

eQTL mega-analysis – a tool for functional assessment of multi-enhancer gene regulation

eQTL has emerged as an important tool for unravelling the relationship between genetic risk factors and disease or clinical phenotypes

Aberrant gene expression can play a large part in complex disease. **Professor Greg Gibson's** study at the Center for Integrative Genomics at Georgia Tech uses advanced statistical methodologies and evolutionary comparisons to refine the identity of a small number of sequence variants that are most likely to regulate expression of genes in the immune system. His team then confirms their role in inflammatory autoimmune disease, in part using genome engineering technologies to experimentally confirm their function.

Professor Gibson's work harnesses the power of expression quantitative trait loci (eQTL), which are regions of the genome containing DNA sequence variants that influence the expression level of one or more genes. eQTL has emerged as an important tool for unravelling the relationship between genetic risk factors and disease or clinical phenotypes. Whereas most sequence variants have no effect on gene expression, there are some

that do. A major current focus in the field is to fine-map the functional sites within loci that have been identified by studying genetically different individuals. In order to investigate these further, it is necessary to compare individual genotypes with the level of gene expression observed. Statistical approaches are subsequently used to test whether a particular sequence variant has a marked effect on the expression of a particular gene. This project also incorporates

evolutionary insight into the fine mapping, since conservation across species is the best signature of functional importance.

THE CHANGES IN GENE EXPRESSION THAT CAUSE DISEASE

Many of the associations between genetics and risk of disease are thought to be the result of regulation of gene expression; that is when, where and at what level the relevant genes are expressed. Particularly in recent years,

whole exome sequencing (sequencing all of the protein-coding genes in a genome), has demonstrated that there is a burden of rare variants in individuals with a variety of neurological and developmental conditions. Given that approximately 90% of disease-associated variants can have regulatory functions, it is reasonable to hypothesise that these variants may be ones that cause misregulation in individuals with common chronic diseases or congenital abnormalities. ▶

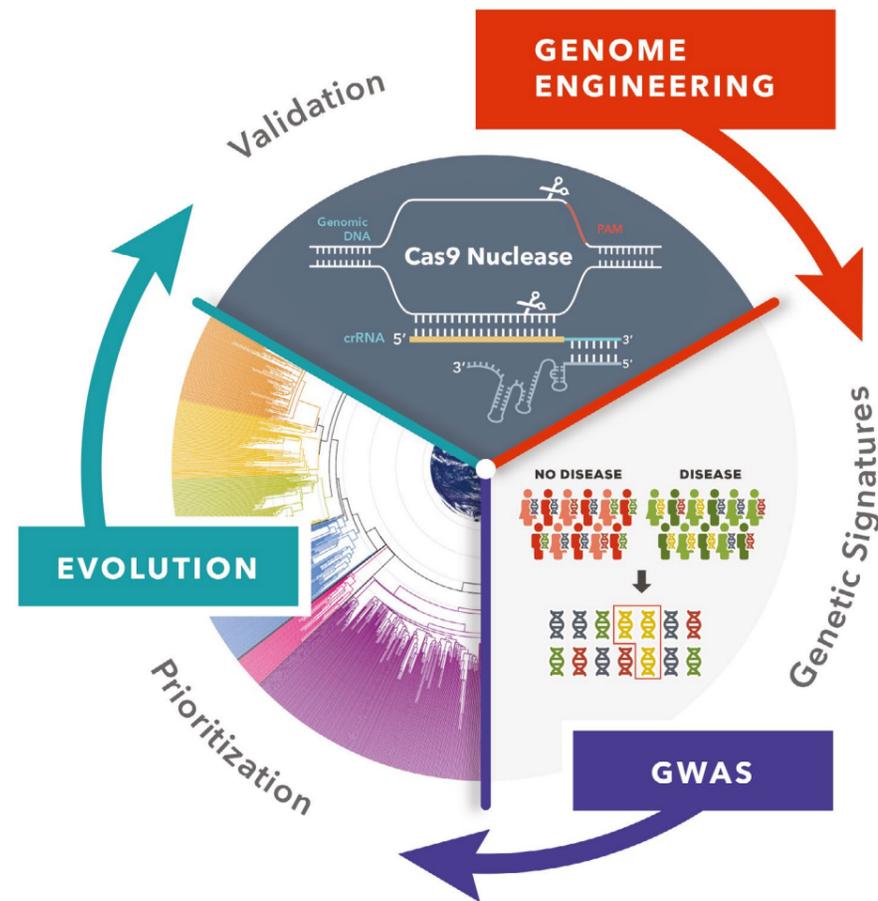
Whilst the work of the Gibson Lab focuses on investigating eQTL in multiple large studies of the genetics of blood gene expression, they collaborate with several other groups. The Kumar Group, based at Temple University, concentrates on insights that can be derived from molecular evolutionary analysis and Gang Bo and Ciaran Lee, at Rice University, provide expertise around CRISPR/Cas9 mediated genome editing that is used to experimentally validate predictions.

Specifically, Prof Gibson's current project searches for regulatory polymorphisms in a unique resource of 10,000 peripheral blood transcriptome profiles linked to whole genome genotypes. Given that peripheral blood contains many cell types, the expectation is that genetic effects are modified in disease by the inflammatory agents, with some variants losing their effect and other novel variants arising only once inflammation is seen.

They will use multivariate regression to fine map (refine the list of all associated variants) the variants that are most likely to be linked to transcriptional regulation in inflammatory autoimmune diseases. Further computational techniques, such as sparse learning models deployed by collaborator Li Liu at Arizona State University, will be used to predict which variants are more likely to influence transcript abundance, and simulations can be used to model what the effects of this may look like.

REGULATION OF GENE EXPRESSION, AND WHY EVOLUTIONARY METHODS ARE SO RELEVANT

Recent work from the Gibson Lab shows that there is a burden of rare sequence variants in the promoter regions (regions of DNA that initiate transcription of a particular gene) of genes with unusually high or low expression. This suggests that it is not simply the effect that mutations may have on protein structure that is important, but also the effect that they have on gene expression. This finding was confirmed by another group in a recent study published in *Nature*. In earlier studies which also looked at multiple blood parameters and whole genome genotypes, they showed that



Professor Gibson's research harnesses the power of expression quantitative trait loci (eQTL), regions of the genome containing DNA sequence variants that influence the expression level of one or more genes. His approach combines quantitative genetics, molecular evolutionary analysis, and genome engineering work.

half the genes with one variant influencing gene expression (so-called cis-eQTL) have a second one, and half of these have a third. This means that genetic regulation of gene expression is much more complex than currently thought, and that linking eQTL to pathologies is not as straightforward, and needs to become a major focus for future studies.

It is generally accepted that nucleotide sites which are highly conserved across species are more likely to be functional. Dr Kumar, one of Prof Gibson's collaborators, has developed a score known as the Evolutionary Probability

(EP) score to investigate this further. The EP score evaluates the probability that at any one position in the genome, specifically which one of the four nucleotides (A, G, C or T) is present. When this is related to protein coding regions, nucleotides with a low EP score are more likely to be pathogenic. The team will evaluate whether the same holds true for regulatory variants.

CRISPR-Cas9 mediated genome engineering is becoming the most commonly used tool to remove, add or alter parts of a DNA sequence. It harnesses an existing natural system used by bacteria in their adaptive immune responses. In conjunction with the group at Rice University, Prof Gibson aims to mutagenise candidate polymorphisms identified by the eQTL and EP analyses, to explore what effect this has on genes that are related to auto-immune disease. CRISPR-Cas9 mediated site-specific genome engineering will be used to experimentally confirm the predictions of

© Michael Li and Sudhir Kumar (Temple University)

This resource can be used to understand the genetic mechanisms behind some of our most rare and complicated diseases



Q&A

If you understand how gene expression is aberrantly regulated in disease, does this mean it could be used as a potential gene therapy target?

Yes, that could certainly be the case. Either you could try to up- or down-regulate the affected gene as appropriate with additional copies or RNAi, or you could use CRISPR/cas9 to revert the mutation to wild-type.

Could you use genotyping and transcriptional profiling to screen individuals and predict how likely they may be to develop a particular disease at some stage in their life?

This is in fact the other major focus of my group. A few years ago I was very interested in using blood profiling to predict risk of common metabolic, immune and possibly other complex diseases with late-middle age onset. However, it is too complex! Lately we have switched to prediction of disease progression at diagnosis, which we think has enormous potential for guiding therapeutic intervention. See <https://www.ncbi.nlm.nih.gov/pubmed/28259484> and <https://www.ncbi.nlm.nih.gov/pubmed/28805827> for Crohn's Disease.

What can people do to help progress your research, i.e. should they volunteer with programmes such as 'The 100,000 Genomes Project' in England?

That's certainly a good suggestion. The equivalent AllOfUs program in the US is inviting participants at JoinAllOfUs. People need to appreciate that there is a long lag

their computational analysis in an *in vitro* lymphoid cell line. Each candidate site will be disrupted, and the impact that the mutation has on gene expression measured using a combination of nanoscale PCR and single cell RNA sequencing.

Prof Greg Gibson's work focuses on ways that advanced statistical methods can be used to refine the identity of a small number of the most likely variants that regulate expression of genes in the immune system, and to confirm their role in inflammatory immune diseases. The computational and

time between the research and changing their life. I have people with autoimmune diseases approach me to have their blood profiled, and I tell them that I am happy to, but there is not much we can do with the information yet. It is disappointing, but then the focus switches to the positive feelings most people get from participating in research (so long as it is appropriately consented, which is difficult since it is so complex).

What are the biggest challenges or limitations that your research currently faces?

As always, ongoing funding! And the changing landscape of methods. Since the project started, we have decided to switch to single cell RNASeq, which has become cost effective and is so much more precise than PCR, but of course there are protocols to be developed which takes time.

What's next for your research?

I really want to spend the next ten years working on various strategies for using RNASeq to go from bench to bedside. I now think that prediction of disease progression from the combination of genetic analysis and transcript profiling can be so powerful in the context of therapeutic intervention. Then of course we are always going to be interested in the evolution of disease risk, which means understanding in fine detail how changes in eQTL frequency influence population differences in disease prevalence.

experimental approaches are expected to be applicable to many other, commonly occurring diseases. All code will be made publicly available, alongside software for evolutionary genome analysis. In the long term, this will build up a resource that can be used to try and understand the genetic mechanisms behind some of our most rare and complicated diseases.

Detail

RESEARCH OBJECTIVES

Professor Gibson's laboratory explores the use of transcriptome profiling to better understand the genetic basis of complex traits and disease susceptibility. He research focuses on three main areas: quantitative evolutionary genetics, immuno-transcriptomics and personalised genomic medicine.

FUNDING

- National Institutes of Health (NIH)

COLLABORATORS

- Sudhir Kumar (Temple University),
- Li Liu (Arizona State University)
- Gang Bao (Rice University)
- Ciaran Lee (Rice University)

BIO

Professor Gibson trained as a geneticist at Sydney University and earned his PhD at the University of Basel. He completed his postdoctoral training at Stanford University.

Greg has been at a Professor in the School of Biology at Georgia Institute of Technology since 2009, where he currently is Director for the Center for Integrative Genomics.

CONTACT

Professor Gregory Gibson
Krone Engineered Biosystems Building
950 Atlantic Drive
Atlanta, GA 30332
USA

E: greg.gibson@biology.gatech.edu

T: +1 404 385 2343

W: <https://gibsongatech.wordpress.com>