membrane	Apicolateral Plasma Membrane	—
GOTERM CC 4	Plasma Membrane Part	Plasma Membrane Part	—
GOTERM CC 4	Plasma Membrane	Axoneme	—
GOTERM CC 4	Axoneme	—	—
GOTERM CC 4	Cell Projection Part	—	—

Table C.2: (continued.)

Annotation	Exon All	Exon Up	Exon Down
GOTERM CC 5	Apical Junction Complex	Apical Junction Complex	—
GOTERM CC 5	Plasma Membrane Part	Apicolateral Plasma Membrane	—
GOTERM CC 5	Apicolateral Plasma Membrane	Cell-Cell Junction	—
GOTERM CC 5	Cell-Cell Junction	Plasma Membrane Part	—
GOTERM CC 5	Axoneme	Axoneme	—
GOTERM CC 5	—	Cell Junction	—
GOTERM CC 5	—	Cytoskeleton	—
GOTERM CC ALL	—	Apical Junction Complex	—
GOTERM CC ALL	—	Apicolateral Plasma Membrane	—
GOTERM CC ALL	—	Cell-Cell Junction	—
GOTERM CC ALL	—	Plasma Membrane	—
GOTERM CC ALL	—	Occluding Junction	—
GOTERM CC ALL	—	Tight Junction	—
GOTERM CC ALL	—	Axoneme	—
GOTERM CC ALL	—	Apical Part of Cell	—
GOTERM CC ALL	—	Cell Junction	—
GOTERM CC ALL	—	Plasma Membrane	—
GOTERM CC ALL	—	Cilium	—
GOTERM CC ALL	—	Cytoskeleton	—
GOTERM CC FAT	—	Apical Junction Complex	—
GOTERM CC FAT	—	Apicolateral Plasma Membrane	—
GOTERM CC FAT	—	Cell-Cell Junction	—
GOTERM CC FAT	—	Plasma Membrane Part	—
GOTERM CC FAT	—	Tight Junction	—
GOTERM CC FAT	—	Occluding Junction	—

Table C.3: (continued.)

Annotation	Exon All	Exon Up	Exon Down
GOTERM CC FAT	—	Axoneme	—
GOTERM CC FAT	—	Apical Part of Cell	—
GOTERM CC FAT	—	Cell Junction	—
GOTERM CC FAT	—	Plasma Membrane	—
GOTERM CC FAT	—	Cilium	—
GOTERM CC FAT	—	Cytoskeleton	—
PANTHER MF ALL	— —	Cell junction protein	—
PIR SUMMARY	GA733	—	—
PIR SUMMARY	Tight Junction Specific Obliteration of the Intercellular Space	—	—
PUBMED ID	12477932	—	—
PUBMED ID	15489334	—	—
PUBMED ID	14702039	—	—
PUBMED ID	16344560	—	—
BIOCARTA	SUMOylation	—	—
UCSC TFBS	75 Entries	64 Entries	—
CGAP SAGE QUARTILE	Brain Ependymoma 3rd	Brain Ependymoma 3rd	—
CGAP SAGE QUARTILE	Brain Ependymoma 3rd	Brain Ependymoma 3rd	—
CGAP SAGE QUARTILE	Brain Ependymoma 3rd	—	—
CGAP SAGE QUARTILE	Thyroid 3rd Thyroid 3rd	—	—
GNF U133A QUARTILE	Trigeminal Ganglion 3rd	Trigeminal Ganglion 3rd	Adrenal Cortex 3rd

Table C.4: (continued.)

Annotation	Exon All	Exon Up	Exon Down
GNF U133A QUARTILE	Skin 3rd	Skin 3rd	—
GNF U133A QUARTILE	Pancreatic Islets 3rd	—	—
GNF U133A QUARTILE	Whole Blood 3rd	—	—
GNF U133A QUARTILE	Liver 3rd	—	—
GNF U133A QUARTILE	Fetalbrain 3rd	—	—
GNF U133A QUARTILE	Spinalcord 3rd	—	—
UNIGENE EST QUARTILE	Uterine Tumor Disease 3rd	Uterine Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Trachea Normal 3rd	Trachea Normal 3rd	—
UNIGENE EST QUARTILE	Uterus Normal 3rd	Uterus Normal 3rd	—
UNIGENE EST QUARTILE	Bladder Normal 3rd	Throid Normal 3rd	—
UNIGENE EST QUARTILE	Throid Normal 3rd	Bladder Normal 3rd	—
UNIGENE EST QUARTILE	Thyroid Tumor Disease 3rd	Thyroid Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Ovary Normal 3rd	Ovary Normal 3rd	—
UNIGENE EST QUARTILE	Pituitary Gland Normal 3rd	—	—
UNIGENE EST QUARTILE	Urinary Bladder Tumor Disease	—	—
UNIGENE EST QUARTILE	Mammary Gland Normal 3rd	—	—
UNIGENE EST QUARTILE	Mixed (Normal and Tumor) disease	—	—
UNIGENE EST QUARTILE	Esophagus Normal	—	—

Table C.5: Complete results from DAVID for the malignant transcript lists (1 of 6).

Annotation	Exon All	Exon Up	Exon Down
SP COMMENT TYPE	Subunit	Subunit	—
SP COMMENT TYPE	Interaction	Interaction	—
SP COMMENT TYPE	PTM	Development Stage	—
SP COMMENT TYPE	Function	PTM	—
SP COMMENT TYPE	Subcellular Location	Function Tissue Specificity	—
SP COMMENT TYPE	Developmental Stage	Subcellular Location	—
SP COMMENT TYPE	—	Catalytic Location	—
SP COMMENT TYPE	—	Induction	—
SP COMMENT TYPE	—	Disease	—
SP COMMENT TYPE	—	Domain	—
SP PIR KEYWORDS	—	Phosphoprotein	—
SP PIR KEYWORDS	—	Nucleus	—
SP PIR KEYWORDS	—	Tight Junction	—
SP PIR KEYWORDS	—	Cell Membrane	—
SP PIR KEYWORDS	—	Cell Junction	—
GOTERM BP 2	Chromosome Segregation	—	—
GOTERM CC 2	—	Apical Part of Cell	Extracellular Matrix

Table C.6: (continued.)

Annotation	Exon All	Exon Up	Exon Down
GOTERM CC 3	Plasma Membrane Part	—	Proteinaceous Extracellular Matrix
GOTERM CC 3	—	—	Extracellular Matrix
GOTERM CC 3	—	—	Extrinsic to Membrane
GOTERM CC 4	—	Apicolateral Plasma Membrane	Proteinaceous Extracellular Matrix
GOTERM CC 5	Cell-cell Junction	Apical Junction Complex	—
GOTERM CC 5	Apical Junction Complex	Apicolateral Plasma Membrane	—
GOTERM CC 5	Apicolateral Plasma Membrane	Cell-cell Junction	—
GOTERM CC 5	Plasma Membrane Part	—	—
GOTERM CC 5	Cell Junction	—	—
GOTERM CC ALL	Cell-cell Junction	Apical Junction Complex	—
GOTERM CC ALL	Apical Junction Complex	Apicolateral Plasma Membrane	Extracellular Matrix
GOTERM CC ALL	Apicolateral Plasma Membrane	Cell-cell Junction	—
GOTERM CC ALL	Plasma Membrane Part	Tight Junction	—
GOTERM CC ALL	Occluding Junction	Occluding Junction	—
GOTERM CC ALL	Tight Junction	—	—
GOTERM CC ALL	Cell Junction	—	—

Table C.7: (continued.)

Annotation	Exon All	Exon Up	Exon Down
GOTERM CC FAT	Plasma Membrane Part	Complex Apical Junction	Extracellular Matrix
GOTERM CC FAT	Cell-cell Junction	Apicolateral Plasma Membrane	—
GOTERM CC FAT	Apical Junction Complex	Cell-cell Junction	—
GOTERM CC FAT	Apicolateral Plasma Membrane	Occluding Junction	— —
GOTERM CC FAT	Cell Junction	Tight Junction	—
GOTERM CC FAT	Occluding Junction	Plasma Membrane Part	—
GOTERM CC FAT	Tight Junction	Apical Part of Cell	—
GOTERM CC FAT	—	Cell Junction	—
GOTERM	—	Chromosome	—
CC FAT	—	Centromeric region	
GOTERM CC FAT	—	Condensed Chromosome	—
GOTERM CC FAT	—	Apical Plasma Membrane	—
GOTERM CC FAT	—	Condensed Chromosome Centromeric Region	—
GOTERM CC FAT	—	Spindle	—
GOTERM CC FAT	—	Plasma Membrane	—
PANTHER MF ALL	—	—	Extracellular Matrix
CHROMOSOME	—	8	—
PUBMED ID	15489334	15489334	1544908
PUBMED ID	12477932	12477932	
PUBMED ID	19019843	18691976	—
PUBMED ID	—	18669648	—
PUBMED ID	—	19019843	—
PUBMED ID	—	17418912	—
PUBMED ID	—	18550469	—

Table C.8: (continued.)

Annotation	Exon All	Exon Up	Exon Down
EC NUMBER	2.7.11.1	2.7.11.1	—
UCSC TFBS	25	36	—
CGAP SAGE QUARTILE	Stem Cell Null 3rd	Stem Cell Null 3rd	Mammary Gland Normal Breast Tissue From a Breast Cancer Patient (Corresponding to IDC7) 3rd
CGAP SAGE QUARTILE	Uncharacterized Tissue Mixture of Human Cancer Cell Lines 3rd	Stem Cell 3rd	Mammary gland Breast Carcinoma Myoepithelium 3rd
CGAP SAGE QUARTILE	Stem Cell 3rd	Retina Bilateral Retinoblastoma, Poorly Differentiated, Left Orbit 3rd	Mammary gland Null 3rd
CGAP SAGE QUARTILE	—	Uncharacterized Tissue Mixture of Human Cancer Cell Lines 3rd	Peritoneum Malignant Peritoneal Mesothelioma 3rd
CGAP SAGE QUARTILE	—	Brain 3rd	Mammary gland Normal Breast Stroma 3rd
CGAP SAGE QUARTILE	—	Stomach Gastric Cancer 3rd	Peritoneum Normal 3rd
CGAP SAGE QUARTILE	—	—	Brain glioblastoma Infected With Mutant EGFR
CGAP SAGE QUARTILE	—	—	Mammary Gland Grade II
CGAP SAGE QUARTILE	—	—	Mammary Gland Grade I ER+ PR+ Her2- Invasive Ductal Carcinoma 3rd

Table C.9: (continued.)

Annotation	Exon All	Exon Up	Exon Down
CGAP SAGE QUARTILE	—	—	Cartilage Dedifferentiated Chondrosarcoma Lung Metastasis 3rd
CGAP SAGE QUARTILE	—	—	Mammary Gland Ductal Carcinoma In Situ Extensive Grade III Her2+ 3rd
CGAP SAGE QUARTILE	—	—	Mammary Gland Null 3rd
CGAP SAGE QUARTILE QUARTILE	—	—	Mammary Gland ER+ PR+ HER2- Grade II 3rd
CGAP SAGE QUARTILE	—	—	Skin Null 3rd
GNF U133A QUARTILE	Trigeminal Ganglion 3rd	Skin 3rd	PB-BDCA4+ Dentric Cells 3rd
GNF U133A QUARTILE	Skin 3rd	Pancreatic Islets 3rd	Fetabrain 3rd
GNF U133A QUARTILE	Pancreatic Islets 3rd	Trigeminal Ganglion 3rd	Adrenal Cortex 3rd
GNF U133A QUARTILE	Fetalbrain 3rd	Caudatenucleus 3rd	Uterus Corpus 3rd
GNF U133A QUARTILE	Heart 3rd	Heart 3rd	Atrioventricular Node 3rd
GNF U133A QUARTILE	Caudatenucleus 3rd	Trachea 3rd	Trigeminal Ganglion 3rd
GNF U133A QUARTILE	Parietal Lobe 3rd	Lymphnode 3rd Lymphnode 3rd	Subthalamcnucleus 3rd
GNF U133A QUARTILE	Trachae 3rd	Leukemialymphoblastic (Molt4) 3rd	— —
GNF U133A QUARTILE	—	Parietal Lobe 3rd	— —
GNF U133A QUARTILE	—	Fetallung 3rd	— —

Table C.10: (continued.)

Annotation	Exon All	Exon Up	Exon Down
UNIGENE EST QUARTILE	Uterine Tumor Disease 3rd	Uterine Tumor Disease 3rd	Adrenal gland normal 3rd
UNIGENE EST QUARTILE	Esophageal Tumor Disease 3rd	Esophageal Tumor Disease 3rd	Kidney Tumor Disease 3rd
UNIGENE EST QUARTILE	Esophagus Normal 3rd	Esophagus Normal 3rd	Connective Tissue Normal 3rd
UNIGENE EST QUARTILE	Ovarian Tumor Disease 3rd	Colorectal Tumor Disease 3rd	Umbilical Cord Normal 3rd
UNIGENE EST QUARTILE	—	Ovarian Tumor Disease 3rd	Adipose Tissue Normal 3rd
UNIGENE EST QUARTILE	—	Urinary Bladder Tumor Disease 3rd	Adrenal Tumor Disease 3rd
UNIGENE EST QUARTILE	—	Abdominal Cavity Normal 3rd	—
UNIGENE EST QUARTILE	—	Gastrointestinal Tumor Disease 3rd	—
UNIGENE EST QUARTILE	—	Stomach Normal 3rd	—
UNIGENE EST QUARTILE	—	Disease 3rd Disease 3rd	—
UNIGENE EST QUARTILE	—	Bladder Normal 3rd	—
UNIGENE EST QUARTILE	—	Mixed (normal and tumor) Disease 3rd	—
UNIGENE EST QUARTILE	—	Tongue Normal 3rd	—
UNIGENE EST QUARTILE	—	Larynx Normal 3rd	—
UNIGENE EST QUARTILE	—	Trachea Normal 3rd	—
UNIGENE EST QUARTILE	—	Laryngeal Cancer Disease 3rd	—
UNIGENE EST QUARTILE	—	Small Intestine Normal 3rd	—

Table C.11: Complete results from DAVID for the benign transcript cluster lists (1 of 2).

Annotation	Gene All	Gene Up	Gene Down
SP COMMENT TYPE	Tissue Specificity	—	—
PUBMED ID	12477932	12477932	—
PUBMED ID	15489334	15489334	—
PUBMED ID	—	16344560	—
UCSC TFBS	48 Entries	57 Entries	—
CGAP SAGE QUARTILE	Brain Ependymoma 3rd	Brain Ependymoma 3rd	—
CGAP SAGE QUARTILE	Brain Ependymoma 3rd	—	—
GNF U133A QUARTILE	Trigeminal Ganglion 3rd	Skin 3rd	—
GNF U133A QUARTILE	Skin 3rd	Trigeminal Ganglion 3rd	—
GNF U133A QUARTILE	Whole Blood 3rd	Pancreatic Islets 3rd	—
GNF U133A QUARTILE	Pancreatic Islets 3rd	Whole Blood 3rd	—
GNF U133A QUARTILE	—	Parietal Lobe 3rd	—
UNIGENE EST QUARTILE	Uterine Tumor Disease 3rd	Uterine Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Trachea Normal 3rd	Trachea Normal 3rd	—
UNIGENE EST QUARTILE	Uterus Normal 3rd	Thyroid Normal 3rd	—
UNIGENE EST QUARTILE	Throid Normal 3rd	Uterus Normal 3rd	—
UNIGENE EST QUARTILE	Thyroid Tumor Disease 3rd	Thyroid Tumor Disease 3rd	—

Table C.12: (continued.)

Annotation	Gene All	Gene Up	Gene Down
UNIGENE EST QUARTILE	Bladder Normal 3rd	Bladder Normal 3rd	—
UNIGENE EST QUARTILE	Mammary Gland Normal 3rd	Mammary Gland Normal	—
UNIGENE EST QUARTILE	Pituitary Gland Normal 3rd	Breast (Mammary Gland) Cancer Disease 3rd	—
UNIGENE EST QUARTILE	Ovary Normal 3rd	Ovary Normal 3rd	—
UNIGENE EST QUARTILE	—	Esophagus Normal 3rd	—
UNIGENE EST QUARTILE	—	Pituitary Gland Normal 3rd	—
UNIGENE EST QUARTILE	—	Esophageal Tumor Disease 3rd	—

Table C.13: Complete results from DAVID for the malignant transcript cluster lists (1 of 6).

Annotation	Gene All	Gene Up	Gene Down
SP COMMENT TYPE	PTM	PTM	—
SP COMMENT TYPE	Subunit	Subunit	—
SP COMMENT TYPE	Interaction	Induction	—
SP COMMENT TYPE	Function	Developmental Stage	—
SP COMMENT TYPE	Domain	Function	—
SP COMMENT TYPE	Developmental Stage	Interaction	—
SP COMMENT TYPE	Induction	Catalytic Activity	—
SP COMMENT TYPE	Catalytic Activity	Domain	—
SP COMMENT TYPE	Tissue Specificity	Tissue Specificity	—
SP COMMENT TYPE	Alternative Products	Subcellular Location	—
SP COMMENT TYPE	Disease	—	—
SP COMMENT TYPE	Subcellular Location	—	—
SP PIR KEYWORDS	Tight Junction	Tight Junction	—
SP PIR KEYWORDS	Cell Membrane	Cell Membrane	—
SP PIR KEYWORDS	—	Phosphoprotein	—
SP PIR KEYWORDS	—	Cell Junction	—

Table C.14: (continued.)

Annotation	Gene All	Gene Up	Gene Down
GOTERM CC 3	Plasma Membrane Part	Plasma Membrane Part	Plasma Membrane Part
GOTERM CC 3	Plasma Membrane	—	—
GOTERM CC 4	Plasma Membrane Part	Apicolateral Plasma Membrane	Extrinsic to Membrane
GOTERM CC 4	Apicolateral Plasma Membrane	Plasma Membrane Part	— —
GOTERM CC 5	Plasma Membrane Part	Apical Junction Complex	Extrinsic to Membrane
GOTERM CC 5	Apical AJunction Complex	Apicolateral Plasma Membrane	—
GOTERM CC 5	Apicolateral Plasma Membrane	Plasma Membrane Part	—
GOTERM CC 5	Cell-Cell Junction	Cell-Cell Junction	—
GOTERM CC 5	—	Chromosome, Centromeric Region	—
GOTERM CC 5	—	Condensed Chromosome	—
GOTERM CC 5	—	Apical Plasma Membrane	—
GOTERM CC 5	—	Condensed Chromosome, Centromeric Region	—
GOTERM CC ALL	Plasma Membrane Part	Apical Junction Complex	Extrinsic to Membrane
GOTERM CC ALL	Apical Junction Complex	Apicolateral Plasma Membrane	—
GOTERM CC ALL	Apicolateral Plasma Membrane	Tight Junction	—
GOTERM CC ALL	Cell-Cell Junction	Occluding Junction	—
GOTERM CC ALL	Occluding Junction	Plasma Membrane Part	—

Table C.15: (continued.)

Annotation	Gene All	Gene Up	Gene Down
GOTERM CC ALL	Tight Junction	Cell-cell Junction	—
GOTERM CC ALL	Plasma Membrane	Chromosome, Centromeric Region	—
GOTERM CC ALL	—	Condensed Chromosome	—
GOTERM CC ALL	—	Apical Plasma Membrane	—
GOTERM CC ALL	—	Condensed Chromosome Centromeric Region	—
GOTERM CC FAT	Plasma Membrane Part	Apical Junction Complex	Extrinsic to Membrane
GOTERM CC FAT	Apical Junction Complex	Apicolateral Plasma Membrane	—
GOTERM CC FAT	Apicolateral Plasma Membrane	Plasmal Membrane Part	—
GOTERM CC FAT	Cell-Cell Junction	Occluding Junction	—
GOTERM CC FAT	Plasma Membrane	Tight Junction	—
GOTERM CC FAT	Occluding Junction	Cell-cell Junction	—
GOTERM CC FAT	Tight Junction	Chromosome Centromeric Region	—
GOTERM CC FAT	—	Plasma Membrane	—
GOTERM CC FAT	—	Condensed Chromosome	—
GOTERM CC FAT	—	Apical Plasma Membrane	—
GOTERM CC FAT	—	Condensed Chromosome, Centromeric Region	—

Table C.16: (continued.)

Annotation	Gene All	Gene Up	Gene Down
GOTERM CC FAT	—	Spindle	—
GOTERM CC FAT	—	Anchored to Plasma Membrane	—
GOTERM CC FAT	—	Cell Junction	—
Panther MF ALL	—	Cell Junction Protein	—
PUBMED ID	15489334	15489334	—
PUBMED ID	12477932	12477932	—
PUBMED ID	18669648	18669648	—
PUBMED ID	—	18691976	—
PUBMED ID	—	17418912	—
PUBMED ID	—	18550469	—
PUBMED ID	—	19008095	—
PUBMED ID	—	16964243	—
PUBMED ID	—	9892664	—
EC NUMBER	2.7.11.1	2.7.11.1	—
REACTOME PATHWAY	Cell-Cycle Mitotic	Cell-Cycle Mitotic	—
UCSC TFBS	35	39	—
CGAP SAGE QUARTILE	Stem Cell 3rd	Stem Cell 3rd	—
CGAP SAGE QUARTILE	Stem Cell Null 3rd	Stem Cell Null 3rd	—
CGAP SAGE QUARTILE	Uncharacterized Tissue Mixture of Human Cancer Cell Lines 3rd	Retina Bilateral Retinoblastoma Poorly Differentiated Left Orbit 3rd	— — —
CGAP SAGE QUARTILE	—	Uncharacterized Tissue Mixture of Human Cancer Cell Lines 3rd	—
CGAP SAGE QUARTILE	—	Liver Poorly Differentiated Adenocarcinoma 3rd	—

Table C.17: (continued.)

Annotation	Gene All	Gene Up	Gene Down
GNF U133A QUARTILE	Trigeminal Ganglion 3rd	Pancreatic Islets 3rd	Fetalbrain 3rd
GNF U133A QUARTILE	Pancreatic Islets 3rd	Skin 3rd	—
GNF U133A QUARTILE	Skin 3rd	Trigeminal Ganglion 3rd	—
GNF U133A QUARTILE	Heart 3rd	Heart 3rd	—
GNF U133A QUARTILE	Fetalbrain 3rd	Caudatenucleus 3rd	—
GNF U133A QUARTILE	Parietal Lobe 3rd	Trachea 3rd	—
GNF U133A QUARTILE	Caudatenucleus 3rd	Parietal Lobe 3rd	—
GNF U133A QUARTILE	Trachea 3rd	Lymphnode 3rd	—
UNIGENE EST QUARTILE	Uterine Tumor Disease 3rd	Uterine Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Esophageal Tumor Disease 3rd	Esophageal Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Esophagus Normal 3rd	Esophagus Normal 3rd	—
UNIGENE EST QUARTILE	Ovarian Tumor Disease 3rd	Urinary Bladder Tumor 3rd	—
UNIGENE EST QUARTILE	Urinary Bladder Tumor Disease 3rd	Ovarian Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Gastrointestinal Tumor Disease 3rd	Oral Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Abdominal Cavity Normal 3rd	Abdominal Cavity Normal 3rd	—
UNIGENE EST QUARTILE	Colorectal Tumor Disease 3rd	Gastrointestinal Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Oral Tumor Disease 3rd	Colorectal Tumor Disease 3rd	—
UNIGENE EST QUARTILE	Mixed (Normal and Tumor) Disease 3rd	Bladder Normal 3rd	—
UNIGENE EST QUARTILE	Bladder Normal 3rd	Mixed (Normal and Tumor) Disease 3rd	—

Table C.18: (continued.)

Annotation	Gene All	Gene Up	Gene Down
UNIGENE EST QUARTILE	—	Stomach Normal 3rd	—
UNIGENE EST QUARTILE	—	Small Intestine Normal 3rd	—
UNIGENE EST QUARTILE	—	Tongue Normal 3rd	—
UNIGENE EST QUARTILE	—	Ovary Normal 3rd	—
UNIGENE EST QUARTILE	—	Trachea Normal 3rd	—