

Research Objectives

Uncovering genetic causes of diseases using a multidisciplinary bioinformatics-driven approach

My research focuses on designing and using bioinformatics techniques to identify causes and mechanisms underlying human diseases such as Systemic Lupus Erythematosus, Glioblastoma Multiforme, and Arteriovenous Malformations followed by designing appropriate diagnostics and treatment methods. Once identified, I work with collaborators to verify the bioinformatics-discovered mechanisms utilizing *in silico*, *in vitro*, and *in vivo* techniques, and develop therapeutic and diagnostic techniques to identify and treat the underlying human disorder. Cell-culture-based methods are utilized as the first step in designing multi-target treatments, followed by appropriate animal models..

Unique Qualifications

Straddling Biology, Computer Science, and Statistics

In addition to being a cellular and molecular biologist, I have extensive experience in algorithm design, computer programming, and statistics. The combination of these areas enables me to handle biological problems which involve large numbers of samples and data and require statistical analysis which can address confounders which are often present in non-laboratory settings. It also gives me a unique perspective which enables me to design novel methods to analyze and interpret large amounts of data.

Multidisciplinary approach

I have published in multiple disciplines, including membrane biophysics, bioinformatics, genetics, and cellular biology. My experiences in these fields and my experiences while transitioning fields has allowed me to apply unique insights garnered from my previous work to new topics leading to novel approaches and breakthrough discoveries.

Previous Research Projects

Genetic Basis of System Lupus Erythematosus (SLE)

I developed novel bioinformatic methods which increase likelihood of identifying reproducible genetic associations using prior knowledge from publicly available databases and expert information [2]. Using these methods, I was able to identify genes previously unassociated with SLE in a trio-based study [4]. These genes were then replicated in a larger case-control study which was funded by the NIH on the basis of the original findings [1, 3]. Among other findings, this larger study identified a missense allele in NCF2 (H389Q, rs17849502) which was associated with SLE. Collaborative work indicated that H389Q altered the binding energy of NCF2 with VAV1 using docking simulations, and *in vitro* experiments confirmed that H389Q altered NADPH oxidase function [5], thus identifying it as a causative SLE mutation.

Cancer Stem Cells in Glioblastoma

Glioblastoma is a almost invariably fatal form of brain cancer¹ which is diagnosed in $\approx 9,000$ people in the US annually. It is typified by high levels of chemotherapeutic-resistant recurrences, some of which is likely caused by cancer stem cells which are insensitive to many chemotherapeutic agents. In collaboration with Florence Hoffman at USC, I have classified glioblastoma-derived cancer stem cells and cancer cell lines into distinct classes using gene microarrays and various clustering approaches, which

¹Only 4% of patients survive to 5 years after diagnosis

will enable the design of a class-specific treatment of this devastating disease.

Regulatory pathways underlying cranial arteriovenous malformations

The mechanisms underlying the formation of arteriovenous malformations² (AVMs) which occur in the brain are unknown. Using gene microarrays, qPCR, and *in vitro* experiments on primary human endothelial cells cultured from resected AVMs, I identified multiple gene regulation pathways, many of them novel, including the Id1/Thsb1 inhibitory pathway. Additional experiments indicated that the *in vitro* pathology of endothelial cells could be partially rescued using extracellular Thsb1 [6].

Future research directions

Identification of causal mutations in SLE

While many regions and SNPs which are associated with SLE have been identified, few of those regions have identified causal alleles with known function. Furthermore, even for regions with identified causal alleles, no systematic searches have been performed to identify additional causal regions. Continuing my existing collaboration with Chaim O. Jacob, I will rectify this by

1. Deep sequencing the associated regions in 500 lupus cases and 1000 controls
2. Identifying which newly found variants are associated with SLE
3. Selecting variants with a high likelihood of producing functional variants
4. Verifying biological relevance of variants by *in vitro* study
5. Verifying functional relevance in human subjects

Developing analysis tools for massive amounts of sequencing data

The ability to cheaply and rapidly sequence large numbers of samples has massively increased the amount of data processing required by researchers. Compounding this increase in data, most existing tools have not been designed to take full advantage of the current advances in computer architecture, which parallel solutions to problems which can be run on architectures with vastly differing computational abilities (such as GPUs).

To resolve this, I am in the process of actively developing new open source tools³ which are capable of running on massively parallel architectures using both multiple computers (openMPI) and multiple GPUs on the same computer (Nvidia's CUDA) to

1. Develop an open source massively parallel imputation method which can combine existing GWAS results with new deep sequencing and genetic profiling results.
2. Extend same tools to call SNPs both incrementally and in parallel.

I am currently working on extending samtools (an Open Source SNP calling suite) to support running on multiple computers.

Gut microbiota alteration in SLE

Many autoimmune disorders are known to be affected by gut microbiota. Preliminary evidence suggests that mice which develop SLE have differences in gut microbiota from mice which do not develop SLE, which suggests that SLE severity may also be affected by differences in human gut microbiota. I am currently developing novel methods to

²Direct artery to vein connection without an intervening capillary bed; leads to high pressure arterial flow in venous tissue and can lead to hemorrhage and death.

³and extending existing open source tools where they exist

1. Determine which gut microbiota differ between mice with and without SLE and using 16S sequencing
2. Determine which gut microbiota differ between humans with and without SLE using 16S sequencing, given #1.

Continuous, incremental analysis

Just as science requires continuous testing of hypotheses, bioinformatics is slowly moving towards continuous incremental analysis of data. Most current tools use iterative analysis, where analyses must be completely re-run with each new piece of data that is obtained. When the amount of data remains small relative to the overall computing power available, this is a feasible approach. However, as the amount of data increases, it stops being feasible to completely reanalyze data as new data is obtained, and incremental analysis approaches are necessary. I will be working to extend existing analysis pipelines to handle the incremental analysis of data.

Research Funding

Funding Opportunities

I anticipate obtaining funding from the following sources to pursue the research goals outlined previously:

1. National Institute of Health
 - (a) [Research Project Grant \(R01\)](#) (NHGRI, NIAID, BISTI)
 - (b) [NLM Career Development Award in Biomedical Informatics \(K01\)](#)
 - (c) [Continued Development and Maintenance of Software \(R01\)](#) (extending existing bioinformatics software)
2. Alliance for Lupus Research
 - (a) [Target Identification in Lupus](#)
3. Institution specific funding

Track Record of Funding

The projects that I am proposing have a strong track record of being funded by both the NIH and the Alliance for Lupus Research.

Wider impact of research agenda

My research will identify genetic variants and pathways underlying SLE and other important human diseases, leading to better methodologies for both the diagnosis and treatment of those diseases, and resulting in significant decreases in patient morbidity and mortality. Secondly, the tools that I develop to effect this work will enable researchers in other fields to more rapidly and cheaply identify relevant factors for economically and environmentally important phenotypes, such as pathogen-resistance in crops or disease-susceptibility in thylocenes. Furthermore, as all of my tools will be released under Open Source licenses, external researchers will be able to build upon and improve my tools without being forced to reinvent them.

References

- [1] D L Armstrong et al. “Identification of new SLE-associated genes with a two-step Bayesian study design”. In: *Genes Immun.* 10.5 (July 2009), pp. 446–456. DOI: [10.1038/gene.2009.38](https://doi.org/10.1038/gene.2009.38).
- [2] Don L Armstrong, Chaim O Jacob, and Raphael Zidovetzki. “Function2Gene: a gene selection tool to increase the power of genetic association studies by utilizing public databases and expert knowledge”. In: *BMC Bioinformatics* 9 (2008), p. 311. DOI: [10.1186/1471-2105-9-311](https://doi.org/10.1186/1471-2105-9-311).
- [3] Chaim O Jacob et al. “Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus”. In: *Proc. Natl. Acad. Sci. U.S.A.* 106.15 (Apr. 2009), pp. 6256–6261. DOI: [10.1073/pnas.0901181106](https://doi.org/10.1073/pnas.0901181106).
- [4] Chaim O Jacob et al. “Identification of novel susceptibility genes in childhood-onset systemic lupus erythematosus using a uniquely designed candidate gene pathway platform”. In: *Arthritis Rheum.* 56.12 (Dec. 2007), pp. 4164–4173. DOI: [10.1002/art.23060](https://doi.org/10.1002/art.23060).
- [5] Chaim O Jacob et al. “Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase”. In: *Proc. Natl. Acad. Sci. U.S.A.* 109.2 (Jan. 2012), pp. 59–67. DOI: [10.1073/pnas.1113251108](https://doi.org/10.1073/pnas.1113251108).
- [6] Christopher J Stapleton et al. “Thrombospondin-1 modulates the angiogenic phenotype of human cerebral arteriovenous malformation endothelial cells”. In: *Neurosurgery* 68.5 (May 2011), pp. 1342–1353. DOI: [10.1227/NEU.0b013e31820c0a68](https://doi.org/10.1227/NEU.0b013e31820c0a68).