**SPECIFIC AIMS need to change**

Inflammatory bowel disease (IBD) which includes **ulcerative colitis (UC)** is emerging as a global health problem. Although the incidence and prevalence of IBD are increasing worldwide, therapeutic options to treat this debilitating condition remain limited with about 30% of UC patients becoming refractory to pharmacological treatment or developing UC-associated dysplasia. These patients undergo a restorative proctocolectomy with ileal pouch-anal anastomosis (IPAA), but half of them will develop a UC-like condition in the pouch (pouchitis) and require more aggressive treatment strategies. Although the pathogenesis of pouchitis is not entirely clear it has been suggested that similar factors that lead to UC in the colon may be replicated within the ileal pouch after the fecal stream is reestablished. This, however, does not explain why only half of patients that undergo IPAA will eventually develop pouchitis, thus, suggesting the existence of additional contributing factors transitioning normal IPAA to pouchitis. Therefore, **the goal of this proposal is to identify factors predictive of pouchitis in UC/IPAA patients** with the underlying ***rationale*** that the knowledge gained from these proposed studies will fill a gap in our basic understanding of pathophysiology of pouchitis and provide a means for successful management of UC patients with IPAA.

IBD is thought to be driven by a convergence of genetic, environmental and microbial factors that results in the dysregulated immune response to luminal *commensal microbes*. Mouse and human studies have demonstrated that functional defects in *mucosal macrophages (Ms)* significantly contribute to the abnormal inflammatory responses in IBD. On the other hand, gut inflammation and microbial stimulation alter homeostasis and function of mucosal Ms. Likewise, **Ms have been identified as a critical component of pathogenesis of pouchitis**.

The gut microbiome is highly influenced by the immune state of the host and inflammation, with dysbiosis being not only a cause but also outcome of IBD. Interestingly, pouchitis much better responds to antibiotic treatment than UC raising a possibility that the microbes that drive pouchitis are distinct from those that drive UC, although some alterations in gut microbiota are shared by UC and pouchitis patients leading to the idea that **the composition of perioperative microbiome can be predictive of developing pouchitis**.

Based on these data, **we *hypothesize* that the perioperative M/microbial phenotype of UC patients undergoing IPAA procedure will be predictive of developing pouchitis**. Specifically, the dynamic interplay between the host immune system, particularly mucosal Ms, and microbiota involved in pathogenies of UC will promote an expansion of a set of microbial *pathobionts*, which will induce the development of pouchitis in UC/IPAA patients in the context of a M-specific defect underlying their UC disease. We will test our hypothesis in two specific aims by combining cellular immunology, microbial and host OMICS with transkingdom network analysis and state of the art computational approaches for predicting disease outcome. Our ***collaborative team***, consisting of the M biologist (MB), microbiome and systems biology experts (AM and NS) and surgeon specializing in IBD (WK), is uniquely positioned to adress these aims.

**Aim 1 will establish perioperative Mmicrobiome transOMIC landscape in patients with UC.** In a prospective study, we will analize intestinal tissue from surgical resections to establish the phenotypes and transcriptomic signatures of intestinal mucosal Ms along with profiles of tissue-associated microbiota. By performing *transkingdom network analysis* of these data sets, we will identify a UC-specific proinflammatory M gene expression signature and determine candidate microbial pathobionts activating Ms.

**Aim 2 will establish mucosa/microbiome transOMIC signature predictive of pouchitis in patients with UC/IPAA.** We will perform a retrospective study on frozen perioperative mucosal specimens isolated from UC/IPAA patients that did or did not develop pouchitis. We will determine their mucosal transcriptome signatures and microbial profiles and develop an unbiased predictive model based on *class prediction analysis* and *survival risk analysis.* We will also perform a supervised analysis based on the inflammatory M transcriptomic signature and pathobiont profiles obtained in Aim1.

The findings of this R21 proposal will establish the feasibility of perioperative host and microbial biomarkers for pouchitis and point to novel candidates for cellular and molecular therapeutic targets, altogether, providing a foundation for a future R01 proposal.

**A. SIGNIFICANCE**

**The importance of the problem.** Inflammatory bowel disease (IBD), which includes Crohn’s disease and ***ulcerative colitis (UC)***, has become a global health problem, as its incidence and prevalence are increasing worldwide 1. IBD is thought to result from the ***dysregulated immune response*** to ***commensal microbes*** driven by a convergence of genetic, environmental and microbial factors. The colonic microbiome is highly influenced by the immune system. This leads to the idea that *possible preexisting immune abnormalities in IBD patients predispose them to dysbiosis (including generation of* ***pathobionts****), which, in turn, exacerbates immune response leading to the disease* {ref}. Despite significant advances in the understanding of the disease mechanisms, IBD treatment options are still limited for a large number of patients 2. For example, among IBD patients affected by UC, 30% become refractory to pharmacological treatment or develop UC-associated dysplasia 3. These patients undergo restorative proctocolectomy with ileal pouch-anal anastomosis (IPAA or pelvic pouch). Nearly 50% of patients undergoing IPAA develop a UC-like condition in the pouch (***pouchitis***) within 5 years after surgery. In up to 20% of patients, pouchitis becomes chronic and requires aggressive (including surgical) treatment strategies 4-6. Therefore, there is a need for novel therapeutic strategies to prevent and decrease the incidence of pouchitis in UC/IPAA patients or even avoid IPAA procedure altogether.

**The critical barriers to progress**

***Barrier 1:* *Poor understanding of the etiology and pathogenesis of UC***

Genetic heterogeneity underlying abnormalities in the immune response to a variety of gut microbes is thought to be responsible for the diversity in clinical presentations and responses to therapies in patients with UC {ref- add some review, }7 8. Despite this heterogeneity, multiple human and mouse studies of IBD have emphasized the importance of intestinal **mucosal macrophages (Mfs)** as initiators and regulators of pathological T cell responses driving IBD 9, and a recent functional genomics network model of IBD has identified a strong M component within human IBD key driver genes {ref}. Furthermore, increase in mucosal Ms found in the pouchitis-affected mucosa of patients with UC 10 suggested that Ms could also be a critical component of pouchitis. Despite these advancesMfstheinflamed mucosa of UC patients, particularly those 11

To better understand the pathophysiology of UC (and IBD), analyzedmucosal as an approach to reconstruct pathological immune responses through dissection of immune cell heterogeneity; this approach, however, is ingdue to small number of cells. the casual link between specific immune and microbial phenotypes, and further studies, particularly with the use of more abundant perioperative tissue, is highly needed.

***Barrier 2: Inability to predict pouchitis after IPAA***

Pouchitis can occur after IPAA construction for either UC or familial adenomatous polyposis. UC-associated pouchitis is much more frequent, thus, suggesting that it likely results from the patients’ immune dysregulation underlying their UC. Indeed, studies of the microbiome and host gene expression in pouchitis detected the presence of dysbiosis (ref.) and revealed that the transcripts of complement cascade and IL12 pathways were the most strongly associated with microbial shifts (ref.) showing its similarity to UC. In contrast to UC, however, prevention of pouchitis is possible and consists of several options ranging from the use of probiotics, antibiotics and anti-inflammatory drugs to the changes in surgical strategy (e.g., fecal diversion, colostomy) {refs}. Because these therapies have their own adverse effects impeding their use in all IPAA patients, identifying patients at risk of pouchitis would help implement its targeted prophylaxis. Thus, one study demonstrated that histological signs of inflammation in the terminal ileum at the time of IPAA procedure were significant predictors of pouchitis (MS, Annals of Surgery, 1998). Other factors, such as polymorphism in key immune genes, presence of pANCA autoantibodies and microbe-specific (yeast *S. cerevisiae*, *Clostridium* flagelin CBir1) antibodies, blood platelet counts, levels of serum albumin and fecal butyrate and bile acid, have been also tested as potential markers predictive of pouchitis, although with mixed results (UN, Am. J. of Gatroenterology, 2009). 12 13 {gut 2017}. Despite some success, *no clinically usable biomarkers have been established so far*.

**Steps toward the improvement of scientific knowledge.**

***Step 1.******Characterizing diversity and functional specialization of Mfs in UC-affected intestine***

Intestinal Ms are heterogeneous, and it becomes evident that their complexity extends beyond two originally described subsets – M1 and M2. For example, our own studies identified three distinct Ms populations in the normal mouse intestine, two of which (CD11chiCX3CR1int and CD11chiCX3CR1hi) are distributed in the intestinal mucosa, and the third one accumulates in the underlying intestinal layers (CD11cloCX3CR1+ muscularis Ms). Similarly, a recent human study (ref.) identified four populations of normal small intestinal Ms, with three of them being enriched in the mucosa (CD14+CD11c+HLA-DRint M1, CD14+CD11c+HLA-DRhi M2 and CD14+CD11c–CD11b– M3). The M diversity becomes even more complex in the inflamed intestine, therefore, it is necessary to define *the biological relevance of M heterogeneity, particularly in the context of UC*. As shown in our preliminary data, we started unraveling the complexity of Mf subsets in inflamed and uninflamed human gut (***Fig. 1***).

***Step 2. Identifying microbial pathobionts of UC driving Mf activation.***

Because of the complex multi-layered mechanisms operating in IBD, integrative analysis of host and microbiota is the state-of-the-art strategy to uncover IBD pathogenesis. Several recent studies have reported encouraging results of such analyses 14-16, however, none of them focused on finding microbiota members driving activation of specific Mf subsets or analyzed perioperative colonic tissues of UC patients. Yet, this type of work has a high potential to uncover critical host and microbial drivers of this devastating disease and point to novel therapeutic strategies alternative to IPAA.

***Step 3. Developing a multi-OMICS approach to predict pouchitis.***

Several studies have attempted to identify factors predictive of pouchitis (see ***Barrier 2***), however, the accuracy of prediction was not high enough to warrant their clinical application. This is likely due to the weaknesses in experimental approaches such as limited number of analyzed parameters and use of fecal material instead of the tissue samples. Integration of multi-OMIC methods has been shown to increase accuracy of classification for many diseases 17,18,19,20 which can be explained by the fact that individual OMIC data cannot fully capture the complexity of a disease process but combined they can create a synergistic effect 21,22. *Therefore, studies using multi-OMIC approaches are highly needed for pouchitis prediction.*

**How the methods, concepts, treatments and preventive interventions will be changed if the proposed aims are achieved.** We propose that performing a detailed immune cell (particularly, Mf) profiling of intestinal tissue specimens combined with host transcriptome and microbiome analysis at the time of IPAA surgery in patients with UC will significantly improve prediction precision sufficient for developing laboratory tests to guide preventive therapy for pouchitis. Furthermore, answering the questions raised in this proposal will establish a basis for new cellular and molecular therapeutic targets of UC. The results of the study will also deepen our knowledge about human M biology, microbiome and mucosal immunology, and fill a gap in our basic understanding of etiology and pathogenesis of both UC and pouchitis.

**Scientific premise for the proposed project** is built on multiple publications supporting role of microbiota and Mfs in UC and indicating that pathogenesis of pouchitis is connected with UC {refs}. In addition, the multi-disciplinary expertise of our team in: **1)** M biology (**MB**); **2)** microbiome (**NS**); **3)** computational biomedicine (**AM**); and 4) clinical IBD **(WK)** is supported by our work published in reputable journals such as Cell 23,24, Science 25, Nature Medicine 26, Nature immunology 27,28, Immunity 29,30, Nature Communications 31,32, Gastroenterology 33, Gut 34 and others.

**B. INNOVATION**

We will test the ***innovative hypothesis*** that ***there are specific M/microbial signatures in perioperative inflamed tissue in patients with UC and that these signatures can be predictive of pouchitis*** with the underlying rational that *the dynamic interplay between the host immune system, particularly mucosal Ms, and microbiome in UC patients will promote an expansion of a patient-specific set of microbial pathobionts, which will induce development of pouchitis in UC/IPAA patients.*

Specifically, *innovations of this project include:* 1) discovery and characterization of new human intestinal Mf populations involved in UC using state of the art cellular immunology techniques; 2) investigation of mechanistic link between Mf and microbiota in UC and pouchitis using a novel computational approach, transkingdom network analysis; 3) integrating transcriptome and microbiome of perioperative tissues for predicting pouchitis after IPAA.

**C. APPROACH**

C.1. EXPERIMENTAL RATIONALE

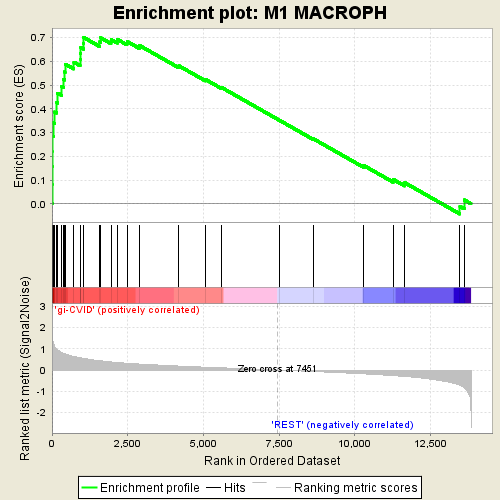
The study is built upon our research expertise and clinical resources available at PSUCoM, particularly, the GI Diseases Tissue Biobank led by the IBD surgeon and Director of IBD Center Dr. Walter Koltun who is a co-investigator on this grant (PSUCoM, letter attached). The study will include prospective and retrospective tissue analyses (***Fig. 1***).



**Figure 1. Schematic Summary of the Proposed Project.**

In the prospective study (Aim 1), blood and fresh intestinal tissue samples (terminal ileum, inflamed and uninflamed regions of the colon) will be obtained from 20 refractory UC and 20 non-IBD control patients (e.g., colorectal cancer, abdominal trauma) undergoing colectomy. M cell subsets will be isolated by FACS from the blood and mucosa of these patients, respectively, and their gene expression signatures will be determined by RNAseq. Bulk tissue gene expression signature and microbial diversity will be determined in the same specimens. Transcriptomic and transkingdom network analyses will be applied to identify UC-specific types of Ms and their transcriptional signatures, and UC-specific microbial pathobionts associated with particular M subsets with the rational to establish the perioperative Mmicrobiome transOMIC landscape in patients with refractory UC. Because intestinal Ms originate from monocytes, purified blood monocyte subsets will be subjected to a similar analysis to pinpoint the genes, which expression is highly influenced by mucosal microenvironment.

**Figure 2. Identifying M Subsets in the Human Colon from UC patients. A.** FACS plots show gating strategy to identify various hematopietic cell subsets in the normal large bowel (LB) mucosa of a patient with colorectal cancer (PMN – polymorphonuclear leucocytes, MP – mononuclear phagocytes, DCs – dendritic cells, Ms – macrophages, CD45+Lin+ – lymphocytes, CD45– – stromal cells). **B.** Verifying manual gating strategy by unsupervised t-SNE analysis. Pregated on Gate 3 shown in (A). **C.** Increased numbers of Ms and PMNs in the inflamed LB mucosa from UC patients. Control (Cont.) – normal mucosa from patients with colorectal cancer and colectomy for non-GI pathology. Most of uninflamed (Uninfl.) and inflamed (Infl.) samples are from unmatched UC patients. **D.** Altered proportions of M subsets in the inflamed LB mucosa of a UC patient with the reduction of M2 and expansion of M1 subsets. Uninflamed and inflamed samples are taken from the LB of the same patient. **E.** Heat map shows relative expression of selected genes measured by 200 gene qPCR array analysis of myeloid cell subsets FACS-purified from human intestine (n=3) and confirms our gatind strategy. Listed genes encode DC-, DC1-, DC2 and M-specific genes.



Enteropathy

Control

0.2

0.4

0.6

0.0

p<0.004

M1 macrophage signature

Enrichment Score

Normalized enrichment score

M1

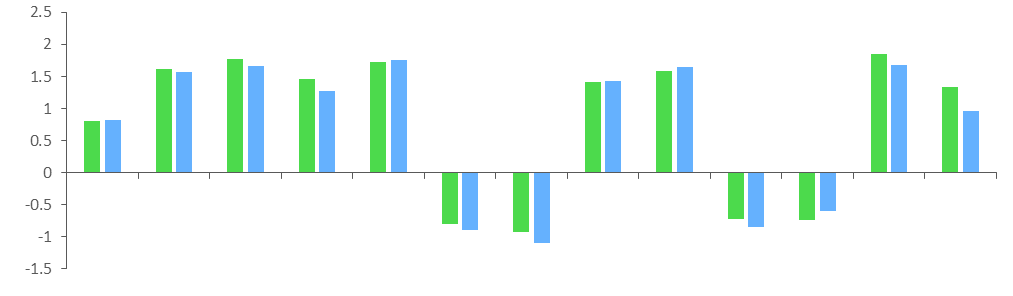
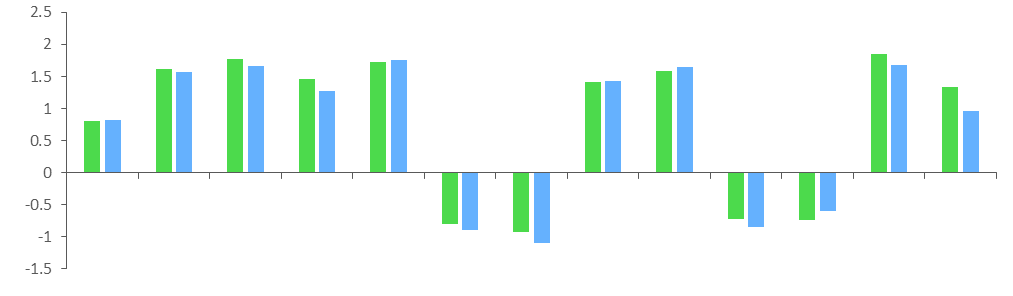
M2

\*

\*

Ctrl vs. E-CVID

noE-CVID vs. E-CVID



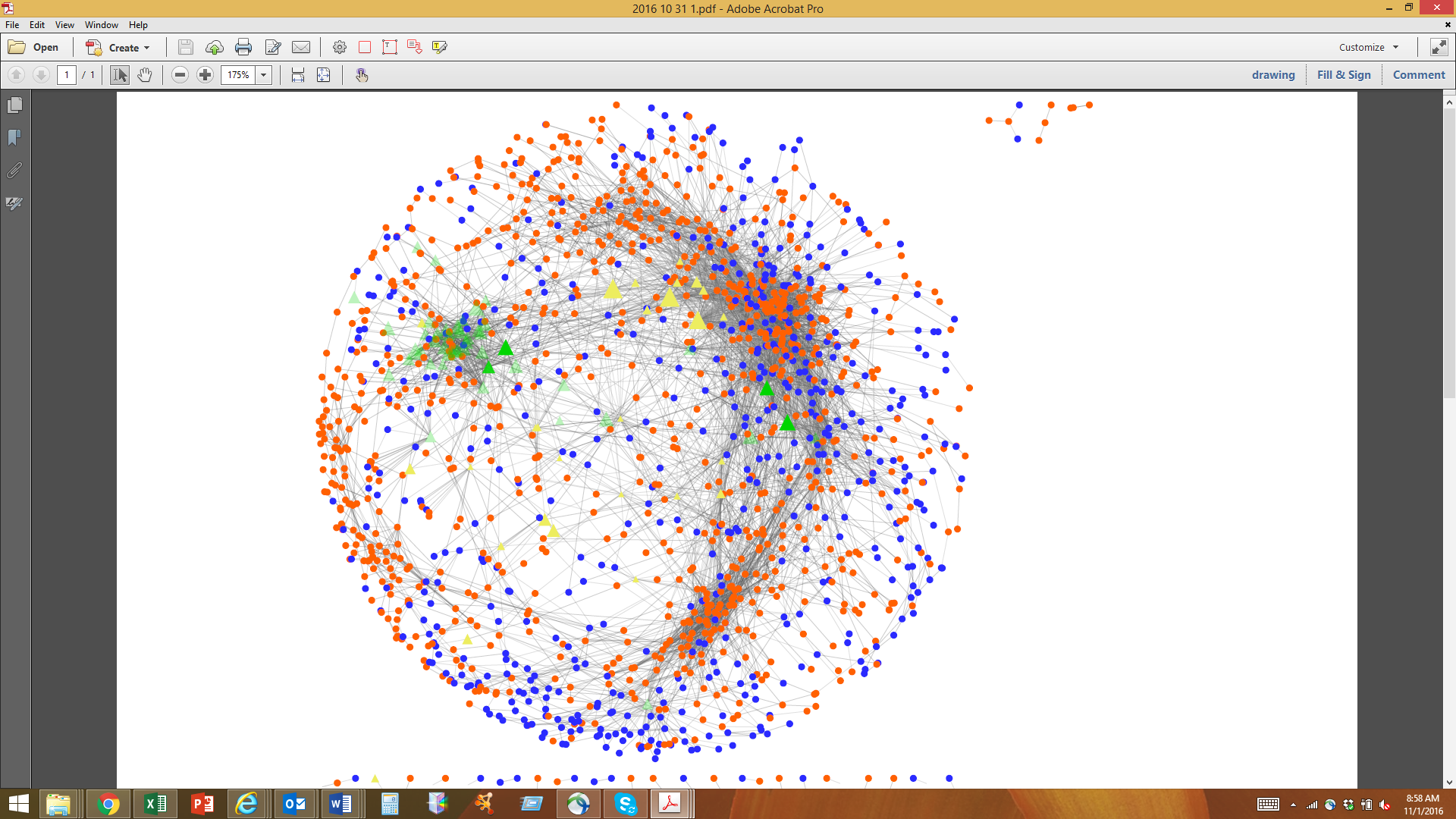
**Fig. 3**. Macrophages M1 but not M2 show gene set enrichment in the signature of CVID enteropathy (E-CVID) compared to controls or no enteropathy CVID (noE-CVID) biopsies. Macrophage signatures used from {17082649}. \*p<0.01.

In the retrospective study (Aim 2), we will take advantage of our frozen intestinal tissue repository, which at the time of submission contained mucosal samples from 96 patients with UC/IPAA collected since 2009, 53 of them had at least one episode of pouchitis. We will isolate host and microbial RNA from these specimens to determine tissue gene expression signatures and microbial diversity and develop diagnostic classifiers focused on a) UC-specific M/microbiome tranOMIC signatures established in Aim 1, and b) host/microbiome transOMIC classifiers using unbiased analyses for prediction of pouchitis in patients with UC/IPAA.

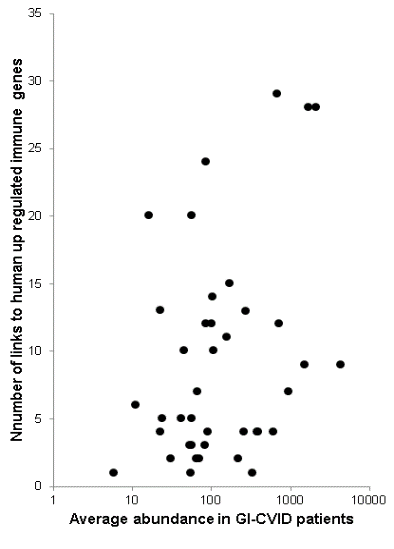
C.2. PRELIMINARY DATA

***C.2.1. Establishing the diversity of mucosal myeloid cells in the human large intestine.*** We have established the heterogeneity of mouse mononuclear phagocytes (MPs) – dendritic cells (DCs) and Ms – in the normal and inflamed mouse intestine (ref., manuscript in preparation). This was achieved by combining several techniques: multi-color FACS analysis and FACS-based cell subset isolation followed by transcriptional profiling. We have applied a similar approach to characterize human myeloid cells and were able to identify human analogues of DC1, DC2, Msubsets, as well as polymorphonuclear leukocyte (PMNs) and other cell types (***Fig. 2***). We have capacity to perform a similar analysis of blood leukocytes (data not shown).

***C.2.2. Transkingdom network reveals microbes and their genes that affect host phenotypes.*** It is thought that inflammatory response in UC is primarily driven by colonic microbiota. Consistent with this idea, several groups have identified changes in the composition and function of fecal and tissue microbiota of UC patients {ref}. However, those studies do not answer the key question of whether those altered microbial components are relevant to the disease phenotype of the host. To answer this question, we have developed and successfully applied an approach called ***Transkingdom Network (TK-NW) analysis*** that links expression of human and bacterial genes { ref}.



*Acinetobacter*



# of links to human up-regulated genes

Abundance in E-CVID biopsies

**Figure 4. Transkingdom network analysis identifies *Acinetobacter* as a candidate microbe driving CVID enteropathy (E-CVID).** **A**. Network reconstructed from levels of expression of human genes from E-CVID signature and candidate microbes; orange and blue – CVID signature and candidate microbes; orange and blue – upregulated and downregulated genes in E-CVID, respectively; green and yellow – microbiota increased or decreased in E-CVID, respectively. **B.** Number of connections to inflammatory genes in E-CVID and average abundance in E-CVID for microbiota taxa.

A

B

*Acinetobacter*

In the published work, we applied this approach to reconstruct TK-NW using metagenomics and host gene expression data focusing on mitochondrial genes dysregulated after antibiotic treatment 34. By analyzing this network, we predicted and experimentally confirmed that *Pseudomonas aeruginosa* negatively affected mitochondria and lead to epithelial cell death uncovering a mechanism of antibiotic-associated comorbidities 34.

Our recent study of common variable immunodeficiency (CVID)-associated enteropathy further validated our approach. Only half of patients with CVID develop enteropathy, however, the factors underlying this predisposition are unknown. Gene expression profiling of duodenal biopsies of CVID patients with and without enteropathy revealed a signature of inflammation enriched for various immune cells including M1 macrophages (***Fig. 3***). Biopsy samples were then analyzed for microbial and host gene expression, based on which a regulatory transkingdom network was reconstructed (Fig. 4). Focusing on Mf-related inflammatory genes upregulated in the group of CVID patients with enteropathy, we found that *Acinetobacter baumannii* was the best candidate microbe driving the inflammatory response, because it displayed high connectivity to inflammatory genes in TK-NW (***Fig. 4***). These findings were confirmed in *in vitro* experiments by activating a Mf cell line THP1 with *Acinetobacter baumannii* {(manuscript under revision and ref meeting abstract}.

Additionally, in a clinical case study of two twins with familial adenomatous polyposis we were able to assess the composition of microbiota in small and large bowel samples harvested during IPAA surgery {ref}., confirming the feasibility of microbiome analysis in frozen perioperative intestinal tissue specimens stored at the PSUCoM Colorectal Biobank.

C.3. EXPERIMENTAL DESIGN

**Aim 1. Establish perioperative Mmicrobiome transOMIC landscape in patients with UC.**

* 1. ***Establish the diversity of phenotypes and transcriptomic signatures of intestinal mucosal Ms in UC patients undergoing colectomy.*** Intestinal tissue will be removed from UC and control patients undergoing colectomy and grossly evaluated by a GI pathologist to identify uninflamed and inflamed regions. Transmural tissie specimens will be resected from terminal ileum, ascending, transverse, discending and sigmoid regions of the colon and transported in tissue medium on ice to our lab (MB). The tissue will be digested by adapting our previously developed protocol (ref) to obtain single cell suspensions, which will be stained with a cocktail of monoclonal antibodies as described in ***Fig. 2.***, and cell types of interest will be purified using BD Aria III cell sorter. To identify myeloid cell populations we will use a combination of 12 antibodies and a marker for dead cells – CD45 to identify hematopietic cells, CD11b and CD11c to identify myeloid cells, MHC Class II to identify antigen-presenting cells, CD14 and CD64 to distinguish between DCs and Ms, CD1c and CD172a/b to separate DC1 and DC2, and CD209 and CD163 to differentiate between four M subsets. By using the gating strategy described in ***Fig. 2***, we will be able to isolate at least 9 cell types (5,000 cells each) inclduing four populations of Ms (CD209–CD163–, CD209+CD163–, CD209–CD163+ and CD209+CD163+), DC1 and DC2, total PMN leukocytes, total lymphocytes, and CD45– stromal cells. We will also analize and compare the frequency of each myeloid cell subset in uninflamed and inflamed regions of the mucosa from UC patients and control subjects. Total RNA will be isolated from each cell type, and RNAseq of human transcriptome will be performed. 3’UTR sequencing {ref-Lexogen} will be used as the most affordable option within limits of R21 budget. We will identify cell type specific and whole tissue gene expression signatures and alterations asscoiated with disease by comparing transcritomes of different cell subtypes from different sites as we have done previously {refs***} fig from prelimn data***. Our controls will consist of myeloid cell populations isolated from matching periferal blood and the intestinal mucosa of non-IBD patients with colorectal cancer and abdomianl trauma.
  2. ***Establish microbiome profiles and candidate pathobionts activating Ms in the gut of UC patients via transkingdom network analysis***

Microbiota composition will be assessed in the same samples described in Aim 1.1 {Schieffer,2017} by employing standard pipelines for 16S rRNA based taxonomic microbiome analysis (e.g. QIIME) routinely used by our lab {ref}. In obtained datasets we will estimate microbiota diversity, richness, and similarity between communities at different sites and states. Pairewise analyses will include comparisons between inflammed and uninflammed colon samples within the same patient. A subset of samples proving to be most informative, will be used for shotgun metagenomics. In this case, we will employ Processing Utility for Metagenomics Analysis (PuMA) for shotgun data, as we have done in our previous studies 34,35. PuMA itself consists of several tools for filtering data, aligning reads to genomes36 37, and assigning reads to genes and taxa38 (<https://github.com/richrr/MorgunShulzhenkoLabs/tree/master/PuMA-toolkit>.

Next, we will apply transkingdom network analysis to pinpoint potential microbial drivers of refractory UC. Our hypothesis is that microbiota members enriched in the inflamed colon contain taxa driving activation of Mf subsets that contribute to UC. Thus, we will reconstruct a human network from Mf subtypes and gene expression upregulated in inflamed colon and a microbial network from taxonomic groups showing increased abundance in affected colon. We then integrate them into a TK-NW as we described 8,[10](#_ENREF_10),11. Next, we will identify microbe candidates predicted to induce inflammation. For this, we will calculate a bi-partite betweenness centrality39 between microbial and host compartments of the network. This metric is highly indicative for “bottleneck” nodes in the network critical for transfer of information between subnetworks corresponding to functionally different compartments31,40[\_ENREF\_10](#_ENREF_10" \o "Morgun, 2015 #6).

The ***outcome of the Aim 1*** will be: a) the identification of new subsets of Mfs involved in UC and their cellular and molecular characterization; b) list of candidate microbes involved in driving individual subtypes of Mfs in the gut of UC patients.

**Aim 2. Establish transOMIC signatures predictive of pouchitis in patients with UC/IPAA.**

Because this study is exploratory, we will establish a proof-of-principle that information obtained from host transcriptome/microbiome signatures in perioperative samples will allow for development of diagnostic criteria to guide therapeutic decisions for UC patients after IPAA. To do that, we will take advantage of frozen perioperative intestinal samples from UC/IPAA patients stored by the Colorectal Diseases Biobank at the PSUCoM. We will obtain tissue gene expression and taxonomic microbiome datasets as described in Aim 1.2 and develop *predictive classifiers* using *transcriptome* and *microbiome* data*.* Wewill perform two types of analysis:

***2.1. focused on Mf gene signatures defined in Aim 1 and taken from the literature*** {refs};

***2.2. unbiased analyses from global microbiome and transcriptomes.***

To achieve high applicability, predictive signatures should be tested and performed with high accuracy in a separate set of patients {ref}. Therefore, we will generate predictors in a training set of patients and validate them in a test set according to established practices of the field {REFs}.

We will separate all patients into two cohorts, namely ‘***Training***’ and ‘***Test***’ sets. As estimated by power analysis (see corresponding section) “training” dataset will require 75 samples to establish robust predictors. We will organize all patients in chronological manner and use as the ‘training’ set samples those that were collected earlier. Next, we will test our classifier/s on the rest of the samples collected from newer cohort of 45 UC patients. Total RNA and DNA will be isolated from the frozen tissue samples of the *inflamed colon*. We will obtain data for global human gene expression using RNAseq and for bacterial 16S rRNA gene content using established methods as described in Aim 1.

To address Aim 1.1., we will use pre-defined gene lists corresponding to various Mfs signatures (obtained in Aim 1 and from the literature) and test their performance for discriminating samples between two groups of patients – with and without pouchitis (42 and 52 respectively, total 96). To address Aim 2.2., we will determine gene sets predictive of pouchitis which are not based on particular biological functions, i.e. unbiased predictors.

For both **Aims 1.1. and 1.2.** we will perform two types of analyses:

1) *class prediction analysis* that aims to find predictors that discriminate between the two groups of patients with and without pouchitis (follow-up at least 5 years). For this, we will use state-of-the-art machine learning and statistical methods previously employed by us and others (Support Vector Machine, Diagonal Linear Discriminant analysis, K-nearest neighbors, Bayesian compound covariate, Compound covariate predictor, Nearest Centroid) 41,42. The consensus classification of a sample provided by these methods will be used to assign patients into one of the two groups.

2) *survival risk analysis* that will account for the amount of time free of pouchitis after surgery in order to develop a predictive classifier. In this case, we will employ standard methods developed for estimation of survival risk for large scale data (Principal Component Method 43,44 and penalized Cox regression 45.

For all analyses performed on ‘Training’ dataset, we will use ten-fold cross validation. All analyses will account for known clinical covariates and results of histological evaluation. We will evaluate the performance of predictive classifiers discovered in the ‘training’ set on OMICS data of the ‘Test’ set.

As first step, all above analyses will be done using transcriptome and microbiome data separately. Next, we will establish transOMIC predictors/classifiers of pouchitis by integrating microbiome and host transcriptome data. Specifically, we will employ a kernel-based integration approach that had demonstrated increased classification by combining multi-OMIC data in cancer settings 46.

The ***outcome of the Aim 2*** will be molecular classifiers that predict pouchitis as well as estimates (i.e. sensitivity/specificity) for the precision of their performance.

Sample size justification: *For Aim1* power calculations showed that assuming differences of at least two-fold, using the median of the variance distribution from human gene expression data and power of 90% and accounting for multiple hypothesis correction, the required sample size for observational studies will be 20 samples in each group (i.e., affected and non-affected tissue). For network inference, because there is no standard methods for sample size estimation, we have based our sample size estimates on our previously published studies in which we inferred networks from gene expression and microbiome data26,31,34 and where sample size of 20 was sufficient for building robust networks. *For Aim 2*, using the approach for calculation of sample size required for developing predictive classifiers 47,48 we estimated sample size of 62 patients for “training” dataset considering prevalence of pouchitis 0.5, the difference between two classes for a given variable at least two-fold, variance 0.7. A safe option for the sample size of the “test” set is at least half of “training” dataset (~35 samples) 49 this the book--- 50.

**Pitfalls and alternative strategies**

*For aim 1:* It is possible that we will identify significant UC-specific transcriptomic changes associated with other myeloid cells types – *DCs and PMNs*. In this case, we will perform the same computational analysis as for M subsets. Also, there is a possibility that our FACS based separation strategy will generate somewhat heterogeneous subtypes of cells detactable in transcriptome analyses. In this case, we will perform single cell *RNA seq from the limited number of samples*, which will allow establishing new subtypes of Mfs in an unbiased manner. If 16S rRNA will not allow to identify species level resolution of candidate pathobionts we will employ shotgun data as we have done previously 34,35.

*For Aim 2:* if developed classifiers will not perform well on training dataset we will: 1) apply a different set of classification algorithms such as random forest, top scoring pairs and others; 2) increase the training set by using all samples in dataset and will use newly collected prospective samples as a test set.

The results of this project will establish the basis for ***our future multi-PI*** R01 proposal containing: a) *prospective multi-centric clinical trial* that will use developed classifiers to guide preventive therapies of pouchitis; and b) testing identified candidate pathobionts and Mf subsets in experimental models aiming to develop new therapies.

**Timeline**

The experiments proposed in Aim 1 and 2 will be performed in parallel. Sample collection and analysis will be completed by 18 months. The last 6 months of this proposal will be dedicated to computational analysis of the generated data sets.

D. REFERENCES

1 Molodecky, N. A. *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142**, 46-54 e42; quiz e30, doi:10.1053/j.gastro.2011.10.001 (2012).

2 Colombel, J. F. Decade in review-IBD: IBD-genes, bacteria and new therapeutic strategies. *Nature reviews. Gastroenterology & hepatology* **11**, 652-654, doi:10.1038/nrgastro.2014.170 (2014).

3 Bernstein, C. N. *et al.* A review of mortality and surgery in ulcerative colitis: milestones of the seriousness of the disease. *Inflammatory bowel diseases* **19**, 2001-2010, doi:10.1097/MIB.0b013e318281f3bb (2013).

4 Hoda, K. M., Collins, J. F., Knigge, K. L. & Deveney, K. E. Predictors of pouchitis after ileal pouch-anal anastomosis: a retrospective review. *Dis Colon Rectum* **51**, 554-560, doi:10.1007/s10350-008-9194-7 (2008).

5 Hurst, R. D., Molinari, M., Chung, T. P., Rubin, M. & Michelassi, F. Prospective study of the incidence, timing and treatment of pouchitis in 104 consecutive patients after restorative proctocolectomy. *Arch Surg* **131**, 497-500; discussion 501-492 (1996).

6 Hahnloser, D. *et al.* Results at up to 20 years after ileal pouch-anal anastomosis for chronic ulcerative colitis. *Br J Surg* **94**, 333-340, doi:10.1002/bjs.5464 (2007).

7 Ananthakrishnan, A. N. *et al.* Gut Microbiome Function Predicts Response to Anti-integrin Biologic Therapy in Inflammatory Bowel Diseases. *Cell Host Microbe* **21**, 603-610 e603, doi:10.1016/j.chom.2017.04.010 (2017).

8 Zhou, Y. *et al.* Gut Microbiota Offers Universal Biomarkers across Ethnicity in Inflammatory Bowel Disease Diagnosis and Infliximab Response Prediction. *mSystems* **3**, doi:10.1128/mSystems.00188-17 (2018).

9 Lugering, A. *et al.* Infliximab induces apoptosis in monocytes from patients with chronic active Crohn's disease by using a caspase-dependent pathway. *Gastroenterology* **121**, 1145-1157, doi:10.1053/gast.2001.28702 (2001).

10 de Silva, H. J. *et al.* Lymphocyte and macrophage subpopulations in pelvic ileal pouches. *Gut* **32**, 1160-1165 (1991).

11 Gren, S. T. & Grip, O. Role of Monocytes and Intestinal Macrophages in Crohn's Disease and Ulcerative Colitis. *Inflamm Bowel Dis* **22**, 1992-1998, doi:10.1097/MIB.0000000000000824 (2016).

12 Araki, T. *et al.* Colonic Histological Criteria Predict Development of Pouchitis after Ileal Pouch: Anal Anastomosis for Patients with Ulcerative Colitis. *Dig Surg*, doi:10.1159/000477268 (2017).

13 Navaneethan, U. & Shen, B. Laboratory tests for patients with ileal pouch-anal anastomosis: clinical utility in predicting, diagnosing, and monitoring pouch disorders. *Am J Gastroenterol* **104**, 2606-2615, doi:10.1038/ajg.2009.392 (2009).

14 Hasler, R. *et al.* Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease. *Gut* **66**, 2087-2097, doi:10.1136/gutjnl-2016-311651 (2017).

15 Tang, M. S. *et al.* Integrated Analysis of Biopsies from Inflammatory Bowel Disease Patients Identifies SAA1 as a Link Between Mucosal Microbes with TH17 and TH22 Cells. *Inflamm Bowel Dis* **23**, 1544-1554, doi:10.1097/MIB.0000000000001208 (2017).

16 Imhann, F. *et al.* Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* **67**, 108-119, doi:10.1136/gutjnl-2016-312135 (2018).

17 Kim, D. *et al.* Knowledge boosting: a graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction. *J Am Med Inform Assoc* **22**, 109-120, doi:10.1136/amiajnl-2013-002481 (2015).

18 Stephenson, A. J. *et al.* Integration of gene expression profiling and clinical variables to predict prostate carcinoma recurrence after radical prostatectomy. *Cancer* **104**, 290-298, doi:10.1002/cncr.21157 (2005).

19 Wu, D., Rice, C. M. & Wang, X. Cancer bioinformatics: a new approach to systems clinical medicine. *BMC Bioinformatics* **13**, 71, doi:10.1186/1471-2105-13-71 (2012).

20 Seoane, J. A., Day, I. N., Gaunt, T. R. & Campbell, C. A pathway-based data integration framework for prediction of disease progression. *Bioinformatics* **30**, 838-845, doi:10.1093/bioinformatics/btt610 (2014).

21 Kim, D., Shin, H., Song, Y. S. & Kim, J. H. Synergistic effect of different levels of genomic data for cancer clinical outcome prediction. *J Biomed Inform* **45**, 1191-1198, doi:10.1016/j.jbi.2012.07.008 (2012).

22 Huang, S., Chaudhary, K. & Garmire, L. X. More Is Better: Recent Progress in Multi-Omics Data Integration Methods. *Front Genet* **8**, 84, doi:10.3389/fgene.2017.00084 (2017).

23 Sisirak, V. *et al.* Digestion of Chromatin in Apoptotic Cell Microparticles Prevents Autoimmunity. *Cell* **166**, 88-101, doi:10.1016/j.cell.2016.05.034 (2016).

24 Muller, P. A. *et al.* Crosstalk between Muscularis Macrophages and Enteric Neurons Regulates Gastrointestinal Motility. *Cell* **158**, 1210, doi:10.1016/j.cell.2014.08.002 (2014).

25 Mortha, A. *et al.* Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. *Science* **343**, 1249288, doi:10.1126/science.1249288 (2014).

26 Shulzhenko, N. *et al.* Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nat Med* **17**, 1585-1593, doi:10.1038/nm.2505 (2011).

27 Johnson, A. L. *et al.* Themis is a member of a new metazoan gene family and is required for the completion of thymocyte positive selection. *Nat Immunol* **10**, 831-839, doi:10.1038/ni.1769 (2009).

28 Miller, J. C. *et al.* Deciphering the transcriptional network of the dendritic cell lineage. *Nat Immunol* **13**, 888-899, doi:10.1038/ni.2370 (2012).

29 Bogunovic, M. *et al.* Origin of the lamina propria dendritic cell network. *Immunity* **31**, 513-525, doi:10.1016/j.immuni.2009.08.010 (2009).

30 Greter, M. *et al.* GM-CSF controls nonlymphoid tissue dendritic cell homeostasis but is dispensable for the differentiation of inflammatory dendritic cells. *Immunity* **36**, 1031-1046, doi:10.1016/j.immuni.2012.03.027 (2012).

31 Greer, R. L. *et al.* Akkermansia muciniphila mediates negative effects of IFNgamma on glucose metabolism. *Nat Commun* **7**, 13329, doi:10.1038/ncomms13329 (2016).

32 Mine, K. L. *et al.* Gene network reconstruction reveals cell cycle and antiviral genes as major drivers of cervical cancer. *Nat Commun* **4**, 1806, doi:10.1038/ncomms2693 (2013).

33 Sehgal, R. & Koltun, W. A. Fibrin glue for the treatment of perineal fistulous Crohn's disease. *Gastroenterology* **138**, 2216-2219, doi:10.1053/j.gastro.2010.04.029 (2010).

34 Morgun, A. *et al.* Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut* **64**, 1732-1743, doi:10.1136/gutjnl-2014-308820 (2015).

35 Dong, X. *et al.* Arsenic exposure and intestinal microbiota in children from Sirajdikhan, Bangladesh. *PLoS One* **12**, e0188487, doi:10.1371/journal.pone.0188487 (2017).

36 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

37 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).

38 Moon, C. *et al.* Vertically transmitted faecal IgA levels determine extra-chromosomal phenotypic variation. *Nature* **521**, 90-93, doi:10.1038/nature14139 (2015).

39 Dong, X. *et al.* Reverse enGENEering of Regulatory Networks from Big Data: A Roadmap for Biologists. *Bioinform Biol Insights* **9**, 61-74, doi:10.4137/BBI.S12467 (2015).

40 Morgun, A. *et al.* Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. *Gut*, doi:10.1136/gutjnl-2014-308820 (2015).

41 Morgun, A. *et al.* Molecular profiling improves diagnoses of rejection and infection in transplanted organs. *Circ Res* **98**, e74-83 (2006).

42 Thakkar, S. G. *et al.* Survival and predictors of outcome in patients with acute leukemia admitted to the intensive care unit. *Cancer* **112**, 2233-2240, doi:10.1002/cncr.23394 (2008).

43 Bair, E. & Tibshirani, R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* **2**, E108, doi:10.1371/journal.pbio.0020108 (2004).

44 Chen, X., Wang, L., Smith, J. D. & Zhang, B. Supervised principal component analysis for gene set enrichment of microarray data with continuous or survival outcomes. *Bioinformatics* **24**, 2474-2481, doi:10.1093/bioinformatics/btn458 (2008).

45 van Houwelingen, H. C., Bruinsma, T., Hart, A. A., Van't Veer, L. J. & Wessels, L. F. Cross-validated Cox regression on microarray gene expression data. *Stat Med* **25**, 3201-3216, doi:10.1002/sim.2353 (2006).

46 Daemen, A. *et al.* A kernel-based integration of genome-wide data for clinical decision support. *Genome Med* **1**, 39, doi:10.1186/gm39 (2009).

47 Dobbin, K. K. & Simon, R. M. Sample size planning for developing classifiers using high-dimensional DNA microarray data. *Biostatistics* **8**, 101-117, doi:10.1093/biostatistics/kxj036 (2007).

48 Dobbin, K. K., Zhao, Y. & Simon, R. M. How large a training set is needed to develop a classifier for microarray data? *Clin Cancer Res* **14**, 108-114, doi:10.1158/1078-0432.CCR-07-0443 (2008).

49 Dobbin, K. K. & Simon, R. M. Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med Genomics* **4**, 31, doi:10.1186/1755-8794-4-31 (2011).

50 Garrett-Mayer, E. Principles of Anticancer Drug Development. *Cancer Drug Discovery and Development*, 674, doi:10.1007/978-1-4419-7358-0 (2011).