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: December 2016

TYPE OF REPORT: Final

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE
December 2016

2. REPORT TYPE
Final

3. DATES COVERED
15 Sept 2013 - 14 Sept 2016

4. TITLE AND SUBTITLE
The Impact of Epithelial-Stromal Interactions on Human Breast Tumor Heterogeneity

5a. CONTRACT NUMBER
W81XWH-13-1-0357

5b. GRANT NUMBER

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 plays a substantial role in the variability of patient response to treatment, especially in triple-negative (TN) breast cancer. A fuller understanding of the molecularly distinct TN subgroups linked to outcome is essential to promote the development of more personalized treatment strategies. The goal of this project was to identify, define and formally test critical pathways mediating tumor epithelial-stromal communication and co-dependency in TN breast cancer. Interestingly, we established that tumor heterogeneity in TN disease could be captured by stromal-specific subtypes - immune infiltration, androgen receptor signaling/invasive epithelia and desmoplastic stroma. These subtypes were associated with distant metastasis free survival, suggesting that outcome in TN breast cancer may be stromal-dependent or even stromal driven. Our project has provided the first integrated in-depth analysis of the contribution of tumor stromal processes to TN disease heterogeneity, and has positioned 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	30	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1. Introduction 4
2. Keywords 4
3. Overall Project Summary 5
   3.1 Project Objectives 5
   3.2 Results, Progress and Accomplishments 5
   3.3 Discussion 11
4. Key Research Accomplishments 13
5. Conclusion 13
6. Publications, Abstracts and Presentations 14
   6.1 Poster presentations 14
7. Inventions, Patents and Licenses 14
8. Reportable Outcomes 14
9. Other achievements 15
   9.1 Training and Professional Development 15
10. References 17
11. Appendices 21
   11.1 Statement of Work 21
   11.2 Figures 24
   11.3 Tables 30

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 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 was 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 hypothesized that by defining the tumor stromal pathways associated with poor outcome in TN tumors, we would uncover mechanisms for co-evolution, biomarkers and potential therapeutic targets. 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. Our project provided the first integrated in-depth analysis of the contribution of tumor stromal processes to TN disease heterogeneity, and has positioned the tumor microenvironment for therapeutic intervention. The results of this project also promise the "next generation" of signatures based on microRNA 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 Project Objectives

This research project had three tasks covering three 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 microRNA (miR) signatures associated with TN breast tumors.

3.2 Results, Progress and Accomplishments

3.2.1 Development of stromal mRNA expression signatures for TN tumors

The first task outlined in our project proposal was to develop coordinate stromal-epithelial mRNA expression signatures for TN tumors (refer to Statement of Work (SOW) in Appendix 1). To accomplish this task, the first step was to isolate epithelial and stromal tissues from TN breast tumor samples. My mentor, Dr. Morag Park, established the Breast Cancer Functional Genomics Group (BCFGG) in 1999. This group has banked fresh-frozen breast cancer tumor (approx. 700) and normal (approx. 500 including matched samples and reduction mammoplasties) tissue samples obtained from surgeries conducted at the McGill University Health Centre (MUHC) 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 immunohistochemistry and *in situ* hybridization analyses. An attending clinical pathologist specializing in breast pathology rescores all banked samples for consistency. HER2 fluorescence *in situ* hybridization (FISH) 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. Using this valuable resource, tumor epithelial and stromal tissues were isolated via laser capture microdissection (LCM, see Figure 1 in Appendix 2) from *ca*. 50 TN patient samples. In addition, adjacent normal epithelial and normal stromal tissues were isolated. LCM was performed as previously described by our group [6]. Briefly, human breast tumor tissue collected from

consenting patients at primary surgery was snap-frozen in Tissue-Tek O.C.T. (Sakura) and stored in liquid nitrogen. Blocks were sectioned on a cryostat as 5 µm sections and reviewed by an experienced attending pathologist specializing in breast cancer to identify representative regions of tumor epithelium (TE), tumor-associated stroma (TS), histologically normal epithelium (NE) and histologically normal stroma (NS) (the latter two distal from the tumor). Sections (10 µm) were cut and stained using the Arcturus Histogene LCM Frozen Section Staining Kit (Life Technologies) and representative areas were isolated by LCM using an Arcturus PixCell IIe instrument (Life Technologies). All collection was performed within 30 minutes of placing the slide on the LCM stage.

Once the epithelial and stromal compartments had been isolated from banked tumor samples, the next step was to extract RNA for microarray-based gene expression profiling. 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]. Briefly, material from LCM caps for each sample compartment was pooled and RNA isolated using the Arcturus PicoPure RNA Isolation Kit (Life Technologies) according to the manufacturer's directions. Following quantification of yield (Nanodrop spectrophotometer) and quality control analysis (Agilent Bioanalyzer), samples judged as of sufficient quantity and quality were subjected to 2 rounds of amplification using the Arcturus RiboAmp HS Plus kit (Life Technologies) according to the manufacturer's directions. Resulting amplified RNA was subjected to quality control assay (Agilent Bioanalyzer), labelled with Cy3, and hybridized to Agilent SurePrint G3 8x60K Human Gene Expression arrays together with a Cy5-labelled common reference. Hybridization and washing were carried out according to the manufacturer's directions. Subsequently, arrays were scanned on an Agilent Microarray Scanner and feature-extracted using Agilent FE software.

To verify 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, our next goal was to identify stromal subclasses of TN tumors (SOW, Appendix 1). Initially, 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. As an alternate approach, we classified groups of genes (gene modules) by the degree to which they co-varied across patient samples. This method, referred to as Weighted Gene Correlation Network Analysis (WGCNA) [6], identified co-modulated groups of genes that had high absolute correlation in patient stromal and epithelial tissues. However, the large number of gene groups and high degree of covariance across these groups due to shared genes necessitated an alternate approach. Therefore, we subjected the most variable genes in TN tumor stromal samples (IQR > 2, n=211 genes) to hierarchical clustering (Ward's algorithm, Pearson correlation distance). Four distinct clusters (named teal, orange, magenta and purple) were observed that contained a significant number of genes with strong pairwise gene-gene correlations of expression (Figure 3A, Appendix 2). These clusters were statistically stable and reproducible (pvclust, AU > 85%). Genes within each cluster that exhibited strong co-expression across the patient cohort were considered to be characteristic gene set.

To measure the level of expression of the gene clusters (termed "stromal properties") in TN tumors, patients were linearly ranked based on the overall amount of observed expression of the characteristic genes for each stromal property independently. A rank-based permutation test (termed ROI_{95}, Paquet *et al.*, *manuscript in preparation*) was applied to each linear ordering to estimate boundaries of regions that delineate samples that are low, intermediate or high for the characteristic gene set (Figure 3B, Appendix 2). Thus, each patient sample was independently measured for each of the four ternary properties (low, medium, high) such that patients could be high for multiple stromal properties (Figure 3C, Appendix 2). This approach differed from traditional subtyping approaches that partition the patient cohort into distinct, non-overlapping subtypes.

We then wanted to characterize the molecular pathways and/or presence of specific cell types associated with each stromal property. Thus, we identified differentially expressed genes between

patients deemed low versus those deemed high for each stromal property by fitting a linear model to each stromal property using the R package limma [7] corrected with Benjamini-Hochberg ($p < 0.05$). Differentially expressed genes lists were examined using QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) and compared against the Molecular Signatures Database (MSigDB) for pathway analysis. This analysis revealed that differentially expressed genes in the "teal" property included keratins (KRT6B and KRT23) and metallothioneins (Table 1, Appendix 3). Because these genes are expressed by tumor epithelial cells [8], this tumor property could represent invasive tumor cells that have retained some of their epithelial characteristics due to tumor plasticity [9]. The "orange" stromal property included multiple collagens (collagens 1A1/2, 3A1, 5A1/2, 8A1/2, 10A1, 12A1, 16A1), platelet-derived growth factor receptor-β (PDGFRB), fibroblast activation protein-α (FAP), and collagen stabilizing/modifying enzymes (Table 1, Appendix 3). All of these are factors associated with a desmoplastic reaction [10]. The differentially expressed genes in the "magenta" property included B-cell markers (CD19, CD79A, CD72), immunoglobulins (IGLL5, IGLL1, IGJ), and transcription factors associated with B-cell activation (POU2AF1, XBP1) (Table 1, Appendix 3). Finally, the differentially expressed genes in the "purple" stromal property included general (CD2, CD3D, IL-2Rα IL-2Rβ, IL-2Rγ), as well as lineage-specific (CD4, CD8A, CD8B) T cell-associated markers, and markers of a Th1-mediated anti-tumor response including IL-15 [11], granzymes (GZMBA, GZMB, GZMK, GZMH) [12], markers of an interferon response (IFI30, IFIT5) [13], transcription factors involved in Th1 differentiation (STAT1, STAT4) [14], and TNFα-induced genes (TNFAIP2, TNFAIP8) [15] (Table 1, Appendix 3). On the basis of these observations, the four stromal properties were labelled:

- E = invasive epithelial cells (teal)
- D = desmoplastic stroma (orange)
- B = B cell (magenta)
- T = T cell (purple)

Having accomplished all of our objectives to define the stromal properties of TN breast cancer, the final goal in our first project task was to use the stromal gene clusters to identify corresponding tumor epithelial gene signatures. Unfortunately, no statistically significant epithelial properties could be defined or correlated with the stromal gene expression properties, emphasizing the high heterogeneity of the tumor epithelium in TN breast cancer. This is consistent with previous reports that TN breast cancer has higher levels of inter-tumoral (patient-to-patient) heterogeneity than other breast cancer subtypes with respect to both gene expression [16], and somatic genomic aberrations [17, 18]. Despite the absence of

a strong correlation between the epithelial and stromal gene expression clusters, we determined that our four TN stromal properties were associated with patient outcome (see section 3.2.2) consistent with our previous work which demonstrated that gene expression signatures in human stroma were sufficiently powerful to predict the outcome of breast cancer patients independently of clinical parameters and molecular subtypes [5].

3.2.2 Stromal properties capture TN heterogeneity and associate with patient outcome

The second task outlined in our project proposal was to identify stromal-epithelial gene interaction networks with the ultimate goal of identifying candidate genes with prognostic and/or interventional applicability (SOW, Appendix 1). In the absence of statistically significant epithelial properties that could be defined or correlated with the stromal gene expression properties (as mentioned in section 3.2.1), we questioned instead how our stromal properties would relate to published subtypes of TN breast cancer derived from bulk tumor gene expression profiles (*i.e.* combined epithelial and stromal gene signatures). Using bulk gene expression data (rather than separate epithelial and stromal gene expression as we have done), Lehmann *et al.* [19] defined six TN subtypes – two basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem-like (MSL), and a luminal androgen receptor (LAR) subtype. We subjected the gene sets of the six Lehmann subtypes to our methodology, estimating their activation as either low, intermediate or high across the TN breast cancer compendium. This method rendered the Lehmann groups in a format for direct comparison with our four stromal properties using Cohen's kappa statistic (fmsb package version 0.5.1). This analysis revealed that our T and B stromal properties captured the inversely-correlated M and IM properties ($p < 1e-10$; Figure 4A&B, Appendix 2), whereas our stromal E property exhibited strong correlation with BL1 and anti-correlation with LAR ($p < 1e-10$; Figure 4C, Appendix 2). Patient samples estimated to be high for the D property were almost always estimated high for the BL2, LAR, and MSL properties and low for the BL1 property ($p < 1e-10$; Figure 4D, Appendix 2). These observations highlighted that the Lehmann TN groups are strongly associated with our stromal properties, and suggests that TN heterogeneity can be succinctly summarized by three distinct (and possibly stromal-dominant) properties related to immune infiltration (B and T), androgen receptor signalling/epithelia (E), and a desmoplastic stroma (D).

Knowing that our stromal properties seemed to accurately distinguish molecular subtypes of TN breast cancer, we wanted to test their association with patient outcome as proposed in our SOW (Appendix 1). Due to the unavailability of TN breast cancer stromal datasets, we developed and tested a statistical method to estimate the status of each stromal property in bulk expression data. This method

was applied to a large cohort of TN patient samples (n=1098) selected from 13 individual non-overlapping publicly available breast cancer datasets [16]. Stromal property assignments were computed independently per dataset, and pooled across the constituent datasets. This enabled us to test if the low, intermediate and high partitions of each property stratified patients by clinical outcome. While the D property (orange) did not demonstrate significant association with outcome, the T, B and E properties (purple, magenta and teal) were significantly correlated with outcome (log-rank test, distant metastasis free survival (DMFS) at 5 years all $p < 0.05$; Figure 5, Appendix 2). This demonstrates that that the T, B and E properties of the stroma inform on clinical outcome for TN patients. We are now in the process of validating outcome predictors by IHC using matched archival FFPE tissue.

3.2.3 Identification and integration of stromal-epithelial miRNA signatures associated with TN breast tumors

The final task outlined in our project proposal was to identify and integrate stromal-epithelial miRNA (miR) signatures associated with TN breast tumors (SOW, Appendix 1). The first step in this objective was to profile the miR expression in normal and tumor epithelia and stroma. We initially proposed to profile the miR expression using the NanoString platform available at the Innovation Centre (McGill University), however, due to technical difficulties, we chose an alternate platform, *i.e.* TaqMan LDA plate assays at the Institute for Research in Immunology and Cancer (IRIC) at Université de Montreal. Despite delays in our analysis, the new platform was successful.

In order to prepare the miR, material from LCM caps for each sample compartment was pooled and resuspended in 300 µL RLT buffer before being loaded on a Qiashredder column (Qiagen) and centrifuged at 14 000 rpm for 2 minutes. Flowthrough was loaded on a Qiagen AllPrep DNA spin column and centrifuged at 10 000 rpm for 30 seconds. Flowthrough was combined with 30 µL 2 M sodium acetate pH 4.0, 330 µL water-saturated phenol and 90 µL chloroform-isoamyl alcohol (23:1). After vortexing, the mixture was incubated on ice for 15 minutes and centrifuged at 12 000 rpm for 15 minutes. The upper phase (200 µL) was transferred to a new tube and 1.5 µL GlycoBlue (Ambion; resuspended at 100 µg/mL in isopropanol) and 200 µL isopropanol were added. After mixing by inversion (10x), the mixture was incubated at -80°C overnight, then centrifuged for 30 minutes at 4°C (12 000 rpm) to pellet RNA. Pellets were washed twice with 400 µL ice-cold 75% ethanol and air-dried for 15 minutes. Air-dried pellets were resuspended in 10 µL ddH20 then thoroughly combined with 250 µL RLT buffer. Ethanol (390 µL of 100%, equating to 1.5 volumes) was added and mixed by pipetting. The entire mixture was loaded onto a RNeasy MinElute Spin Column (Qiagen) in a 2 mL collection tube and centrifuged for 15 seconds at 10 000 rpm.

The column was washed twice with 500 µL Buffer RPE (Qiagen) and dried by centrifugation at 14 000 rpm for 5 minutes. RNA was eluted from the column with 20 µL RNAse-free ddH20. The RNA was re-applied to the column and centrifuged again at 14 000 rpm for 1 minute to elute any remaining RNA. Extracted RNA was quantified using a spectrophotometer (Nanodrop) and subjected to BioAnalyzer to assay for quality (Agilent Technologies). Total RNA (150-200 ng) was subjected to a pre-amplification step using the TaqMan MegaPlex PreAmp primer pool and the pre-amplified products were assayed for miR levels using TaqMan LDA 384-well plates (Pools A and B) on an ABI 7900HT Fast Real-Time system.

Relative miR expression was determined as per Puigdecanet *et al.* [20]. Briefly, cycle threshold (C_t) values were calculated and relative gene expression levels were expressed as the difference in C_t values (ΔC_t) of the target gene and the geometric mean of the housekeeping genes. $\Delta\Delta C_t$ values were calculated for each sample using the mean of its ΔC_t subtracted from the mean ΔC_t value measured in the calibrator. Gene expression quantification was achieved using the comparative C_t method for relative quantification, in which the amount of target is expressed as $2^{-\Delta\Delta Ct}$. ΔC_t, $\Delta\Delta C_t$ and $2^{-\Delta\Delta Ct}$ were determined using the "HTqPCR" R package, and batch effects were corrected using ComBat function implemented in the sva package.

Having achieved our milestone objective of collecting miR expression profiles from TN tumor and normal epithelia and stroma, we wanted to investigate the miR signatures for their prognostic value using linked patient outcome data (SOW, Appendix 1). Given our previous observations that the stromal properties T, B and E inform on clinical outcome for TN patients (see section 3.2.2), we clustered the miR results for tumor versus normal epithelia and stroma, and compared the output to relative T, B and E gene expression for each patient (Heatmap complete linkage for clustering and Euclidean for the distance using ComplexHeatmap R Bioconductor package; Figure 6, Appendix 2). Because no clear association between miR clusters and stromal properties was observed using this method, we are now using alternate methods to identify and validate miR of interest based on their prognostic value.

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 [21-24] 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 [23, 25, 26]. However, overall outcome remains poor in TN disease [23] and no mechanisms exist to determine which patients will respond to chemotherapy. We proposed that interrogating the gene expression of the epithelium and surrounding stroma in TN tumors would 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 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. This study represents the first large-scale effort to investigate the tumor microenvironment specifically in TN patients. Previous studies have focused on gene expression profiling [19, 27, 28] or DNA sequencing [29] of bulk material enriched for tumor epithelial cells in TN breast cancer. Efforts to study the tumor microenvironment, including our own [5, 30, 31] have used LCM to isolate stromal elements across all breast tumors, *i.e.* not restricted to TN. By looking specifically at TN breast stroma, we identified four properties by gene expression – T cell enriched (T), B cell enriched (B), invasive epithelial cell (E) and desmoplastic stroma (D). Despite being discovered in LCM-derived material, these stromal properties are consistent even when applied to bulk tumor gene expression profiles. In addition, the T, B and E properties associate with patient survival. This strongly suggests that TN heterogeneity can be succinctly summarized by three distinct properties related to immune infiltration (B and T), androgen receptor signalling/epithelia (E) and a desmoplastic stroma (D). Moreover, the ability of these properties to stratify large combined epithelial-stromal gene expression cohorts, and our inability to correlate our stromal gene expression with epithelial gene expression, indicates that patient prognosis in TN breast cancer may stromal-dominant, or even, stromal-driven. This is very exciting and ongoing work in our lab is dedicated to validating these results.

Another focus of this project was the stromal and epithelial miR expression profiles of TN breast cancer. Changes in miR expression have been documented in breast cancer [32-36], and several of these have been shown to be associated with clinical features [32, 37-42] including response to therapy [43-46]. However, little is known regarding the prognostic value of miR sets in tumor stroma, particularly in TN breast cancer. We wanted 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, now that we have the normalized miR data and have begun analysis, we expect to be able to continue our work and identify and integrate stromal-epithelial miRNA (miR) signatures associated with prognosis in TN breast tumors.

4. Key Research Accomplishments

We achieved the following key research accomplishments in this project:

- Identified four stromal-specific properties in TN breast cancer by gene expression: T cell enriched (T), B cell enriched (B), androgen receptor/invasive epithelial cell (E) and desmoplastic stroma (D)
- Determined that the stromal properties were associated with patient survival and therefore, could provide biomarkers of therapeutic intervention or outcome
- Profiled the miR expression of tumor and normal epithelia and stroma to examine tumor-associated miR expression changes responsible for patient outcome

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 was to identify, define and formally test critical pathways mediating tumor epithelial-stromal communication and co-dependency in TN breast cancer. Interestingly, we established that tumor heterogeneity in TN disease could be captured by stromal-specific subtypes that were associated with distant metastasis free survival. Unlike traditional subtyping approaches that partition the patient cohort into distinct non-overlapping subtypes, our stromal properties were not exclusive and patients could belong to multiple groups concurrently. We believe this method of stratification more accurately describes TN heterogeneity and patient prognosis, and our results suggest that outcome in TN breast cancer may be stromal-dependent or even stromal driven. Our project has provided the first

integrated in-depth analysis of the contribution of tumor stromal processes to TN disease heterogeneity, and has positioned the tumor microenvironment for therapeutic intervention.

6. Publications, Abstracts and Presentations

6.1 Poster presentations

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.

C. Thompson, S.M. Saleh, N. Bertos, M. Gigoux, T. Gruosso, M. Souleimanova, H. Zhao, M.T. Hallett & M. Park. Novel prognostic stromal subtypes in triple negative breast cancer. American Association for Cancer Research annual meeting, New Orleans LA, April 2016.

7. Inventions, Patents and Licenses

Nothing to report.

8. Reportable Outcomes

During this project, we successfully profiled the mRNA and miR of tumor epithelia and stroma in TN breast cancer. We identified four stromal-specific properties, *i.e.* T cell enriched, B cell enriched, androgen receptor/invasive epithelial cell and desmoplastic stroma, that are associated with patient survival. Unlike traditional subtyping approaches that partition the patient cohort into distinct non-overlapping subtypes, our stromal properties were not exclusive and patients could belong to multiple groups concurrently. This method of stratification more accurately describes TN heterogeneity and patient prognosis, and suggests that outcome in TN breast cancer may be stromal-dependent or even stromal driven. Our project has provided the first integrated in-depth analysis of the contribution of tumor stromal processes to TN disease heterogeneity, and has positioned the tumor microenvironment for therapeutic intervention.

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 course of this project. My collaborators, Dr. Nicholas Bertos and Dr. Hong Zhao, are key members of the BCFGG with experience in tissue banking, tissue microdissection, expression profiling and target validation. With their guidance, I have learned how to perform LCM, how to extract RNA and miR from LCM isolates, and how to perform microarray-based gene profiling as proposed in the Statement of Work (Appendix 1). 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 training 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.

Trainees at the Goodman Cancer Research Centre, through the McGill Integrated Cancer Research Training Program (MICRTP) as well as the Systems Biology Training Program, have access to workshops such as development of hypothesis and grant writing, time management, effective oral and visual communication, advanced statistical analysis, ethics, knowledge translation and bioinformatics. Through this program, I attended a course in the bioinformatics programming language, R. This was very valuable as it facilitated my understanding of the gene expression and miR profile data, and help me to communicate more effectively with the bioinformatics personnel on this project.

Throughout the course of this project, I have participated in the training and mentorship of undergraduate and graduate students in Dr. Park's lab. This has been 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.

I have also had the opportunity to attend international conferences and workshops. In November 2013, I attended the Translational Cancer Research for Basic Scientists Workshop offered by the American Association for Cancer Research (AACR) in Boston MA. This workshop covered topics such as diagnostics, clinical trials, regulatory requirements, personalized medicine and translational collaborations. In addition to lectures and small group discussions, this workshop offered the unique opportunity to observe and interact with health professionals in various clinical-related settings. All participants visited a surgical pathology laboratory, a diagnostic radiology laboratory, patient clinics, and an Institutional Review Board (IRB) meeting at Massachusetts General Hospital or the Dana-Farber Cancer Institute. These on-site sessions included shadowing doctors meeting with their patients. In order to participate in this workshop, trainees had to pass an online course in ethics (working with human subjects) offered by the Dana-Farber Cancer Institute. This was very informative and relevant to my project as I am working with patient samples.

I also 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.

Finally, I attended the AACR annual meeting in April 2016 in New Orleans, LA. This meeting is regarded as one of the largest cancer research meetings and, as such, is known for the vast number of scientific sessions, educational sessions, methods workshops, career fair, professional advancement meetings, exhibits and posters. The topics range from oncogenes to heterogeneity to systems biology to regulatory science offering a dynamic and informative environment to expand one's vision and consider research in a different or bigger context. I attended several educational workshops, *e.g.* "Cancer Metabolism and Immunometabolism", and presented a poster entitled "Novel prognostic stromal

subtypes in triple negative breast cancer". In addition to answering questions from people who were interested in my poster, I was able to see the related posters during my session about the pro-tumorigenic microenvironment. I also spoke with some of the exhibitors, including Cell Signaling and Abcam, about antibodies for immunohistochemistry which is something we are actively performing on slides from our patient samples to validate our *in silico* results. Overall, I found the conference to be very interesting, and it has definitely shaped my perspective on my project and its future direction.

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. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.

7. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

8. Tai SK, Tan OJ, Chow VT, et al. Differential expression of metallothionein 1 and 2 isoforms in breast cancer lines with different invasive potential: identification of a novel nonsilent metallothionein-1H mutant variant. Am J Pathol 2003;163:2009-19.

9. Cheung KJ, Ewald AJ. A collective route to metastasis: Seeding by tumor cell clusters. Science 2016;352:167-9.

10. Gandellini P, Andriani F, Merlino G, D'Aiuto F, Roz L, Callari M. Complexity in the tumour microenvironment: Cancer associated fibroblast gene expression patterns identify both common and unique features of tumour-stroma crosstalk across cancer types. Semin Cancer Biol 2015;35:96-106.

11. Jabri B, Abadie V. IL-15 functions as a danger signal to regulate tissue-resident T cells and tissue destruction. Nat Rev Immunol 2015;15:771-83.

12. Cullen SP, Brunet M, Martin SJ. Granzymes in cancer and immunity. Cell Death Differ 2010;17:616-23.

13. Zitvogel L, Galluzzi L, Kepp O, Smyth MJ, Kroemer G. Type I interferons in anticancer immunity. Nat Rev Immunol 2015;15:405-14.

14. Takeda K, Akira S. STAT family of transcription factors in cytokine-mediated biological responses. Cytokine Growth Factor Rev 2000;11:199-207.

15. Kumar D, Whiteside TL, Kasid U. Identification of a novel tumor necrosis factor-alpha-inducible gene, SCC-S2, containing the consensus sequence of a death effector domain of fas-associated death domain-like interleukin- 1beta-converting enzyme-inhibitory protein. J Biol Chem 2000;275:2973-8.

16. Tofigh A, Suderman M, Paquet ER, et al. The prognostic ease and difficulty of invasive breast carcinoma. Cell Rep 2014;9:129-42.

17. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61-70.

18. Bergamaschi A, Kim YH, Wang P, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 2006;45:1033-40.

19. 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.

20. Puigdecanet E, Espinet B, Lozano JJ, et al. Gene expression profiling distinguishes JAK2V617F-negative from JAK2V617F-positive patients in essential thrombocythemia. Leukemia 2008;22:1368-76.

21. 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.

22. 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.

23. 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.

24. 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.

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

26. 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.

27. Rody A, Karn T, Liedtke C, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res 2011;13:R97.

28. Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res 2015;21:1688-98.

29. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395-9.

30. 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.

31. Pepin F, Bertos N, Laferriere J, et al. Gene expression profiling of microdissected breast cancer microvasculature identifies distinct tumor vascular subtypes. Breast Cancer Res 2012;14:R120.

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

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

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

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

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

48. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160-7.

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.			

<table>
<tr><th>Task Description</th><th>Year 1</th><th>Year 2</th><th>Year 3</th></tr>
<tr><td colspan="4">2. Identify stromal-epithelial gene interaction networks.</td></tr>
<tr><td>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.</td><td></td><td>Months 6-12</td><td>Months 1-3</td></tr>
<tr><td colspan="4">Milestone: Development of a new bioinformatics tool STR-EPI to identify stromal-epithelial gene signatures.</td></tr>
<tr><td>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.</td><td></td><td></td><td>Months 3-6</td></tr>
<tr><td>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.</td><td></td><td></td><td>Months 7-12</td></tr>
<tr><td colspan="4">Milestone: Identification and validation of candidate genes, pathways and interaction pairs with prognostic and/or interventional applicability.</td></tr>
</table>

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 $P < 1 \times 10$-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

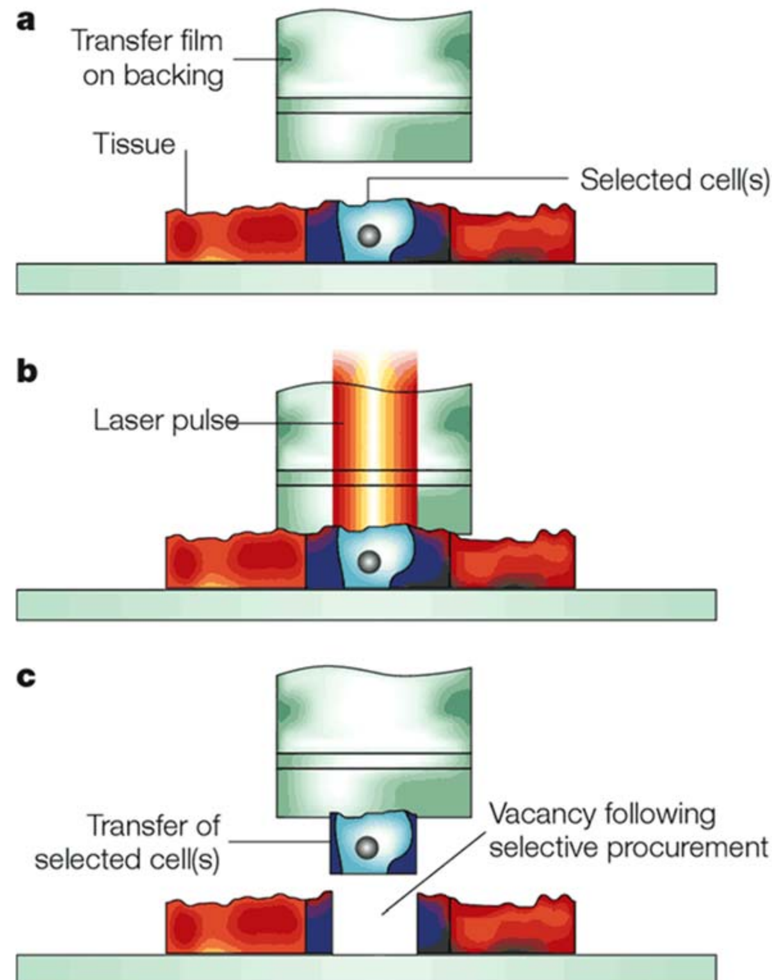


Figure 1: Laser capture microdissection is a technology for rapid and easy procurement of a microscopic and pure cellular subpopulation away from its complex tissue milieu, under direct microscopic visualization. The starting material can be frozen, or fixed, and stained. A thin polymer film is placed 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 [47].

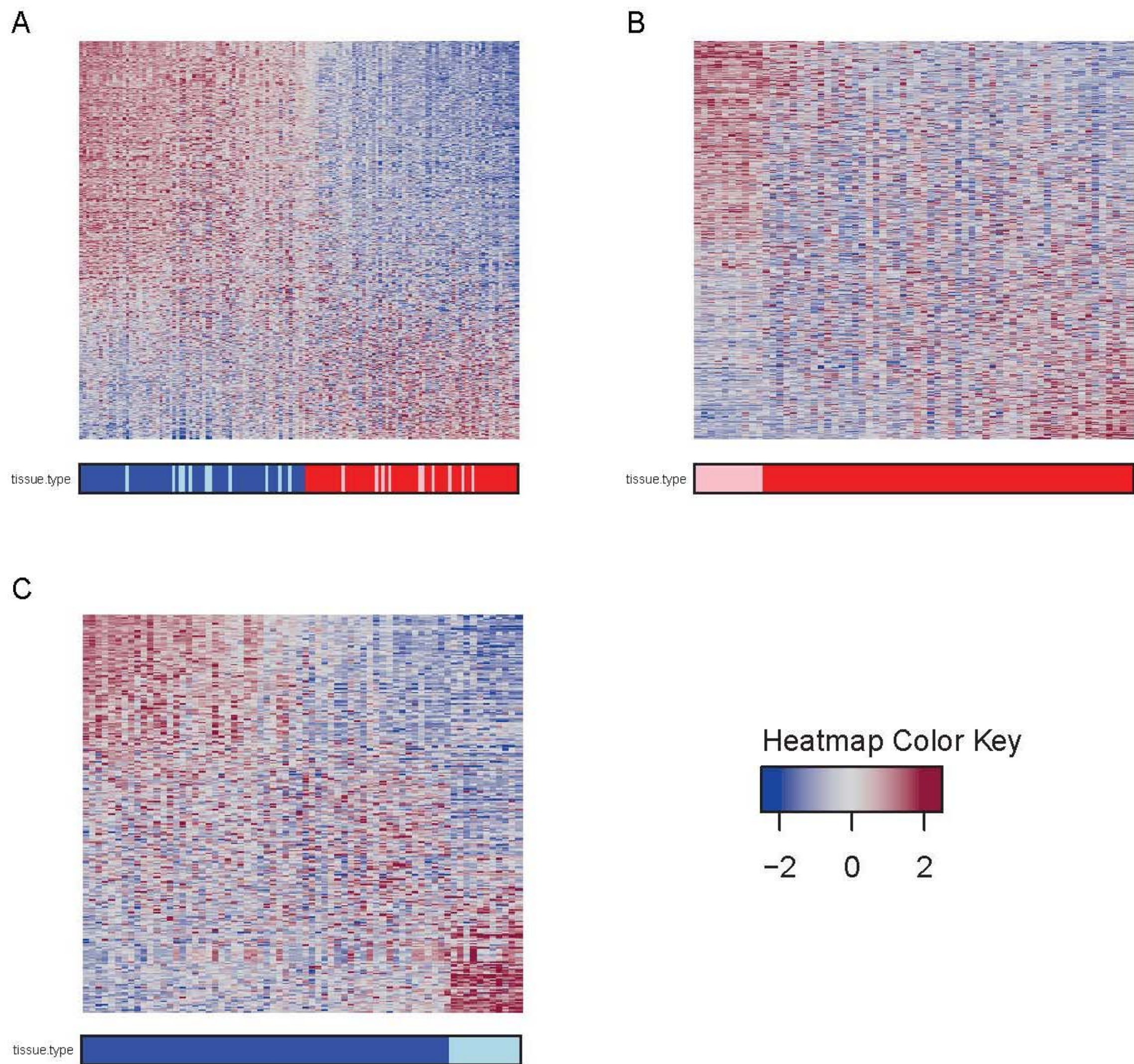


Figure 2: LCM successfully isolates distinct compartments of the tumor. Separation of the most variable genes (IQR > 2) unbiasedly into two opposing directions using the Partitioning Around Medoids function and subsequent ranksum ordering of gene expression profiles distinguishes epithelial from stromal 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.

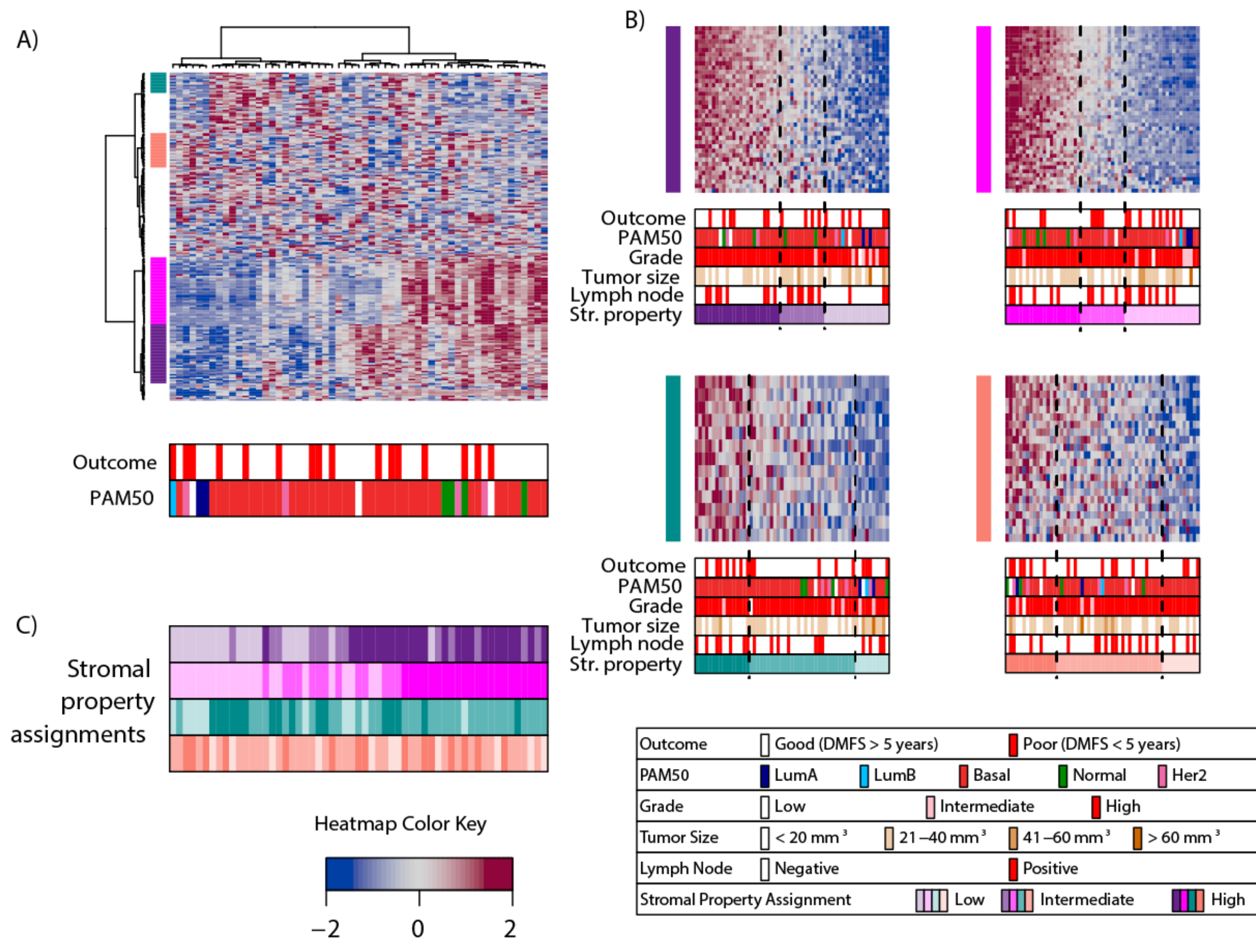
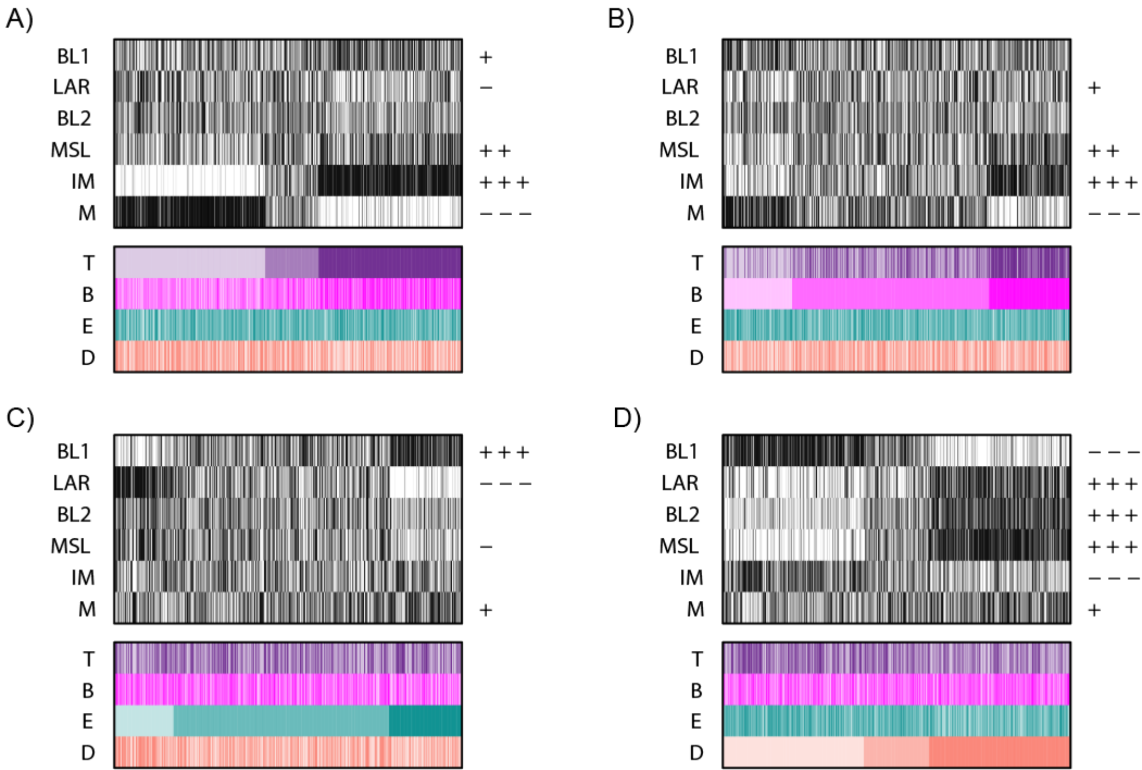


Figure 3: Hierarchical clustering identifies four stromal gene clusters in TN tumors. A) Hierarchal clustering of tumor stromal gene expression profiles using genes with IQR > 2. Stable clusters with AU > 0.85 and > 12 genes are indicated by colored bars at left (teal, orange, magenta, purple). B) Assignment of samples into 3 classes (high, intermediate, or low) for each property using ROI_{95} (classes demarcated by dashed lines in heatmaps). Patients with the smallest sum of expression are ranked lowest and depicted in lightest color (at right) and those with the largest sum are ranked highest and depicted in the darkest color (at left). Vertical colored bars at left of each heatmap correspond with the color assigned to samples high for that subtype. C) Relationships between the assignments for each stromal property. Patient rankings for each cluster are denoted by colors as in panel B. Note that samples can be high for multiple stromal properties. For all heatmaps – rows, transcripts; columns, samples; values are centered and scaled per transcript across all samples and represented by the color key. DMFS, distant metastasis free survival at 5 years; PAM50, an intrinsic subtyping classifier that measures expression of 50 genes selected as characteristic of five breast cancer intrinsic subtypes - luminal A (LumA), luminal B (LumB), basal, normal and Her2-positive [48].



Association

	Positive	Negative
p < 0.01	+	−
p < 1e −05	+ +	− −
p < 1e −10	+ + +	− − −

Figure 4: Comparison of our four TN stromal properties with the Lehmann TN subtypes. Lehmann et al. [16] defined six TN subtypes – two basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem-like (MSL), and a luminal androgen receptor (LAR) subtype. We subjected the six Lehmann subtypes to our methodology, estimating their activation as either low, intermediate or high across the TN breast cancer compendium (ROI_{95}). This method rendered the Lehmann groups in a format for direct comparison with our four stromal properties using Cohen's kappa statistic (fmsb package version 0.5.1; Association table at the bottom). Heatmaps summarize ROI_{95} assignments for each Lehmann group. Samples are colored white, grey and black to represent low, intermediate and high subtype assignments, respectively. Our stromal properties are colored as in Figure 3. Patients are ordered by the: **A)** T cell property (T), **B)** B cell property (B), **C)** invasive epithelial cells property (E) or **D)** desmoplastic reaction property (D).

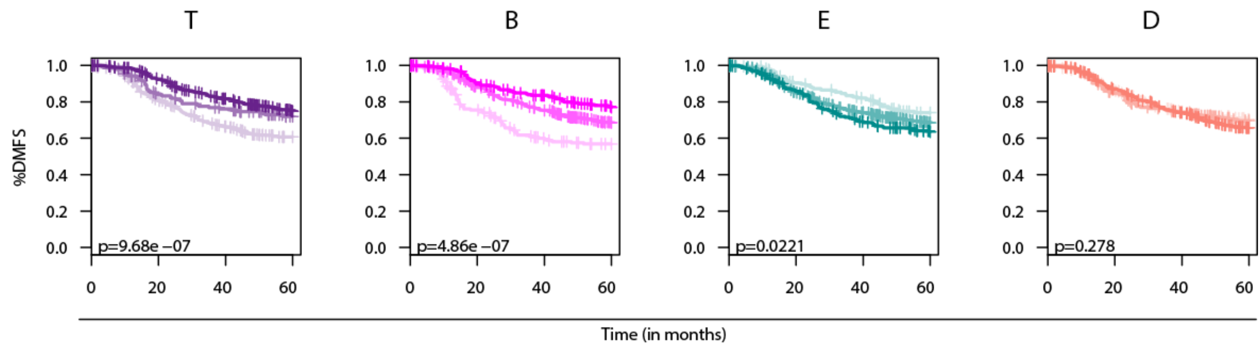


Figure 5: Our stromal properties associate with TN patient outcome. Kaplan-Meier survival analysis of the stromal properties for distant metastasis free survival (DMFS) of TN breast cancer patients in external TN bulk expression datasets (n=1,098). Log-rank test p-values are indicated at bottom left for each graph. Each stromal property is colored and partitioned into low (light color), intermediate (medium color) or high (dark color) as per Figures 3&4. The four stromal properties: T cell (T), B cell (B), invasive epithelial cells (E) and desmoplastic stroma (D).

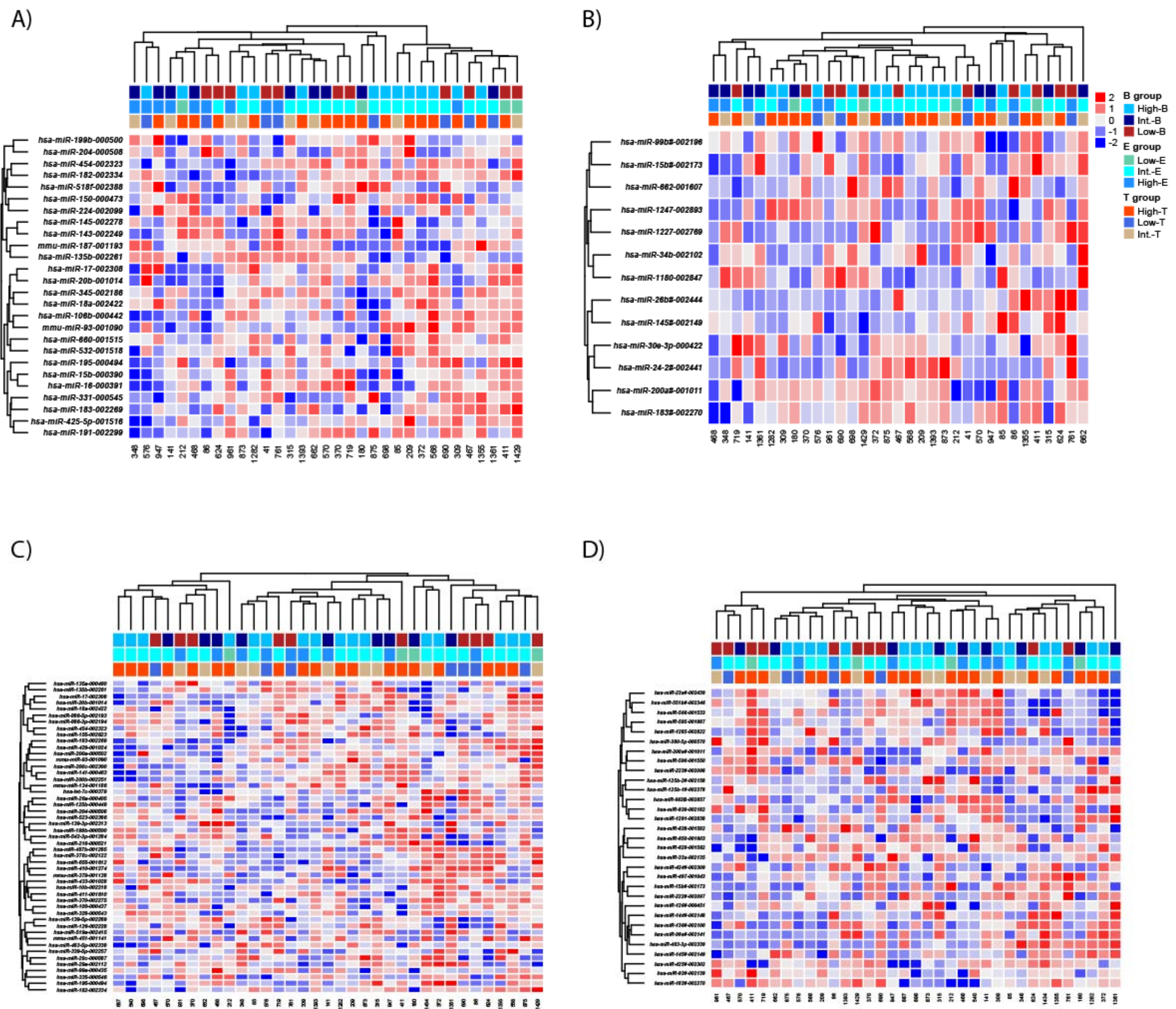


Figure 6: Association of miR expression with stromal properties. miR whose expression most significantly differed between tumor and normal epithelia and stroma were clustered and compared to relative T, B and E gene expression for each patient as an indication of disease outcome (Heatmap complete linkage for clustering and Euclidean for the distance using ComplexHeatmap R Bioconductor package). A) Tumor versus normal epithelia for miR pool A. B) Tumor versus normal epithelia for miR pool B. C) Tumor versus normal stroma for miR pool A. D) Tumor versus normal stroma for miR pool B. Pool A was comprised of more commonly known miR, whereas pool B contained less well-known miR.

11.3 Tables

Table 1: Each stromal property is associated with distinct cell types/processes. Differentially expressed gene lists from the four stromal properties identified in Figure 3 were examined using QIAGEN's Ingenuity® Pathway Analysis (IPA®). On the basis of these observations, the four stromal properties were labelled B-cells (B), Invasive epithelial cells (E), T-cells (T) and Desmoplastic stroma (D).

Stromal Property	Representative Significant Pathways from Ingenuity Pathway Analysis	Representative Genes	Property
	cell viability of B lymphocytes, quantity of B lymphocytes, differentiation of B lymphocytes, maturation of B lymphocytes	CD79A, POU2AF1, PDK1, PRDM1, TNFRSF13C, TNFRSF17, CD38, CD72, IGHM, IGLL1	B-cells (B)
		KRT6B, KRT23, Metallothioneins	Invasive Epithelial Cells (E)
	quantity of T lymphocytes, T cell development, activation of T lymphocytes, cytotoxicity of leukocytes	CD2, CD3D, IL-2Rα IL-2Rβ, IL-2Rγ, CD4, CD8A, CD8B, GZMBA, GZMB, GZMK, GZMH, STAT1, STAT4, TNFAIP2, TNFAIP8	T-cells (T)
	Hepatic Fibrosis / Hepatic Stellate Cell Activation, Adhesion of connective tissue cells	COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, COL8A1, COL8A2, COL10A1, COL12A1, COL16A1, PDGFRB, FAP, P4HA2, MMP2, LOXL1	Desmoplastic stroma (D)