

The following is an excerpt from the Project Plan developed by the AMP RA/SLE Steering Committee in late 2013. However, the Committee recognized that the relevant technologies are evolving, and the Project Plan explicitly encouraged innovative ideas from applicants. *Potential applicants responding to the NIH AMP RA/SLE Funding Opportunity Announcements should NOT take the Project Plan as a blueprint. It is NOT meant to proscribe the research.* Rather, it should be seen as a working draft, and both NIH and the Steering Committee welcome novel proposals to modify and improve the Project Plan.

**Accelerating Medicines Partnership Detailed Research Plan  
RA, SLE and Related Autoimmune Disorders Steering Committee**

**Contents**

Section 0: Disease Context and Case for Action ..... 3

Section I: Project Overview ..... 5

Section II: Scientific Design ..... 8

Section III: Project Management ..... 26

Section IV: Timeline, Milestones and Deliverables ..... 28

## **Section 0: Disease Context and Case for Action**

### **Disease Context and Case for Action**

Rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) are relatively common, autoimmune diseases which can have severe manifestations. These two diseases are representative examples of a larger number of diseases with underlying autoimmune pathology such as multiple sclerosis, Crohn's disease, ulcerative colitis, juvenile diabetes, and psoriasis, among many others. 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 adaptive (B and T cell)-related autoimmunity combined with innate (macrophage, neutrophil, innate lymphocyte)-mediated inflammation.

Despite significant research progress on the fundamental mechanisms of autoimmunity that operate in experimental mouse models, our basic understanding of human autoimmune disorders remains rudimentary, in part due to the limited availability of clinical samples needed to study the complex nature of immune regulation. Research efforts focused on disease tissue in humans have not, in general, been well coordinated. The field could benefit greatly from a coordinated integration of basic research with the study of clinically well-defined patient cohorts, at the same time bringing cutting edge analytic and bioinformatic approaches to bear on the problem. This is among the most important research goals that could enable advances in the treatment and management of these disorders.

In recent years, the ability to target specific immunological cells or inflammatory mediators (e.g. cytokines) has resulted in the first real advances in treatments for these diseases in decades. However, the clinical benefit achieved is limited. In some conditions like RA, current biotherapeutic drugs reduce disease activity by approximately 50% in about half of the patients, with 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 reasons that are not well understood. In SLE, no effective targeted therapies exist for the most severe forms of the disease including CNS lupus and lupus nephritis. So while currently available therapies that target the immune can lead to successful treatments, a major challenge remains to find new targeted therapies that can achieve a high degree of reduction in disease activity (i.e., remission) or even cure the disease. There is also need for new treatments that have fewer immunosuppressive side effects and/or offer oral alternatives to biologics. 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. 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. As a result, there is a critical need to study tissue from humans with autoimmune diseases directly to identify the aberrant immunological pathways and their regulators in order

to reveal new, directly implicated drug targets in humans, and provide a framework for how best to apply the existing drugs to the patients most likely to respond.

This proposal outlines an approach to achieve that goal in RA and SLE. 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 overarching vision is that these detailed studies would identify key targets (with initial *in vitro* validation) that regulate the pathways that drive the diseases.

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. The approach developed here can be subsequently applied to other autoimmune or inflammatory disorders. 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 abnormalities in cell functions clearly differ between the diseases. This project should clarify 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.

## Section I: Project Overview

The goal of this partnership is to molecularly deconstruct and compare two major autoimmune diseases, rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). Autoimmune diseases offer a special opportunity for molecular deconstruction because the majority of the complex cellular components of the immune system are available for study in the peripheral blood. In addition, the 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 overarching vision is that these detailed studies will identify and validate (*in vitro*) key targets that regulate the pathways that drive the diseases.

Autoimmune diseases share a common set of genetic risk factors. In addition, abnormalities have been described in many of the same sets of cell types (e.g., T cells) across these diseases. Furthermore, a treatment targeting a particular pathway that is developed for one autoimmune disease often has proved effective in another. Thus, in both pathogenesis and treatment, there are commonalities across several autoimmune diseases. However, the patient characteristics and organ involvement vary dramatically, and certain of the autoimmune diseases respond better to treatments targeting different pathways. Our ability to predict which pathways to target to cure which disease – who will respond to what therapy - remains limited because our understanding of differences in pathogenic mechanisms across diseases remains rudimentary.

The primary hypothesis is that molecular analyses of gene expression and signaling in very specific subsets of leukocytes and resident cells in the disease tissue predicts 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 will define the molecular heterogeneity that will 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 will allow individualized patient assessment and enrichment. 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 will 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 organ. This will 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 will be conducted in an integrated approach which incorporates individual patient genotype, microbiome characteristics, and clinical data from carefully selected, annotated cohorts. This data will facilitate targeted drug development as

well as inform the application of existing and future therapies to appropriate patient populations.

Careful consideration has been 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 these 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 will 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 also will be examined and compared. Lupus nephritis patients will 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 will be compared with single cell analysis from kidney biopsies. A second approach to lupus would be to examine a separate cohort without nephritis but with skin involvement, to allow simultaneous examination and comparison of leukocytes from blood and skin.

The combined analyses of blood and involved tissue in RA and lupus patients, employing genome-wide analytics performed in highly refined relevant cell subsets or single cells, will represent the first time such molecular disease deconstruction has been performed. The analyses of pathways, signatures and networks will provide insight into key nodes and regulators of expression and signaling that are disease or treatment response related. Candidate nodes and the products that regulate them will be tested in vitro to validate their function in the critical, disease relevant pathways.

A consortium of researchers (whether assembled before application by investigators or integrated after funding by the sponsor) 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 will depend on extensive collaborative efforts both within and across disciplines. This project will study cohorts of both RA and SLE patients, but is designed in part to facilitate investigation of additional populations of interest in the future. Potential cutting edge experimental techniques would 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 CyTOF, 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 would be mined using a systems approach. Platforms will be developed to allow data sharing and will permit the incorporation of additional data such as microbiome, metabolomic and proteomic studies on biological samples that will 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 could affect gene expression. Thus, a critical goal of the initial phase 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 partnership will 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 potential 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 will likely advance the effectiveness of therapeutic targeting strategies in different diseases.
4. Identification of changes in circulating cells in blood that reflects 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 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 inflammatory diseases of interest.
8. Initial validation of the molecular targets and networks that are dysregulated as disease progresses or that correlate with sensitivity and response to treatment using RNAi methods applied to *in vitro* functional assays.

The Accelerating Medicines Partnership's RA/SLE proposal greatly exceeds the scope of a project that could be undertaken by any single entity on its own. The unique aspects of this project that require a partnership are as follows:

1. This project will utilize technologies that are not generally available. The focus on 'omic level data from single cells (in tissue) or highly purified subsets (in blood) is fundamentally different than previous systems approaches (e.g., expression profiling of PBMC or whole tissue) in autoimmunity.
2. No other mechanism could combine and integrate the different technologies as proposed here.

3. The partnership is required to collect sufficient numbers of cohorts for the proposed studies. The acquisition of tissue (synovium and kidney) for molecular analysis would be extremely difficult otherwise.
4. The partnership will develop the computational tools to permit the systematic approach to integrating the datasets into pathways, which will not otherwise be available.
5. The partnership approach allows all participants access to the primary datasets, including the detailed results of the initial in vitro target validation.

## Section II: Scientific Design

### Rationale

The goal of this project is to apply emerging technologies to sites of tissue injury in autoimmune diseases so as to develop a systematic understanding of the regulatory pathways that lead to damage. The partnership will develop a new approach to understanding the similarities and differences in pathogenic pathways, and hence valid targets, in autoimmune diseases. Disease deconstruction will be achieved by systematic molecular analysis of highly refined subsets of cells that are responsible for disease inflammation and pathology in the blood and in the affected end organ tissues.

Conducting these studies in RA has practical advantages. The primary factor is that the site of inflammation, the synovium, is accessible. By judiciously applying the technologies outlined above, we can readily identify the activated pathways in specific cells in the inflamed tissue. Disease-associated circulating cells are also accessible in the peripheral blood compartment. By applying the same techniques to certain carefully defined subsets of cells in the circulation, we can identify both changes in the blood that are associated with specific activated pathways in the tissue (biomarkers for the mechanisms driving synovitis), as well as other changes that might be linked to causative mechanisms. By studying patients with recent disease onset, we can focus on pathways that are likely causative, rather than a result of prolonged tissue damage. By studying patients with established disease before and after starting a new therapy, we can identify molecular signatures that distinguish responders from non-responders at the level of specific cell types.

Thus, the partnership will deconstruct RA as a prototypical autoimmune disease, with well-defined responders and non-responders, and a significant unmet need. This will start with an initial phase to identify which assays best detect altered pathways, which will then be analyzed in the full cohorts. However, the goal is not to understand RA *per se*, but to serve as a footprint for the modular deconstruction of other autoimmune diseases as well as non-autoimmune diseases that have underlying inflammatory mechanisms. Therefore, we include a parallel analysis of a second autoimmune disease, systemic lupus erythematosus (SLE). Many patients currently suffer greatly or die from disease complications (such as renal disease) or from the toxic effects of the available drugs, such as high dose corticosteroids. Thus, SLE represents a major unmet medical need. The scientific advantages are that disease manifestations are quite apparent in the blood, and involved tissue is routinely biopsied (i.e., medically indicated kidney



biopsy in patients with renal involvement or skin biopsies) and available for molecular evaluation.

SLE and RA have common genetic risk alleles and abnormalities in both B and T cell functions. Thus, some disease-associated pathways are likely to be shared. However, some changes in these cell functions will differ between the diseases (even taking into account different tissues). The partnership will identify which pathways are shared, and which differ. The concept is that these pathways will act as modules that are differentially activated in autoimmune diseases.

**Strategy**

This project will require increasing the capacity to obtain target tissue and development of protocols for applying new analytic technologies to study the cells in the tissue. Once a tissue acquisition partnership is organized (Phase 0), two scientific questions will be addressed. In Phase 1, the goal is to develop a map of the molecular pathways active in specific cells in the tissue (and related cells in the circulation). Such a systems analysis does not require large numbers of samples if they are relatively homogeneous. At the same time, this Phase will also test the feasibility of the methodology. Once Phase 1 is complete, the focus will shift to identifying differences between subsets of patients (“stratification”), which will include comparisons of early vs. late disease, responder vs. non-responder, and subsets with different autoantibodies. How many different comparisons can be made will depend on feasibility, power calculations derived from the data acquired in Phase 1, and budget, as described in **Decision Making** in Section III.

**Patient cohorts**

The patient cohorts utilized in each of the three phases are described in **Table 1**. Several existing resources have been identified as outlined in Appendix A.

**Table 1. Patient Cohorts in Phase 0, 1, and 2**

RA/SLE	Stage	Cohort	Tissues	Goal	# Subjects
RA	Phase 0	Homogenous cohort (1 <sup>st</sup> DMARD or TNF IR)	Synovium & Blood	<ul style="list-style-type: none"> <li>Define biopsy method Shaver vs. Needle</li> <li>Demonstrate feasibility of RNA seq</li> </ul>	3
RA	Phase 1	Beginning first DMARD	Synovium & Blood	<ul style="list-style-type: none"> <li>Demonstrate disease specific pathway maps</li> </ul>	~15
RA	Phase 2	1. Beginning anti-TNF therapy 2. Early RA	Synovium & Blood	<ul style="list-style-type: none"> <li>Identify pathways and targets in non-responder populations</li> </ul>	TBD
SLE	Phase 0	Nephritis	Kidney & Blood	<ul style="list-style-type: none"> <li>Demonstrate feasibility of RNA Seq</li> </ul>	
SLE	Phase1	Nephritis	Kidney & Blood	<ul style="list-style-type: none"> <li>Demonstrate disease specific pathway maps</li> </ul>	
SLE	Phase 2	1. New onset Nephritis 2. Non-nephritis (skin?)	Kidney & Blood	<ul style="list-style-type: none"> <li>Identify pathways and targets in non-responder populations</li> </ul>	

## Phase 0

The goal of phase 0 is to establish the method of biopsy collection (in RA) and to demonstrate the feasibility of single cell analysis by RNA-seq in tissue (synovium and kidney). The biggest hurdles of the protocol are the quantity of tissue and homogeneity of tissue procurement, as most US rheumatologists do not perform biopsies. Assuming that US investigators are willing or have some capacity to do biopsy, these investigators should be able to prove that they can provide sample that will provide high quality RNA for downstream analysis. An alternative to this is to use investigators in Europe who do biopsies more frequently. The current state of RA biopsies is moving away from arthroscopic techniques of the knee into ultrasound guided biopsy of small joints. In fact several investigators have generated biobanks that have already confirmed the integrity of the RNA from these samples. The pathology of RA may not be fully represented with the large weight being joints (i.e. knee).

Ideally, one would like to compare RA to healthy control, but that may only be possible with matched tissue from surgery, which implies some sort of damage (e.g. ACL, OA). The project would allow for other types of samples for comparison to help elucidate how the data may vary based on technical differences. This might include comparison of sample collection techniques (e.g. shaver vs. guided needle biopsy), different types of samples (e.g. biopsy vs. surgical), and various whole tissue processing techniques (e.g. histology/laser capture microdissection (LCMD) vs. disaggregated tissues).

Single cell RNA-seq will be utilized as the leading technique to inform which method of biopsy will be incorporated into later stages of the project plan. These approaches along with the relative priority (highest being A, lowest being C) are summarized in **Table 2**.

**Table 2. Summary of Phase 0 samples and analytics**

RA/SLE	Tissue	Source	Processing	Analytics	# Subjects	Comments	Priority
RA	Synovium	Biopsy	Shaver Block	LCMD RNA-seq	3	Combine with immunostain	A
RA	Synovium	Biopsy	Shaver: Disaggregate	RNA-seq	3	Combine with CyTOF	A
RA	Synovium	Biopsy	Needle: Block	LCMD RNA-seq	3	Combine with immunostain	A
RA	Synovium	Biopsy	Needle: Disaggregate	RNA-seq	3	Combine with CyTOF	A
RA	Synovium	Surgical	Disaggregate	RNA-seq	3	Combine with CyTOF	B
RA	Blood	Matched w/ biopsy	PBMC	CyTOF	3		A
RA	Blood	Matched w/ biopsy	Sorted subsets <sup>3</sup>	RNA-seq	3		A
SLE	kidney	Biopsy	Block	LCMD RNA-seq	3	Combine with immunostain	A

SLE	kidney	Biopsy	Disaggregate	RNA-seq	3	Combine with CyTOF	A
SLE	Blood	Matched w/ biopsy	PBMC	CyTOF	3		A
SLE	Blood	Matched w/ biopsy	Sorted subsets	RNA-seq	3		A
Healthy	Synovium	Surgical	Disaggregate	RNA-seq	5	Combine with CyTOF	B
Healthy	Blood		PBMC	CyTOF	3		B
Healthy	Blood		Sorted subsets	RNA-seq	3		B
SLE	Kidney	Biopsy		Phosphoproteome/epigenetics			C
RA	Synovium	Biopsy <sup>1</sup>		Phosphoproteome/epigenetics			C

### Phase 1

The systems analysis of RA and SLE in Phase 1 would benefit from restricting the analysis to cohorts that are narrowly defined so as to be as homogeneous as possible. The criteria must be designed so that there are sufficient numbers of patients with active disease who are willing and eligible. However, this phase does not require large numbers of participants. A target of 15 each for RA and SLE was adopted as a sufficient number to determine whether a meaningful systems analysis could be produced with this approach. The analysis for tissue and blood samples for Phase 1 along with relative priority is described in table 3 and described in greater detail (page 16).

**Table 3. Summary of analytic evaluation from Phase 1 cohorts**

Analytics	Synovium	Bio-fluids			Comments
		Blood	Serum	Urine	
RNA-seq	100+ cells	8+ cell types		?	Disaggregated or LCMD?
Cytometrics (Cy-TOF)		✓			
Phosphoproteome	✓	✓			
Metabolome			✓	✓	
Microbiome				✓	Oral, GU, lung, stool
Cytokines			✓		
Proteome			✓	✓	
Genotype		✓			
Epigenetics	✓	✓			
Ig Repertoire		plasmablasts			Autoantibody profiling

### RA

In order to create a disease map, we need to identify a small core cohort that will serve as the initial look at single cell data, with the idea that a more homogeneous group of patients will reduce variables, so that likelihood of meaningful systems analysis is increased. The population

of patients with active disease in an accessible joint starting therapy on their first DMARD would be relatively homogeneous population. If this proves too restrictive, the criteria would be broadened to include those requiring a change in DMARD (as a sicker cohort, this group is also seen as more likely to agree to participate). Additional criteria might be anti-citrullinated protein autoantibody (ACPA+), a maximum steroid dose at the time, and a minimum number of swollen joints.

### *SLE*

Similarly, in the first phase for SLE, the plan is to produce a map of the disease in involved tissue as well as blood. The initial cohort would include SLE patients with verified disease and nephritis where a kidney biopsy is medically indicated. The major hurdle here is to demonstrate that we can obtain sufficient tissue of adequate quality under this circumstance. If recruitment is insufficient, alternatives include seeking lymph node biopsy, or simply analysis of peripheral blood, because causative disease mechanisms are likely active in these compartments. Cutaneous lupus should not be included because of the difficulty with heterogeneity. Issues for sample processing and analytic techniques are similar to RA.

The analyses of the participants in these initial cohorts in RA and SLE would be evaluated at the end of Phase 1 of the project. As detailed below, this interim phase would be used to validate the analytic technologies, as applied to blood and tissue cells, and to identify specific areas of focus for the analyses conducted on additional cohorts in Phase 2. Advancement into the additional cohorts will be contingent upon demonstrating that the quality of the data is sufficient to produce a meaningful systems analysis.

### Phase 2

For the scale-up phase, the question shifts from disease-specific analysis to patient stratification analysis. Whereas homogeneity within a cohort was important in phase 1, in the second phase the project analyzes the heterogeneity, at the molecular and cellular level, between patients with different characteristics. The work in Phase 1 will also define the signal-to-noise parameters for the assays from the results of the initial phase. Power calculations would be used to define the size of the cohorts, and the number of variable allowed within the cohort, needed to get meaningful between-patient data. The scope would be on the order of 200 patients each in RA and SLE.

Careful patient phenotyping would be essential for proper data analysis and interpretation. Templates and SOPs for collecting data and possibly extracting information directly from the electronic medical record would be needed and linked to individual samples.

### *RA*

Based on lessons learned from the first phase, including availability of subjects willing to undergo biopsy, the second phase would scale up under the broad category of patients starting anti-TNF therapy, with modifications to this based on lessons learned and available budget. Eligible patients could be treated with an anti-TNF inhibitor and then subject to a repeat biopsy at an exact time point to be determined (preferably at an early date, ~3 weeks, after initiation

of treatment) and then followed for ~6months to evaluate the clinical response. RA patients with established disease going on a new DMARD may have a recruitment benefit because patients are in need of the treatment and may also be more amenable to the 2nd biopsy. This approach would provide a comparison of molecular signatures between patients with established RA that do respond or do not respond to anti-TNF therapy.

A potential second cohort would be early RA. The data set in early RA could 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.

Differences in disease pathways could also be investigated in following patient variables, as power calculations and budget allows:

- Response or no response to DMARD
- Effect of prior DMARD
- Duration of disease
- ACPA positive vs. negative
- Need to initiation of biologic therapy

### *SLE*

For SLE, patients with new onset nephritis would comprise a cohort with homogeneous and life threatening disease and would be a group where therapeutic intervention could have a most profound effect. Patients with established SLE who present with nephritis will be examined just prior to starting therapy for nephritis. Renal biopsies, obtained as medically indicated, would be subjected to single cell expression analyses. Blood from these patients will be subjected to the same leukocyte subset separation and molecular interrogation as above for the RA cohorts.

A potential second SLE cohort would consist of patients with skin disease studied at baseline and 6 weeks into therapy. This cohort is chosen for several reasons. First, there is a medical need for better, less immunosuppressive treatments for these patients. Second, the availability of tissue from both this cohort and the SLE nephritis cohort offers an exceptional opportunity to learn if SLE-mediated tissue inflammation and scarring engage the same immune cells and mediators in two different tissue compartments. Third, there is no impediment to a second skin biopsy to understand the molecular and histopathologic correlates of resolving inflammation in SLE skin. Precisely which type of skin disease in SLE would be most informative, and presumably homogeneous, remains to be defined.

- Skin vs. 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)

### **Patient Phenotype**

Careful patient phenotyping will be essential for proper data analysis and interpretation. Templates and SOPs for collecting data and possibly extracting information directly from the electronic medical record will be needed and linked to individual samples. The clinical parameters that will be determined for each patient include:

Clinical datasets for RA patients:

- 1) Disease activity index for the disease (e.g. DAS28)
- 2) Disease duration
- 3) Previous and current medications
- 4) Presence of other autoimmune disease signs and symptoms
- 5) Presence of co-morbid conditions such as cardiovascular disease, pulmonary disease
- 6) Radiographic imaging
- 7) History of epidemiologic factors such as smoking, periodontal disease, environmental exposures
- 8) CRP and presence of RF assessed by nephelometry and anti-CCP antibodies assessed by ELISA using anti-CCP3.1 and anti-CCP2 antibody assays; anti-DNA antibodies
- 9) Demographic determinants of outcomes such as socioeconomic status, education.

Clinical Datasets for SLE patients:

- 1) Disease duration
- 2) Organ-specific disease activity: in the kidney (e.g., glomerular filtration, proteinuria, and biopsy characteristics) and skin (CLASI score and biopsy characteristics)
- 3) Overall disease activity: Systemic lupus erythematosus disease activity index (SLEDAI) and British Isles Lupus Assessment Group (BILAG) scale
- 4) Previous and current medications
- 5) Previous SLE criteria
- 6) Co-morbid conditions
- 7) Autoantibodies, complement levels
- 8) Demographic determinants of outcomes such as socioeconomic status, education.

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 disease and normal, would facilitate identification of both shared immune modules and 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 hurdle. Time and resources would be required to establish this capacity. During this initial period, 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 study of tissue biopsies, the small amount of tissue and cell numbers are limiting. Therefore implementing the emerging single cell RNA sequencing technology would be important. Since at the time of this writing, the technique is just becoming available and has not yet been applied to tissue biopsy samples, a pilot study should be conducted on the initial participants in the Early RA cohort just starting DMARD therapy to assess technical feasibility of its use in tissue biopsies before employing it on a large scale across all biopsy samples.

In the initial phase, a variety of tissue collection and analytic techniques would be applied and compared. The proposed plan would be to compare the results of RNA-seq analyses from fresh, frozen, or fixed tissue and of cells isolated by laser capture microdissection vs. disaggregation of synoviocytes. 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 being considered for synovial biopsy of consenting RA patients to undergo arthroscopy using a small-bore short arthroscope 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.

Early RA biopsies would only be done once. For established RA patients, when possible, samples would 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 weeks after starting therapy.

In some cases, patients with established RA will have total joint replacements, which will serve as a source of synovial tissue. The advantage is that a much greater number of cells can be isolated and studied. The disadvantage is that the samples are exclusively from patients with end stage disease, which is still useful as a comparator for the biopsy samples. Osteoarthritis, avascular necrosis, and other destructive arthropathies can be collected as controls, and normal samples can be obtained from patients with femoral fractures or at the time of autopsy from tissue banks. The tissues would be processed as described below.

For the SLE nephritis cohort, after consent, kidney material would be obtained as part of biopsies done for medical indications, presumably before initiation of therapy directed at the nephritis. Blood would be obtained at the time of the biopsy and at follow up.

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

## **Tissue Analysis**

### Molecular analysis of cell subsets in pathological tissues

Since biopsies from synovium (RA) and kidney (SLE) are small and will contain limited numbers of cells, separation of the main subpopulations of leukocytes would not be carried out. Instead, RNA sequencing to determine gene expression profiles would be performed at the single cell level as this technology is now emerging. The analysis would be carried out in two stages, first as a pilot study to perform single cell RNA sequencing on 1000 cells from each of 3 different synovial samples isolated by different technique (shaver vs. needle biopsy and surgical isolation). Analyses would be performed to determine if the main findings are revealed from the first 100 cells sequenced and if so, analysis of all subsequent samples would utilize 100 cells. Besides the single cell analysis, synovial fibroblasts (synoviocytes) are highly implicated as major contributors to joint damage in RA. These mesenchymal cells are readily separated from leukocytes by expression profiling and would be examined separately to elucidate molecular phenotypes (as outlined below). Multiplex, 3D immunofluorescence staining of tissue from the same tissue sample would be used to characterize the geographic distribution of cells. Additional methods that would be tested for feasibility would include phosphoproteome analysis and genome-wide approaches for analysis of epigenetic marks.

### Molecular analysis of cell subsets in blood

The RA/SLE project would systematically examine relevant circulating cells with a far higher degree of refinement than has been done previously in patients with autoimmune disorders. In order to identify the activities and alterations in the key pathways and regulators of these pathways, the partnership analysis would include targeted analysis of the most important adaptive and innate leukocyte subpopulations in the peripheral blood. Highly multiplexed CyTOF allows examination of many subpopulations of leukocytes, including but not limited to 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). Small populations of rigorously defined, highly purified subpopulations would be sorted by traditional cytometry for RNA-seq, to allow comparisons between expression profiles of circulating cells and those in the tissue, with the goal of identifying differences in expression profiles in circulating cells that correlate with specific changes in the tissue. This will also permit single-cell immunoglobulin repertoire analysis of individual circulating plasmablasts. At this level of resolution, it should be feasible to identify new biomarkers and mechanisms that would not have been detectable in prior profiling studies that are associated with pathways of disease in the tissue obtained from the same individual. The precise cell types to be analyzed will be defined at later stages of the project with investigator input but will be limited to 8 or 12



subtypes. Finally, in addition to the analysis of highly purified cells, whole-blood (or PBMC) sequencing will be included as it will capture global changes that could be used as biomarkers to identify treatment responders/nonresponders, and will facilitate comparison of the results, obtained here with purified cells, with previous studies that included whole blood.

#### Overview of molecular analyses to be conducted

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. This analysis is focused on the state of cells *in vivo* in blood and in the active pathological tissue. While each individual measurement (e.g. DNA sequence, DNA methylation, RNA sequence, immunoglobulin repertoire, protein levels and modifications) is valuable, an integrated analysis would be even 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. As one step in characterizing the molecular targets and networks that are dysregulated as disease progresses or that correlate with sensitivity to treatment, RNAi-mediated knockdown and RNA overexpression methods will be used to evaluate gene targets in *in vitro* cell culture systems. RNA silencing 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) would reveal changes in the expression program implicating the role of the target in relevant expression pathways.

The methods described below represent the leading-edge technologies available at the moment, with the expectation that the results of rapid advances in the coming few years can also be adopted in the proposed RA/SLE studies.

#### RNA sequencing in purified populations

While microarrays have been the workhorse of expression profiling in the past decade, RNA sequencing provides significantly more information, including unannotated genes, spliced isoforms and non-coding RNAs. These new entities, together with more accurate calling of known genes, would lead to entirely new observations of differential transcript expression within specific cells in disease states.

#### CyTOF mass cytometry

This new technology is analogous to flow cytometry but allows >30 simultaneous, quantitative measurements of cellular proteins at the single cell level. This technology will soon become a

standard profiling approach in human immune studies and should be included in the proposed studies.

#### Single cell RNA-seq

Very recently, RNA-seq of single cells has become possible. While this is currently economically feasible only for hundreds of cells, it would allow unprecedented profiling of tissue-derived cells, such as biopsies of synovium, kidney and skin where sample is limited. In addition this technique can be applied to circulating peripheral blood plasmablasts to identify the immunoglobulin repertoire in RA and SLE to determine the type and nature of the autoimmune response presented in each patient.

#### Global phosphoproteomics

The precision, speed and dynamic range of mass spectrometers now enables a global view of phosphorylation events with relatively small cell numbers. The opportunity is to discover unanticipated regulatory pathways in tissue resident cells (e.g., synoviocytes). The challenge here is in isolating and processing cells in a sufficient manner to maintain tissue-specific phosphorylation profiles. This will require careful method development and QC, and can be verified for some known proteins by anti-phospho flow cytometry. This is a rapidly evolving technology, and deployment will depend on the state of the art in producing meaningful data from limited numbers of cells. Another approach that could be used is to stimulate specific cellular pathways *ex vivo* to induce coherent responses that can help stratify individuals. These data would provide a basis for explaining the gene expression profiles found in subsets of cells, and in turn, candidate targets in these pathways.

#### Serum/urine proteomics and metabolomics

Historically, it has been very challenging to quantify components within the complex mixture of serum proteins due to highly abundant proteins dominating the signal in the mass spectrometer. Recent advances in removing abundant proteins and fractionating complex mixtures now allows detection of thousands of proteins. This provides the first opportunity to properly profile serum and other fluids.

#### Epigenetic profiling

An important question in analyzing gene expression is what explains stable programs of gene expression. Advances in methylation, DNase hypersensitivity and histone mark measurements have enabled genome-wide profiling, as demonstrated extensively by the ENCODE project. These studies would complement RNA-seq by providing additional clues to understand how gene expression patterns are controlled

#### Functional validation of nodes predicted to be critical in disease

The proposed analyses would reveal molecular 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. *In vitro* cellular systems could be used to study the functions of the top nodes, both known and novel, that are predicted to drive disease. RNAi-mediated knockdown and overexpression using wild type, dominant-negative, activated alleles when possible, of each gene would 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 program following activation of the relevant cells in the presence or absence of RNA inhibition, should confirm the role of the target in relevant expression pathways. 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 would 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 disorders. Now, there are hundreds of alleles that contribute to these diseases, with empirical evidence that hundreds, if not more, 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 may 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. The cost of whole genome sequencing remains moderately expensive (~\$4,000 per genome), although costs continue to drop at a remarkable rate.

We propose to perform comprehensive genotyping on all patient samples using commercial genotyping arrays. DNA samples from peripheral blood and synovial tissue would be stored so that whole genome sequencing, or even targeted sequencing of T cell or B cell receptors, can be performed in the future, as the technology improves and costs continue to drop. Genotyping would be performed on unfractionated cells derived from peripheral blood, as there is little evidence that somatic mutations in individual subsets of cells are a major determinant of disease.

## Microbiome

There are data suggesting that the autoimmunity of RA is may be initiated outside of the joints. In particular, recent evidence suggests that the disease may originate at a mucosal surface prior to the first signs and symptoms. The primary sites that have been suggested are the oral cavity, the gut, the GU tract and the lung. Mucosal surfaces, including the lung, which was once thought to be sterile, are home to multiple organisms that can modulate inflammation and immunity through multiple mechanisms (e.g. metabolome changes, molecular mimicry, disruption of barrier function). Therefore, it is possible that these microorganisms are playing a role in triggering autoimmunity. Microorganisms that have been implicated in RA pathophysiology include *Porphyromonas gingivals*, which appears to lead to citrullination of human tissue at sites of inflammation, as well as *Prevotella* and *Leptotrichia* species. Because of this, assessing microorganisms at the mucosal surface using established collection and culture-independent techniques may lead to identification of an organism or organisms associated with the causal initiation or propagation. Once identified, such organism(s) can be further studied to understand how they are mechanistically related to RA and perhaps ultimately lead to the identification of novel targets for the treatment of RA and SLE.

## **Sample collection and processing**

The initial phase of this project will determine the feasibility of applying newly emerging analytic techniques to study tissue, blood and other samples from participants with RA or SLE. The final protocol for Phase 2, which will analyze differences between patients with a disease, with the goal of stratification for improved treatment responses, will be designed once the feasibility, signal-to-noise, and expense are known. The following provide options that detail what samples could be collected. For some of these, the expectation is that there might not be adequate resources to complete all desired analyses, and some sample material may be banked for future study.

## Tissue specimens

The tissues will be processed as described below.

- 1) A sufficient amount of tissue will be disaggregated, fixed, and shipped to a central analytic laboratory.
- 2) A limited number of fragments will be embedded in OCT and snap frozen in liquid nitrogen for immunohistological studies.
- 3) The remaining tissue samples will be used for additional analyses (e.g. phosphoproteome analysis, CyTOF, etc.) as prioritized by the Steering Committee.

## Blood and Urine

- 1) The following samples will be collected on all subjects: DNA for genotyping and genomic sequencing at enrollment only– one 10cc EDTA tube. This will yield 200-300ug of purified genomic DNA. Whole blood can be shipped overnight, or if needed frozen and batched shipped to the processing laboratory
- 2) One 10cc serum separator tube for serum aliquoting and storage at all time points. Centrifugation and aliquoting and freezing of samples at site of blood draw would be

optimal. Spinning of tube with overnight shipping to processing lab also a reasonable option.

- 3) One 10cc EDTA tube for plasma aliquoting and storage at all time points. NOTE: buffy coat from these tubes could be used as a source of neutrophils. Feasibility will likely require overnight shipping to central processing lab, although processing and aliquoting at collection site is ideal.
- 4) Three 10 cc heparinized tubes for preparation and storage of viable PBMC – all time points. PBMCs should be frozen using protocols optimized for phosphoflow/CyTOF assays. This will also permit future cell sorting/separation for epigenetic studies of targeted cell subsets as well as more comprehensive phosphoprotein analyses as technology advances. While isolation of cells on site is optimal, overnight shipping to a central processing facility will likely result in more uniform QC.
- 5) One 2.5cc PAXgene Tube for total RNA – all time points – store locally at -20C and batch ship.
- 6) 3 cc whole blood tube for fixing cells for single cell RNA expression analysis. All time points.
- 7) 20 cc urine collection at all time points, freeze -20 on site, aliquot if feasible, and ship to central lab. NOTE, first morning collection may have some advantages for metabolomic studies, to be determined.

#### Microbiome samples

- 1) Fecal sample collection at baseline within 24 hours prior to baseline and follow-up visits. Sampling kits and methodologies for sample collection instructions will be provided to subjects for home collection. Details are provided in sections 7-12 to 1-17 of the Human Microbiome Project Manual of Procedures (Appendix 1)
- 2) Sputum collection at each study visit under supervision of recruitment coordinators. Briefly, induced sputum using inhalation of 5% hypertonic saline for 15-30 minutes, with oral wash prior to inhalation, and expectoration of sputum after oral rinse and drying to reduce oral contamination. Typically sputum is homogenized, evaluated for cell count to determine if contaminated (<10 epithelial cells per HPF = not contaminated), then snap frozen and stored at minus 80 C.
- 3) Oral/periodontal sampling at each study visit. The approach to this sampling will involve swabs of subgingival tissue and storage in standard medium as recommended (Appendix 1, 7-3- to 7-7). The specific sites to be sampled remain to be determined after further discussion with outside experts and the members of the partnership.

#### Storage and biorepositories

RA patients who consent to this procedure will undergo synovial biopsy by one of two techniques, which will be compared in the initial phase of the project. One option is arthroscopy using a small-bore short arthroscope (1.9 mm to 2.7 mm) with a motorized shaver. This will provide approximate 0.5-1 gm of tissue in small pieces. The other option is ultrasound-guided needle biopsy. This will yield less tissue, but has the benefit of obtaining tissue from small joints. With the needle aspirate, there is greater chance of sampling bias than with the shaver, but the ultrasound should guide the biopsy to inflamed tissue. When possible, samples

will 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 weeks after starting therapy, when changes in targeted pathways, as opposed to global changes in response to therapy, are expected to be evident. Companies participating in the AMP would agree to provide a portion of the sample for the biorepository in addition to the samples that they require for the study specific protocols.

In some cases, patients with established RA will have total joint replacements, which will serve as a source of synovial tissue. Osteoarthritis, avascular necrosis, and other destructive arthropathies can be collected as controls, and normal samples can be obtained from patients with femoral fractures or at the time of autopsy from tissue banks. All samples could be stored in a single central biorepository or at an academic site or industry site performing the relevant analysis.

- 1) Genomic DNA and PAXgene RNA processing and storage, will be batch shipped from collection sites. DNA will be stored in at least 10 aliquots to facilitate distribution for genotyping and sequencing. PAXgene RNA is prepared and stored in at least 4 aliquots of approximately 1ug, again to facilitate distribution for multiple assays.
- 2) Heparinized tubes for preparation and viable storage of PBMCs, plasma (with neutrophil isolation and storage), serum and urine will be stored at an experienced facility for high quality viable cell storage in LN2 and -80C freezers. Preparation and storage of fixed PBMCs for RNA-seq analysis could also be done at this site.
- 3) Microbiome samples will be shipped to a site specialized experienced in the preparation and analysis of these types of samples.
- 4) Synovial tissue samples will be stored at institutions with laboratories devoted to the analysis of synovial biology.

### Sample tracking

Regardless of the number of facilities involved in sample processing and storage, a centralized web-based system for tracking of bar coded samples will be required for this project. Standard bar coded sample kits will be distributed to collection sites. The unique bar coded ID of samples along with appropriate metadata will be entered into this system and this will occur at all stages, from initial sample collection, local storage, shipping, receiving and processing, and utilization and distribution at each collection site and storage site. In this manner, real time tracking of all samples will be available to the partnership's members. All collecting sites and storage will need to undergo training on this system.

### **Data collection, storage and analysis**

A crucial component to the success of this partnership is a highly effective center for data analysis and bioinformatics. Data analysis should be coordinated and concentrated in a single dedicated center, modeling the Genome Data Analysis Centers (GDACs) components of the National Institutes of Health funded Cancer Genome Atlas (TCGA) Project. The success of this center would depend critically on its ability to develop new computational approaches for data integration for model systems analysis. The center would work collaboratively with tissue

acquisition centers and other investigators within the partnership in order to be responsive to the clinical questions as they arise, and with groups outside of the partnership to make data analysis and interpretation available.

The center will have the technical ability to:

1. Securely store and process data produced by the partnership
2. Compile and curate germane publically available external data sets and literature
3. Conduct genome-wide data analysis
4. Conduct analysis of immunological data
5. Support advanced statistical data analysis
6. Conduct innovative high-level integrative analysis with careful consideration of statistical biases and confounders.
7. Data interrogation portal to provide access to raw data, to the systems-level pathway analysis, and to integrative analyses of genetics and function.

### 1. Secure Storage

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

### 2. 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 partnership 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 the partnership. Genetic data for rheumatoid arthritis and SLE from genome-wide association studies and direct sequencing will be compiled to permit linkage of allelic variants with function. Data from the ENCODE and Roadmap Epigenomics projects will be imported to allow linking of non-coding allelic variants to expression data from critical cell types. These data sets will be obtained, curated, and harmonized with data sets generated through this partnership to facilitate more complex queries and integrative analyses as needed. In addition the center will also compile relevant literature from within the field, and employ statistical text-mining as needed to interpret generated data and to link public data to original publications in which it is described in detail.

### 3. Conduct genome-wide data analysis

The center would need to have expertise in analyzing genomic data, and in particular next-generation sequence data. Many of the technologies used in this proposal, including 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.

#### 4. Conduct low-level immunological data analysis

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

#### 5. Advanced statistical data analysis

The center will need to have the capacity to conduct advanced statistical analyses of the high-throughput data sets generated through this partnership. 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 the center should be adequately qualified to this end.

#### 6. Conduct innovative high-level integrative analysis

A major challenge for the center will be to integrate the data for integrated analyses. Aims of this proposal will require a global view of genotype, transcription, function, and immunological and clinical parameters of individual subjects. The goal will be 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. The center will need to be familiar with these issues and adept at handling them.

#### 7. Data access and interrogation portal

The decision to make data sets more broadly available to the public will be coordinated with sponsoring agencies and partnership members, and then coordinated through the center via the public portal. Ultimately, these are the key deliverables for this project.

- 1) Access to low-level data (e.g., RNA-seq reads), and also analyzed data sets (e.g., cell-specific expression profiles) will need to be freely accessible to partnership members through secure intra-net. Models for such a portal for primary data (e.g., <https://tcga-data.nci.nih.gov/tcga/>) and for integrating with published data (e.g., <http://www.broadinstitute.org/icbp/home>) already exist. The center will be responsible for setting standard data analysis pipelines for data sets generated by the partnership, and will generate final quality controlled data sets.
- 2) Connection of multiple layers of information will be needed to provide access to pathway analyses. Extracting the key findings with sound statistical models and connecting the results to prior knowledge of networks and pathways currently cannot be done in an automated fashion. Thus, the best computational groups will be brought



into the project for sophisticated analyses. Once performed, existing tools allow visualizing and making comparisons (e.g., Connectivity Map). This will provide insight into which networks should be interrogated in order to understand whether a target should be inhibited or enhanced in function. The availability of tissue-derived pathway information (e.g., synovial biopsy with single cell expression analysis of inflammatory cells, as well as detailed epigenetic analysis of synovial cells) will provide an unprecedented level of detailed information about whether the pathway in question is active at the site of disease. Given knowledge of cell-type specific pathways in patients, one also can ask whether drug interventions might have effects on cells that are not considered central to pathogenesis, thereby highlighting potential unexpected untoward effects, and considering whether they will constitute a barrier to drug development.

- 3) Access to integrated expression and pathway analysis (produced here) with genetic (GWAS and direct sequencing) and epigenetic (analysis of synoviocytes here, plus ENCODE and Roadmap) data will provide an extraordinary tool for target validation. After data acquisition and pathway analysis are complete, in year 5 the data interrogation portal will be constructed to allow an investigator from industry or academia to type in the name of a potential target and see the genetic association data, the expression profile in crucial cell types, and whether the target is part of a network that is active in the disease. . Unpublished data now unequivocally establishes over 100 genes as involved in the pathogenesis of RA. These genes are enriched across many diseases that include autoimmunity, immunodeficiency, inflammation and hematopoietic neoplasms. Many of these genes are obvious targets of current therapies and a few suggest re-purposing of current drugs. For coding variants, this project will provide both direct evidence of involvement of targets in disease-associated pathways, including expression and activation profiles, and evidence of causality (from integrated genetics). Most genetic variants are non-coding and influence regulation. Clearly, it is critical to establish which direction these regulatory changes go in order to understand how therapeutics should be targeted. Therefore, understanding the cell types, stages and functional perturbations associated with these genetic variants is essential. For non-coding loci, evidence of increased or decreased expression of a gene regulated by the loci (as identified by ENCODE and Roadmap Epigenetics), as demonstrated by the expression profile in specific cells (single-cell RNA-seq), and the association with altered signaling associated with disease (by network analysis) will provide a means to de-risk potential targets. We can ask how the networks derived from GWAS overlap with the networks derived from cellular profiling. Finally, this portal will allow the analysis of how risk alleles that are shared between autoimmune diseases lead to active immune modules that are shared or different between RA and lupus, and within these diseases, to responder/non-responder predictions and to organ involvement. This data interrogation portal will be an extension of current models that incorporate clinical data (e.g., ITN – TrialShare: <https://www.itntrialshare.org/project/home/begin.view>), with tentative plans to use the infrastructure for the knowledge portal proposed by the AMP-T2D group.

## **Section III: Project Management**

### **Policies**

This project will operate under the general policies of the partnership; the following are drawn from a document shared between projects. The RA/SLE project does not at this time anticipate any deviation from the shared policies.

### **Anti-trust**

The project participants agree that all research activities funded by the partnership fall into the pre-competitive space. There is to be no discussion of marketing activities. FNIH personnel will sit on all Executive Committee meetings to monitor this policy.

### **Confidentiality**

The project participants agree that there is to be no sharing of confidential information as a "blanket rule". If sharing is required, CDA will be established by relevant parties & FNIH.

### **Solicitations**

Solicitations will be open where practicable. Sole source solicitations are permissible but require justification.

### **Conflict of interest**

Any conflicts of interest that arise are to be documented and reviewed with FNIH and the Executive Committee, who will jointly develop a mitigation strategy.

### **Publications**

This project will operate under a "team science" approach, and publications will have joint authorship. Specific publication strategies will be developed by project teams prior to project start, including proposal for lead authors and co-authors.

### **Data sharing**

Findings will be shared broadly and quickly, in the interest of patients and the public health; partnership participants will have access to findings during assessment of data quality (up to 6 months of QA/QC). Protection of confidentiality may require limitations on sharing of raw genomic sequence, including raw RNAseq data.

### **Intellectual property**

Pre-existing IP can be used by partnership; FNIH legal determines if free of encumbrances. All research discoveries to fall in public domain, with no pre-emptive patenting. In rare instances

when this is not possible, FNIH will determine fair strategies for distributing IP to encourage broad commercialization and balanced public health benefit and review them with the Steering Committee and Executive Committee.

### **Decision Making**

#### **Joint management for separate streams of NIH and industry funding**

The Steering Committee will define the joint NIH/industry project plan, but the funds provided by NIH and industry flow through separate streams. NIH funds must be disbursed according to NIH procedures for solicitation of applications, review of applications, and decision-making. However, industry may have input as to those who serve on initial review panels (“study sections”) for these grants. In addition, industry participants may serve on sub-committees of the Advisory Committees that provide a second level of review and advise the Institute Director of the strengths and weaknesses of applications that have been recommended as a result of study sections.. Industry funds will be contributed through and managed by FNIH. Such funds may be dispersed either directly by FNIH through grants or contracts, or transferred by FNIH to NIH for disbursement through NIH grant procedures described above. The Steering Committee will review and select proposals made directly to FNIH for funding. After awards are made, the Steering Committee will provide project oversight for all studies, whether funded by NIH or industry/FNIH.

The RA/SLE Steering Committee for AMP will operate under the direction of the overall AMP Executive Committee (EC), comprised of 3-4 leaders each from industry and NIH, as well as a representative each from FDA, academia, and the patient advocacy sector. The EC is in turn advised by an Extended Executive Committee comprised of R&D heads of companies involved in the partnership. The RA/SLE Steering Committee is responsible for defining the research agenda and project plan, for review of ongoing projects, and for the detailed assessment of milestones. The project plans are submitted by the Steering Committee to the EC for review and approval. The EC will also review the assessment of milestones and any revision to the project plan that results from a “No-go” assessment that some element of the current plan is not feasible.

The Steering Committee is currently comprised of representatives from participating companies as well as members from government. After projects are launched, regular Steering Committee meetings will be held quarterly to guide progress and resolve issues. The frequency of meetings will be adjusted as the scientific agenda requires. The decisions of the Steering Committee will be made by simple majority. Each participating company will have one vote and industry's cumulative vote will remain constant at 50% of the total votes. If additional industry members are added to the partnership, votes for all industry participants will be scaled appropriately.

NIH, academia, and non-profit organizations will have votes that will not exceed 50% of total. Non-profit organizations can join the Steering Committee. They will serve as a voting member should their financial contribution be \$100,000 or more annually with their vote being proportional to their contribution. Non-profit organizations donating less than \$100,000 annually can join the Steering Committee as non-voting members. Academic investigators,

whether provided funding by the project or not, may be added at the Steering Committee's discretion. However, due to potential conflicts of interest, such academic members would not be voting members of the committee.

The Steering Committee may elect to form working groups or sub-committees to manage specific aspects of the project or to advise the full Steering Committee on particular issues.

### **Governance**

The Steering Committee for the AMP RA/SLE operates under the direction of the Core Executive Committee for the partnership, which is advised by the Extended Executive Committee. The Steering Committee is responsible for defining the research agenda and project plan, for review of ongoing projects, and for the assessment of milestones. The project plans are submitted by the Steering Committee to the Executive Committee for review and approval. The Executive Committee will also review the assessment of milestones and any revision to the project plan that results from a “No-Go” assessment that some element of the current plan is not feasible.

## **Section IV: Timeline, Milestones and Deliverables**

The set of activities for this project can be divided into the following segments:

### Phase 0

- 1) Start-up activities and organization
- 2) Initial testing of techniques for preparation and analysis of tissue and development of SOP

### Phase 1

- 3) Enrollment and analysis of blood and tissue of the initial clinical study cohorts
- 4) Systems analysis of data from initial cohorts to define pivotal disease pathways

### Phase 2

- 5) Defining within disease comparative cohorts and analytic approaches for scale-up
- 6) Generation of all the profiling data on the fully enrolled cohorts
- 7) Bioinformatics analysis for molecular stratification of patients and potential target identification
- 8) RNAi screening to validate targets for drug discovery research
- 9) Reporting and roll-out of data access through a data interrogation portal

Phase 1 will be critical to ensure feasibility of enrollment and that application of the technical methodologies are sufficiently robust for generation of high quality data. This will include analysis of the initial comprehensive data sets. Results will be used to guide the design of studies to generate the highest value data sets and analysis plan in Phase 2 and then leverage any lessons learned / efficiencies for the final analyses for the fully enrolled RA and lupus cohorts. The interim data will be used by members of the team for initial pathway mapping and network analysis. However, as the initial deliverable, these data, including analyses of protein expression and activation and cell characterization by CyTOF, will also be made available to members of the partnership. Thus, during year 2, members of the partnership will

be able to start analyzing these data for the expression and function of specific potential targets of interest in well-defined populations of cells.

An outline of the key activities timeline, milestones and deliverables is summarized in the table below.

A 5-year program is proposed starting in late 2014. Start up activities will require approximately 6 months to reach the point of IRB approved protocols, selection and contracting with clinical centers, development of SOPs for sample collections and sample processing, analytical methods validation, selection and contracting with vendors. Similarly, the final 6 -12 months (2019) will be focused largely on completion of bioinformatics analyses, target validation by RNAi screening and report generation. The 4 years from 2015 through 2018 are the critical years for execution of this program. Based on the proposed activities and resources required during these years, this period will also consume the bulk of the program budget.

### **Operational Risks and Mitigation**

There are several critical aspects that could challenge the successful completion of this project. Reaching the target enrollment for the clinical cohorts may be the key risk. This risk could be mitigated by enlisting academic centers that have already established large RA and SLE patient cohorts for long-term evaluation. These centers must be well networked to capture referral of patients from community practices who meet key study inclusion criteria. Increasing the number of study centers and/or decreasing the target enrollment (assuming power calculations indicate this will not compromise the study objectives) are the two most direct ways to address challenges in hitting enrollment targets.

Single-cell analyses (e.g., single cell RNA-seq) are emerging technologies with most previous applications focused on cancer. Production of meaningful data from tissue will require initial validation, and comparative approaches to better understand signals that are due to the technique rather than biology (e.g., effects of sample processing). Tissue acquisition, processing, storage and shipping will be tightly standardized by SOP. Data quality control will be critical.

Sorting of cell subsets by flow cytometry and CyTOF phosphoproteomic profiling are both very time sensitive and require highly trained technical staff to process the bio-specimens promptly (within 2-4 hrs from collection). Ideally, clinical centers involved with the program will already have the necessary equipment and experienced staff. Matching potential for high enrollment with pre-existing technical expertise may be a challenge. One mitigation approach would be to set up 2-3 new centers with a suitable equipment budget and access to local talent. Shipment of blood samples to a central processing center is an alternative option, but may not be optimal.

The other technical aspects of the program are better established. Whole genome sequencing, genotyping, mRNA profiling are all well-precedented techniques. As technical barriers are

crossed, RNA-seq is will become more cost efficient, a development which is expected to occur during this project period. Nevertheless, the interim analysis will employ RNA-seq at the single cell level on synovial biopsy tissue (n=20) to assess the analytical and cost efficiency of the technology, and to support a decision on using RNA-seq in the full cohort analyses.

Finally, this project will generate a substantial amount of biological data, which will stress the bioinformatics approach to data analysis and multivariate integration with the clinical data. In fact, the bioinformatics tools may not yet be sufficiently robust to support the necessary analyses in this proposal. However, this is a developing field, and the timeline and budget are constructed to allow select team members to build the tools and expertise.

### **Threshold Criteria for Project Progression**

Although most of the technology envisioned for this research plan is not novel, many refinements of existing applications will need to be optimized and standardized for use in clinical and laboratory settings in support of this project. Also critical to the success of the plan is the identification, consent and recruitment of rheumatoid arthritis and lupus patients that meet the inclusion criteria of the designed studies. Due to the time limited nature of this partnership additional challenges include executing this study within the timeframe and budget provided. To ensure project feasibility the project will need to meet specific criteria prior to the transition from Phase 1 to the scaled up and resource intensive Phase 2 portion. Thus the following criteria must be met:

1. The rate of sample acquisition in Phase 1 demonstrates feasibility of proceeding to Phase 2 within the proposed budget and timelines
2. The tissue-derived data is sufficiently robust to:
  - a. Identify cells of the same type within and between samples
  - b. Generate a significant disease-specific signature derived from single cells
  - c. Establish the number of samples required to successfully execute Phase 2

The successful attainment of these criteria will be adjudicated by the Steering Committee prior to progression to Phase 2, approximately 24 month from the initial funding award. The findings of the RA/SLE Steering Committee will be binding on participants; in other words, if the committee determines that the "go" criteria have been met, individual members will not have an opt out option.

**Table 4: Summary of Project Timelines, Key Milestones & Deliverables**

<b>Timeline</b>	<b>Activities</b>	<b>Milestones</b>	<b>Deliverables</b>
1H 2015	<p>Draft Protocols</p> <ul style="list-style-type: none"> <li>RA starting anti-TNF; Lupus nephritis</li> </ul> <p>Identify potential clinical sites</p> <p>Define clinical data collection process</p> <p>Draft SOPs</p> <ul style="list-style-type: none"> <li>Sample collections</li> <li>Flow sorting &amp; subset (n=8) cryopreservation</li> <li>RNA, DNA profiling</li> <li>Synovial biopsies</li> </ul> <p>Identify vendors, analytical laboratories</p> <ul style="list-style-type: none"> <li>CyTOF profiling</li> <li>Biorepository</li> <li>Genomics profiling</li> </ul>	<p>Protocols finalized</p> <p>10-12 sites selected per disease</p> <p>CRFs drafted</p> <p>SOPs finalized</p> <p>Vendor / collaborator labs selected</p>	<p>IRB approved protocols</p> <p>Contracts established with clinical sites</p> <p>Data management system validated</p> <p>Analytical methods validated in pilot studies</p> <p>Contracts established with vendors &amp; analytical labs</p>
2H 2015	<p>Initiate enrollment (Phase 0)</p> <ul style="list-style-type: none"> <li>RA (biopsy, surgical)</li> <li>SLE (kidney biopsy)</li> </ul> <p>Initial RNA-seq, CyTOF, PBMC analysis</p> <p>Refine SOP</p>	<p>Patients enrolled</p> <p>Cells identifiable by RNA-seq</p>	<p>Interim statistical analysis plan drafted</p>
1H 2016	<p>Systems Analysis (Phase 1)</p> <ul style="list-style-type: none"> <li>RA &amp; lupus cohorts</li> </ul> <p>Initiate 'omics profiling</p> <p>Assemble bioinformatics team Initial data integration</p>	<p>Enrollment: 5 each RA &amp; SLE</p> <p>Analytical data sets generated</p> <p>Bioinformatics analysis started</p>	<p>QA/QC of pilot data completed</p> <p>Initial bioinformatics analysis strategy</p>

			defined
2H 2016	Phase 1 completes enrollment  Phase 1 data integration and systems analysis  Assess RNA-seq data and other 'omics data for scale up	10 each SLE & RA (total of 15 each)  Interim analysis completed	QA/QC completed for Phase 1 RA & SLE annotated clinical data set  Interim draft report complete  Final Phase 2 project plan
1H 2017	Patient Stratification (Phase 2)  'Omics profiling continues  Expand bioinformatics team  Initiate RA/SLE comparison	Phase 2 Enrollment 15%  Refine bioinformatics strategy / methods based on interim analysis experience	Final statistical analysis plan complete
2H 2017 – 1H 2018	Patient Stratification (Phase 2) cont  Initiate design and construction data interrogation portal	Phase 2 Enrollment 60%	QA/QC completed for early RA & lupus 'omics data set  Data interrogation portal operational and accessible to partnership member
2H 2018	Patient Stratification (Phase 2) cont  Initiate RA & lupus cohort bioinformatics analyses	Phase 2 Enrollment 100%	Networks altered in disease state defined; Key networks that distinguish patient populations identified (early from established RA; Lupus nephritis from lupus skin involved & change over time for skin cohort). Pathways selected for target validation RNAi screen
1H 2019	Integrated RA / lupus cohort	Integrated RA /	Signaling networks &



	<p>analyses completed</p> <p>Integrated data analysis (pathways, expression GWAS, epigenetics, etc) initiated via interrogation portal</p> <p>Initiate RNAi target validation screen</p>	<p>lupus cohort analyses completed</p>	<p>pathways mapped that are similar or distinct across the study cohorts</p> <p>Targets ranked for biological relevance &amp; suitability defined for phenotypic or biochemical assay based library screening</p> <p>Top 20-50 targets across RA and lupus. Targets common / distinct among cohorts identified.</p>
2H 2019	<p>Complete RNAi target validation</p> <p>Initiate reports</p>	<p>RNAi screen completed</p> <p>Draft reports reviewed</p>	<p>Initial target validation</p> <p>Reports finalized</p>