Massively Parallel Reporter Assays (MPRAs) Identify Allelic Transcriptional Dysregulation in Atopic Dermatitis

Amy A. Eapen*, Xiaoming Lu*, Xiaoting Chen*, Carly Forney, Daniel Miller, Sreeja Parameswaran, John P. Ray, Carl G. de Boer, Matthew T. Weirauch*, Leah C. Kottyan*; *Contributed Jointly; #Co-corresponding

**Background**
- 20% of children develop atopic dermatitis (AD)
- 10-30% AD children have persistent disease as adults
- CD4+ T cells produce inflammatory cytokines and contribute to dysfunction of the skin barrier
- There is a strong genetic component to AD development as documented by twin concordance studies
- Genome-wide association studies have identified 29 independent genetic risk loci associated with AD (p<5x10^-8)
- 95% of these genetic risk variants are non-coding

**Experimental Design**
- 3,143 SNPs and insertions/deletions accounting for 29 independent loci.
- Loci were identified from 122 reported “tag” variants reaching genome-wide significance.
- 150 bp around each SNP.
- Calibration oligos that are from genetic variants that have been previously shown to be allelic or not in previous MPRAs assays.
- 300-500 tags per allele (oligo)*
- Replicates – 5 transfections/condition
- Sequencing → 30-fold coverage

**Schematic of MPRA design and library preparation**

**Massively Parallel Reporter Assays**
- MPRAs allow us to assess gene regulatory activity of DNA surrounding 1000's of genetic variants at the same time.
- Any cell type
- Any condition (e.g., stimulation, therapeutics)
- A pool of reporter constructs is assembled.
- DNA “tags” are used instead of luciferase as a reporter.
- Each reporter is made many times with many different tags.
- Pool is transfected into cells.
- Measures enhancer activity and genotype-dependent enhancer activity.

**Enhancer activity at AD risk variants**
- 619 AD oligos → 302/3143 variants have enhancer activity (p<0.05, >50% change)

**Genotype-dependent Enhancer Activity at AD-disease risk variants**
- Ongoing work
  - Different cell lines + stimulation – e.g., HaCaT (immortalized human keratinocytes) and Jurkat (immortalized human T cells) +/- TNFα
  - Integration with other functional genomics data to nominate causal allelic transcriptional regulatory mechanisms at AD disease risk variants.
  - Complement with eQTL data, chromatin conformation assay data, genome-editing to identify genes that are regulated by SNPs.

**Conclusions**
Our results are consistent with a model in which AD genetic risk variants regulate gene expression in a genotype-dependent manner. We nominate 108 plausibly causal variants for AD.

**Genotype-dependent Enhancer Activity at AD-disease risk variants: Example locus**

*There are no STAT6 "motifs" in HOMER, but all STATs bind similar DNA sequences.

**Effect of IL13 on enhancer activity at AD risk variants**

The DNA sequence around variants with IL13 dependence were enriched for STAT* and IRF motifs.

**GTEx – skin biopsy**

There are 98 variants in LD with rs6729638 (r2>0.8), and we know the single variant with allelic activity (MPRA) and the gene affects (eQTL)!