

Concept Development for Tissue Analytics in Rheumatoid Arthritis and Lupus

Background

Rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) are relatively common, severe autoimmune diseases. These two diseases represent examples of a larger number of autoimmune diseases such as multiple sclerosis, Crohn's Disease, ulcerative colitis, juvenile diabetes, and psoriasis, among many others that collectively affect millions of Americans. Basic and clinical studies have shown that these diseases share in common abnormalities in adaptive and innate immune function and regulation, resulting in inflammation that destroys end organ function. The specific immunological aberrations and the location of inflammation differ for each disease but all include B and T cell related autoimmunity combined with innate leukocyte (macrophage, neutrophil) mediated inflammation.

The ability to target specific immunological cells or inflammatory mediators (cytokines) has resulted in the first real advances in treatments for these diseases in decades. However, the clinical benefit achieved so far is limited. In some conditions like RA, current biotherapeutic drugs reduce disease activity by approximately half in half the patients, and a majority of the remaining patients respond poorly to all subsequent drugs. In addition, many patients that show initial response to therapy can lose response over time for unknown reasons. In other diseases, like SLE, no effective targeted therapies exist for the most severe forms including CNS lupus and lupus nephritis. The experience in RA and several other disorders shows that targeting immune system inflammation and immune-regulation can lead to successful treatments. However, a major challenge remains to find new targeted therapies that can achieve a high degree of disease activity reduction (i.e., remission) or cure disease, have fewer immunosuppressive side effects and/or offer oral alternatives. There has been a very high failure rate among drug targets identified from studies in mouse models; for example, all targets tested so far for lupus nephritis have failed in human trials. In addition, we need to understand the underlying disease pathobiology in patient subsets in order to determine a logical and rational way to tailor the specific therapeutic mechanism to the proper patient subset. To this end it is critical to study tissue from humans with autoimmune diseases directly to identify the relevant immunological pathways and their regulators in order to reveal new, directly implicated drug targets in humans, and to provide a framework to use the existing drugs in the patients who are most likely to respond.

This concept paper outlines a team science approach to achieve that goal in RA and SLE. The overarching vision is that detailed, comprehensive and integrated studies in human samples would identify key targets (with initial *in vitro* validation) that regulate the pathways that drive the diseases. Major target tissues for RA (the synovial tissue) and SLE (the kidney and skin) can be biopsied, and emerging technologies allow for a detailed analysis of even small amounts of tissue, down to the single cell level.

The approach developed here can be subsequently applied to other autoimmune or inflammatory disorders. The ability to compare across autoimmune diseases, in which similar

immune cells are involved but where their functions and interactions result in distinct inflammatory outcomes, is a central feature of the proposed approach. SLE and RA have a set of common genetic risk alleles and abnormalities in both B and T cell functions. Thus, some disease-associated pathways are likely to be shared. However, distinct genetic risk factors are also present for each disease, and certain changes in cell functions clearly differ between the diseases. This project should identify which pathways are shared, and which differ. The concept is that these pathways function as modules that are differentially regulated in autoimmune diseases. Drugs developed for a particular module are likely to have efficacy in other diseases where similar perturbations of that module are found.

The approach being considered is to establish a consortium of researchers to molecularly deconstruct and compare RA and SLE. The rationale is that molecular analyses of gene expression and signaling in very specific subsets of leukocytes and resident cells in the disease tissue could predict pathological processes that lead to end-organ damage. These gene expression and signaling programs are modular in the immune system and their patterns of activity define molecular phenotypes of autoimmune disease. These analyses should define the molecular heterogeneity that would stratify patients, leading to improved application of existing drugs, as well as identification of targets for new drugs. Identifying immune modules that are active in subsets of patients with a disease should allow individualized patient assessment. Conversely, certain modules may be active in several autoimmune diseases, allowing rapid expansion of drug application across indications.

Past analyses of whole blood or tissue have revealed generalized inflammation, but it is now appreciated that these mixed cell preparations typically fail to reveal the pathological state of particular subpopulations of cells. Therefore, disease deconstruction could be achieved by systematic molecular analysis of highly refined subsets of immune and resident cells that are responsible for disease inflammation and pathology in the blood and in the affected end organs. This would include purified functional subsets of T cells, B cells, dendritic cells and macrophages from blood and single cell analyses of cells in tissue biopsies. Expression profiling, phosphoproteomic analysis, DNA methylation, and other epigenetic interrogation of leukocytes from blood and relevant tissue cells could be conducted in an integrated approach which incorporates individual patient genotype, microbiome characteristics, and clinical data from carefully selected, annotated cohorts.

Careful consideration should be given to the issue of which patient populations to study. Early disease in RA offers insight into mechanisms initiating disease. On the other hand, fully developed disease constitutes the major clinical problem, and there is striking heterogeneity in the response of individual patients to current biologic therapies. These differences in therapeutic response offer an opportunity to understand differences in disease mechanism, as well as to develop more personalized approaches to treatment. The molecular modular analysis of these patients could identify the differences between responders and non-responders relevant to predicting response and defining pathways that may be targeted in non-responders.

Highly informative lupus cohorts could also be examined and compared. For example, lupus nephritis patients could be examined just prior to the onset of therapy for nephritis in order to define the molecular pathways in relevant blood leukocyte populations in this severe condition. The findings from blood could be compared with single cell analysis from kidney biopsies. A second approach to lupus could be to examine a separate cohort without nephritis but with skin involvement, to allow simultaneous examination and comparison of leukocytes from blood and skin.

A consortium of researchers would be essential for these studies because they require a broad range of clinical, experimental, technical and analytical expertise. Indeed, success in each of these areas would depend on extensive collaborative efforts both within and across disciplines and should facilitate investigation of additional populations of interest. Potential cutting edge experimental techniques could include 1) deep RNA-seq of multiple distinct immune cell subsets, including at the single cell level, from blood and disease tissue, 2) single cell RNA sequencing of tissue-resident cells from biopsies, 3) evaluation of epigenetic changes including methylation, histone modifications, DNase hypersensitivity in specific cell subsets of interest, 4) multiparameter cell phenotyping and phospho-flow analysis using mass spec approaches such as Cy-TOF, 5) global phosphoproteome analysis of selected cell types, and 6) immunoglobulin repertoire analysis by application of single cell RNA-seq from circulating plasmablasts. These data, along with clinical laboratory measures and data on the microbiome could then be mined using a systems approach. Platforms would need to be developed to allow data sharing and permit the incorporation of additional data such as microbiome, metabolomic and proteomic studies on biological samples that would be stored for future use on study subjects.

The success of this project will depend on careful standardization of procedures to minimize technical variability. This is particularly important in the analysis of tissue. For example, the method of tissue biopsy and the disaggregation into single cell suspension for high-throughput single cell analysis will alter gene expression. Thus, a critical initial goal will be to understand the sources of non-biological variability by comparing different approaches and minimize them by standardization. The lessons learned will pave the way for future studies of tissues from patients with other autoimmune diseases.

Thus, this consortium would deliver:

1. An integrated data set of changes at the molecular level by extensive profiling of gene expression and signaling in immune and tissue-resident cells in RA and SLE, available for exploration of specific potential targets.
2. An in-depth analysis of pathways active in target tissues, as well as blood, in RA synovium and SLE kidney tissue and skin, including identification of likely causative pathways in RA through the analysis of early disease populations.
3. Characterization of immune modules and how they can be used to understand differences between autoimmune diseases, between early and established disease and between responders and non-responders. Such data would likely advance the effectiveness of therapeutic targeting strategies in different diseases.

4. Identification of changes in circulating cells in blood reflecting activation of specific pathways in the tissues that can be used to improve targeting and serve as surrogate biomarkers.
5. Identification of changes in circulating cells that predict response to specific therapies, using the responder/non-responder comparison, as an enrichment strategy.
6. Development of the computational tools to permit the systematic approach to integrating the datasets into pathways, which would not otherwise be available.
7. A roadmap for how to apply contemporary molecular technology to similarly assess therapeutic strategies in additional autoimmune diseases of interest.
8. Initial “de-risking” of the molecular markers and networks that are dysregulated as disease progresses or that correlate with sensitivity and response to treatment by *in vitro* functional assays and RNAi methods for validation.

The research strategy would need to employ emerging technologies to sites of tissue injury in autoimmune diseases to develop a systems-level understanding of the regulatory pathways that lead to damage. This will require increasing the capacity to obtain target tissue and development of protocols for applying new analytic technologies to study resident cell from end organs and tissues.

Patient cohorts

The initial systems analysis of RA and SLE would benefit from restricting the analysis to a more homogeneous cohort in order to reduce variables and increase the likelihood of obtaining meaningful data. For example, the population of patients starting TNF neutralization therapy would be the most abundant cohort whereas analysis of patients with early RA would reveal the initial, inciting pathways. Identifying these early causative pathways would provide the opportunity to identify targets involved in mediating or regulating the disease initiating or disease progression pathways. Indeed, current studies demonstrate that early intervention is more likely to induce a durable remission, but currently that only occurs in a small number of patients. A better understanding of the pathways in early disease is likely to result in therapies that enable one to prevent disease progression or cure the disease at its presentation in most patients. For SLE, a cohort that is homogeneous yet with life threatening disease, and a group where therapeutic intervention could have a most profound effect, would be new onset nephritis. Patients with established SLE who present with nephritis could be examined just prior to starting therapy for nephritis. Renal biopsies, obtained as medically indicated, could be subjected to single cell expression analyses. Blood from these patients could be subjected to the same leukocyte subset separation and molecular interrogation for the RA and the SLE cohorts. However, while not requiring large numbers of samples, the criteria must be designed so that there are sufficient numbers of patients with active disease who are willing and eligible.

The analyses of the participants in these initial cohorts in RA and SLE would be evaluated in order to validate the analytic technologies, as applied to blood and tissue cells, to define the signal-to-noise parameters for the assays and to identify specific areas of focus for the analyses conducted on additional cohorts in the next phase. Once the quality of the data is sufficient to

produce a meaningful systems analysis, the analysis could be advanced to other cohorts or patient subsets.

Exploring the heterogeneity between patients would require shifting from disease-specific analysis to patient stratification analysis. The initial studies and completed standardization of the methodologies would enable power calculations used to define the size of the cohorts, and the number of variables allowed within the cohort, needed to get meaningful between-patient data. For example, differences in disease pathways could be investigated in the following RA patient variables, as power calculations and budget allow:

- Response or no response to new DMARD
- Early vs. established disease
- Previous DMARD or not
- The new DMARD being started

For SLE patients, variables that could be investigated could include:

- Skin versus kidney disease
- responder/non-responder analysis (with longitudinal follow-up)
- ethnicity
- antibody status
- nephritis vs. no nephritis (blood only)
- Disease flare vs. quiescent disease (blood only)

A control cohort should be also be analyzed as a point of reference. Ideally, subjects with no autoimmune diseases should be matched by demographic and HLA Class II status. Blood, urine and stool samples could be obtained at one time point (no biopsies would be performed).

Once standardized methodology is established and the analysis platforms employed, then the approach to deconstruct autoimmune disease as modular can readily be applied to other autoimmune and inflammatory conditions. Further, the comparisons across diseases may reveal treatments that are effective in one disease and likely to work in other diseases and correspondingly those targets and therapeutics predicted to be different among diseases. The cross-disease comparison, like the comparison between diseases and normal, would facilitate identification of the distinguishing molecular features of each disease state.

Tissue Acquisition

Developing the infrastructure to capture tissue from patients, before end-stage damage, is a central feature of this project. However, this also represents the project's greatest challenge. Time and resources would be required to establish this capacity. Tissue samples from different sources might be analyzed to provide an indication as to the relative quality and as to

differences between, for example, active disease, end-stage disease (surgical), and post-mortem tissue. These alternates might also be the only source of control tissues.

In the initial phase, a variety of tissue collection and analytic techniques would be applied and compared. For RA, this might include synovial sample collection techniques (shaver vs. guided needle biopsy), types of samples (biopsy vs. surgical), and different processing techniques that use whole tissue (histology or laser capture RNAseq) vs. disaggregated (high-throughput single cell RNAseq or CyTOF). Tissue processing would be standardized for subsequent scale-up of a few analytics. As part of establishing a baseline for this approach reference or “control” tissues would be obtained and processed for comparison.

Two methodologies are currently being used at some institutions for synovial biopsy of consenting RA patients: arthroscopy using a small-bore short arthroscope (1.9 mm to 2.7 mm) and synovial biopsies obtained using a motorized shaver. This requires extensive training. Needle aspirates may be suitable for initial analyses during a training and standardization period. Alternatively ultra-sound guided biopsies will be considered as a preferred source of tissue from small joints in patients with early or established rheumatoid arthritis.

A strategy to define the number and frequency of biopsies that would be done in early RA will have to be defined. For established RA patients samples could be collected in the context of a clinical trial where pre- and post-biopsy samples can be analyzed to evaluate the molecular effect of a therapeutic agent. The precise timing of the second biopsy depends on the agent, but would generally be 3-4 weeks after starting therapy.

For the lupus patients with skin involvement, the expectation is that two skin biopsies could be done with concurrently collected blood samples.

Tissue Analysis

The fundamental challenge in studies of human immunity is the extensive inter-individual variation of immune responses. To correctly classify patients into sub-groups with similar immune properties, future studies need to take advantage of more comprehensive and unbiased profiling strategies. The combined expression and signaling analysis would reveal in detail the state of activation and the expression programs in the relevant cell subsets as a means to predict their functional activities. Abnormal, excessive or polarized functions are likely to underlie the cellular states that correlate with autoimmune and inflammatory pathology.

A critical and essential feature of this project is the focus on the state of cells *in vivo* in blood and in the affected tissues and organs. For example, targeted analysis of the most important adaptive and innate leukocyte subpopulations in the peripheral blood would identify the activities and alterations in the key pathways and regulators of these pathways. Some of the main subpopulations of leukocytes may include, CD4+ T cells (Treg, Th17, naïve, effector memory, central memory), CD8+ T cells (naïve, memory), B cells (naïve, transitional, IgM and switched memory, marginal zone-like, regulatory, plasmablasts/plasma cells),

monocytes/macrophages (M1, M2), and dendritic cells (CD103, CD11c, CD11b). Besides leukocytes, tissue resident cells such as synovial fibroblasts (synoviocytes) in RA are highly implicated as major contributors to joint damage. These mesenchymal cells can be readily separated from leukocytes by expression profiling and may be examined separately to elucidate molecular phenotypes. Since biopsies from synovium (RA) and kidney (SLE) are typically small and contain limited numbers of cells, separation of leukocytes into major subpopulations would be difficult to carry out. Therefore, the molecular analysis may require the use of emerging technologies such as RNA sequencing at single cell level to determine gene expression profiles.

Currently, a number of cutting-edge technologies are available for analysis of molecular phenotypes; for example, RNA sequencing in purified cell populations, Cy-TOF mass cytometry, single cell RNA-seq, global phosphoproteomics, serum/urine proteomics and metabolomics, epigenetic profiling (methylation and histone modification), and microbiome analysis. It is anticipated that with the rapid advances of technologies, newer methodologies in the coming few years may also be adopted in the RA/SLE studies. The analyses may need to be carried out in stages, with an initial pilot phase performed on fewer cells from fewer tissue samples to test the feasibility and establish the standardized protocol.

While each of the individual molecular measurements is valuable, an integrated analysis would be more useful for inferring connected pathways and building networks of gene expression and signaling in the relevant cell subsets. The genome-wide level data sets that reflect pathways would define modules active in particular cell types. Such analyses from individual cells or cell subsets when integrated with clinical data on disease state would facilitate the discovery of more reliable predictors of disease pathogenesis and response to therapy, and identify new pathways and targets for drug development.

Functional validation of nodes predicted to be critical in disease

The molecular analyses would reveal markers and networks that are dysregulated as disease progresses or that correlate with sensitivity to treatment. However, additional evidence would be needed to identify critical nodes for therapeutic targeting. Some of the nodes inferred from disease networks will have a known function in the immune system (from mouse or human genetic studies) that helps explain their role in disease. In contrast, the functions of unannotated nodes would need to be studied *in vitro* and in animals. As a first step in validating the molecular markers and networks that are dysregulated, an *in vitro* system could be used to study the functions of the top ranked nodes (known and novel) that are predicted to drive disease. For example, RNAi-mediated knockdown and overexpression (using wild type, dominant-negative, activated alleles when possible) of each gene may be performed in the immune cell types and activation conditions (e.g., activated T cells, B cells or monocytes/macrophages/DCs) predicted to be relevant for each gene's function, followed by expression profiling (e.g., RNA-seq). Changes in the expression profiles following activation of the relevant cells in the presence or absence of RNA inhibition, should allow confirmation of the role of the target in relevant expression pathways. A more novel approach would be to apply similar methods for reversing abnormal phenotypes of tissue resident cells that are activated in

a disease. For example, RA synoviocytes maintain a hypertrophic phenotype *in vitro*. RNAi-mediated knockdown could be used to test whether pathways identified by this project as activated in these cells in RA synovium *in vivo* regulate the hypertrophic phenotype. New functions would thus be discovered for each node, allowing a more mechanistic understanding of the node within the human immune response. Finally, the *in vitro* derived node-specific signatures could be compared to differential expression patterns observed in patient cells to assess whether the node is likely dysregulated in patients.

Genotyping of patients

There has been a revolution in our understanding of the genetic basis of complex traits such as RA and SLE. Just a few years ago, only a handful of genetic factors were known to contribute to the risk of RA, SLE and other autoimmune diseases. Now, there are hundreds of alleles that contribute to these diseases, with empirical evidence that hundreds (if not more) of alleles remain to be discovered. On their own, these associations provide little insight into disease pathogenesis. However, when integrated with detailed molecular profiles – such as those described in this proposal – highly informative patterns emerge. For example, integrating SNP associations from genome-wide association studies (GWAS) with gene expression or epigenetic profiles of immune cell subsets could implicate specific cell types (e.g., CD4+ T cells in RA, CD19+ B cells in SLE) and identify the causative pathways altered by at-risk or protective alleles.

Today, it is cost-effective to genotype large sample collections with commercial arrays (e.g., GWAS + exome chip) to capture the vast majority of alleles that are present at low-frequency (~1%) or are common (>5%) in the general population. Advances in sequencing technology will expand genetic analysis to include all allele frequency classes – including rare variants that are private to individual families – as well as other types of genetic variants (e.g., indels) that are not captured by contemporary genotyping arrays.

Data collection, storage and analysis

A crucial component to the success of this project is a highly effective infrastructure for data analysis and bioinformatics. The success of this component will depend critically on the ability to develop new computational approaches for data integration for model systems analysis. It is expected that this component would work collaboratively with other components of the consortium in order to be responsive to the clinical questions as they arise, and with groups outside of the consortium to make data analysis and interpretation available.

Secure Storage

The infrastructure needs to have the hardware capacity to provide secure multi-site storage of data, accessible to consortium members via secure intra-net. Data access should be feasible through secure intra-net portals for data upload and download. It should have the capacity for storing the most low level forms of raw data (e.g. sequence reads) as they are generated, and should be the last line of defense in data loss or corruption.

Compile and curate public data

A substantial quantity of highly relevant public data is already available that might be germane and has the potential to augment analysis and interpretation efforts of consortium data. For example, RNA-seq data generated by GTEX might serve as an important reference data set to calibrate single-cell RNA-seq data generated by this project. Genetic data for rheumatoid arthritis and SLE from genome-wide association studies and direct sequencing could be compiled to permit linkage of allelic variants with function. Data from the ENCODE and Roadmap Epigenomics projects could be imported to allow linking of non-coding allelic variants to expression data from critical cell types. These data sets could be obtained, curated, and harmonized with data sets to facilitate more complex queries and integrative analyses as needed. In addition, there could also be effort to compile relevant literature from within the field, and employ statistical text-mining as needed to (1) interpret generated data and (2) to link public data to original publications in which it is described in detail.

Conduct genome-wide data analysis

There should also be expertise in analyzing genomic data, and in particular next-generation sequence data. Many of the technologies, such as single cell RNA-seq and whole-genome sequencing, will require expertise with managing short sequence reads generated through next-generation sequencing data. While these technologies are continuously evolving, intimate familiarity with read-mapping technologies, peak-calling for epigenetic peaks, variant calling for genome-sequencing data is critical to the success of this center.

Conduct low-level immunological data analysis

Given the critical role of immunological assays to this project, it will be crucial that there is capacity to analyze data to quantify immunological variables of individual human subjects. This would entail analyzing cytometric data with automated analytical methods. In addition, equally important, would be the capacity to analyze data generated through Cy-TOF and other next-generation cytometric data sets, that has the capacity to look at >100 parameters simultaneously.

Advanced statistical data analysis

There is also the need to have the capacity to conduct advanced statistical analyses of the high-throughput data sets generated through the project. This includes established techniques for statistical genetics and transcriptomics (e.g. principal components analysis, data normalization, smoothing, clustering, association testing, eQTL mapping, etc.). In many instances novel statistical methods may need to be devised considering the novel nature of this data set, and there should be adequately qualified personnel to this end.

Conduct innovative high-level integrative analysis

A major challenge for this project would be to integrate the data for analyses. This will require a global view of genotype, transcription, function, and immunological and clinical parameters of individual subjects. The goal is to define immune modules active in specific cell types and how these differ between diseases. Conducting these analyses in aggregate will require high-performance computing, as well as familiarity with the nuances of each of the generated data

sets. Importantly, bioinformatics analyses can be influenced by failure to consider important confounders or by dependency on inaccurate parametric models.

Data access and interrogation portal

It is expected that the data sets will be made more broadly available to the public through coordination with sponsoring agencies and consortium members via a public portal.