

# A Computational Approach to Find Cases of R21 as a Stepping-Stone to R01 in Immunology Research

Fenglou Mao, Dawei Lin

Division of Allergy, Immunology and Transplantation (DAIT), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, Bethesda, MD

Email: fenglou.mao@nih.gov

## Abstract

R21 is an NIH grant mechanism that supports short term exploratory and developmental research. Immunology researchers increasingly use it for their high risk and high reward ideas in recent years. In order to understand the impact of R21 mechanism for generating a regular long term R01 research project grant, we have developed a text based computational approach to find matching pairs, in which R01 applicants are developed directly based on the outcomes from R21 awards.

The manual approach to find such pairs is challenging. First, some investigators may have several R21 and R01 grants, or independent R21 and R01 on different research topics. It is a time consuming process to check all the possible relationships. Second, a manual process has subjective biases if a R21 and a R01 grants are related. Third, the manual approach is not scalable to examine a large number of grants. To address those challenges, we use "Research, Condition, and Disease Categorization (RCDC)" signature similarity, or a LIKE-score to serve as a content similarity measure. The LIKE-score is an objective and quantitative metrics. To further reduce false positive, we introduced a neighborhood metrics, e.g. if a R21 grant is related to a R01, the grants that are similar to the R21 should have a significant overlap with those that are similar to the R01 grant. A few other criteria are also applied to remove false positives, eg., the R21 and the R01 should be from same PI, and the R21 should be funded earlier than the R01 application submission.

Using the aforementioned approach, we searched through all the funded R21 and R01 applications from 2008 to 2014. There are total 8,975 pairs based on the same author. In 221 pairs both the R21 and the R01 was funded through DAIT, NIAID. After applying a time filter, which is that the R01 is one year after the R21, there are 92 pairs left. Using a LIKE-score, a similarity score (threshold 500) between the R21 and the R01 grants, we further reduced the number of pairs to 32. The final step is to find all the pairs that the overlap above 50% of the their first level neighbor based on LIKE score. We found 9 pairs as the most probable R21 and R01 pairs. Among those pairs, there are at least 4 cases have been validated by the Program Officers who have knowledge of these applications.

## R21 and R01 data

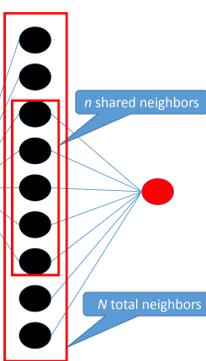
	FY	Type	Number
R21	2007-2015	1	16500
R01	2007-2015	1	34741

## RCDC LIKE-Score Calculation

RCDC Terms	Base Publication Weight	Scale Factor	Base Publication Weight(Scaled)	Matching Publication Weight	Dot Product
Acetates	229	2.71	621		0
Appendix	655	2.31	1513		0
Asthma	492	1.78	876		0
Basophils	79	3.03	239	212	507
Bone Marrow	93	1.59	148	122	180
Boxing	354	2.05	726		0
Cations	34	2.21	75	123	92
Cell Line	750	1.19	893	123	1098
Cell Nucleus	559	1.49	833	321	2674
Cells	524	0.55	288		0
Chromosomes	222	1.74	386	456	1761
Complement	829	1.33	1103	345	3804
Cytosol	925	2.11	1952		0
Diglycerides	861	2.79	2402		0
Disease	41	0.54	22		0
Eicosanoids	186	2.72	506	345	1745
			Like Score		11862

Base Publication Weight(Scaled)=Base Publication Weight\*Scale Factor  
Dot Product=Base Publication Weight(Scaled)\*Matching Publication Weight/100  
Like Score=Sum(Dot Product)

## Shared First Level Neighbors



● R21/R01  
● R21  
● R01

Shared first level neighbors is defined as:  
Shared\_FLN=n

Shared first level neighbors ratio is defined as:  
Shared\_FLN\_Ratio= $\frac{n}{N}$

## Three Examples of True Positives

### 1R21AI089810-01 In vitro differentiation of RAG1-mutated iPS cells and correction by meganuclease

Severe combined immunodeficiency (SCID) comprises a group of heterogeneous genetic disorders that are fatal, unless treated by hematopoietic cell transplantation (HCT). Mutations of RAG1 and RAG2 genes are the most common cause of T-B-NK+ SCID in humans. Hypomorphic defects in the same gene may cause leaky SCID or Omenn syndrome (OS), a severe immunodeficiency associated with multigenetic etiology. One of the major limitations of the preclinical studies that aim at exploring the efficacy of gene transfer in humans with immunodeficiency is the limited availability of patient-derived target cells. However, fibroblasts can be reprogrammed in vitro into induced pluripotent stem cells (iPSCs) through virus-mediated transduction of a combination of transcription factors. These iPSCs can be targeted multilineage differentiation (including T lymphocytes) in vitro. We have generated a repository of fibroblast cell lines from patients with SCID or OS, carrying different RAG1 mutations. We now intend to generate iPSCs from RAG1-mutated patients, and to characterize their stemness and pluripotency profile, chromosomal integrity and patient-specific derivation. In collaboration with Dr. Zuniga-Pflucker, we will investigate the ability of these patient-derived iPSCs to proceed along T-cell differentiation in vitro. We will also explore the ability of the RAG1-specific HE to correct the mutant RAG1 locus in patient-derived iPSCs, and to support V(D)J recombination and T cell differentiation in vitro. We will compare the ability of gene-corrected and uncorrected patient-derived iPSCs to support T cell development in vitro. This study will permit to define the specific ability of various RAG1 mutations to support T-lymphocyte differentiation, and may thus help explain the basis of the phenotypic diversity of RAG defects in humans. Furthermore, this project will also explore the efficacy and safety of a novel approach to gene therapy of SCID, based on gene-specific endonuclease-mediated homologous recombination. PUBLIC HEALTH RELEVANCE: Defects of the RAG1 and RAG2 genes in humans cause a spectrum of severe immunodeficiencies that range from complete absence of lymphocytes (SCID) to a condition characterized by immunodeficiency and autoimmunity (Omenn syndrome). The basis for this diversity remains poorly defined. To address this issue, we will induce skin cells (fibroblasts) from patients with RAG1 gene defects, to become pluripotent stem cells (iPSCs), and we will then study their ability to differentiate into lymphocytes in a test tube. Moreover, we will use a novel protein to repair the RAG1 gene defect in patient-derived stem cells. This may pave the way to novel and safer approaches to correct genetic diseases by gene repair.

### 1R21AI089963-01 Alums adjuvanticity

Protein based receptor ligand interactions are universally regarded as the initiating point of immune activation. However, it is questionable if it is applicable to immune recognition of solid structures. Binding of particulate antigens by antigen presenting cells (APC) is a critical step in immune activation. Previously, we demonstrated that uric acid crystals are potent adjuvants, initiating a robust adaptive immune response. However, the mechanisms of activation are unknown. Using atomic force microscopy as a tool for real time single cell activation analysis, we have collected evidence that uric acid crystals can directly engage cellular membranes, particularly the cholesterol components, with a force substantially stronger than protein based cellular contacts. Binding of particulate substances activates Syk kinase-dependent signaling in dendritic cells (DCs). These observations suggest a mechanism whereby immune cell activation can be triggered by solid structures via membrane lipid alteration without the requirement for specific cell surface receptors, and a testable hypothesis for crystal-associated arthropathies, inflammation and adjuvanticity. In this proposal, we extend our work to study how alum interacts with the immune system and to reveal if such a lipid based mechanism is applicable in alum's adjuvanticity. We will also study the association between cell surface lipid sorting and Nalp3 inflammasome activation, a critical step in uric acid crystal mediated cell activation. We will further study a set of immune activation events unrelated to inflammasome regulation in order to establish a complete picture of alum's immune regulating capacities. The outcome of this work will impact vaccine development and our understanding of crystal related diseases. PUBLIC HEALTH RELEVANCE: This project deals with the basic mechanism for the immune recognition of alum. It has high relevance in the vaccine development and crystal related diseases. Its outcomes will lead to better understanding of the immune system and suggest new methods for population based immunizations.

### 1R21AI075176-01 Regulation of TCR signaling by a novel enzyme activity.

The T cell auto-reactivity that lies at the heart of many autoimmune diseases arises from the very properties that allow T cells to mount an effective immune response. Thus, to understand the etiology of different autoimmune diseases will require a thorough and integrated understanding of the mechanisms that control T cell activation. The long-term objective of my research is to understand the biochemistry of T cell reactivity. Currently, my laboratory is studying how two related proteins, Sts-1 and Sts-2, act in concert to negatively regulate T cell signaling pathways. The role of Sts-1 and -2 in controlling TCR signaling pathways was revealed by a strain of mice engineered to lack the Sts genes. T cells from Sts-1/2-/- mice dramatically hyper-proliferate in response to TCR stimulation. This hyper-proliferative phenotype is accompanied by increased activation of signaling pathways downstream of the TCR, elevated levels of cytokine production, and increased susceptibility of Sts-1/2-/- mice to autoimmunity in a mouse model of multiple sclerosis. In our ongoing effort to characterize the functions of the Sts proteins, we recently discovered that Sts-1 has a novel and potent enzyme activity. This activity derives from an evolutionarily conserved region within the protein and our recent data indicates that it plays an essential role in the ability of Sts-1 to regulate TCR signaling pathways. Sts-1 is sufficiently dissimilar to other known enzymes to make it a novel and unique phosphatase. In addition, the connection between Sts-1 catalytic activity and regulation of T cell activation in unknown and unexplored. The experiments outlined in this proposal are designed to shed light on these areas. Our Specific Aims are: 1. To determine the structural features that define and regulate Sts-1 catalytic activity. 2. To determine the role of Sts-1 phosphatase activity in regulating Zap-70 signaling. We will use some recently developed protocols to accomplish our goals. Completion of the studies described herein will help us build a model of how Sts-1 cooperates with Sts-2 and other intracellular signaling mechanisms to control T cell reactivity. It is our hope that a broad, integrated understanding of all the mechanisms that control T cell activation will translate into therapies that prevent the onset of a variety of autoimmune diseases. Autoimmune diseases arise when normal regulatory mechanisms within the immune system fail. Developing therapies that will prevent or cure autoimmune diseases will require a thorough and integrated understanding of the mechanisms that control the immune response. This project focuses on understanding the function of a novel protein that participates in the regulation of T cell activation, population based immunizations.

### 1R01AI100887-01 Modeling and correcting human SCID using patient-derived iPS cells

This proposal seeks to elucidate the mechanisms involved in the defective T cell development observed in patients with Severe Combined Immunodeficiency (SCID), and to develop a novel approach for the preclinical investigation of efficacy and safety of gene therapy for SCID. Severe combined immunodeficiency (SCID) comprises a group of genetic disorders that affect development and function of T lymphocytes. Some of these defects also compromise development of B and/or natural killer (NK) lymphocytes. Patients with SCID are extremely susceptible to infections and die early in life, unless immune reconstitution is achieved by means of hematopoietic cell transplantation (HCT) and - in selected cases - gene therapy or enzyme replacement therapy. Although the study of SCID animal models has been essential to unravel key mechanisms that govern lymphoid development, significant differences have emerged when comparing the immunological phenotype in SCID patients and in mice carrying defects in orthologous genes. These species-specific differences limit the use of animal models in the study of human lymphoid development. Other approaches to study pathological human T cell development include examination of thymic biopsies from patients with SCID, and use of patient-derived hematopoietic progenitor cells to investigate T cell differentiation in vitro (upon co-culture with OP9-DL1 or OP9-DL4 cells) and in vivo, upon injection into immunodeficient mice. However, the rarity of human SCID conditions limits access to patient cells and hence represents a significant obstacle to applying these tools to investigate pathological human T cell development. Induced pluripotent stem cells (iPSCs) are a novel and practical tool for human disease modeling and correction, and may therefore help overcome these limitations. Our preliminary data demonstrate that: a) we have generated a repository of fibroblast cell lines from 50 patients with SCID and related disorders, representative of mutations in 14 different genes; b) we have successfully initiated generation and characterization of iPSCs from patients with SCID; and that, c) human iPSCs can be differentiated into T lymphocytes in vitro, using the OP9-DL4 system. The proposed experiments are designed to test the overall hypothesis that SCID patient-derived iPSCs can be used to recapitulate pathological human T cell development both in vitro and in vivo (upon injection into NOD-Scid-Il2rg-/-, NSG, mice), and thus represent a novel tool for human disease modeling and corrective therapies. To test this hypothesis and to evaluate whether iPSCs carrying distinct SCID-causing mutations will differ in their ability to progress along T-lineage differentiation both in vitro and in vivo, we propose to 1) analyze pathological human T cell development from patients whose genetic defects are presumed to block T cell differentiation at different stages, and 2) use lentivirus-mediated gene transfer into patient-derived iPSCs to assess correction of defective T-cell development and function both in vitro and in vivo. Specific Aims SA 1. To model human SCID in vitro using patient-derived iPSC cells (iPSCs). 1) To generate and characterize iPSCs from patients with SCID. 1.2 To analyze T cell differentiation capacity of patient-derived iPSCs in vitro. 1.3 To analyze diversity and function of T lymphocytes generated in vitro from patient iPSCs SA 2. To correct human SCID in vitro using patient-derived iPSCs. 1) To correct the genetic defect in patient-derived iPSCs using lentivirus-mediated gene transfer. 1.2 To analyze reconstitution of T cell differentiation in vitro using gene-corrected iPSCs. 1.3 To analyze diversity and function of T lymphocytes generated in vitro from gene-corrected iPSCs SA 3. To model human SCID and analyze gene correction in vivo with patient-derived iPSCs. 1) To analyze in vivo T cell development and function in NSG mice using patient-derived iPSCs. 1.2 To analyze reconstitution of T cell development and function in vivo in NSG mice with gene-corrected iPSCs. 1.3 To perform integration site analysis in gene-corrected iPSCs and their T lymphocyte progeny if successful, this project will represent a novel approach to studying pathological human T cell development, and will provide a new tool for preclinical assessment of efficacy and safety of gene transfer for human SCID.

### 1R01AI098995-01A1 Adjuvants mechanisms

Protein based receptor ligand interactions are universally regarded as the initiating point of immune activation. However, it is questionable if it is applicable to immune recognition of solid structures. Binding of particulate antigens by antigen presenting cells (APC) is a critical step in immune activation. Previously, we demonstrated that uric acid crystals are potent adjuvants, initiating a robust adaptive immune response. However, the mechanisms of activation are unknown. Using atomic force microscopy as a tool for real time single cell activation analysis, we have collected evidence that uric acid crystals can directly engage cellular membranes, particularly the cholesterol components, with a force substantially stronger than protein based cellular contacts. Binding of particulate substances activates Syk kinase-dependent signaling in dendritic cells (DCs). These observations suggest a mechanism whereby immune cell activation can be triggered by solid structures via membrane lipid alteration without the requirement for specific cell surface receptors, and a testable hypothesis for crystal-associated arthropathies, inflammation and adjuvanticity. In this proposal, we extend our work to study how alum interacts with the immune system and to reveal if such a lipid based mechanism is applicable in alum's adjuvanticity. We will also study the association between cell surface lipid sorting and Nalp3 inflammasome activation, a critical step in uric acid crystal mediated cell activation. We will further study a set of immune activation events unrelated to inflammasome regulation in order to establish a complete picture of alum's immune regulating capacities. The outcome of this work will impact vaccine development and our understanding of crystal related diseases. PUBLIC HEALTH RELEVANCE: This project deals with the basic mechanism for the immune recognition of alum. It has high relevance in the vaccine development and crystal related diseases. Its outcomes will lead to better understanding of the immune system and suggest new methods for population based immunizations.

### 1R01AI080892-01 Regulation of TCR Signaling by Sts-1 and Sts-2

The same properties that allow T cells to mount an effective immune response can be turned against host tissue, provoking an autoimmune response. Intriguingly, T cell auto-reactivity is often the result of deregulated signaling pathways within T cells. Thus, understanding the etiology of different autoimmune diseases will require a thorough and integrated understanding of the mechanisms that control T cell activation. The long-term objective of my research is to understand the regulation of T cell activation. Currently, my laboratory is studying how two related proteins, Sts-1 and Sts-2, act in concert to negatively regulate T cell signaling pathways. The Sts proteins have a unique modular structure, with an N-terminal ubiquitin-interacting domain (UBA), a central protein-protein interaction domain (SH3), and a C-terminal region with homology to the enzyme phosphotyrosine kinase (PTK). The realization that the Sts proteins have a role in regulating TCR signaling pathways emerged from an analysis of mice lacking Sts-1 and -2. T cells from Sts-1/2-/- mice dramatically hyper-proliferate in response to TCR stimulation. This hyper-proliferative phenotype is accompanied by increased activation of signaling pathways downstream of the TCR, elevated levels of cytokine production, and increased susceptibility of Sts-1/2-/- mice to autoimmunity in a mouse model of multiple sclerosis. In our ongoing effort to characterize the functions of the Sts proteins, we recently discovered a novel protein tyrosine phosphatase activity associated with the Sts PGM domain. Intriguingly, the Sts phosphatase domain shows no similarity to classical PTPs in primary amino acid sequence and important catalytic features. Coupled with the fact that no known PTP contains a UBA or SH3 domain, our results suggest that the Sts Proteins operate in an intracellular signaling niche within T cells that is separate and distinct from classical PTPs. The experiments outlined in this proposal are designed to address the underlying function(s) and mechanism(s) of action of the Sts-1 and Sts-2. Our Specific Aims are: 1. Determine how the Sts Proteins regulate T cell activation thresholds via regulation of Zap-70. 2. Determine how the modular domains of Sts-1 and Sts-2 cooperate to negatively regulate TCR signaling pathways. 3. Determine the features that define and regulate Sts-1 phosphatase activity. We will use a combination of biochemical, biophysical, and cell biological approaches to accomplish our goals. Completion of the studies described herein will help us build a model of how the Sts Proteins cooperate to negatively regulate TCR signaling pathways. PUBLIC HEALTH RELEVANCE: Autoimmune diseases arise when normal regulatory mechanisms within the immune system fail. Developing therapies that will prevent or cure autoimmune diseases will require a thorough and integrated understanding of the mechanisms that control the immune response. This project focuses on understanding the function(s) and mechanism(s) of action of a closely related pair of negative regulators of T cell receptor signaling, Sts-1 and Sts-2, suggest new methods for population based immunizations.

## Final List of 9

R21				R01				PI	Like Score	Shared FLN	Share_FLN Ratio
FY1	PROJECT1	IRG1	PCC1	FY2	PROJECT2	IRG2	PCC2	PL_NAME_CO NTACT			
								NOTARANGEL O, LUIGI			
2010	AI089810-01	IMM	I4E	2012	AI100887-01	IMM	I4E	DANIELE	588	6068	0.5489
2010	AI089963-01	IMM	I2A	2012	AI098995-01	IMM	B;P145AM	SHI_YAN	1308	5331	0.5188
2011	AI094009-01	IMM	I2H	2012	AI100082-01	AI	I2H B	MCCUNE, JOSEPH M	963	7449	0.5433
2008	AI078449-01	AI	I2J	2010	AI090742-01	AI	I2M	SCHROEDER, HARRY WILLIAM	1417	6868	0.6123
2008	AI078459-01	AI	I2M	2009	AI084808-01	AI	I2M	SANZ, IGNACIO E.	561	3961	0.6348
2007	AI077079-01	AI	I2K B	2011	AI01A1	IMM	I2G B	SALEK-ARDAKANI, SHAHRAM	907	4904	0.538
2007	AI075176-01	IMM	I2E	2010	AI080892-01A2	IMM	I2E	CARPINO, VIRGINIA NICHOLAS A	770	7156	0.6007
2007	AI069031-01A2	IMM	I2E	2010	AI01A1	IMM	M	SHAPIRO, VIRGINIA SMITH	698	6232	0.6268
2007	AI077012-01	AI	I2K B	2010	AI085043-01	IMM	I2G	BROOKS, DAVID G.	651	4428	0.5165

\*Yellow color pairs are true positives. Grey colors are false positives. While colors are to be evaluated.

## Candidate R21-R01 Pairs

- 8975 • R21-R01 with same contact PI
- 4382 • R01 is at least one year later than R21 (FY: project start year)
- 865 • R21 or R01 is awarded by NIAID
- 287 • R21 or R01 is awarded by DAIT
- 92 • R21 and R01 is awarded by DAIT
- 32 • Like Score >= 500
- 23 • Shared\_FLN >= 3000
- 9 • Shared\_FLN\_Ratio >= 0.5

## Acknowledgement

We appreciate the stimulating discussions from NIAID/DAIT colleagues - Thomas Esch, Lara Miller, Wolfgang Leitner, Conrad Mallia, Cheryl Lapham, Ashley Xia, Quan Chen and Susan Cooper, from NIH/OD/QVR team - Brian Haugen and Donald Tiedemann, and computing support from NIAID/OCICB, and grant access help from NIAID/DEA - Virginia Pool.

## Conclusion

In this study, we have developed an algorithm that is able to find real cases of R21-R01 pairs, in which the R21 grant is a step stone of the R01 grant in the same pair. Various filters including a network based measurement and RCDC LIKE score were able to improve the true positive ratio to >40%. The accuracy of the prediction was validated by the NIAID/DAIT Program Officers whose program portfolio support the relevant grants.

We have applied the same algorithm to all NIH R01/R21 grants from 2008-2014 and need help from other ICs to identify additional true positive cases as well as true negative cases for further study. Please contact [fenglou.mao@nih.gov](mailto:fenglou.mao@nih.gov) if you are interested in collaborating with us.