Genetic regulation of gene expression variation

Barbara E. Stranger
Section of Genetic Medicine
Institute for Genomics and Systems Biology
Center for Data Intensive Science

ENCODE Users group meeting, Stanford University 6/8/2016

Institute for Genomics & Systems Biology

Center for Data Intensive Science

THE UNIVERSITY OF CHICAGO
DEPARTMENT OF MEDICINE

Outline

- The Immunological Variation (ImmVar) Project
  - Baseline eQTLs in adaptive and innate immunity
    - Context specificity
    - Role in disease
  - Activation eQTLs

Outline

- The Immunological Variation (ImmVar) Project
  - Baseline eQTLs in adaptive and innate immunity
    - Context specificity
    - Role in disease
  - Activation eQTLs

The Immunological Variation Project (ImmVar)

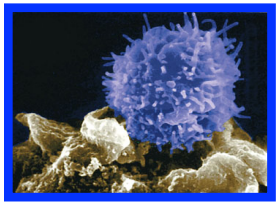
Goal: Understand how genetic variability translates into gene expression variability in adaptive and innate immunity and contributes to higher order phenotypes

ImmVar Study Design

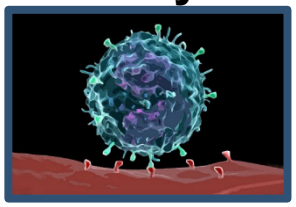
Boston-based healthy donors (n=415)
European American (EU, n=215)
African American (AA, n=115)
East Asian (EA, n=85)

Phenogenetic Project (Phillip De Jager)

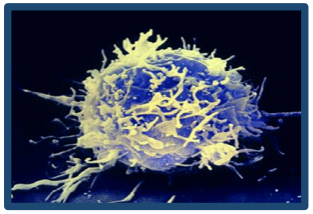
Naïve CD4+ T-cells



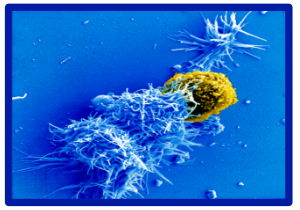
CD14+16- monocytes



T-cell activation



Dendritic cell activation

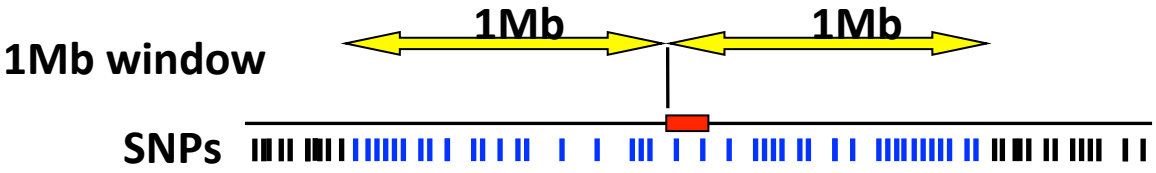


Transcriptome profiling (Affy Exon 1.0 ST)
+
Genome-wide genotyping & imputation

→

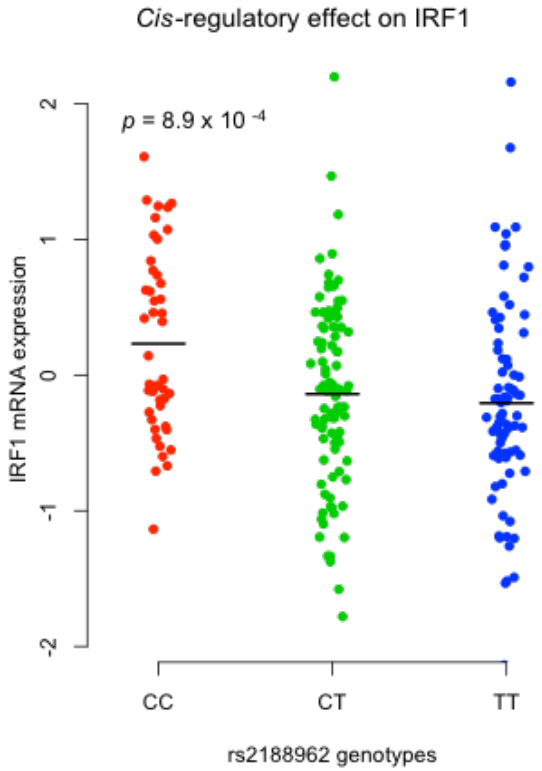
eQTL mapping
cis- and *trans*-

cis-eQTL analysis (baseline)



Model: residual gene expression profile
(control for age, sex, PCgt, PCge)
+
Genotyped and imputed SNPs

Significance assessed by 10,000 permutations per gene:
Nominal $p < 0.01$ tail of the minimal permutation *p-value*



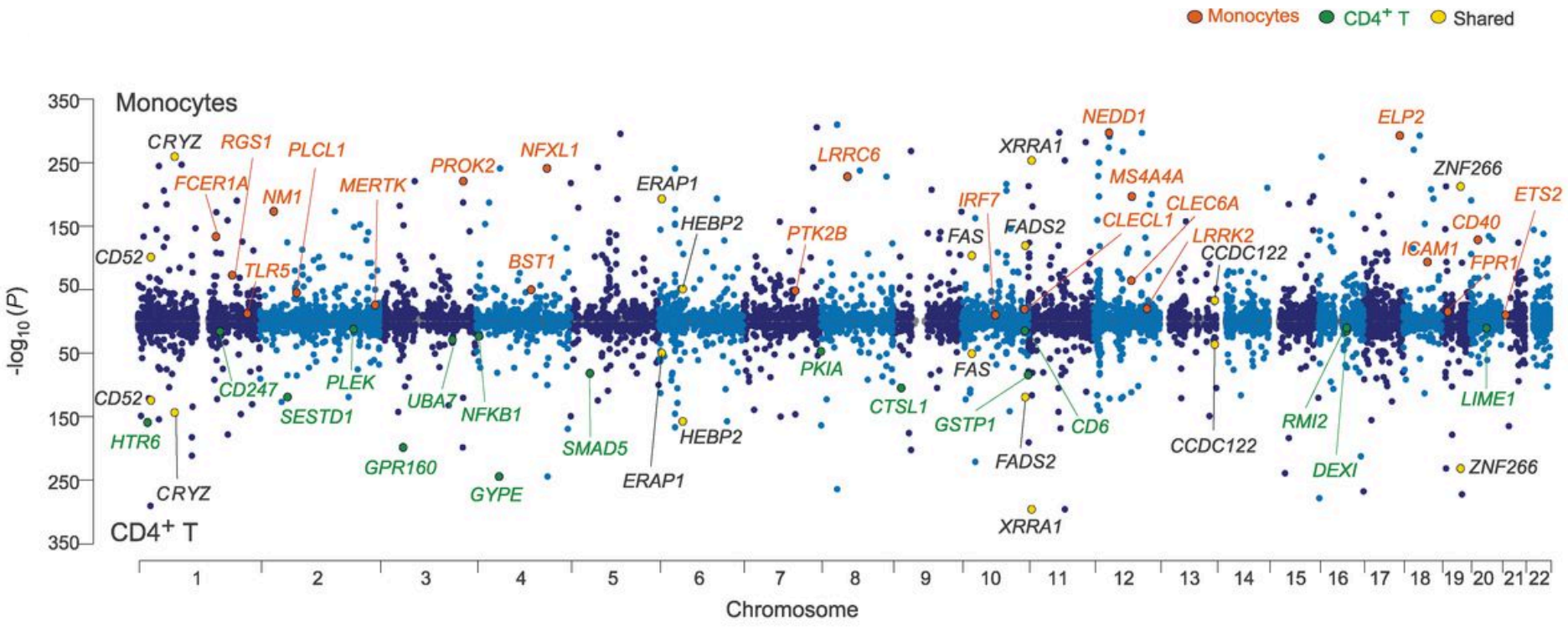
1) Analysis performed ***separately*** for each cell-type and population
2) Multi-ethnic ***Meta-analysis*** for each cell type

T. Raj et al 2014 *Science*

Detected eGenes

Population	Monocyte		$CD4^+$ T cell		Shared
	No. of participants	No. of genes	No. of participants	No. of genes	No. of genes
European American	211	3090	213	2352	1178
African American	112	1318	112	722	259
East Asian	78	1181	82	592	215
≥Two populations	N/A	1352	N/A	739	328
Three populations	401	537	407	255	102
Nonredundant	401	3703	407	2672	1372
Meta-analysis	401	6210	407	4975	2789

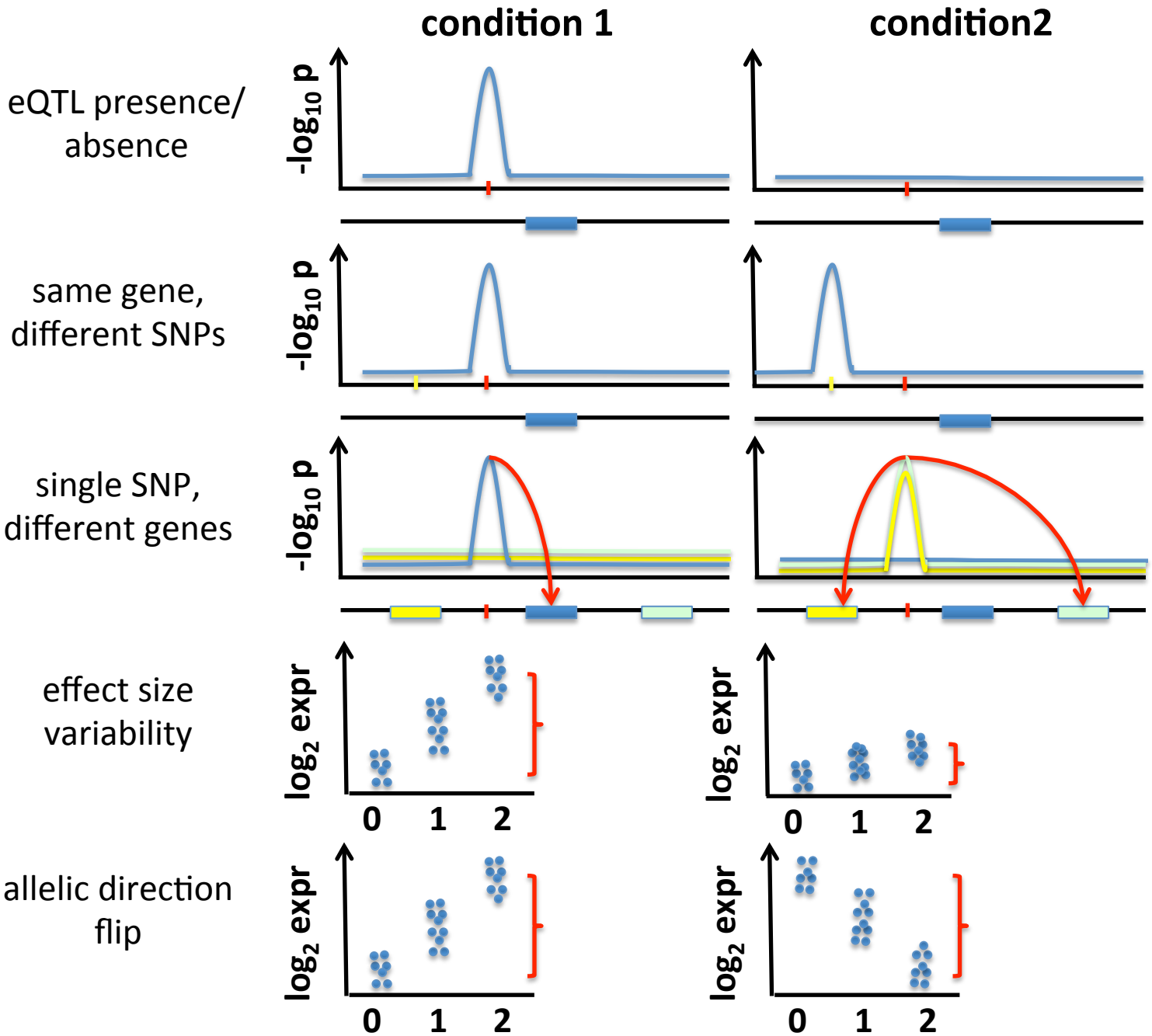
Baseline eQTLs in CD4+ T and CD14+16- 30% of tested genes have *cis*-eQTL associations



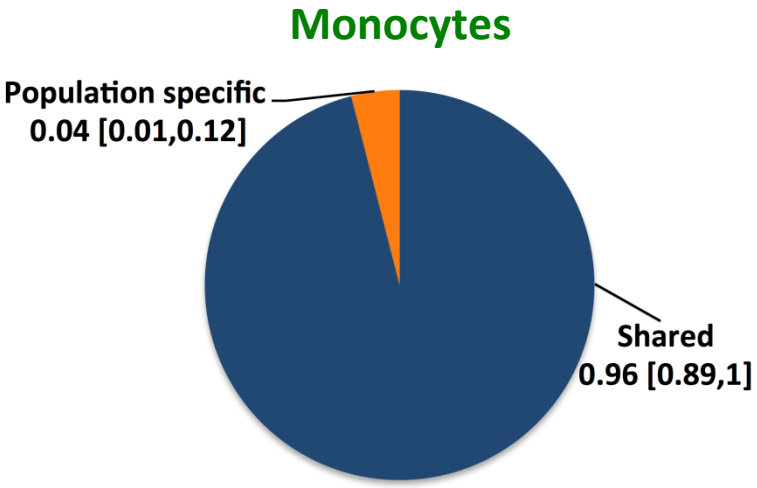
Up to 17% of genes with a cis-eQTL, have >= 2 independent signals

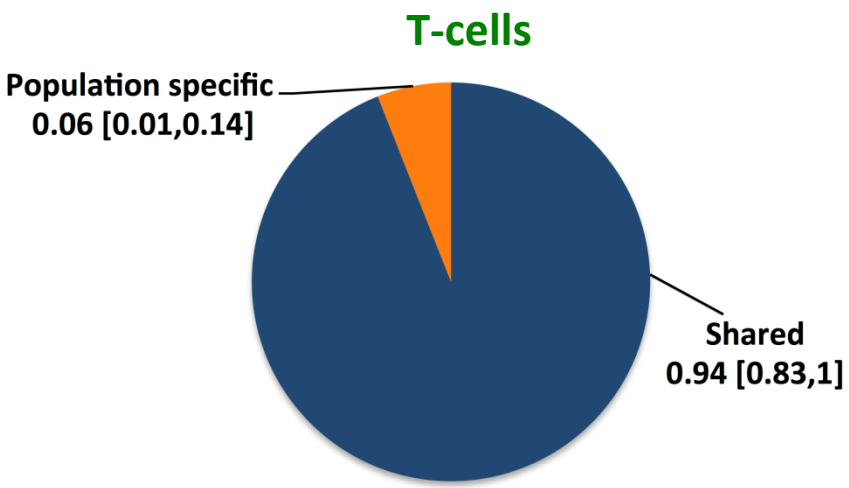
Outline

- The Immunological Variation (ImmVar) Project
  - Baseline eQTLs in adaptive and innate immunity
    - Context specificity
    - Role in disease
  - Activation eQTLs


condition 1
condition2
eQTL presence/ absence
same gene, different SNPs
single SNP, different genes
effect size variability
allelic direction flip
-log10 p
log2 expr
0
1
2

Population specificity of *cis*-eQTLs?





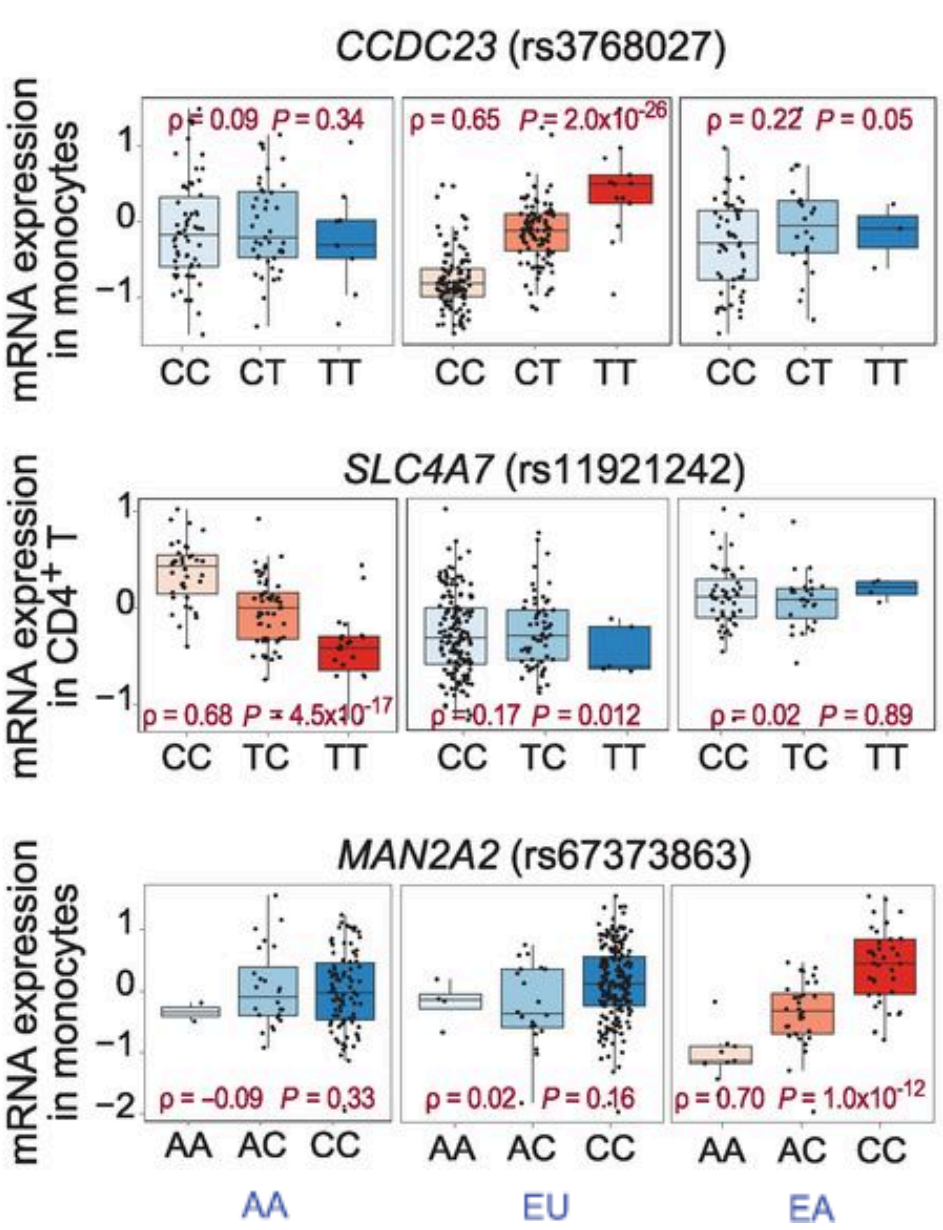
For shared-population eQTLs: Fold change across populations highly correlated; Pearson's r: 0.85 -0.95

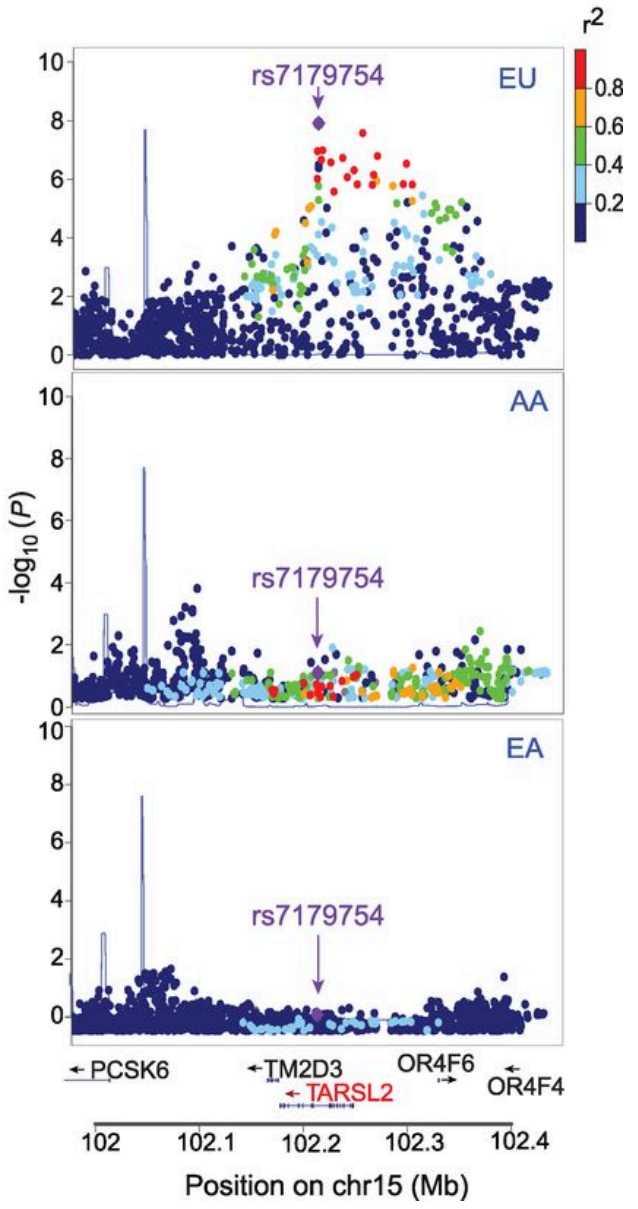
Little population-specificity of presence/absence or fold-change modulation

BUT those differences might be **highly relevant** for population-diverged phenotypes

Bayesian model average and a hierarchical model BF_{BMA}; Flutre et al. 2013

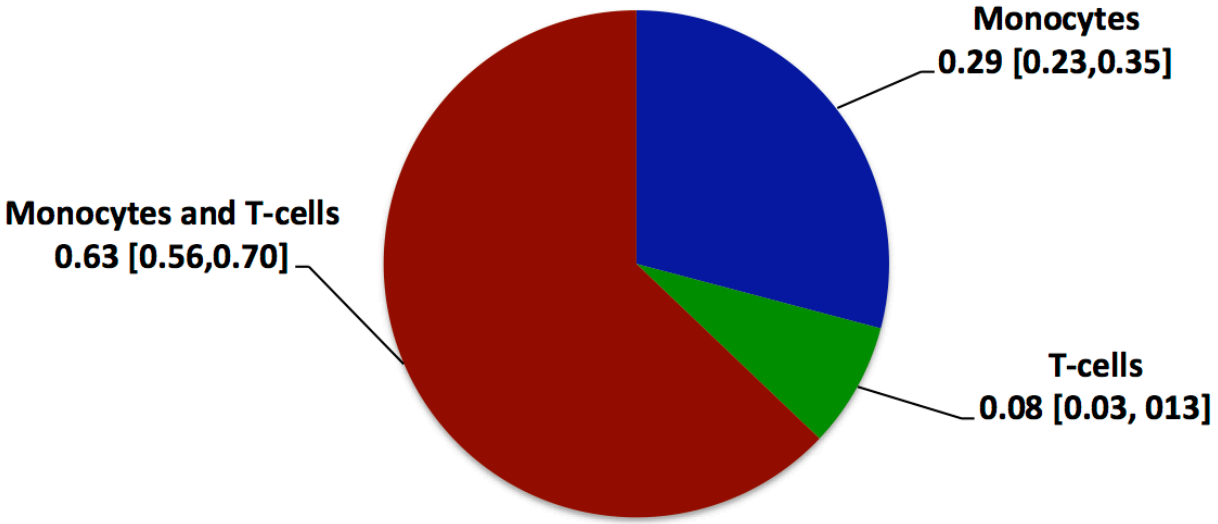
Population specific *cis*-regulatory effects





TARSL2: threonyl-tRNA synthetase-like 2

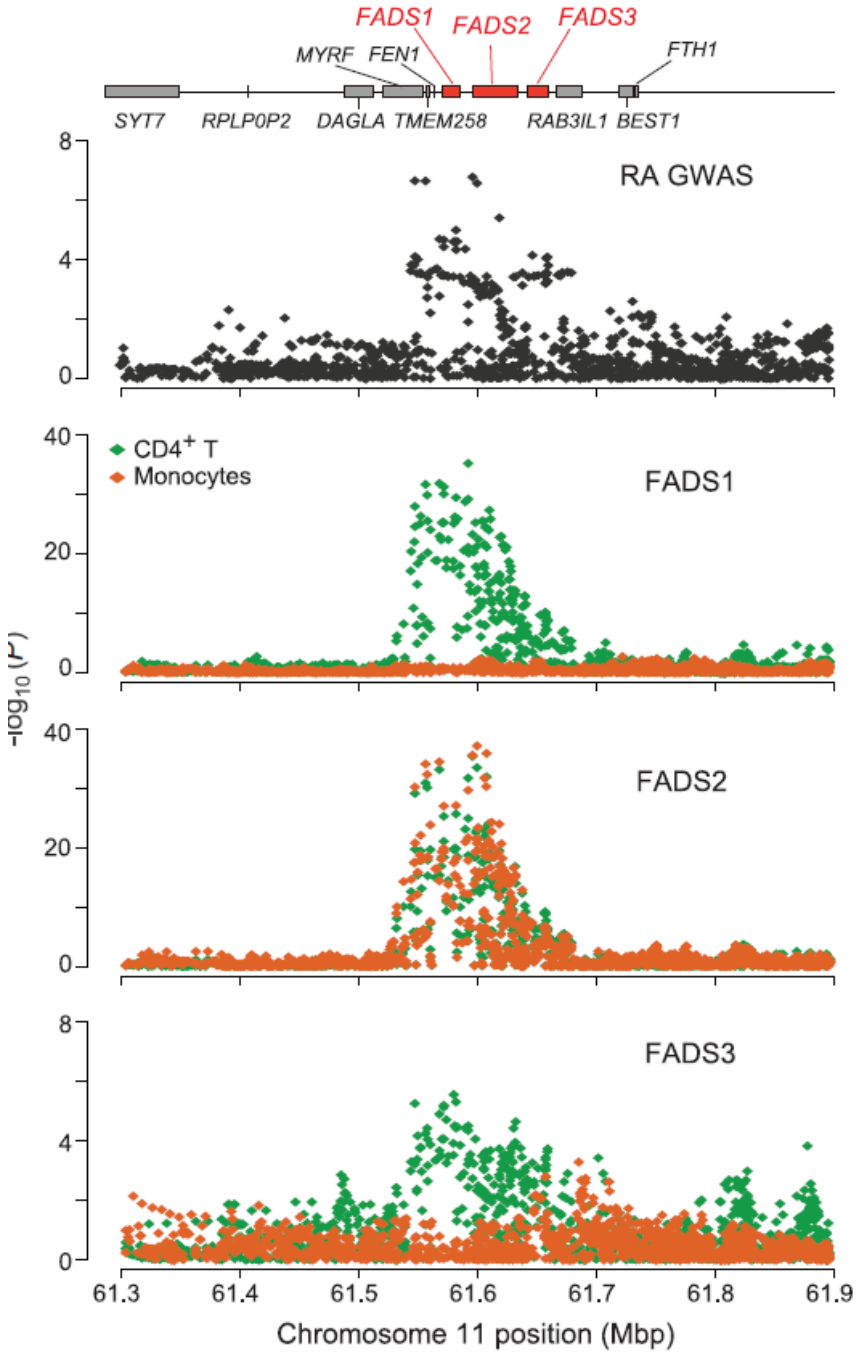
~40% *cis*-eQTLs are cell-type specific



For eQTLs shared across cell types: Fold change across cell types less highly correlated; r: 0.49-0.64

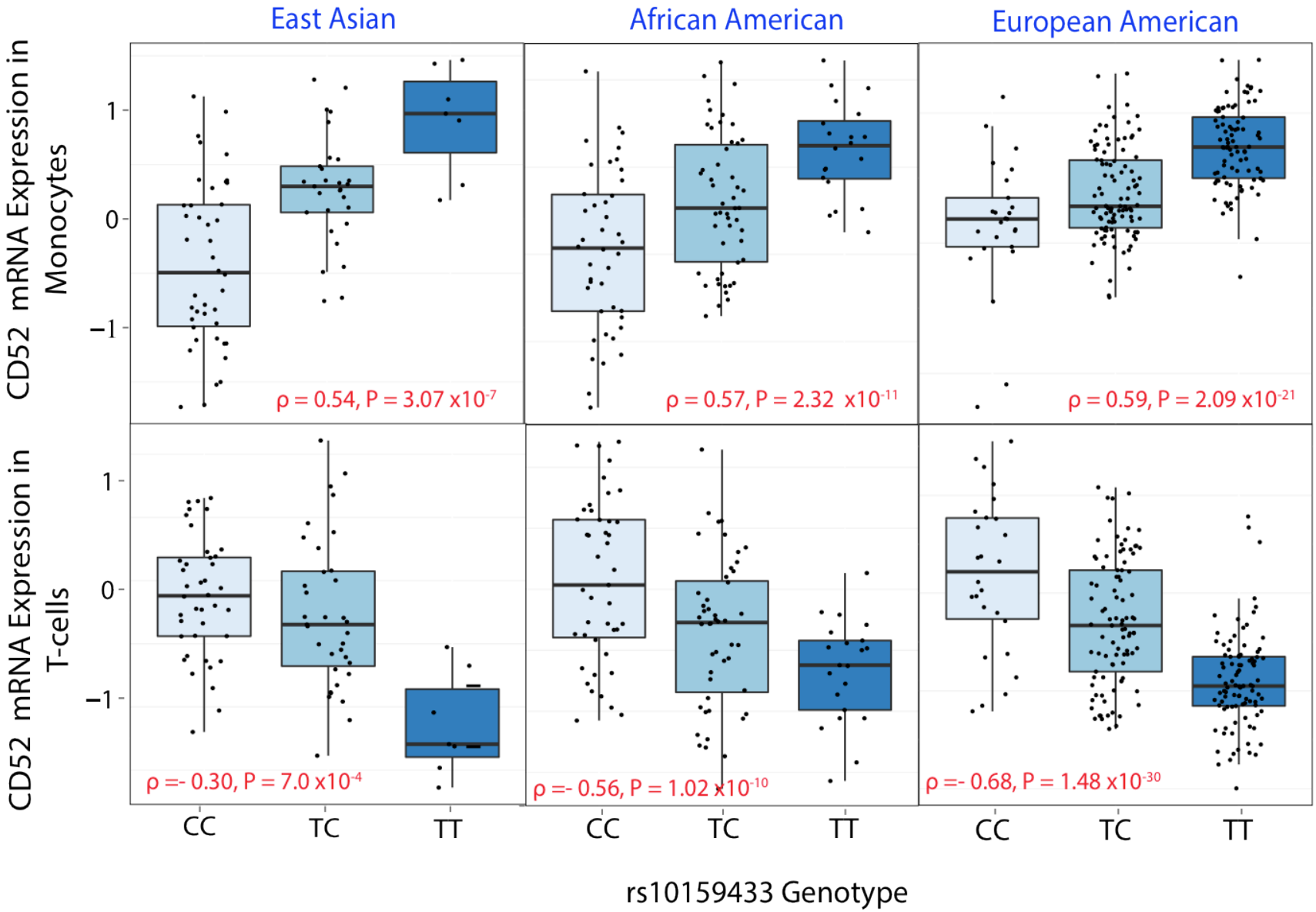
Evidence for cell – type specificity in presence/absence AND fold change

Cell-type specificity of *cis*-eQTLs in a rheumatoid arthritis risk locus



FADS genes: 11q12-q13 fatty acid desaturase involved in atopic disease

CD52 *cis*-eQTL shows directional regulatory effects across cell types



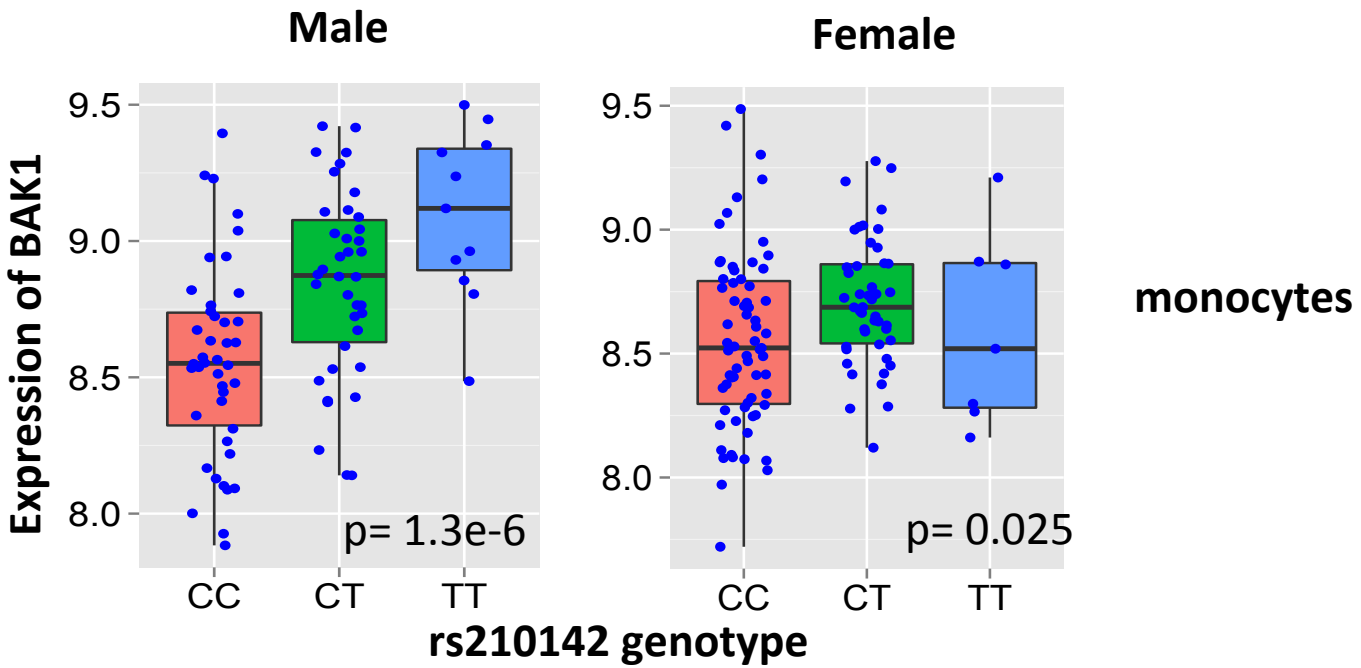
CD52 lymphocyte cell-surface glycoprotein, function in anti-adhesion, role in lymphoma. It is the protein targeted by alemtuzumab, a monoclonal antibody used for the treatment of chronic lymphocytic leukemia

Sex-differentiated eQTLs

- Autosomal genetic variation contributes to sexual dimorphism (Ober *et al.* 2008; Heid *et al.* 2010)
- Sex-specific eQTLs have been detected in mice (Yang *et al.* 2006) and humans (Dimas *et al.* 2012, Kent *et al.* 2012, Yao *et al.* 2014, Kukurba *et al.* 2015)

X% of *cis*-eQTLs exhibit sex-bias

CLL risk locus. chronic lymphocytic leukemia: Male risk is 2X female risk (Slager *et al.* 2012)



BAK1: BCL2-Antagonist/Killer 1; role in apoptosis, interacts with the tumor suppressor P53 after exposure to cell stress

Outline

- The Immunological Variation (ImmVar) Project
  - Baseline eQTLs in adaptive and innate immunity
    - Context specificity
    - Role in disease
  - Activation eQTLs

GWAS-associated SNPs are more likely to be eQTLs

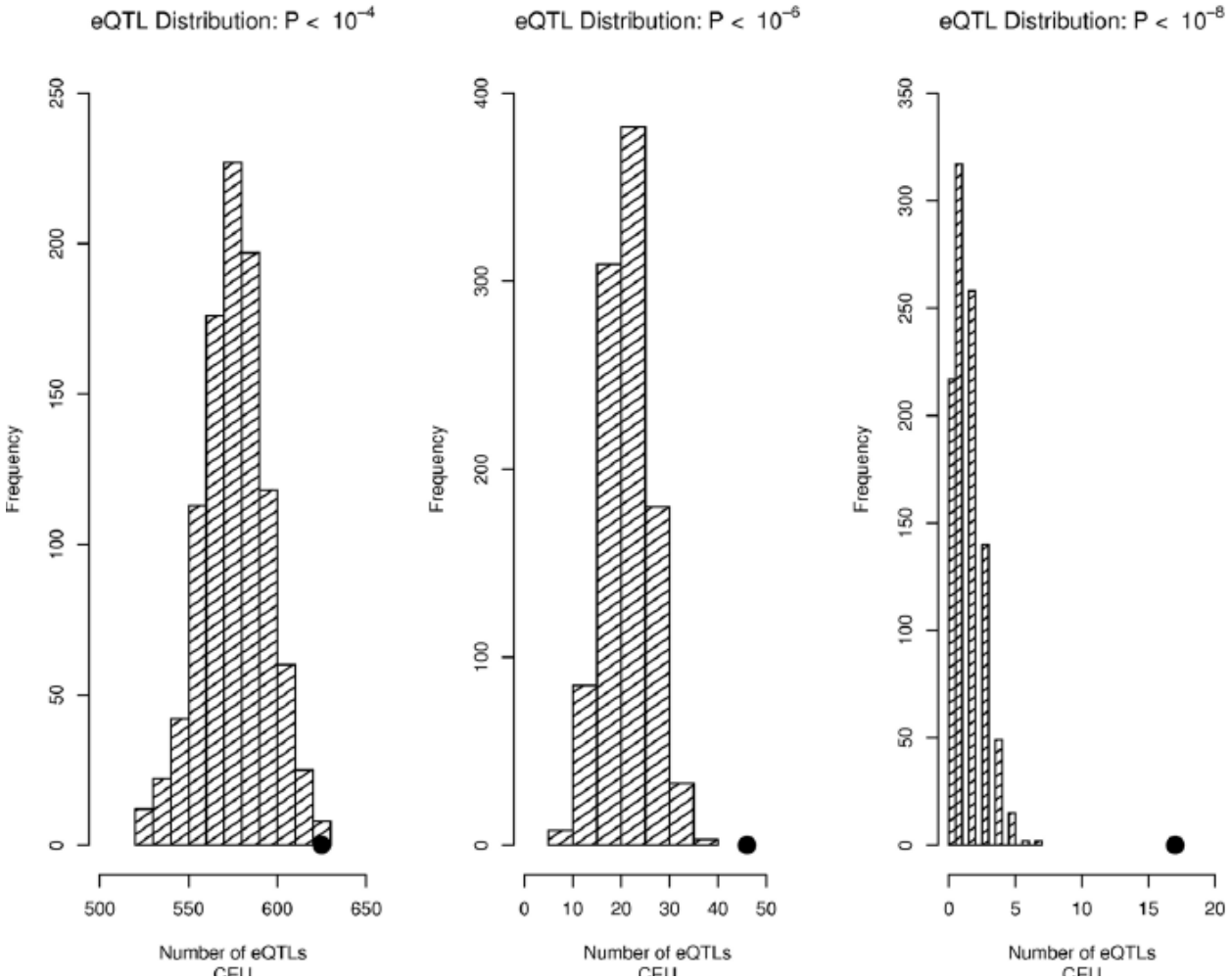
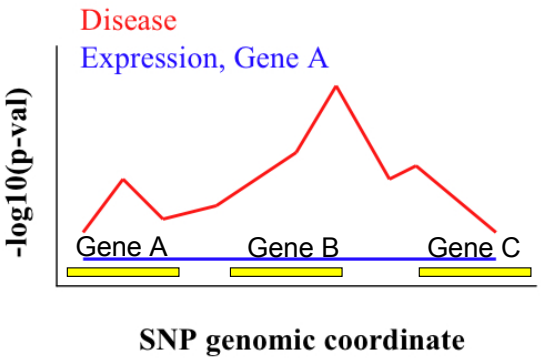


Figure 1. Trait-associated SNPs are more likely to be eQTLs. The distribution of the number of eQTLs (defined as $p<10^{-4}$ left panel, $p<10^{-6}$ middle panel, and $p<10^{-8}$ right panel) observed for each of 1,000 draws of 1,598 SNPs from bins matched for minor allele frequency to the 1,598 SNPs downloaded from the NHGRI catalog (bins include all SNPs in the Illumina 1M and Affymetrix 6.0 products) is shown in the bar graphs, with the actual number of eQTLs observed in the 1,598 SNPs from the NHGRI catalog shown as a solid circle.
doi:10.1371/journal.pgen.1000888.g001

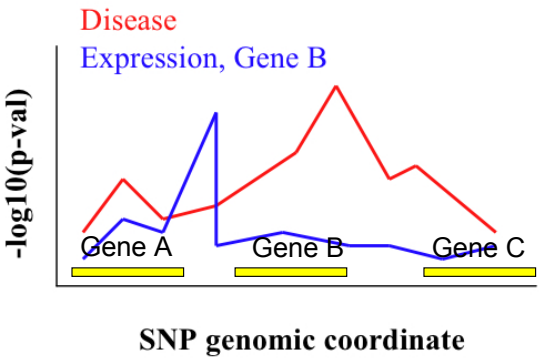
Nicolae *et al.* 2010 *PLoS Genetics*

GWAS: Where is the causal disease variant and what does it do?

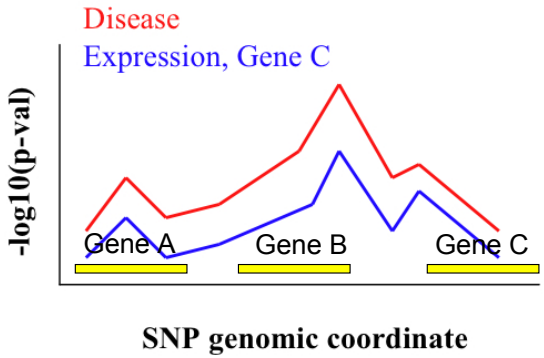
a)



b)



c)



Nica and Dermitzakis *Hum Mol Genet* 2008

Common disease variants are *cis*-eQTLs in ImmVar data

Traits	# SNPs	Monocytes	T-cells	Shared
GWAS Curated (LD-pruned)	1,068	94	53	29
Cancers	121	7	3	1
Neurodegenerative diseases	55	19	0	1
Neuropsychiatric	27	3	4	3
Metabolic diseases	161	13	2	7
Height	180	12	5	1

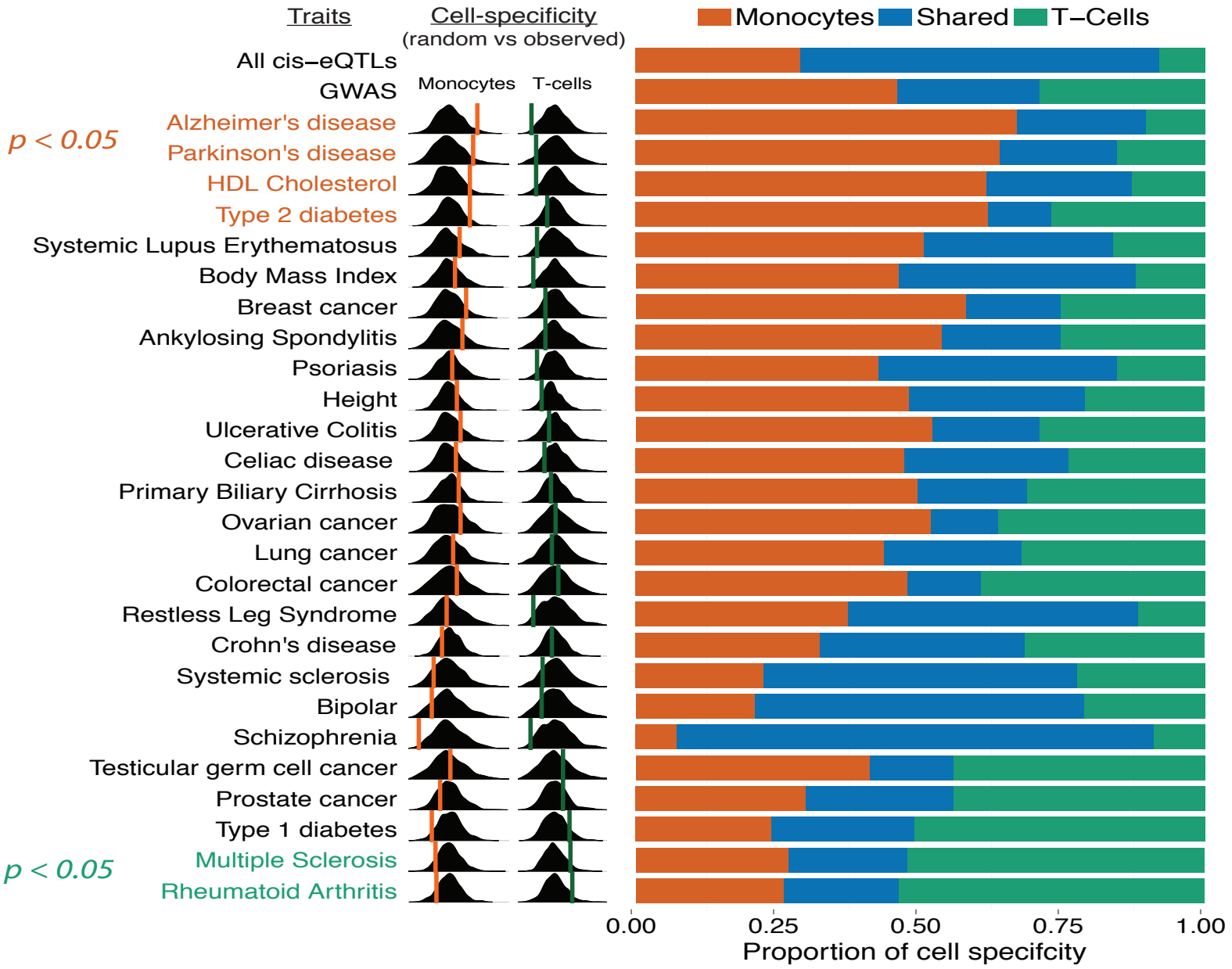
Autoimmune disease associated SNPs

Disease	# of GWAS SNPs	# of SNP-gene(s) (eQTL)	# of GWAS SNPs (eQTL)	# of Genes (eQTL)
Ankylosing spondylitis (AS)	16	17	10	14
Crohn's disease (CD)	90	54	29	52
Ulcerative colitis (UC)	58	34	19	34
Celiac disease (CeD)	80	15	13	13
Multiple sclerosis (MS)	83	22	20	22
Type 1 diabetes (T1D)	53	30	17	29
Rheumatoid arthritis (RA)	70	22	17	22
Primary biliary cirrhosis (PBC)	19	5	5	5
Systemic lupus erythematosus (SLE)	27	12	7	12
Systemic sclerosis (SS)	18	5	4	5
Psoriasis (PS)	54	25	15	22
Total	**568**	**241**	**156**	**230**
Total (LD-pruned, top SNP per LD-block, n.r.)	**425**		**143**	**164**

n.r: Non-redundant.

Regulatory Trait Concordance, RTC score > 0.9 (Nica et al 2010):
T-cells 106 genes, monocytes 123 genes

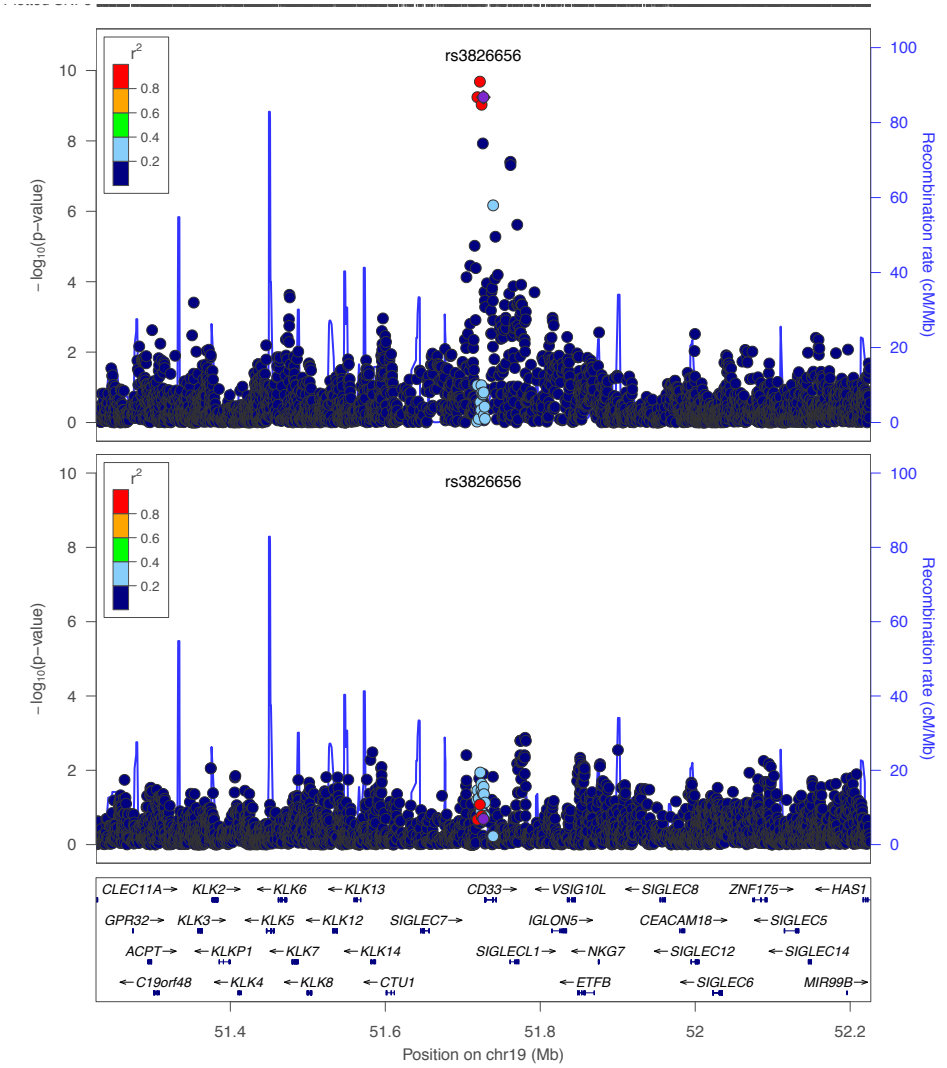
Polarization in the regulatory effects of neurodegenerative and inflammatory disease variants

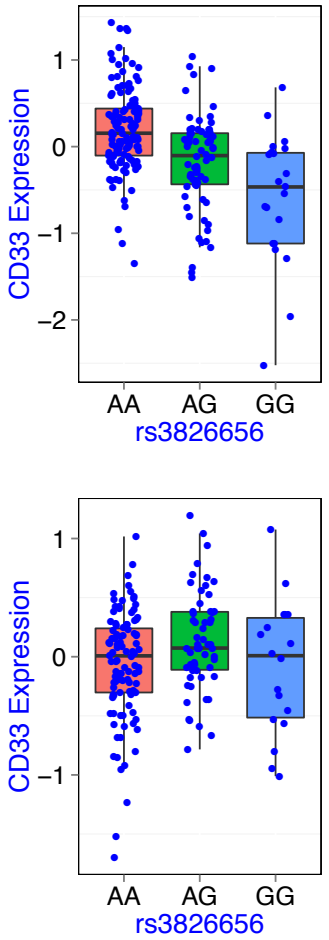


Monocyte-specific *cis*-eQTL for *CD33* associated with Alzheimer's Disease

Monocytes

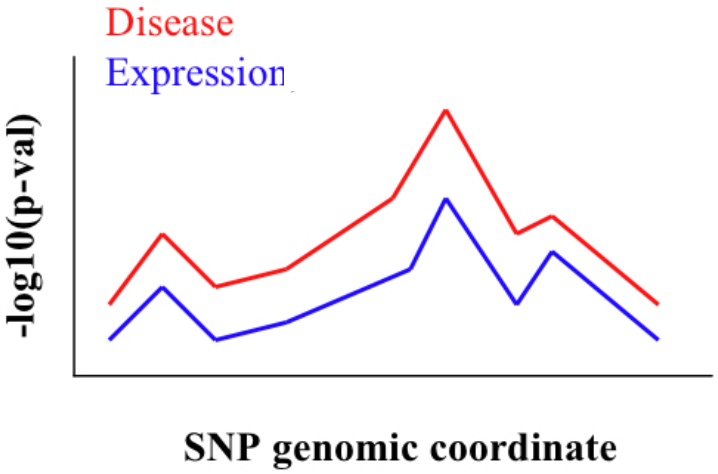
T-cells



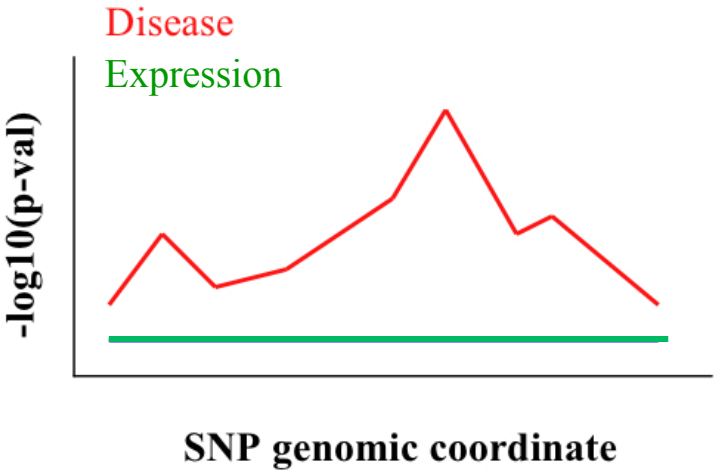


Previous studies report over-expression of CD33 on cell surface of microglia in postmortem brains of late-onset AD patients. CD33 expression level correlated with beta-amyloid protein and plaque accumulation.

Exploring eQTLs in the relevant cell type is important for disease association studies



relevant cell type for disease



cell type not relevant for disease

Importance of cataloguing regulatory variation in multiple cell types

GTEx

modified from Nica and Dermitzakis *Hum Mol Genet 2008*

Outline

- The Immunological Variation (ImmVar) Project
  - Baseline eQTLs in adaptive and innate immunity
    - Context specificity
    - Role in disease
  - Activation eQTLs

Cell stimulation motivation

- Immune cells respond to different stimuli through different circuits
  (Amit *et al.* 2009, *Science*, Gat-Viks *et al.* 2013 *Nat Biotech*)
- Do healthy individuals vary in their immune response?
- Is there a genetic basis to this response?
- Does the variation in immune response relate to clinical disease?
- Can we leverage naturally occurring variation to reconstruct regulatory relationships?

Immune cell activation

Innate Immunity

Dendritic cells (DCs)

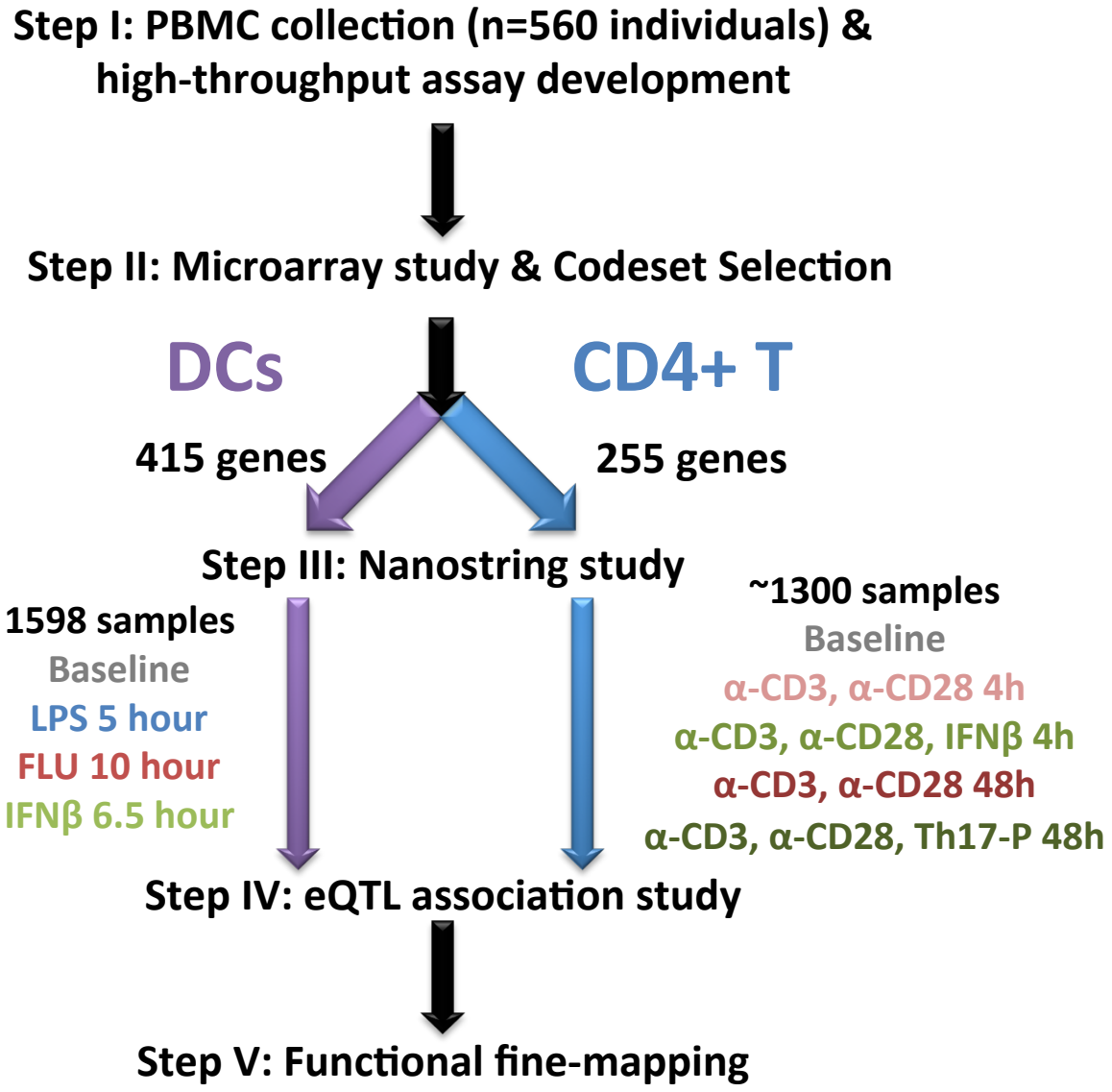
- Lipopolysaccharide (LPS)
  - bacteria
- Influenza
  - Virus
- Interferon-beta (IFN-β)
  - Virus

Adaptive Immunity

T-cells

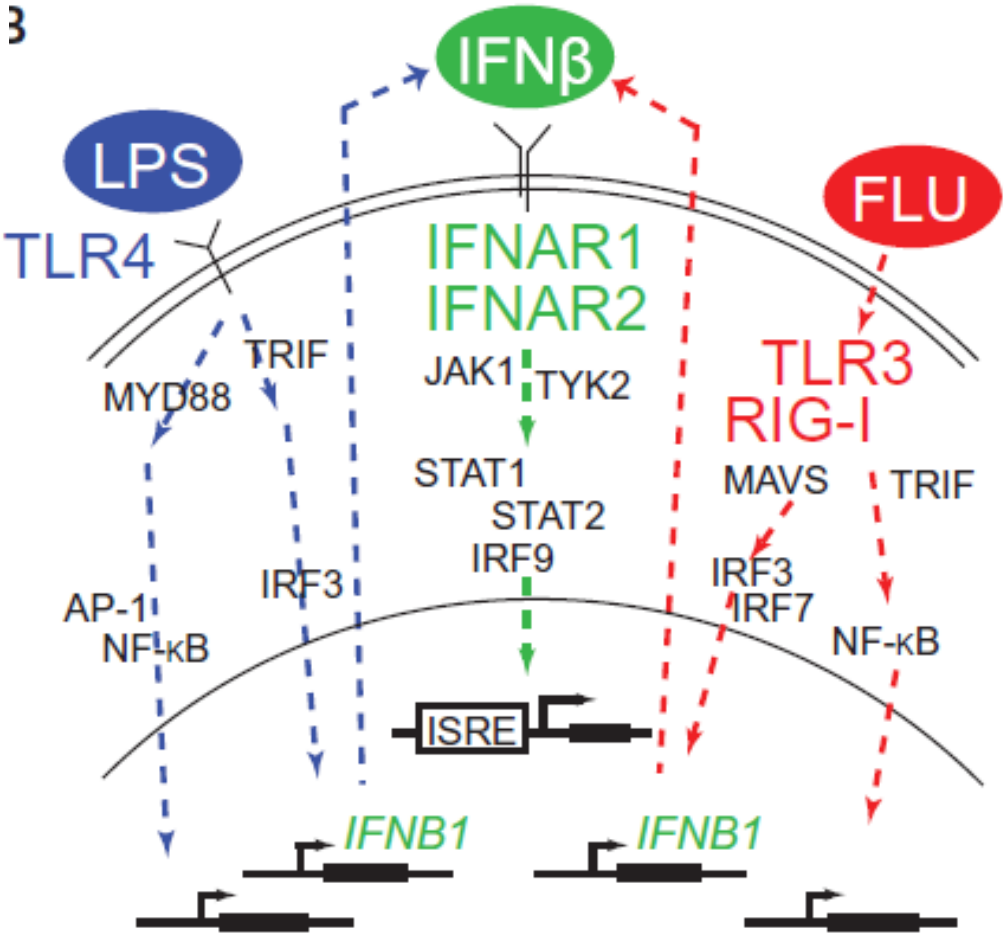
- α-CD3, α-CD28
- α-CD3, α-CD28, IFN-β
- α-CD3, α-CD28, TGF-β
  - Th 17

Study pipeline



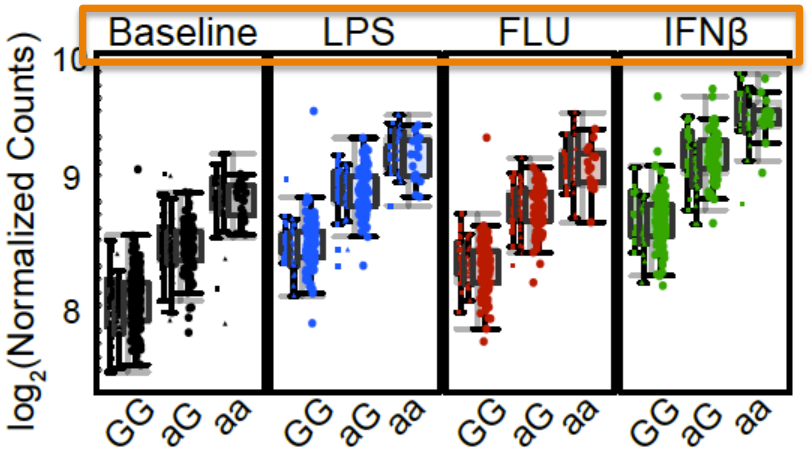
Immune pathways activated by stimuli

Receptor engagement activates signal transduction cascades that regulate expression of inflammatory genes, IFNs and IFN-stimulated genes

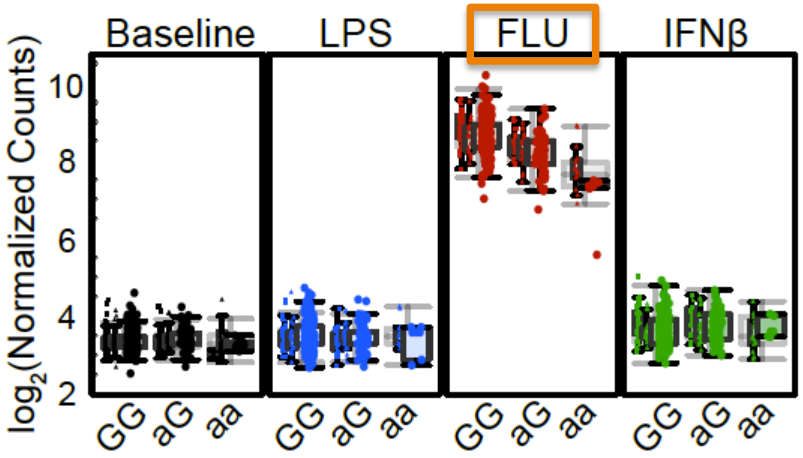


Different categories of *cis*-eQTLs inform mechanism

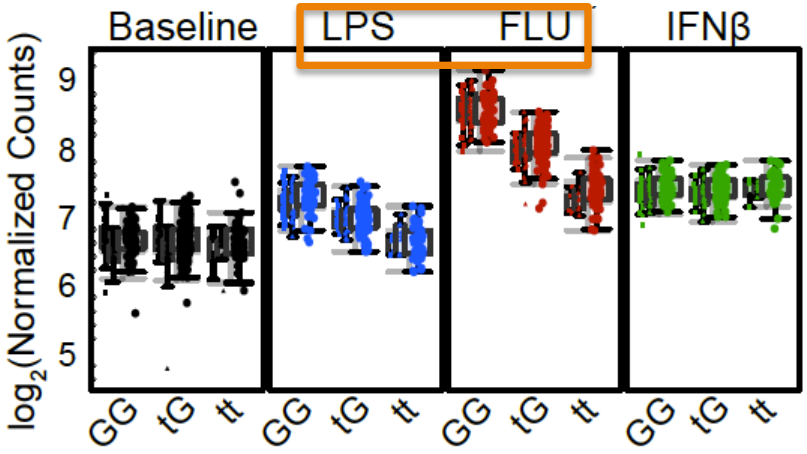
1- Common to ALL conditions (n≈67)



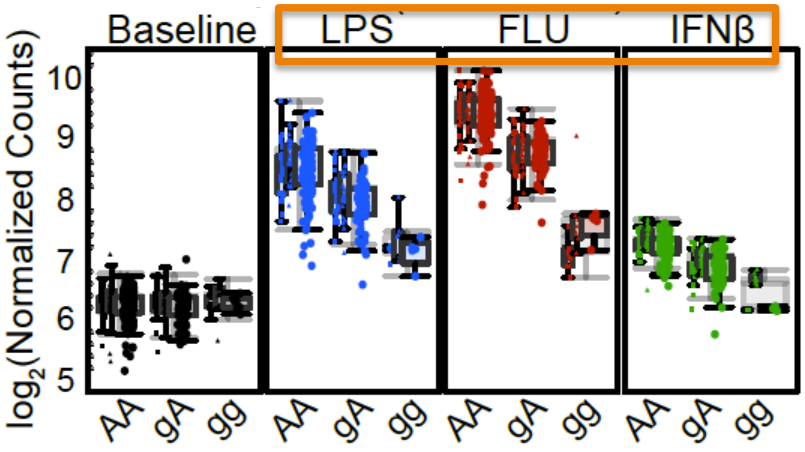
2- Specific to FLU stimulation (n≈11)



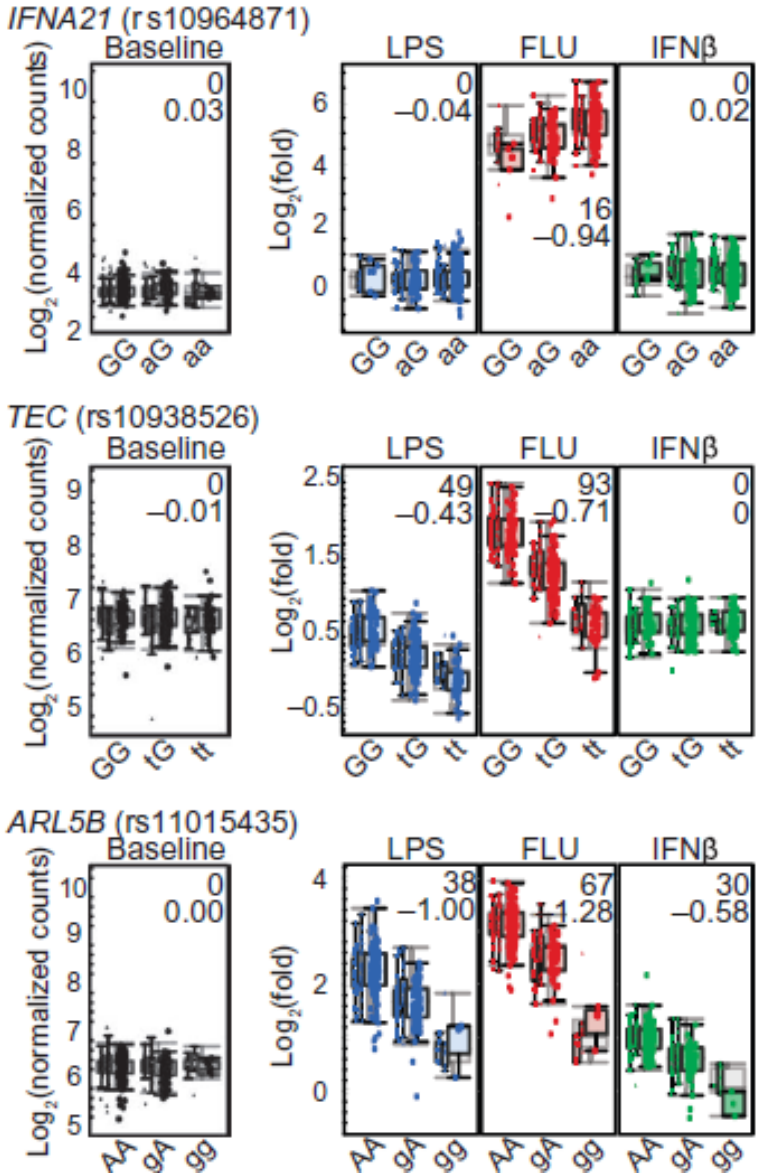
3- Specific to LPS/FLU stimulation (n≈21)



4- Specific to LPS/FLU/IFNβ stimulation (n≈40)



cis-response QTLs (*cis*-reQTLs): 121 genes

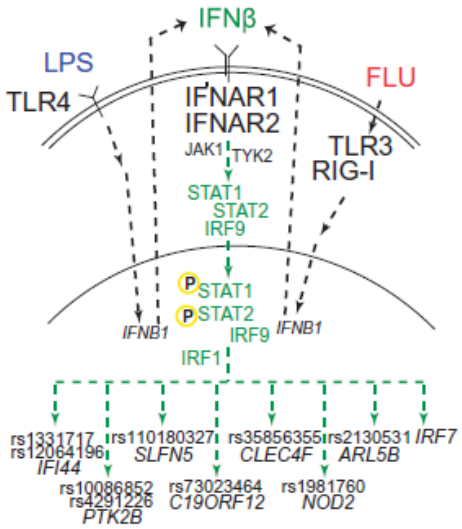


7: FLU only

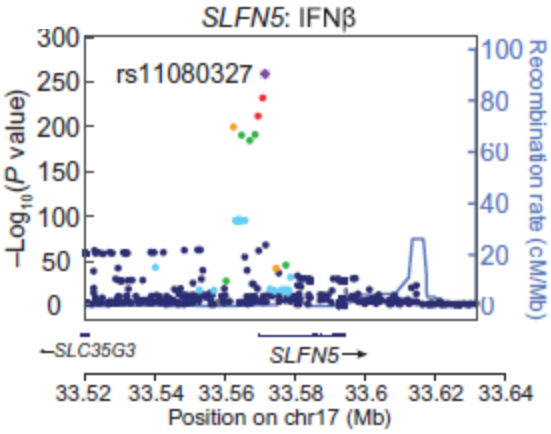
15: LPS + FLU

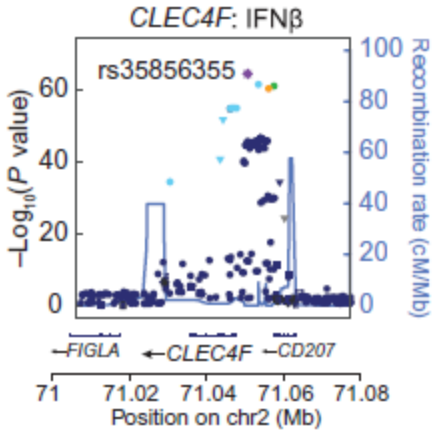
57: ALL

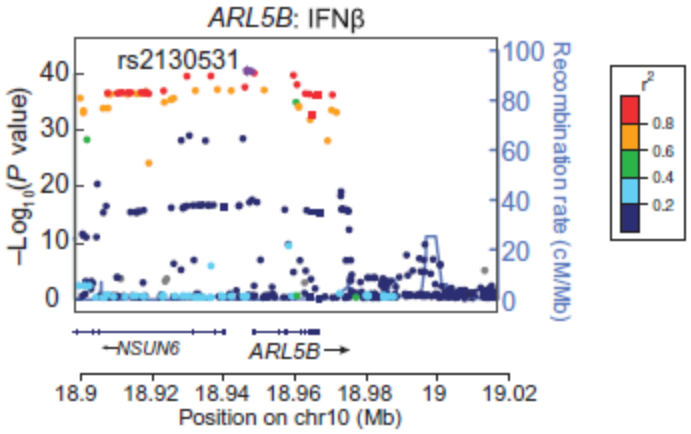
cis-reQTLs that alter sequence of TF binding sequences

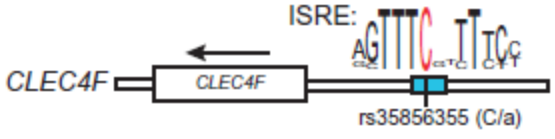


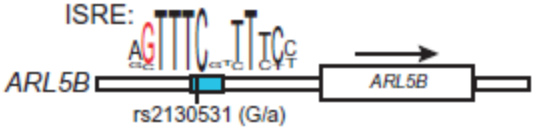
Enrichment of known binding sites for TFs from the STAT family (ENCODE ChIP-Seq)
STAT2: 116-fold, $p < 2.55 \times 10^{-21}$
STAT1: 126-fold, $p < 2.98 \times 10^{-13}$



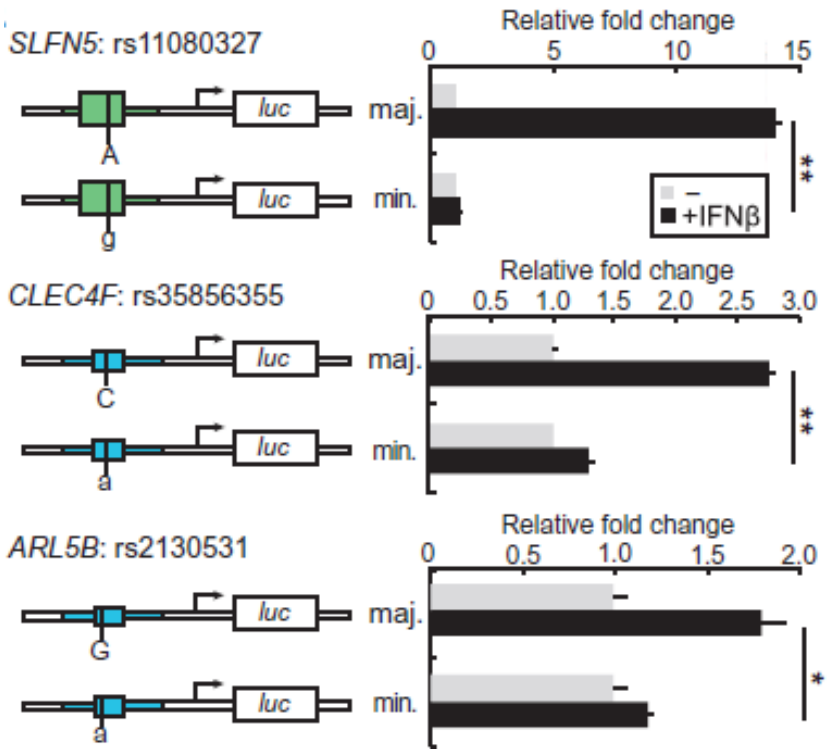




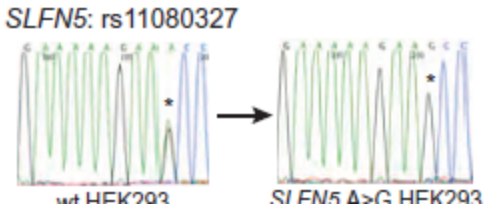




Validation



Luciferase assay in HEK 293
Stimulate with IFN-β

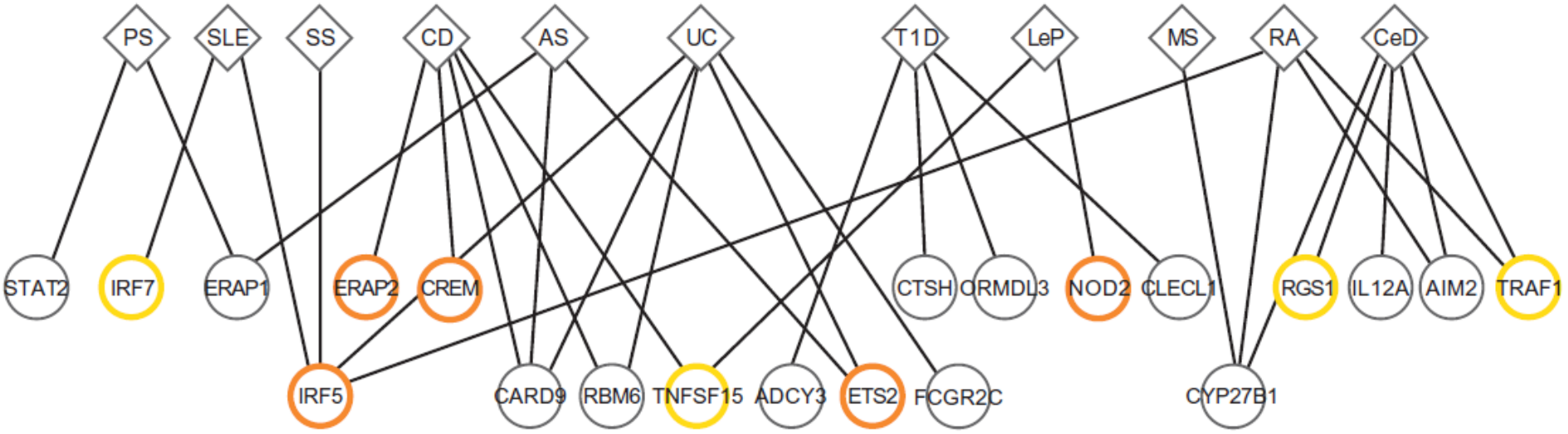


CRISPR alteration of het to homo

+ IFN-β

Fold induction *SLFN5* changed from 3.64 to 1.03

Autoimmune and Infectious disease SNPs from GWAS



PS = Psoriasis
SLE = Systemic lupus erythematosus
SS = Systemic sclerosis
CD = Crohn's disease
AS = Ankylosing spondylitis
UC = Ulcerative colitis

T1D = Type 1 diabetes
LeP = Leprosy
MS = Multiple sclerosis
RA = Rheumatoid arthritis
CeD = Celiac's disease

orange: *cis*-reQTLs, yellow: stimulus-specific *cis*-eQTLs

Summary

- Reference of genetic basis of transcriptome variation in innate and adaptive immune cells of a healthy multi-ethnic cohort
- Characterization of context specificity of eQTLs (population, cell-type, sex, activation state) with real implications for medical phenotypes, foremost in elucidating disease mechanisms.
- Population 'specific' signals are largely explained by allele frequency differences across populations, little effect size differences.
- Approximately 60% of *cis*-eQTLs are shared across adaptive and innate immune cell types, though effect sizes vary.
- Inflammatory disease alleles over-represented in T-cell regulatory effects, whereas neurodegenerative disease alleles are enriched in monocyte effects.
- Genetic effects on response to immune cell activation

Acknowledgements

Christophe Benoist
Katherine Rothamel
Ting Feng
Michael Wilson
Natasha Asinovski
Scott Davis

Diane Mathis

Phillip De Jager
Michelle Lee
Cristin McCabe
Selina Imboywa
Avery Davis
Manik Kuchroo
Alina Von Korff
Nikos Patsopoulos
Irene Frohlich

Nir Hacohen and Aviv Regev
Chun (Jimmie) Ye
Mark Lee
Chloe Villani
Weibo Li

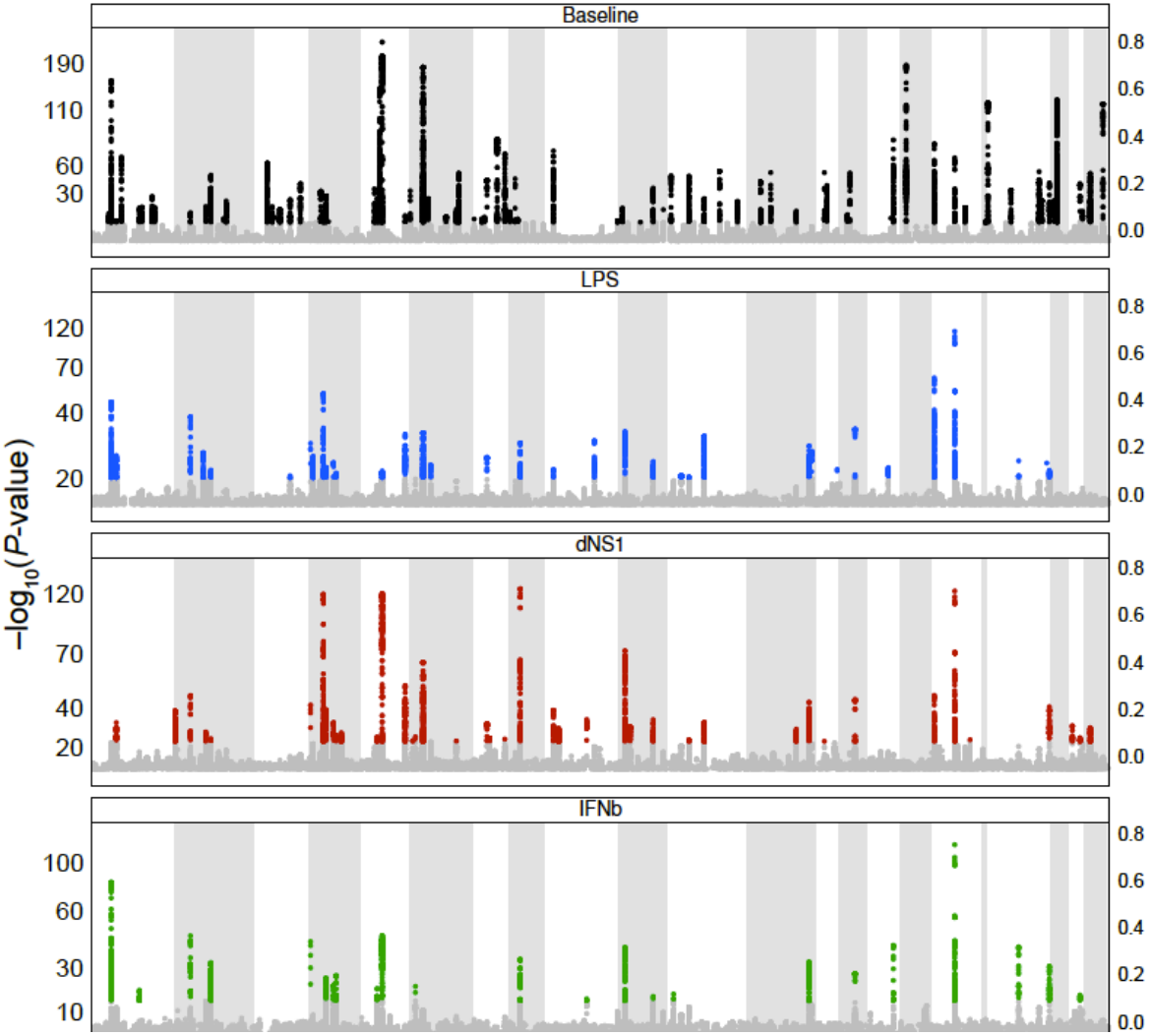
Daphne Koller
Sara Mostafavi

David Haffler
Soumya Raychaudhuri

Barbara Stranger
Towfique Raj*
Joseph Replogle
Eric Gamazon
Patrick Evans
Meritxell Oliva
Charles Czysz
Daisy Castillo
Katya Khramtsova
Nancy Cox
Dan Nicolae

Stranger lab is recruiting postdocs!!

Condition specific *cis*-eQTLs in DCs



Baseline Results

LPS Results

Flu Results

IFNb Results

264/415 genes have cis-eQTL in at least one condition

Lee et al. *Science* 2014