AWARD NUMBER: W81XWH-13-1-0357

TITLE: The Impact of Epithelial-Stromal Interactions on Human Breast Tumor Heterogeneity

PRINCIPAL INVESTIGATOR: Dr. Crista Thompson

CONTRACTING ORGANIZATION: Royal Institution for the Advancement of Learning
Montreal QC H3A 2T5

REPORT DATE: October 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2015	Annual	15 Sept 2014 - 14 Sept 2015

4. TITLE AND SUBTITLE

The Impact of Epithelial-Stromal Interactions on Human Breast Tumor Heterogeneity

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-13-1-0357

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Crista Thompson

E-Mail: crista.thompson@mail.mcgill.ca

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Royal Institution for the Advancement of Learning
845 Sherbrooke St. W. suite 531
Montreal QC
H3A 2T5

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Heterogeneity is a key factor underlying the variability in patient response to treatment, especially in Triple-Negative (TN) breast cancer cases. In addition to the intrinsic molecular characteristics of the tumor epithelium, the stroma can influence tumor progression and response to therapy. The purpose of this project is to identify, define and formally test critical pathways mediating tumor epithelial-stromal communication and co-dependency in TN breast tumor subtypes associated with poor outcome. To date, we have profiled the gene and miRNA expression of distinct epithelial and stromal compartments from ca. 50 TN tumors, identified epithelial and stromal gene expression modules, and created an interaction network for these gene modules. These are critical steps towards understanding the relationship between the tumor and its microenvironment. Our project will provide the first integrated in-depth analysis of the contribution of tumor stromal processes to disease heterogeneity, and will position the tumor microenvironment for therapeutic intervention.

15. SUBJECT TERMS
Breast cancer, Triple-Negative, epithelium, stroma, gene expression, microRNA, laser capture microdissection, heterogeneity, molecular profiles, tumor microenvironment

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	26	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

1. Introduction

Breast cancer is a heterogeneous disease in terms of presentation, morphology, molecular profile and response to therapy. Gene expression profiling has identified six molecular subtypes, *i.e.* luminal A, luminal B, normal breast-like, HER2+, basal-like and claudin-low, that are associated with clinical markers as well as prognosis and survival [1-4]. However, it is well established that the intrinsic molecular profiles of breast tumors are not sufficient to perfectly predict disease outcome. Increasing evidence indicates that characteristics of the breast stroma and, perhaps more specifically, interactions between the tumor epithelium and stroma influence breast cancer progression and response to therapy. Previous work in our lab has demonstrated that gene expression signatures in human stroma can predict outcome of breast cancer patients independently of clinical parameters and molecular subtypes [5]. To expand on these results, the goal of this project is to identify, define and formally test critical pathways mediating tumor epithelial-stromal communication and co-dependency in Triple-Negative (TN) breast cancer (defined as tumors lacking expression of estrogen receptor, progesterone receptor and human epidermal growth factor receptor-2 (HER2)), a subtype associated with poor outcome. We hypothesize that by defining the tumor stromal pathways associated with poor outcome in TN tumors, we will uncover mechanisms for co-evolution, biomarkers and potential therapeutic targets. Our specific aims are to develop coordinate stromal-epithelial expression signatures for a cohort of *ca.* 50 TN breast cancers for which outcome and follow-up are available, to identify stromal-epithelial gene interaction networks, and to identify and integrate stromal-epithelial gene expression and microRNA (miR) signatures associated with TN breast tumors. It is well recognized that heterogeneity is a key factor underlying the variability in patient response to treatment, especially in TN cases. There is a need for a fuller understanding of the molecularly distinct TN subgroups linked to outcome and the development of more personalized treatment strategies for members of this subgroup. This project will provide the first integrated in-depth analysis of the contribution of tumor stromal processes to disease heterogeneity, and will position the tumor microenvironment for therapeutic intervention. This project also promises the "next generation" of signatures based on miR that are stable in clinical materials and can be developed for non-invasive tests suitable for stratification of patients for chemotherapy, monitoring disease progression and, in the long term, for early detection and screening for metastatic disease.

2. Keywords

Breast cancer, Triple-Negative, epithelium, stroma, gene expression, microRNA, laser capture microdissection, heterogeneity, molecular profiles, tumor microenvironment

3. Overall Project Summary

3.1 Current Objectives

This research project has 3 tasks covering 3 years (refer to Statement of Work in Appendix 1):

1. Develop coordinate stromal-epithelial mRNA expression signatures for TN tumors.
2. Identify stromal-epithelial gene interaction networks.
3. Identify and integrate stromal-epithelial miR signatures associated with TN breast tumors.

The objectives for this project in its second year (2014-2015) were as follows:

- Identify stromal subclasses of TN tumors based on gene expression.
- Develop a *de novo* bioinformatics tool to identify genes modulating cross-talk between tumor epithelium and tumor-associated stromal compartments.
- Investigate miR signatures for their prognostic value by using linked patient outcome data.

3.2 Results, Progress and Accomplishments

During the first year of this project, we successfully isolated epithelial and stromal tissue from banked TN tumor samples via laser capture microdissection (LCM) (see Figure 1 in Appendix 2 for a depiction of LCM; methods as per [6]). RNA was extracted from epithelial and stromal LCM isolates and subjected to microarray-based gene expression profiling via Agilent SurePrint G3 8x60K chips using methods based on those previously described by our group [5, 6]. In the second year of this project, our first goal was to use the gene expression data to identify stromal subclasses of TN tumors. We then wanted to identify the genes modulating cross-talk between the tumor epithelium and associated stromal compartments. Before beginning our analyses, we verified the integrity (tissue specificity) of the normalized gene expression data. We selected the most variable genes (interquartile range (IQR) > 2) across all samples and separated this geneset unbiasedly into two opposing directions using the Partitioning Around Medoids (pam) function from the cluster package in R [Maechler, 2015; version 2.0.1;

http://wis.kuleuven.be/stat/robust/papers/2005/maechleretal-rpackagecluster-cran-2005.pdf]. We then ranked the patients from lowest to highest in terms of expression of the overall geneset (Figure 2, Appendix 2). This approach orders all patient samples by first ranking them on the basis of expression of these characteristic genes, followed by summing the ranks. Thus, patients with the smallest sum of expression are ranked lowest (right) and those with the largest sum are ranked highest (left). This approach revealed that the epithelial and stromal tissue samples are distinct, and that adjacent normal tissue can be distinguished from tumor tissue (Figure 2, Appendix 2). Therefore, the LCM procedure was successful in separating epithelial from stromal tissue, as well as tumor from adjacent normal tissue.

After confirming the integrity of the gene expression data, we attempted to identify stromal subclasses using a clustering-based class discovery approach [1]. We defined subtypes as groups of patients with similar gene expression profiles that cluster closely together. However, due to the complexity of the gene expression profiles, this method of subtyping masked certain clusters of genes, *i.e.* patient clusters were predominantly driven by immune-related genes and this strong immune signal masked the effect of weaker gene clusters. Therefore, we adopted an alternate approach that classifies groups of genes (gene modules) by the degree to which they co-vary across patient samples. This method, referred to as Weighted Gene Correlation Network Analysis (WGCNA) [7], identified co-modulated groups of genes that had high absolute correlation in patient stromal and epithelial tissues. The gene modules were given color names to prevent overt assumptions about the biology of the module, and they were associated with disease recurrence in the patients (Figure 3, Appendix 2). In total, there were 25 epithelial modules and 24 stroma modules identified (Figure 4, Appendix 2).

In addition to intra-tissue correlation (Figure 4, Appendix 2), the epithelial and stromal gene modules showed inter-tissue correlation as determined by Spearman correlation permutation and hypergeometric Fisher's exact test (Figure 5, Appendix 2). Using both intra- and inter-tissue correlations (Tables 1-3, Appendix 3), we generated a correlation network of gene modules (Figure 6, Appendix 2). This correlation map revealed two main networks with five network hubs, and an additional four independent epithelial-stromal interactions. Because shared genes between the epithelial and stromal modules could result in a high correlation score, we calculated the percentage of common genes between the modules (Table 1, Appendix 3). We defined distinct but correlative modules as having <20% common genes. With the identification of the epithelial and stromal modules, as well as the evaluation of how they relate to one another, we are ready to characterize these modules according to biological functions and disease outcome.

The third goal for the second year of this project was to investigate miR signatures for their prognostic value using linked patient outcome data. During the first year of this project, we were delayed in profiling the miR expression in tumor and normal epithelium and stroma as a result of technical difficulties. While the profiling is now completed, data normalization is still ongoing. Therefore, we have not yet assessed the miR signatures for their prognostic value. We anticipate that the normalized miR data will be available by December 2015 in order to investigate and validate miR of interest as outlined in the Statement of Work (Appendix 1).

3.3 Discussion

We have previously demonstrated that gene expression signatures in human stroma can predict the outcome of breast cancer patients independently of clinical parameters and molecular subtypes [5]. Moreover, these stromal subclasses have been shown to segregate human breast tumors by disease outcome and contribute significantly to tumor heterogeneity. Thus, it is clear that further investigation into epithelial-stromal interactions is imperative to our understanding of breast tumor heterogeneity and, as such, has significant implications in positively influencing patient stratification, treatment and survival. This is especially true for TN cases which represent approx. 15% of all breast cancers [8-11] and, as a result of no targetable clinical markers, are generally treated by combined surgery, radiotherapy and non-targeted chemotherapy. Many TN tumors display a good response to anthracycline- and taxane-based chemotherapy, especially in the neo-adjuvant setting [10, 12, 13]. However, overall outcome remains poor in TN disease [10] and no mechanisms exist to determine which patients will respond to chemotherapy. We predict that interrogating the gene expression of the epithelium and surrounding stroma in TN tumors will provide insight into the co-evolution and/or co-dependency of these tissues, and reveal which gene signatures are associated with poor outcome as well as foster the development of more personalized treatment strategies for patients with TN breast cancer.

In the first year of this project, we successfully profiled the gene expression of tumor epithelium and associated stroma as well as matched normal epithelium and stroma of *ca.* 50 TN tumors. During the second year of this project, we initially proposed to apply a "class discovery" bioinformatics approach to the gene lists to identify stromal subgroups. However, as a result of the complexity of the gene expression profiles, this method of subtyping masked certain clusters of genes. Therefore, we adopted an alternate approach (WGCNA) that classifies groups of genes (gene modules) by the degree to which they vary concurrently across patient samples. The identification of epithelial and stromal gene modules combined with an evaluation of their correlation allowed us to generate a module interaction map (Figure 6,

Appendix 2). With this interaction network and an understanding of which modules are "distinct" (*i.e.* have a low percentage of shared genes), we are beginning to define the biological functions of our network hubs and related modules. To do this, we are performing gene set enrichment analysis using tools such as QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) and the Molecular Signatures Database (MSigDB) [14]. As an example, preliminary IPA® analysis of one of our network hubs, Epithelial (Epi) Turquoise, indicates that one of the canonical pathways represented by this gene list is Oxidative Phosphorylation, also referred to as mitochondrial respiration (Figure 7, Appendix 2). The fact that patients with high expression of the Epi Turquoise genes have lower reported recurrence (Figure 3, Appendix 2) suggests that TN tumors with high levels of oxidative phosphorylation (versus glycolysis) are more sensitive to therapy resulting in an improved outcome. This is consistent with reports that glycolytic cancer cells are resistant to chemotherapeutic agents and that uptake of functional mitochondria by cancer cells increases drug sensitivity [15]. In addition, mitochondrial respiration is considered an important source of oxidative stress in cancer cells [16] and it has been shown that sensitivity to chemotherapy is increased in cancer cells with elevated oxidative stress [17]. These preliminary results are very encouraging and we anticipate that future analyses will contribute to our understanding of breast cancer heterogeneity by clarifying epithelial-stromal gene signatures and interactions in TN breast cancer.

Changes in miR expression have been documented in breast cancer [18-22], and several of these have been shown to be associated with clinical features [18, 23-28] including response to therapy [29-32]. However, little is known regarding the prognostic value of miR sets in tumor stroma, particularly in TN breast cancer. The third objective for this project in its second year was to investigate miR signatures for their prognostic value using linked patient outcome data. Because of a delay caused by technical difficulties, this objective has not been met. However, we anticipate that the normalized miR data will be available by December 2015 and this will allow us to investigate and validate miR of interest as proposed.

4. Key Research Accomplishments

The following key research accomplishments have contributed to our major objectives of developing coordinate stromal-epithelial mRNA expression signatures for TN tumors and identifying stromal-epithelial gene interaction networks:

- Identification of the epithelial and stromal gene modules in TN patient samples
- Generation of the gene module correlation network

5. Conclusion

Heterogeneity plays a substantial role in the variability of patient response to treatment, especially in TN cases. A fuller understanding of the molecularly distinct TN subgroups linked to outcome is essential to promote the development of more personalized treatment strategies. We have previously demonstrated that gene expression signatures in human stroma can predict outcome of breast cancer patients independently of clinical parameters and molecular subtypes [5]. To expand on these results, the goal of this project is to identify, define and formally test critical pathways mediating tumor epithelial-stromal communication and co-dependency in TN breast cancer. Our specific aims are to develop coordinate stromal-epithelial expression signatures, to identify stromal-epithelial gene interaction networks, and to identify and integrate stromal-epithelial gene expression and miR signatures associated with TN breast tumors. To date, we have profiled the gene and miR expression of distinct epithelial and stromal compartments from *ca*. 50 TN tumors, identified epithelial and stromal gene expression modules, and generated an interaction/correlation map of these gene modules. These are critical steps towards accomplishing our major objectives and understanding the relationship between the tumor and its microenvironment. Future experiments will include identifying the biological functions represented by our gene modules, characterizing the epithelial-stromal interactions associated with good or poor response to chemotherapy, and validating gene and associated miR signatures as predictors of outcome using patient samples. Although other groups have subtyped TN breast cancer (*e.g.* [33], [34]), these molecular studies have been performed on whole tumor samples with >70% epithelial content. Our project will provide the first integrated in-depth analysis of the contribution of tumor stromal processes to disease heterogeneity, and will position the tumor microenvironment for therapeutic intervention.

6. Publications, Abstracts and Presentations

6.1 Poster presentation

C. Thompson, N. Bertos, T. Gruosso, G. Finak, R. Lesurf, S. Saleh, H. Zhao, M. Souleimanova, S. Meterissian, A. Omeroglu, M.T. Hallett & M. Park. A new breast cancer classification scheme based on

novel classes of tumor stroma. San Antonio Breast Cancer Symposium® annual meeting, San Antonio TX, December 2014.

7. Inventions, Patents and Licenses

Nothing to report.

8. Reportable Outcomes

During this reporting period, we generated an interaction network describing the relationships between epithelial and stromal gene expression modules in TN tumors. This research tool is an essential achievement in our strategy to identify, define and formally test critical pathways mediating tumor epithelial-stromal communication and co-dependency and, ultimately, understand the impact of epithelial-stromal interactions on triple negative breast tumor heterogeneity.

9. Other achievements

9.1 Training and Professional Development

As the Principal Investigator on this project, I have had the opportunity to train in new techniques and improve my professional skills over the past year. With the assistance of collaborators with expertise in bioinformatics (*e.g*. S. Saleh), I have learned about class discovery, WGCNA and gene set enrichment analysis. In addition, I have met routinely with Dr. Bertos and my mentor, Dr. Park, to discuss technical and theoretical aspects of the project as well as budgetary concerns. I have learned how to use the financial systems in place at McGill University to monitor and control my research funds. These meetings/training have contributed to my abilities in project management.

My project location, the Goodman Cancer Research Centre at McGill University, runs a weekly seminar series at which Principal Investigators, graduate students and postdoctoral fellows present their work. In addition, invited external speakers present their current research at regular seminars. Many of these researchers are working on breast cancer projects and these seminars are keeping me abreast of current trends in the field. They also provide opportunities for collaborations or additional training.

This year, I trained and mentored a doctoral candidate in her research project. This is an important piece of my training because as a professor with my own lab, I will be training and directing/mentoring undergraduate and graduate students.

Finally, because I am interested in connecting basic research with the clinic, I attended the San Antonio Breast Cancer Symposium® annual meeting in December 2014. The Symposium's mission is to provide state-of-the-art information on breast cancer research. The five-day program is attended by a broad international audience of academic and private researchers as well as physicians from over 90 countries and aims to achieve a balance of clinical, translational, and basic research. In addition to attending presentations covering a range of topics such as patient-derived xenografts as models of metastasis, the reliance of HER2 pathology on HER3 and the most recent advances in immunotherapy, I attended a career development forum for young investigators and presented a poster entitled "A new breast cancer classification scheme based on novel classes of tumor stroma." There was a great deal of interest in the poster presentation and I was able to interact with students, post-doctoral fellows, Principal Investigators, clinicians and breast cancer survivors. It was a great opportunity to discuss the project, highlight the progression of the research and brain-storm future directions and applications of our results. Overall the experience was very inspiring and expanded my vision as a breast cancer researcher.

10. References

1. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.
2. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418-23.
3. Herschkowitz JI, Simin K, Weigman VJ, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 2007;8:R76.
4. Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.
5. Finak G, Bertos N, Pepin F, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nat Med 2008;14:518-27.
6. Finak G, Sadekova S, Pepin F, et al. Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. Breast Cancer Res 2006;8:R58.
7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.

8. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer 2007;109:25-32.

9. Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res 2007;13:4429-34.

10. Carey LA, Dees EC, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res 2007;13:2329-34.

11. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 2007;109:1721-8.

12. Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review. Histopathology 2008;52:108-18.

13. Liedtke C, Mazouni C, Hess KR, et al. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. J Clin Oncol 2008;26:1275-81.

14. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.

15. Elliott RL, Jiang XP, Head JF. Mitochondria organelle transplantation: introduction of normal epithelial mitochondria into human cancer cells inhibits proliferation and increases drug sensitivity. Breast Cancer Res Treat 2012;136:347-54.

16. Balaban RS, Nemoto S, Finkel T. Mitochondria, oxidants, and aging. Cell 2005;120:483-95.

17. Mateescu B, Batista L, Cardon M, et al. miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. Nat Med 2011;17:1627-35.

18. Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065-70.

19. Lowery AJ, Miller N, McNeill RE, Kerin MJ. MicroRNAs as prognostic indicators and therapeutic targets: potential effect on breast cancer management. Clin Cancer Res 2008;14:360-5.

20. Hu J, Xu J, Wu Y, et al. Identification of microRNA-93 as a functional dysregulated miRNA in triple-negative breast cancer. Tumour Biol 2014.

21. Avery-Kiejda KA, Braye SG, Mathe A, Forbes JF, Scott RJ. Decreased expression of key tumour suppressor microRNAs is associated with lymph node metastases in triple negative breast cancer. BMC Cancer 2014;14:51.

22. Li Y, Xu Y, Yu C, Zuo W. Associations of miR-146a and miR-146b expression and breast cancer in very young women. Cancer Biomark 2015.

23. Van der Auwera I, Limame R, van Dam P, Vermeulen PB, Dirix LY, Van Laere SJ. Integrated miRNA and mRNA expression profiling of the inflammatory breast cancer subtype. Br J Cancer 2010;103:532-41.

24. Lowery AJ, Miller N, Devaney A, et al. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. Breast Cancer Res 2009;11:R27.

25. Gasparini P, Cascione L, Fassan M, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget 2014;5:1174-84.

26. De Rinaldis E, Gazinska P, Mera A, et al. Integrated genomic analysis of triple-negative breast cancers reveals novel microRNAs associated with clinical and molecular phenotypes and sheds light on the pathways they control. BMC Genomics 2013;14:643.

27. Li H, Zhou X, Zhu J, et al. MiR-4728-3p could act as a marker of HER2 status. Cancer Biomark 2015.

28. Fkih M, I, Privat M, Ponelle F, Penault-Llorca F, Kenani A, Bignon YJ. Identification of miR-10b, miR-26a, miR-146a and miR-153 as potential triple-negative breast cancer biomarkers. Cell Oncol (Dordr) 2015.

29. Rodriguez-Gonzalez FG, Sieuwerts AM, Smid M, et al. MicroRNA-30c expression level is an independent predictor of clinical benefit of endocrine therapy in advanced estrogen receptor positive breast cancer. Breast Cancer Res Treat 2011;127:43-51.

30. Miller TE, Ghoshal K, Ramaswamy B, et al. MicroRNA-221/222 confers tamoxifen resistance in breast cancer by targeting p27Kip1. J Biol Chem 2008;283:29897-903.

31. Ouyang M, Li Y, Ye S, et al. MicroRNA profiling implies new markers of chemoresistance of triple-negative breast cancer. PLoS One 2014;9:e96228.

32. Wang Q, Li C, Zhu Z, et al. miR-155-5p antagonizes the apoptotic effect of bufalin in triple-negative breast cancer cells. Anticancer Drugs 2015.

33. Prat A, Adamo B, Cheang MC, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. Oncologist 2013;18:123-33.

34. Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750-67.

35. Liotta L, Petricoin E. Molecular profiling of human cancer. Nat Rev Genet 2000;1:48-56.

11. Appendices

11.1 Statement of Work

Statement of Work

Note: All work will be performed at the Goodman Cancer Research Institute, 1160 Des Pins Avenue West, Montreal, Quebec, Canada, H3A 1A3 unless specified. The Principal Investigator (PI) is Dr. Crista Thompson and the Mentor is Dr. Morag Park.

Task Description	Year 1	Year 2	Year 3
1. Develop coordinate stromal-epithelial mRNA expression signatures for Triple-negative (TN) tumors. • Resource: Dr. Park established the Breast Cancer Functional Genomics Group. This group has banked fresh-frozen breast cancer tumor (approx. 400) and normal (approx. 500 including matched samples and reduction mammoplasties) tissue samples obtained from surgeries conducted at the McGill University Health Centre under strict quality control guidelines. Blood samples collected at the time of surgery have been processed as serum and plasma and stored. Matched formalin-fixed paraffin-embedded (FFPE) samples from the clinical pathology archive can be obtained when feasible and tissue microarrays for banked samples have been constructed to aid large-scale IHC and *in situ* hybridization analyses. An attending clinical pathologist specializing in breast pathology rescores all banked samples for consistency. HER2 Fluorescence *in situ* hybridization is performed to confirm HER2 status in equivocal cases and p53 mutation analysis is conducted for all samples. All experimental data is linked to information regarding pathology analysis, therapy and disease course. Tissue and blood collection and participant follow-up providing outcome is conducted with Research Ethics Board approval.			
1a. Conduct laser capture microdissection (LCM) to isolate separate epithelial and stromal compartments from banked tumor samples, both tumor-associated and adjacent normal tissues. • Collaborator/Personnel: Dr. Nicholas Bertos / Hong Zhao • Samples from 30 TN patients with distant recurrence within 5 years and 20 TN patients with no recurrence in 5 years will be analyzed. Therefore, there will be a total of 200 analyses (50 samples × 4 tissue compartments/sample). • PI Training: Learn how to perform LCM.	Months 1-8		
1b. Extract RNA from epithelial and stromal LCM isolates and subject to microarray-based gene expression profiling. • Collaborator/Personnel: Dr. Nicholas Bertos / Hong Zhao • Profiling will be performed with Agilent Whole Human Genome 4x44K chips • PI Training: Learn how to extract RNA from LCM isolates. • PI Training: Learn how to perform microarray-based gene profiling.	Months 6-12		
1c. Identify stromal subclasses. • Collaborator/Personnel: Dr. Michael Hallett / Sadiq Saleh • Methods: Genes defining stromal subclasses will demonstrate homogeneous expression within the corresponding cluster, as well as heterogeneous expression outside the cluster as determined by variance component analysis. The biological functions over-represented in each stroma class will be identified by performing gene set enrichment analysis and testing for enrichment against multiple ontological databases including Gene Ontology (GO), the Kyoto encyclopedia of genes and genomes (KEGG) and List2List (L2L). • PI Training: Learn about class discovery and gene set enrichment analysis.		Months 1-6	
Milestone: Complete characterization of profiles in matched normal and tumor stroma and corresponding epithelia to reveal relevant tumor-associated changes and epithelial-stromal gene expression networks.			

Task Description	Year 1	Year 2	Year 3
2. Identify stromal-epithelial gene interaction networks.			
2a. Develop a *de novo* bioinformatics tool, STR-EPI, to identify genes modulating cross-talk between tumor epithelium and tumor-associated stromal components. • Collaborator/Personnel: Dr. Michael Hallett / Sadiq Saleh • Resources: A comprehensive database of > 1600 breast cancer specific gene signatures (BreastSigDB). These include both signatures from the literature as well as those contained in public databases such as MsigBD. • Methods: We will develop a stromal-epithelial interaction map for each prominent subtype combination identified in task 1 using a variety of established and new informatics tools.		Months 6-12	Months 1-3
Milestone: Development of a new bioinformatics tool STR-EPI to identify stromal-epithelial gene signatures.			
2b. Characterize epithelial-stromal subtypes specifically associated with good or poor response to chemotherapy. • Collaborator/Personnel: Dr. Michael Hallett / Sadiq Saleh • Resource: We have generated a human gene expression data compendium derived from 22 publicly available datasets that contained patients diagnosed with invasive ductal carcinoma with associated clinical information, including recurrence status (defined as distant metastasis within 5 years), survival, and immunohistochemistry results (currently n = 5175 patients containing 619 TN patients). • Methods: Within the stromal and epithelial datasets, each gene present will be ranked as a univariate predictor of recurrence using a parametric test. These predictors will be trained using a Naïve Bayes Classifier and crossvalidated under a leave-one-out cross-validation scheme. The signature will be re-trained in our data and validated using the same procedure in new and existing gene expression datasets with outcome following treatment to an anthracycline- and/or taxane-based regimens utilizing our breast cancer gene expression compendia mentioned above.			Months 3-6
2c. Validate STR-EPI outcome predictors. • Collaborator/Personnel: Dr. Nicholas Bertos / Hong Zhao • Methods: Outcome predictors will be validated by reverse transcriptase PCR and IHC/*in situ* hybridization using available matched frozen and/or archival FFPE tissue • Methods: Results will also be validated with a tissue microarray (TMA) composed of samples from ~500 patients treated at the McGill University Health Centre with 5-year follow-up information.			Months 7-12
Milestone: Identification and validation of candidate genes, pathways and interaction pairs with prognostic and/or interventional applicability.			

Task Description	Year 1	Year 2	Year 3
3. Identify and integrate stromal-epithelial miRNA (miR) signatures associated with TN breast tumors.			
3a. Profile the miR expression in tumor and normal epithelium and stroma. • Collaborator/Personnel: Dr. Nicholas Bertos / Hong Zhao • Methods: miR will be isolated from our LCM samples specified in Task 1. The concentration will be assessed and quality control performed by Nanodrop spectrophotometer and Bioanalyzer analyses. The miR expression will be profiled using the NanoString platform available at the Innovation Centre (McGill University). Reproducibility will be assessed by quantile normalization of biological replicates and the mean normalized signal from biological replicates will be used for comparative expression analysis. • PI Training: Learn how to extract miR from LCM isolates.	Months 6-12		
Milestone: Collection of miR expression profiles in tumor and normal epithelium and stroma.			
3b. Investigate miR signatures for their prognostic value by using linked patient outcome data. • Collaborator/Personnel: Dr. Michael Hallett / Sadiq Saleh • Methods: Differentially expressed miR between normal and tumor tissues (epithelium- or stroma-derived) will be identified using one-way analysis of variance (ANOVA, $p<0.5$) and hierarchical clustering with Pearson correlation using the top 50 most variably expressed miR. Differentially expressed miR between stromal or epithelial samples will be identified at a threshold of $P < 1 \times 10\text{-}5$, using the LIMMA package in Bioconductor. The miR signatures will be evaluated for their prognostic value using linked patient outcome data. • PI Training: Learn how to link miR signatures to patient outcome.		Months 6-12	
3c. Validate miR of interest. • Collaborator/Personnel: Dr. Nicholas Bertos / Hong Zhao • Methods: miR of interest will be validated via *in situ* hybridization on FFPE sections specified in Task 1. • Methods: PCR-based assays for any miR that correspond with tumor subtypes we previously identified will be established such that the miR can be used as biomarkers in TN breast cancer patients. • PI Training: Learn how to quantify miR using PCR-based tests or *in situ* hybridization.			Months 1-12
Milestone: Identification and validation of miR signatures with prognostic value.			

11.2 Figures

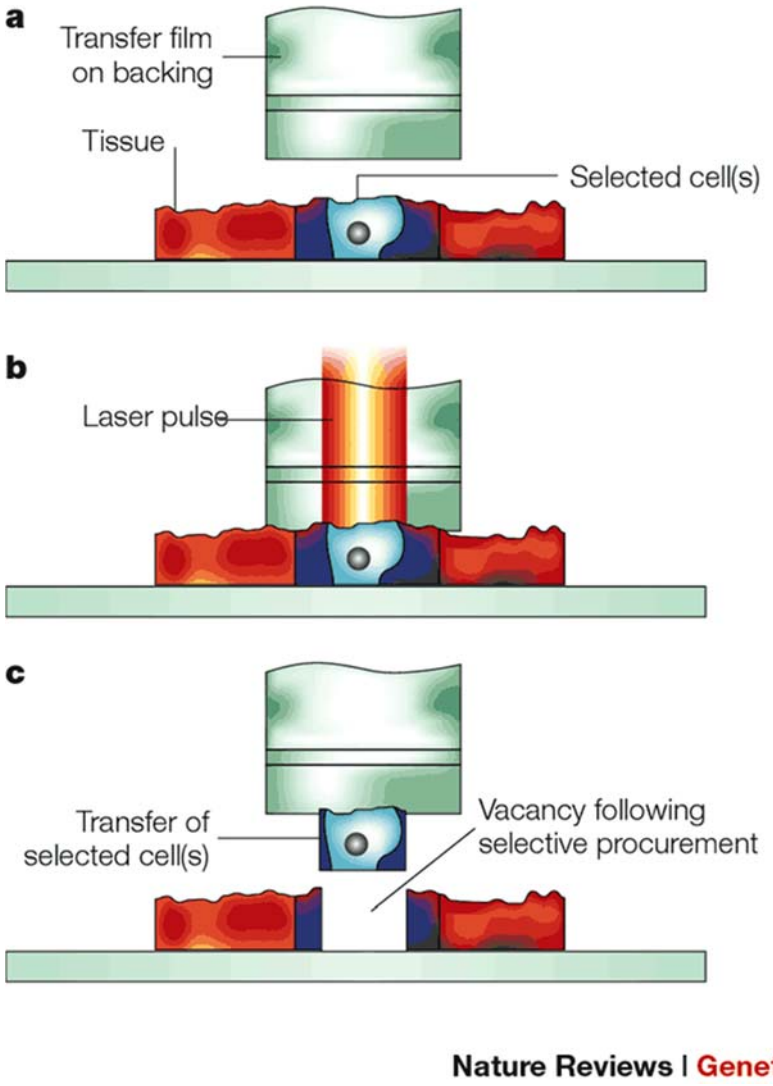

Figure 1: Laser capture microdissection (LCM) is a technology for rapid and easy procurement of a microscopic and pure cellular subpopulation away from its complex tissue milieu, under direct microscopic visualization. The starting material can be frozen, or fixed, and stained. A thin polymer film is placed in direct contact with a frozen or fixed tissue section and a laser beam activates the polymer and so transfers the selected cell(s) out of the tissue and onto the polymer film. This positive selection method is done repeatedly until all of the desired tissue is embedded onto the polymer film. An extraction buffer is applied to the polymer film so that DNA, RNA or proteins can be solubilized from the captured tissue cells. LCM fully preserves the state of the cell's molecules for quantitative analysis. Adapted from [35].

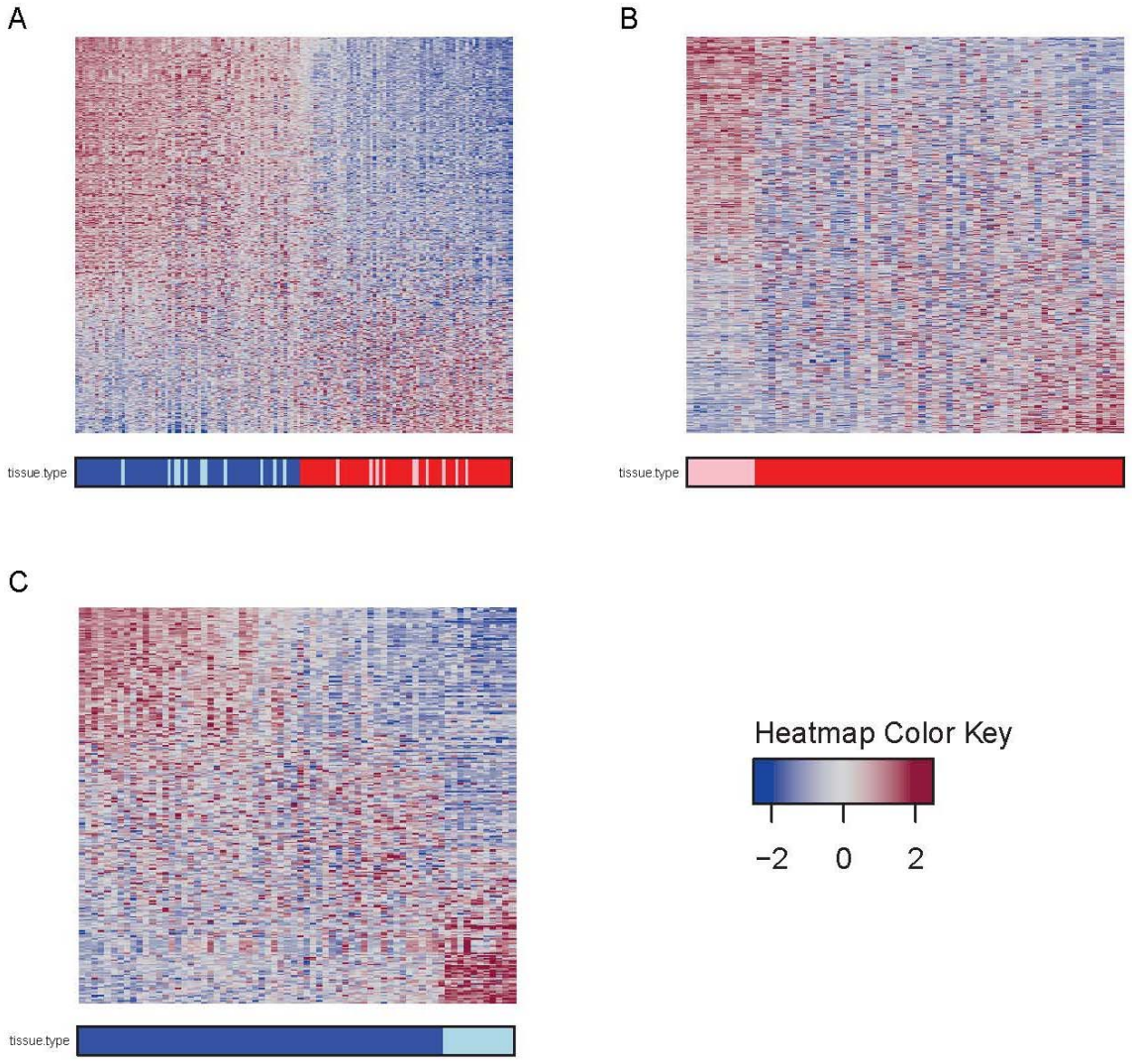

Figure 2: LCM successfully isolates distinct compartments of the tumor. Separation of the most variable genes (interquartile range, IQR > 2) unbiasedly into two opposing directions using the Partitioning Around Medoids (pam) function and subsequent ranksum ordering of gene expression profiles distinguishes epithelial from stromal tissue (A), and normal from tumor tissue (B, C). Tissue types: Red (tumor epithelium), Pink (normal epithelium), Dark blue (tumor stroma), Light blue (normal stroma). Rows represent transcripts and columns represent patient samples. Values are centered and scaled per transcript across all samples and represented by the color key.

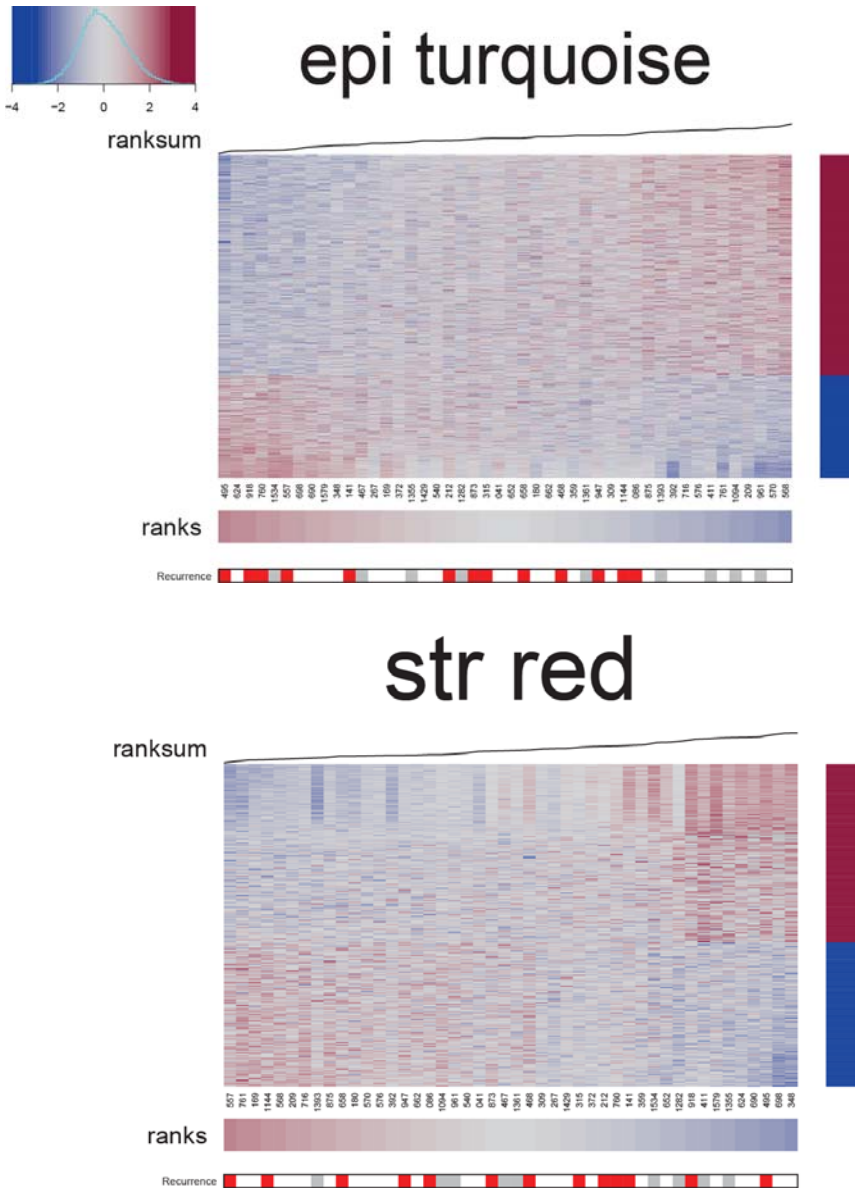

Figure 3: Examples of gene modules for epithelial and stromal components of TN tumors. Weighted Gene Correlation Network Analysis (WGCNA) [7] identified co-modulated groups of genes that had high absolute correlation in patient epithelial (epi) and stromal (str) tissues. The gene modules were given color names (*e.g.* turquoise and red) to prevent overt assumptions about the biology of the module, and they were associated with disease recurrence in the patients. Rows represent genes and columns represent patient samples. Values are centered and scaled per transcript across all samples and represented by the color key. Patients are ordered by the ranksum of module gene expression. Five-year recurrence is scored as red (disease recurrence), white (no recurrence) and grey (no recurrence, but less than 3 years of follow up).

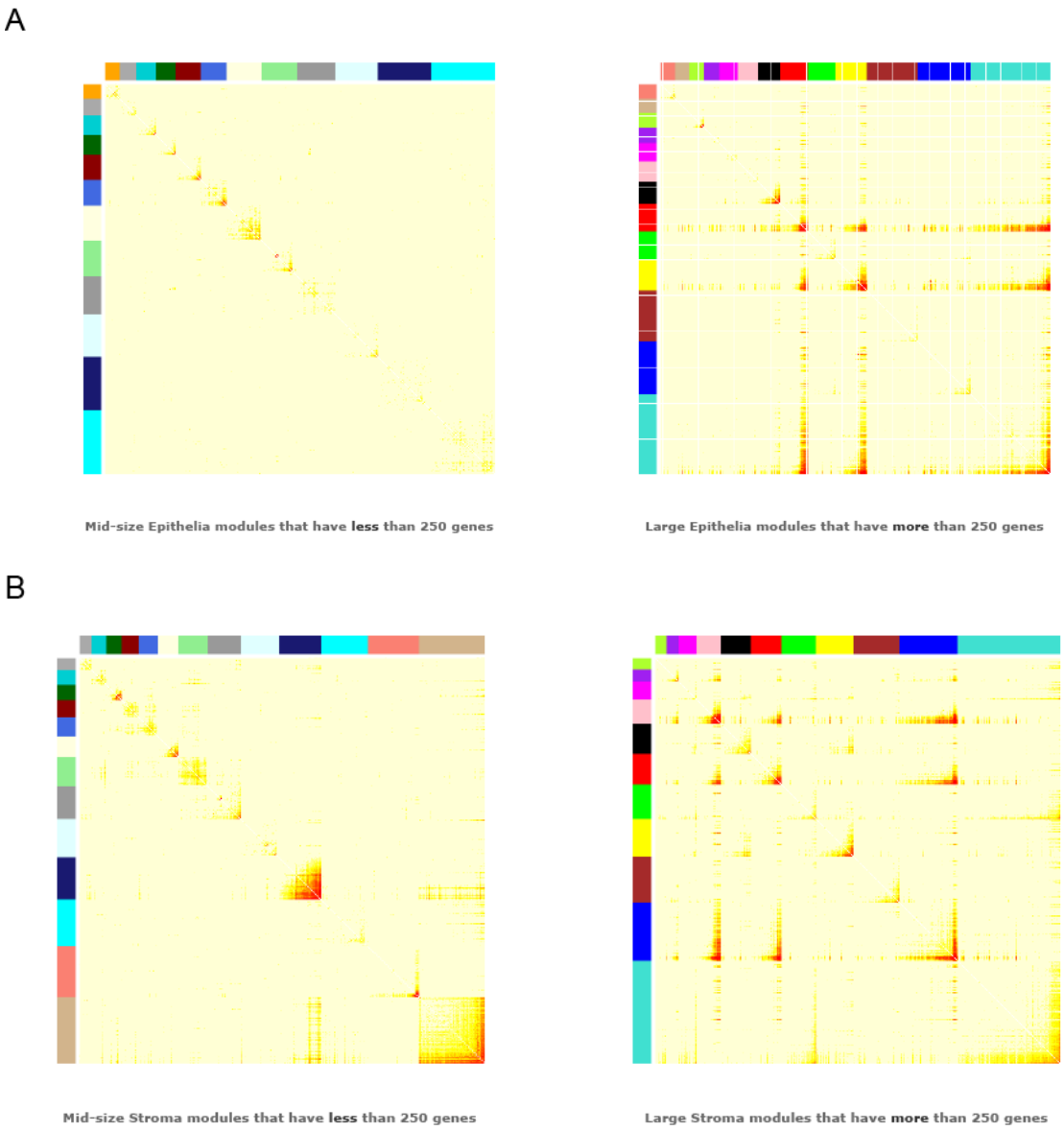

Figure 4: Intra-tissue correlation of epithelial and stromal gene modules from TN tumors. WGCNA identified 25 epithelial and 24 stroma gene modules. The gene modules are named as colors to prevent overt assumptions about the biology of the module. The figure shows correlation heatmaps of all the genes in the epithelial and stromal tissues. Genes are ordered by their modules (represented by different colors on the axes) and separated into mid-size (less than 250 genes) and large (greater than 25 genes) modules for clarity only.

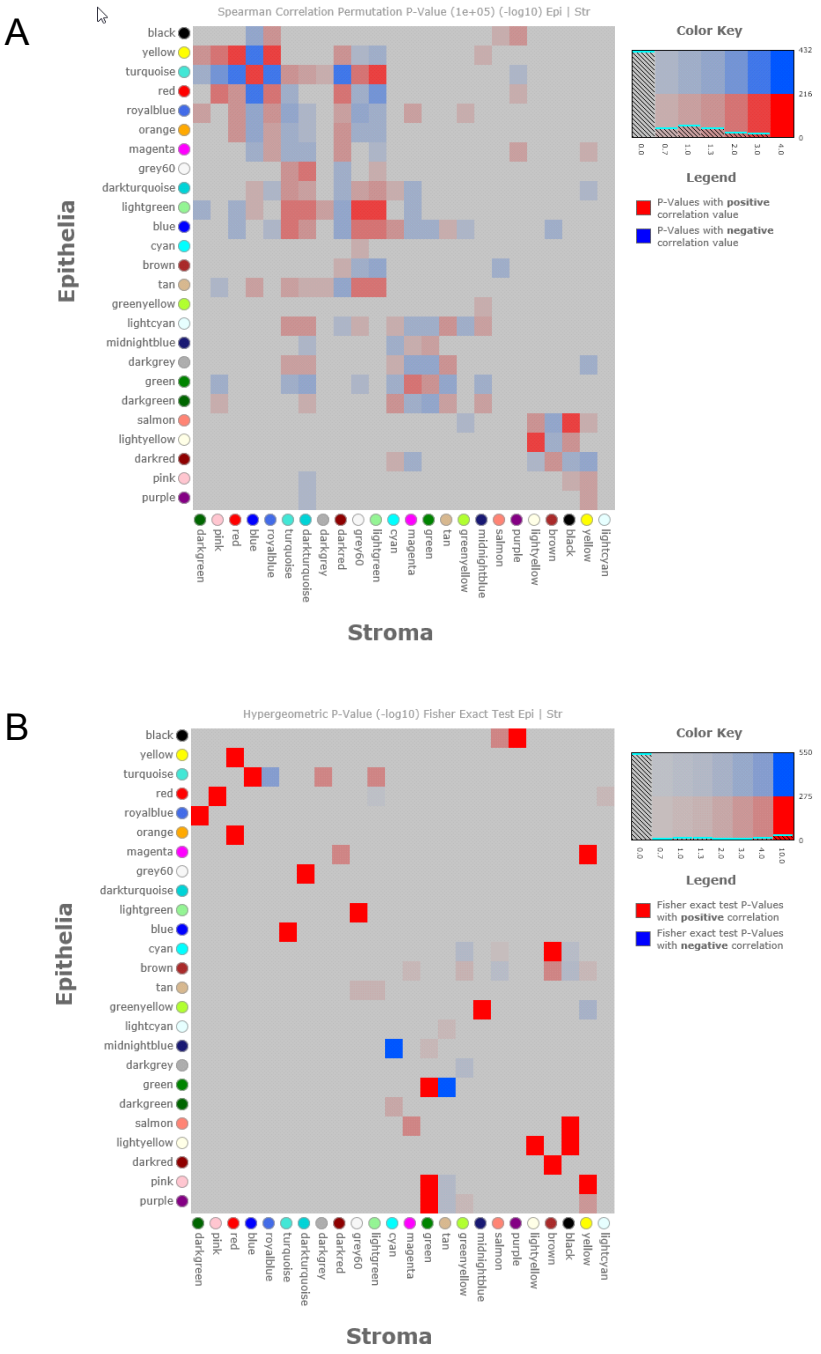

Figure 5: Inter-tissue correlation of epithelial and stromal gene modules from TN tumors. Heatmaps of the correlation of the gene modules (named as colors) between the two tissue types as determined by Spearman correlation permutation (A) and hypergeometric Fisher's exact test (B).

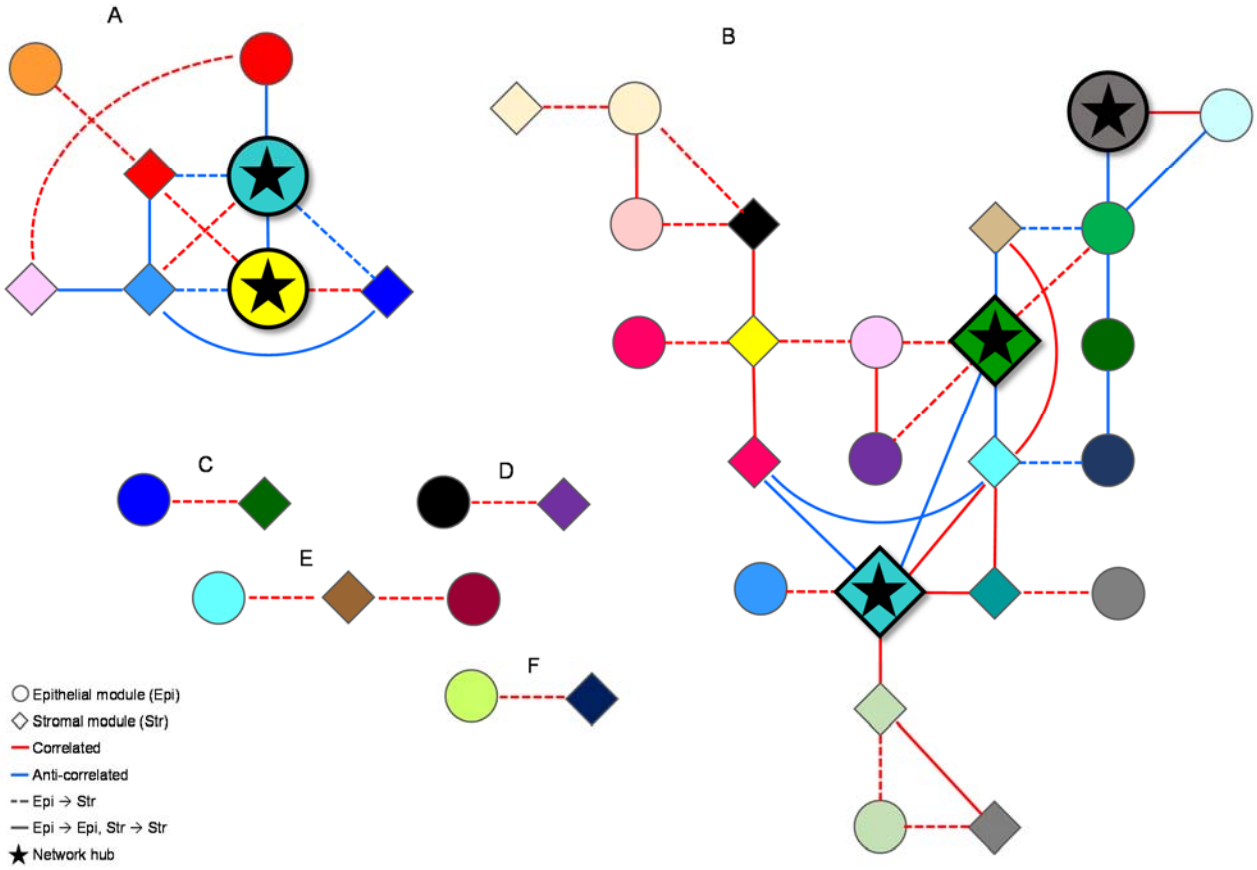

Figure 6: Correlation network of the epithelial and stromal gene modules. Map depicting the correlations between epithelial and stromal gene modules. The modules are colored according to their name. Two main networks with 5 hubs (A, B) as well as four independent epithelial-stromal interactions (C, D, E, F) are shown.

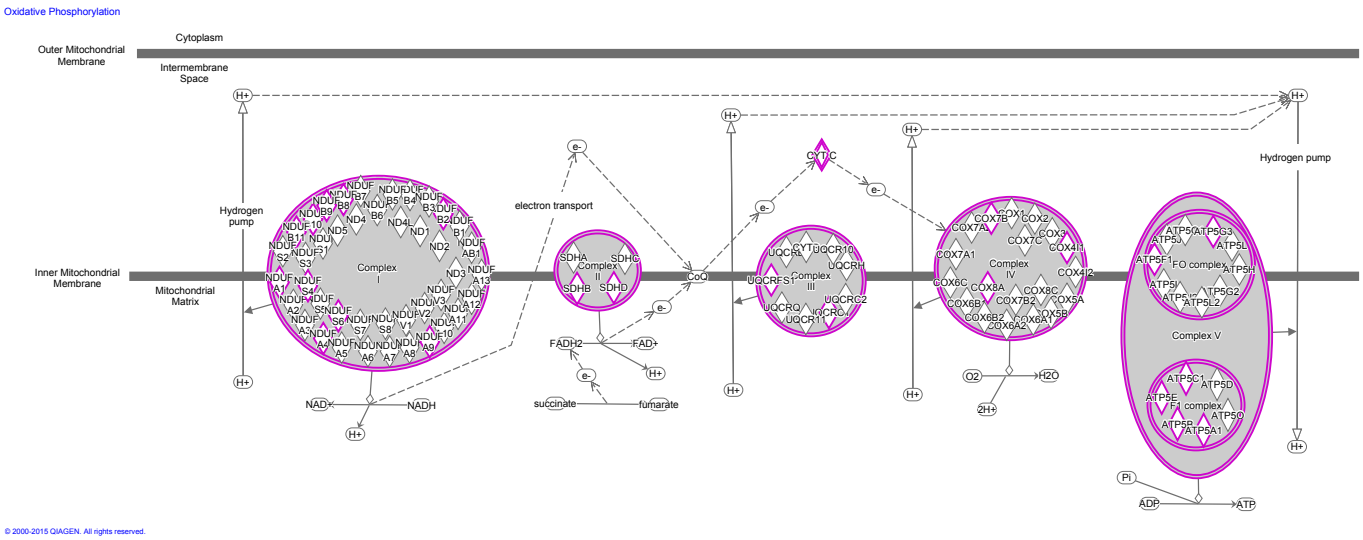

Figure 7: Epi Turquoise contains genes involved in oxidative phosphorylation. Gene set enrichment analysis of Epi Turquoise using QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) indicates that one of the canonical pathways represented by the gene list is Oxidative Phosphorylation. Genes from Epi Turquoise implicated in the pathway are outlined in pink.

11.3 Tables

Table 1: Correlation and common genes between epithelial and stromal gene modules. The correlation between epithelial (epi) and stromal (str) modules as determined by hypergeometric Fisher's exact test (p value < 10^{-10}, + positive correlation, - negative correlation) and Spearman correlation permutation (correlation > absolute 0.6, + positive correlation, - negative correlation). The total number of common genes between epithelial and stromal modules are shown as well as the % of total genes in common. We defined distinct but correlative modules as having <20% common genes (highlighted in yellow).

	Hypergeometric		Spearman		Common genes				
Epi	Str	+/-	Str	Correlation	No. common genes	Total Epi genes	Total Str genes	% Epi	% Str
black	purple	+			131	493	314	26.6	41.7
blue	turquoise	+			412	1186	2587	34.7	15.9
cyan	brown	+			38	240	1150	15.8	3.3
dark red	brown	+			44	93	1150	47.3	3.8
green	tan	-			49	635	229	7.7	21.4
green	green	+			84	635	852	13.2	9.9
green-yellow	midnight blue	+			65	337	144	19.3	45.1
grey60	dark turquoise	+			23	139	48	16.5	47.9
light green	grey60	+	grey60	0.81	33	136	110	24.3	30.0
light green			light green	0.6	1	136	103	0.7	1.0
light yellow	light yellow	+	light yellow	0.73	26	128	69	20.3	37.7
light yellow	black	+			34	128	752	26.6	4.5
magenta	yellow	+			62	406	956	15.3	6.5
midnight blue	cyan	-			19	197	160	9.6	11.9
orange	red	+			21	53	787	39.6	2.7
pink	green	+			67	446	852	15.0	7.9
pink	yellow	+			57	446	956	12.8	6.0
purple	green	+			42	350	852	12.0	4.9
red	pink	+			172	613	614	28.1	28.0
royal blue	dark green	+			26	98	54	26.5	48.1
salmon	black	+	black	0.64	137	311	752	44.1	18.2
turquoise			red	-0.63	119	1777	787	6.7	15.1
turquoise	blue	+	blue	0.75	515	1777	1475	29.0	34.9
turquoise			royal blue	-0.66	26	1777	66	1.5	39.4
yellow	red	+			245	689	787	35.6	31.1
yellow			blue	-0.67	73	689	1475	10.6	4.9
yellow			royal blue	0.6	2	689	66	0.3	3.0

Table 2: Correlation among epithelial gene modules. The correlation among epithelial modules as determined by Spearman correlation permutation (correlation > absolute 0.7, + positive correlation, - negative correlation).

Epi 1	Epi 2	Correlation
dark grey	light cyan	0.76
dark grey	green	-0.82
green	dark green	-0.85
light cyan	green	-0.73
midnight blue	dark green	-0.82
pink	purple	0.7
salmon	light yellow	0.72
turquoise	red	-0.89
yellow	turquoise	-0.81

Table 3: Correlation among stromal gene modules. The correlation among stromal modules as determined by Spearman correlation permutation (correlation > absolute 0.7, + positive correlation, - negative correlation).

Str 1	Str 2	Correlation
black	yellow	0.74
blue	royal blue	-0.75
blue	red	-0.75
blue	pink	-0.84
cyan	turquoise	0.83
cyan	dark turquoise	0.71
green	tan	-0.95
green	cyan	-0.83
green	turquoise	-0.72
grey60	light green	0.75
magenta	cyan	-0.72
magenta	turquoise	-0.75
tan	cyan	0.76
turquoise	dark turquoise	0.87
turquoise	light green	0.72
yellow	magenta	0.7