Functional Genomics and Integrative Analysis

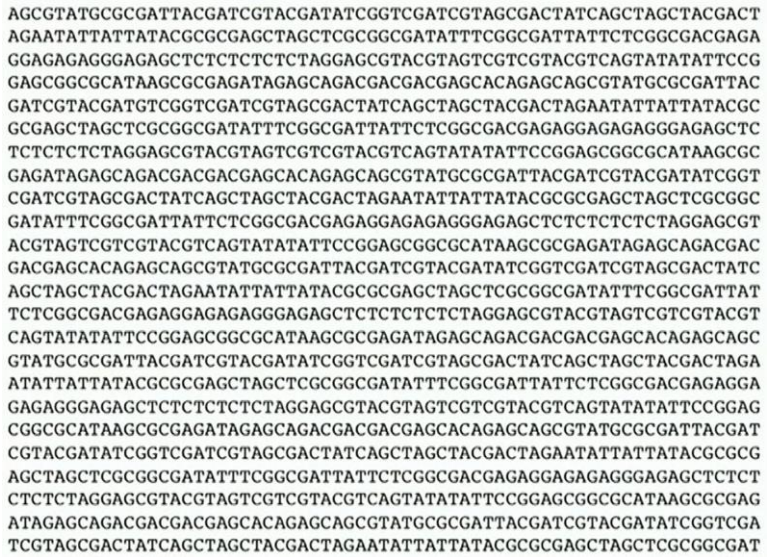
Charlie Fulco
Target, Disease, and Systems Biology, Sanofi, Cambridge MA, USA

FOCIS June 2026

Charles Fulco is a Sanofi employee and may hold shares and/or stock options in the company.

What is Functional Genomics?

Assigning a causal role to genes, epigenetics, or genetic variation at scale



Investigating biology with Functional Genomics

Complementary approaches to build the case

Variant to Function

Surveilling biology in the wild

Anchor on the causal biology of disease from human genetics

Genetic Screening

Focused in vitro interrogation

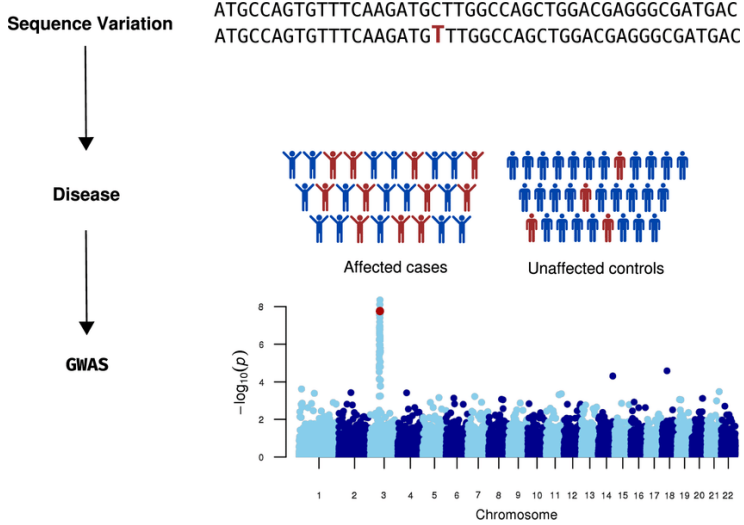
Mechanistic clarity with CRISPR-based genetic screens

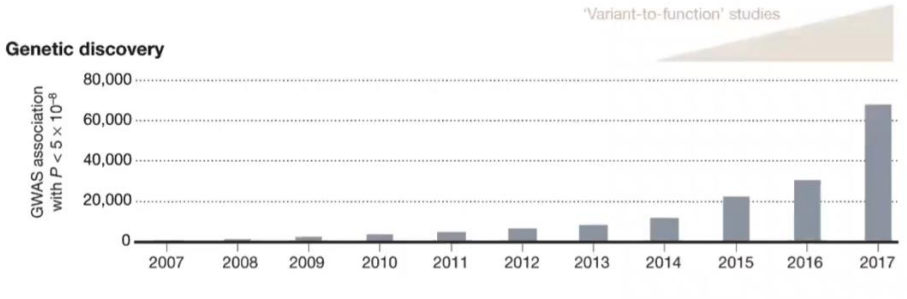
Analogy concept: Sean M Carrol

Part 1: Variant to Function

Human genetics discovers causal associations. Mechanism has lagged behind

Human genetics has delivered hundreds of thousands of associations between genotype and disease, each a clue into casual mechanisms of disease





Palsson, R., Indridason, O.S., Edvardsson, V.O. *et al. Urolithiasis.* (2019). Claussnitzer *et al.* Nature (2020)

Human genetics : Variant to Gene to Function

How do genetic variations in the human genome effect genes to effect cellular functions to effect traits

Human genome variations (V)

Gene (G)

Function (F)

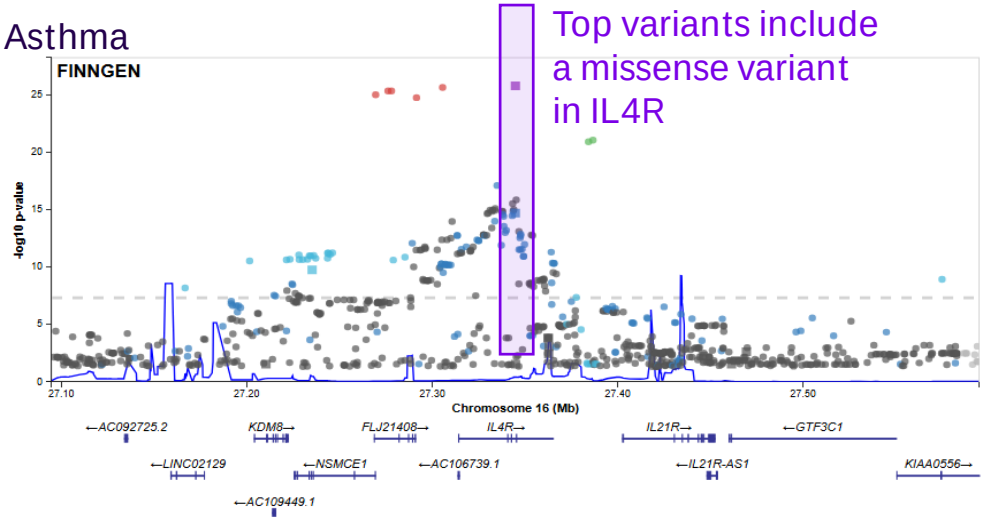
Human traits

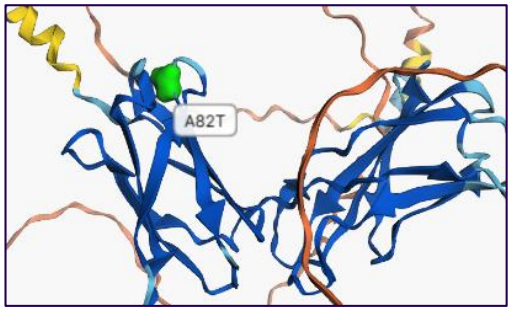
Credit: S. Lessard, C. Chatelain (Sanofi)

Variant to gene can be straightforward for protein-coding variants

About 10% of GWAS variants effect protein sequence

Asthma





Example in IL4R:

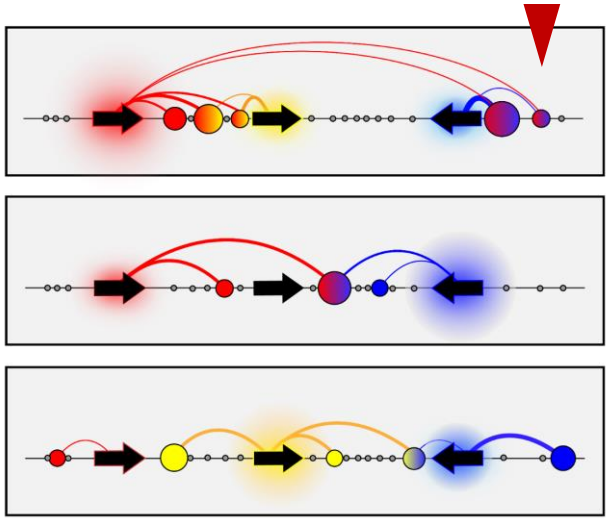
- Variant is protective against type 2 inflammation (allergic asthma, allergic rhinitis)
- Increases risk for type 1 diseases (psoriasis, hypothyroidism, IBD)

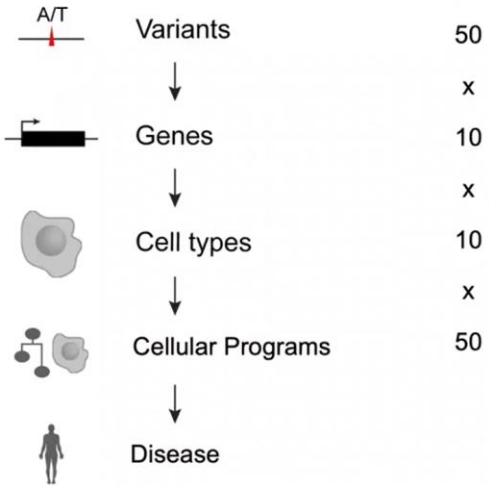
Credit: S. Lessard (Sanofi)

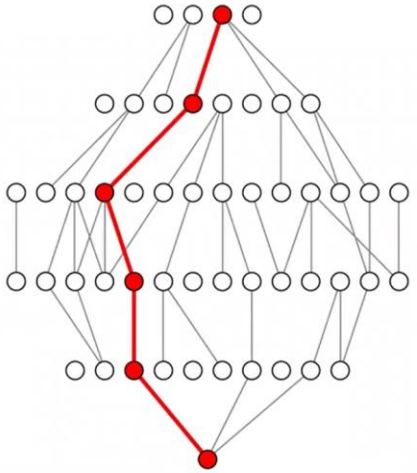
Identifying target genes of non-coding variants is challenging

Non-coding variants effect target genes through cell-type specific gene regulatory elements

Many GWAS SNPs effect non-coding regulatory elements rather than genes directly

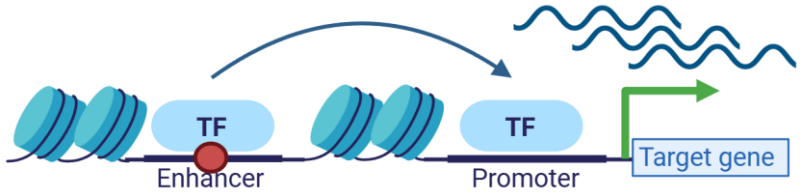




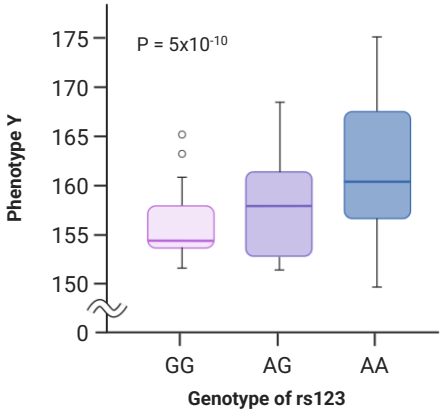


V2F Figure adapted from M. Claussnitzer, J. Engreitz (Broad)

Identifying the causal gene in non-coding GWAS loci



eQTLs
correlate genetic and gene expression changes



"E2G" Predictive Models
trained/tested on CRISPR, eQTL, 'omics data predict regulatory interactions

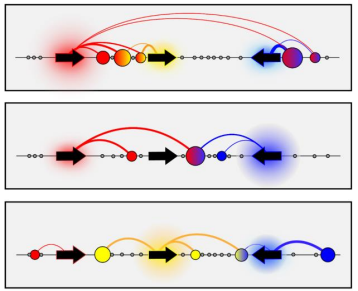
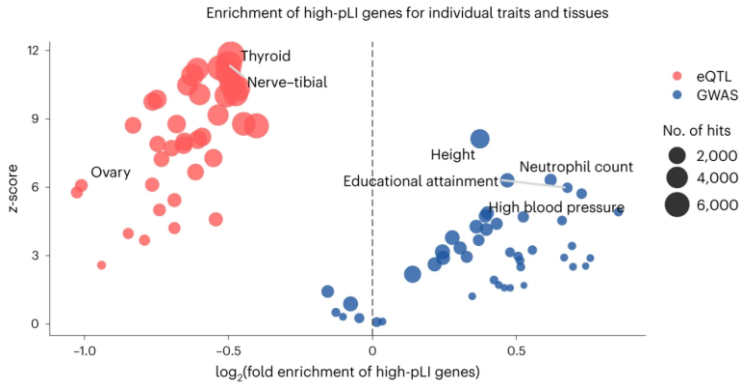


Image Credit: S. Lessard, C. Chatelain (Sanofi)

Leveraging eQTLs is challenging in practice

eQTL : GWAS colocalization generally doesn't identify causal genes, and is less informative than proximity

eQTLs involve different genes than implicated by GWAS



Selective Constraint

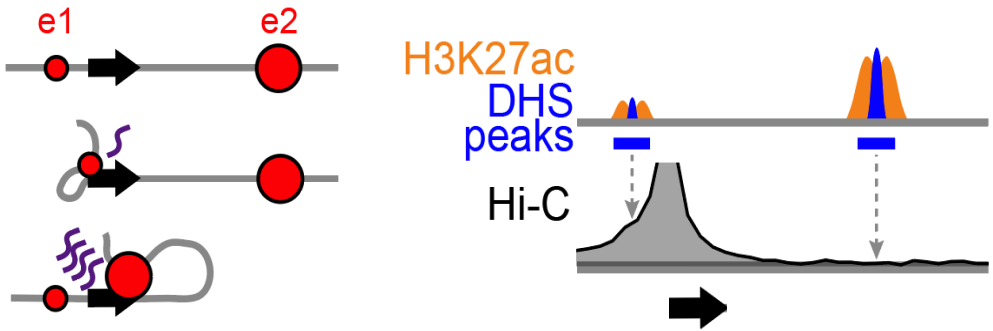
- eQTLs are depleted at constrained genes; GWAS enriched at constrained genes (Mostafavi et al (2023))
- eQTLs are enriched at unconstrained genes; GWAS enriched at constrained genes (Lek, Karczewski et al (2016))

eQTL signals are highly cell-type/state specific → large scale single-cell eQTL maps may be needed to identify GWAS effects

Predictive models are emerging that identify enhancer targets genes

Activity by Contact and its improvements predict enhancer targets with low data needs

Activity by Contact Model (ABC)



Predicted % effect of enhancer on gene =

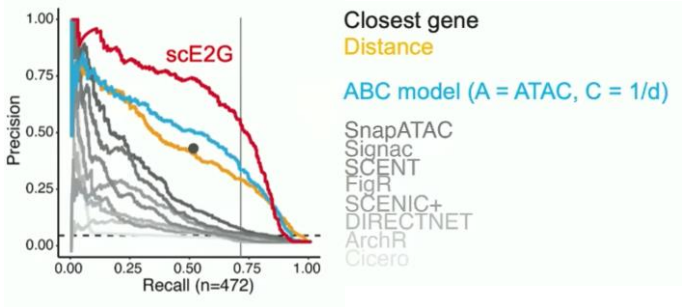
$$\frac{\text{Activity} \times \text{Contact}}{\Sigma\ (\text{Activity} \times \text{Contact})}$$

Fulco, Nasser, *et al*. Nature Genetics (2019)

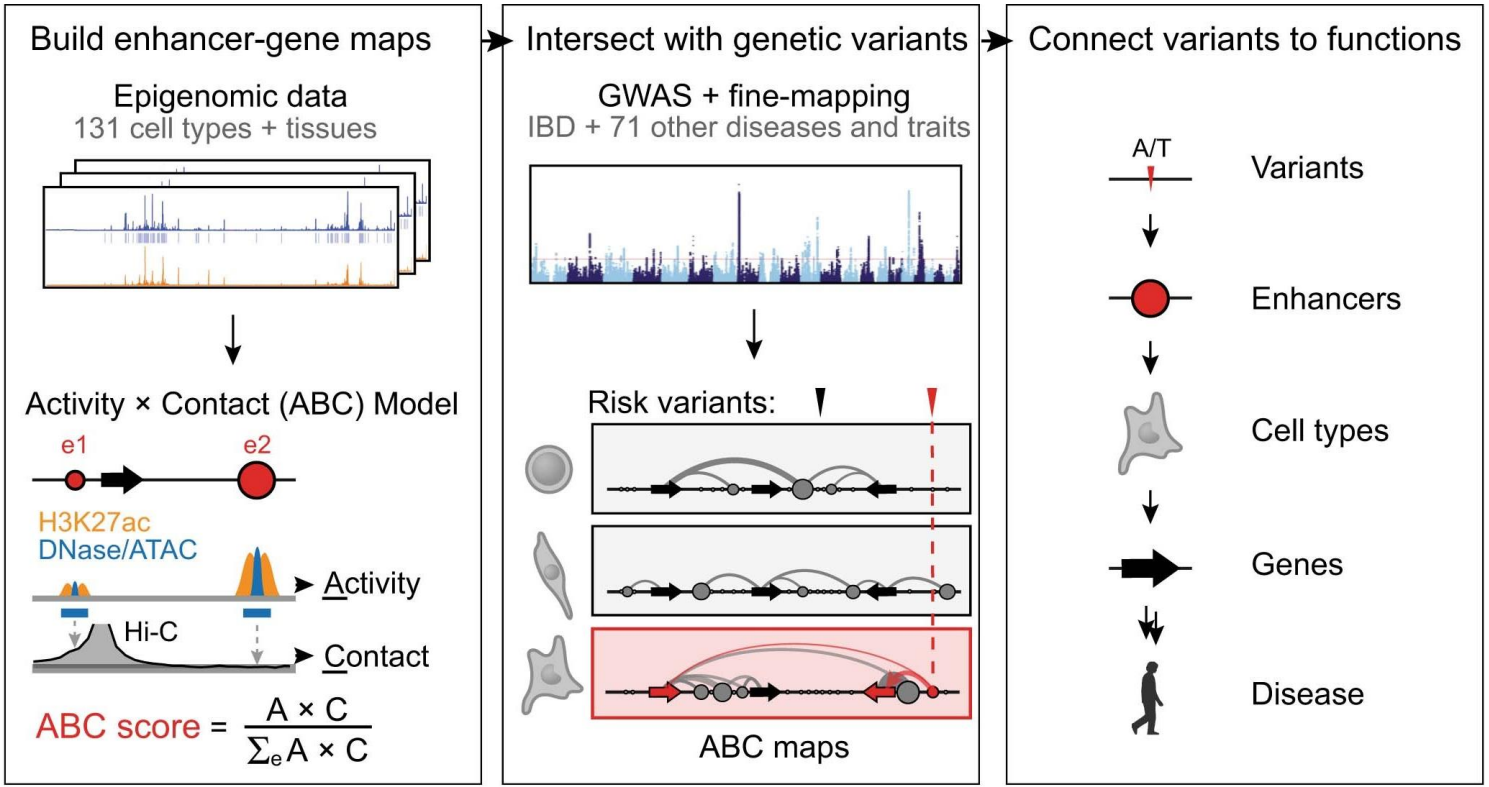
Implementations

ABC + additional features

- **ENCODE-rE2G** (Gschwind et al 2023 bioRxiv; available on Open Targets Platform)
- **scE2G** (Sheth, Qiu et al 2024 bioRxiv; available at e2g.stanford.edu). Only requires scATAC

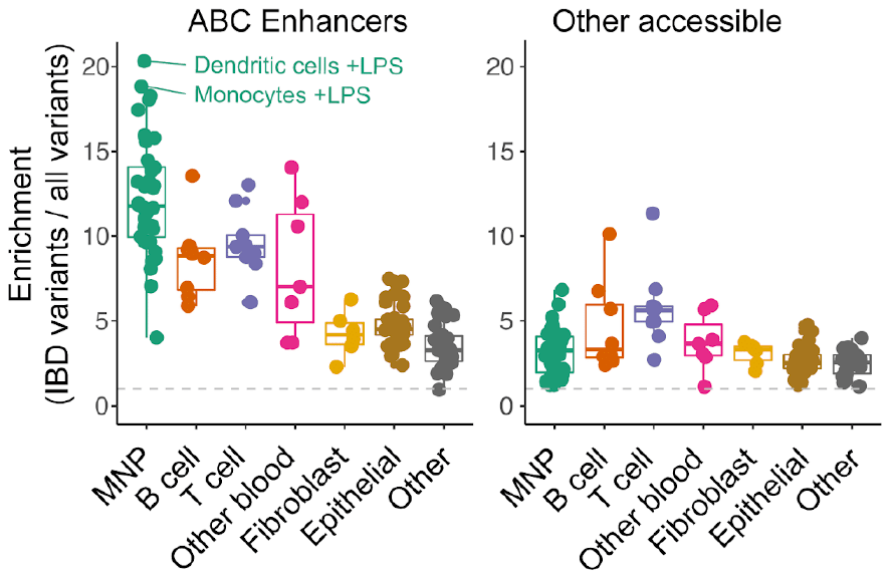


Example of application at scale: Nasser 2021

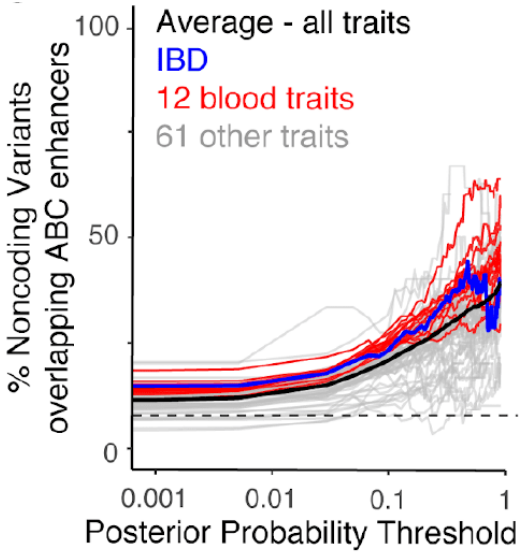


Nasser, Bergman, Fulco, Guckelberger, and Doughty *et al.* Nature (2021)

GWAS variants are strongly enriched in ABC enhancers



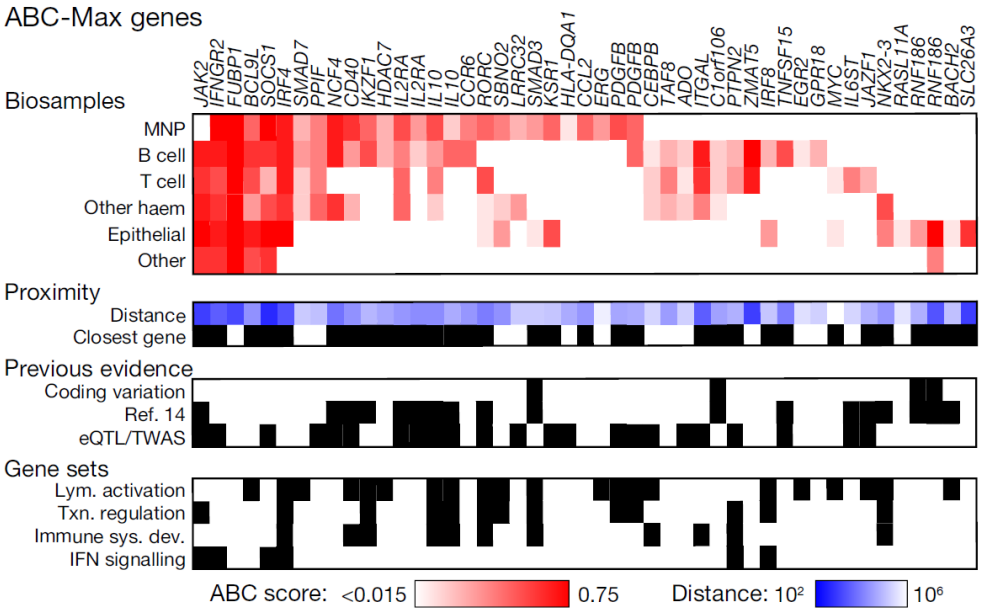
ABC enriches for IBD variants in relevant cell types

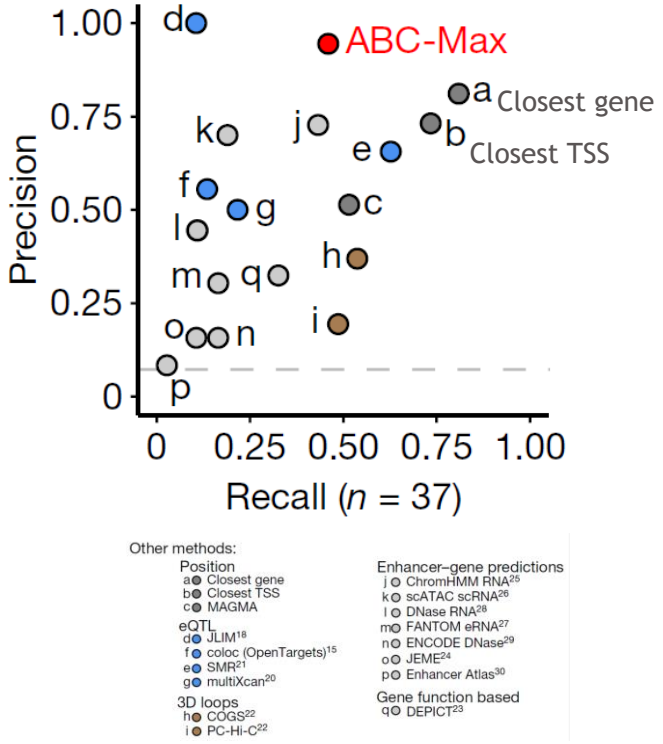


ABC enriches for likely causal variants across traits

Nasser, Bergman, Fulco, Guckelberger, and Doughty *et al.* Nature (2021)

ABC nominates known and novel IBD risk genes

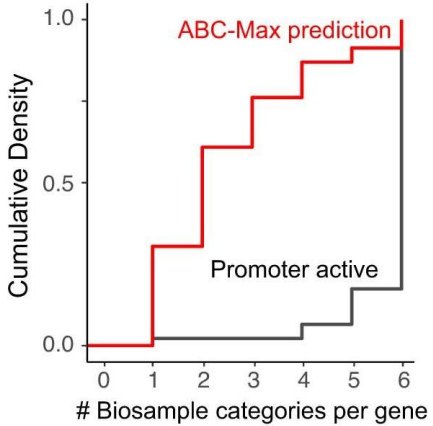




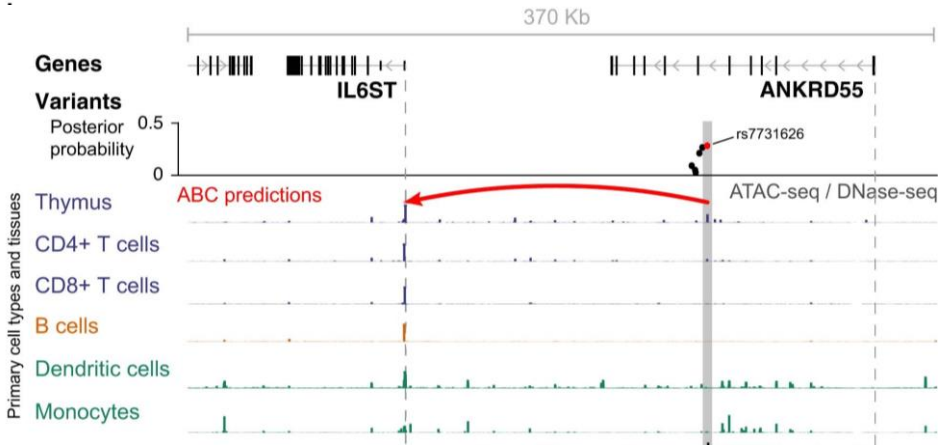
Nasser, Bergman, Fulco, Guckelberger, and Doughty *et al.* Nature (2021)

ABC zeros in on cell types

ABC-max predictions are more cell type specific than is gene expression



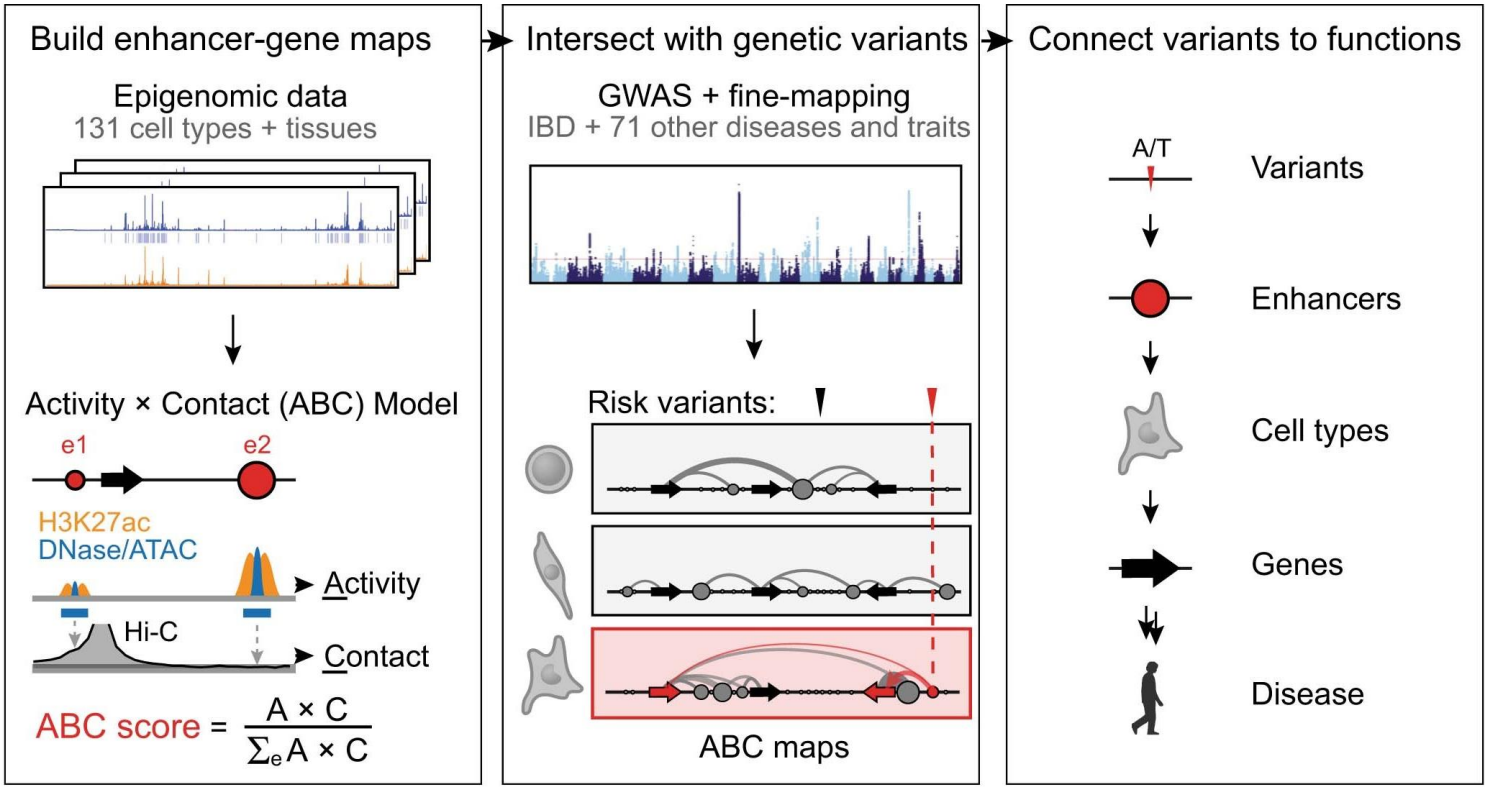
Promoters are active (gene expressed) in more lineages than are ABC predicted enhancers



IBD risk variant overlaps ABC enhancer predicted to regulate IL6ST only in Thymus, even though IL6ST is expressed in most cell types.

Nasser, Bergman, Fulco, Guckelberger, and Doughty *et al.* Nature (2021)

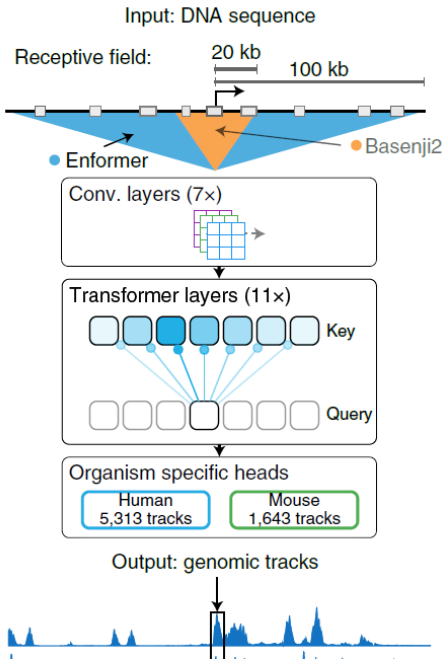
Example of application at scale: Nasser 2021



Nasser, Bergman, Fulco, Guckelberger, and Doughty *et al.* Nature (2021)

Variant effect prediction can inform on direction of effect

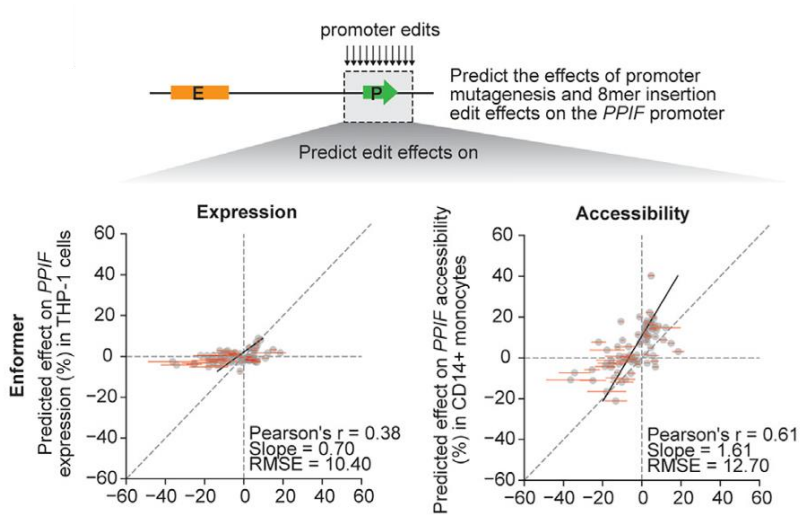
AI models are emerging that predict epigenetic or transcription data from gene sequence



Evaluate Performance: Promoter Variants

1. Edit promoter sequence in vitro
2. Compare RNA expression or DNA accessibility to AI prediction

→ Decent Performance



Avsec et al (2021). Nat Methods; Martyn, Montgomery et al. (2025). Cell

Variant effect prediction can inform on direction of effect

In practice, best performance observed when predicting local effects and using ABC to propagate to genes

Evaluate Performance: Enhancer Variants

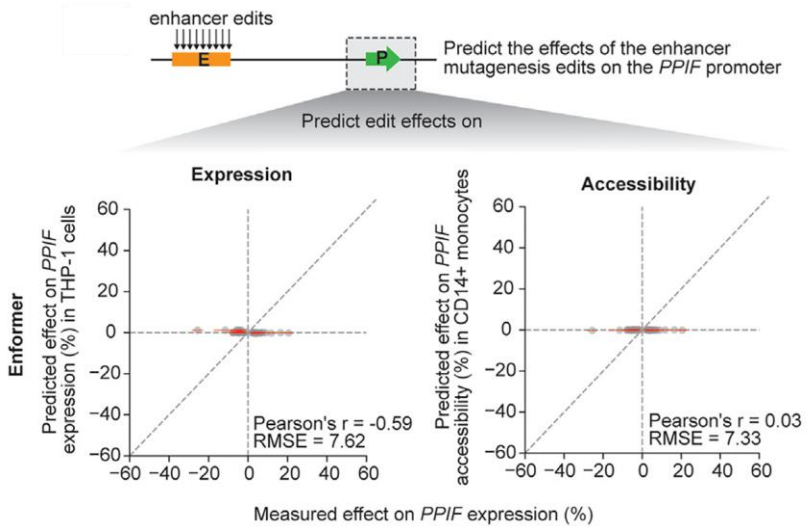
1. Edit enhancer sequence in vitro
2. Compare RNA expression or DNA accessibility to AI prediction

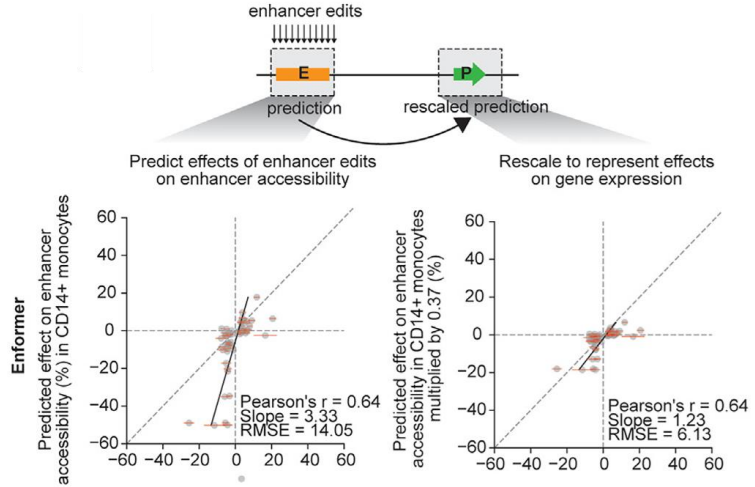
→ Poor Performance

Evaluate Updated Predictor: Enhancer Variants

1. Edit enhancer sequence in vitro
2. Predict accessibility at the enhancer
3. Use ABC to propagate prediction to RNA levels

→ Decent performance





Martyn, Montgomery et al. (2025). Cell

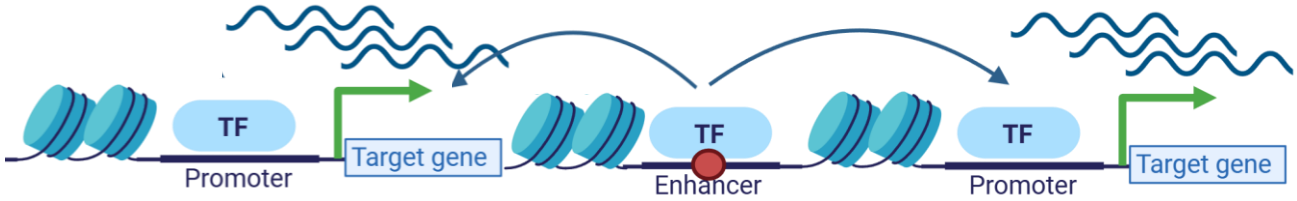
Part 1. Variant to Function Conclusions and Cautions

- Interpreting the mechanism driving GWAS associations is a challenge lagging behind association discovery
- Many GWAS associations act through regulatory mechanisms that are difficult to interpret directly
- eQTL approaches struggle to identify causal mechanisms in GWAS
- Predictive models based on ABC guide variant-to-function in non-coding GWAS

Part 1. Variant to Function Conclusions and Cautions

- Interpreting the mechanism driving GWAS associations is a challenge lagging behind association discovery
- Many GWAS associations act through regulatory mechanisms that are difficult to interpret directly
- eQTL approaches struggle to identify causal mechanisms in GWAS
- Predictive models based on ABC guide variant-to-function in non-coding GWAS

- Word of caution: many non-coding GWAS variants likely effect the expression of multiple genes! Identifying the causal driver requires additional functional investigation



Part 2: Genetic Screens

Genetic Screens Reveal Function at Scale

What does each gene / set of genes / regulatory element do?

Break them and see what happens!

The art and science of ‘break it and see what happens’

Cell Model

- Immortlizaed cell lines
- Primary cells
- iPSC-derived
- Organoids
- In vivo

Balance of tractability and translatability

Perturbation

Scale:
- Genomewide
- Focused Library
- Combinatorial

Modality:
- Knockout (Cas9)
- Knockdown (CRISPRi)
- Overexpression (CRISPRa)

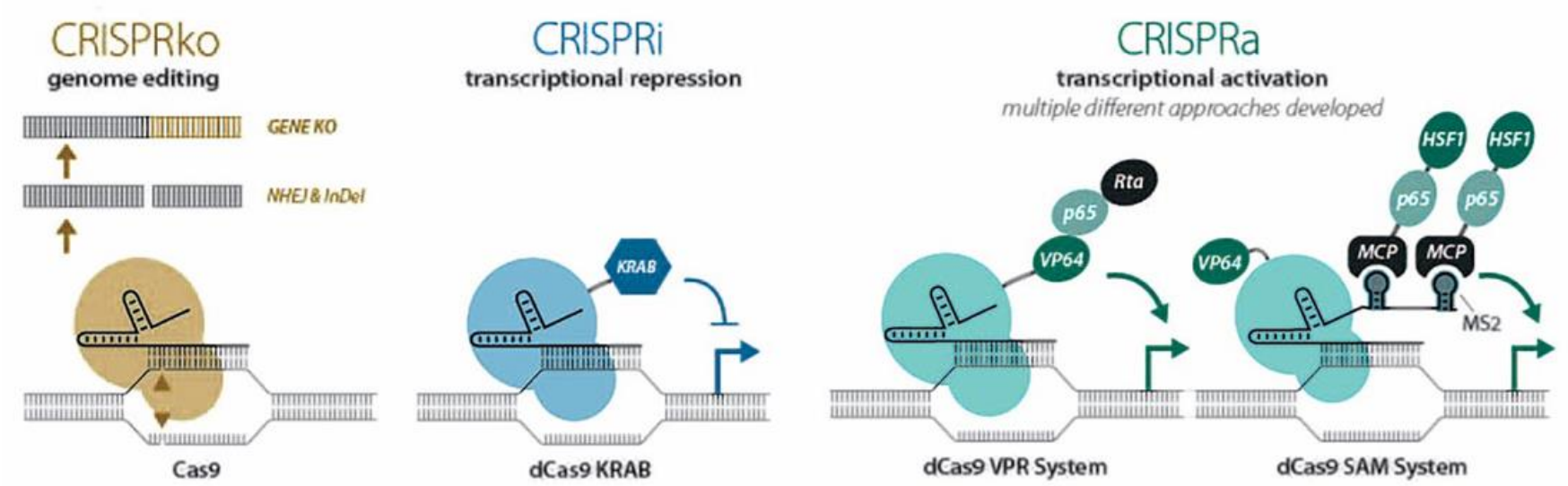
Format:
- Arrayed
- Pooled

Adapted to Cell Model and Assay at hand

Assay

- Proliferation / Survival
- FACS / Flow Cytometry
- Imaging
- Transcriptomics
- Supernatant / Cytokine assays

Cas9-based tools enable genetic screening



Sage, Lawo, Cross B. SLAS Discovery (2020)

Key distinction: Arrayed vs Pooled

The main practical

Arrayed

- Cells for each perturbation are grown separately
- Flexible with respect to cell model, assay
- Scale requires robotics
- Well-to-well variability can be problematic

Pooled

- Cells for all perturbations are grown together
- Highly scalable
- Only assays cell autonomous effects
- Typically use viral integration to deliver perturbations

Genomescale Primary T-cell Perturb-seq

Cell Model

Primary Human CD4+ T-cells

Perturbation

Genome-scale, Pooled, CRISPRi

Assay

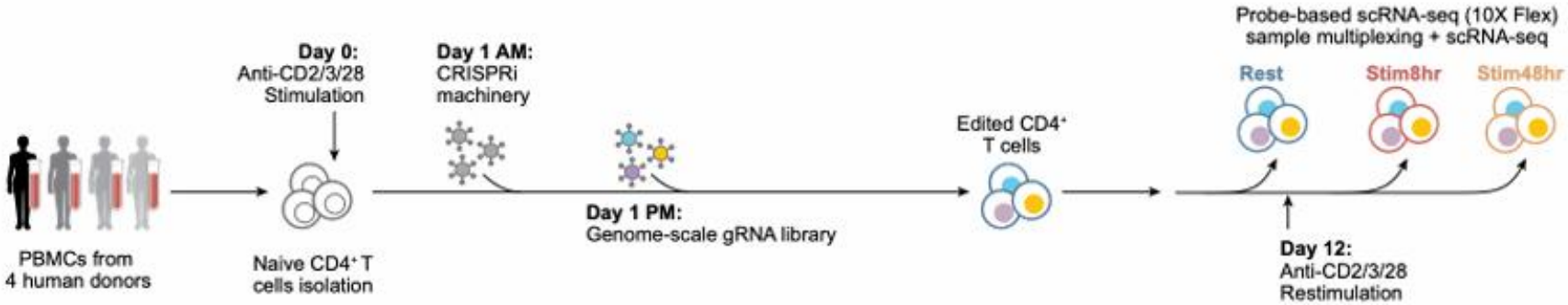
scRNA-seq (“Perturb-seq”)

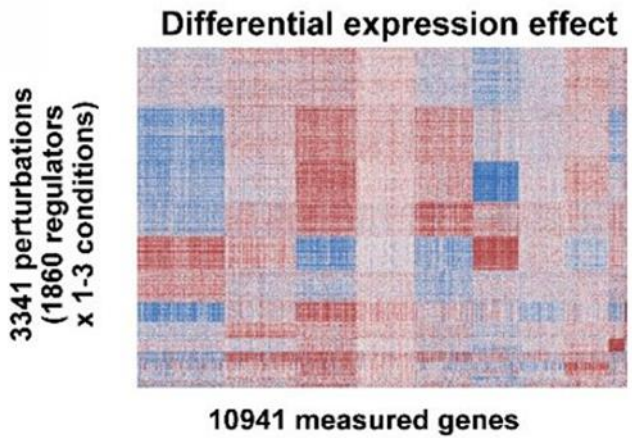
bioRχiv

Genome-scale perturb-seq in primary human CD4+ T cells maps context-specific regulators of T cell programs and human immune traits

Ronghui Zhu[1,2,*,†], Emma Dann[1,2,*,†], Jun Yan[1], Justine Reyes Retana[1], Ryunosuke Goto[3], Reese C. Guitche[1,4], Lillian K. Petersen[2], Mineto Ota[1,2,5], Jonathan K. Pritchard[2,6,†], Alexander Marson[1,7,8,9,10,11,12,†]

Genomescale Primary T-cell Perturb-seq

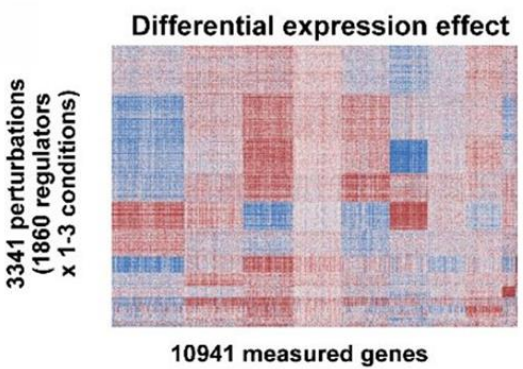




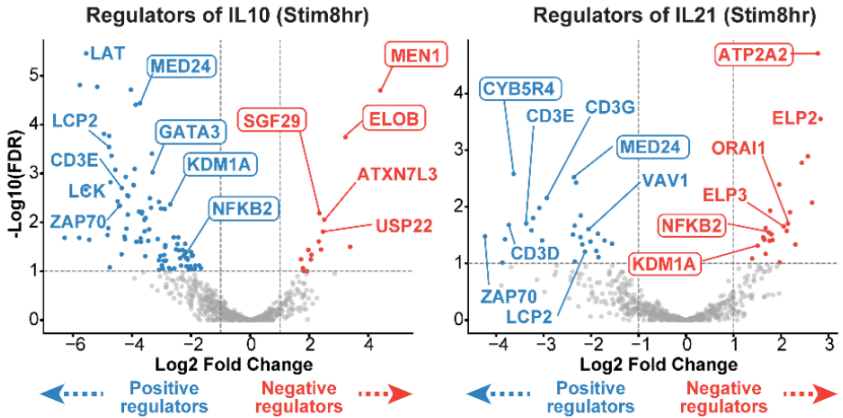
Zhu, Dann et al. BioRxiv (2025)

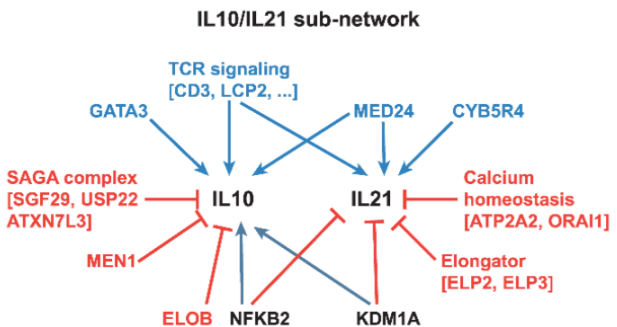
Perturb-seq maps gene regulatory networks

Example of digging into a few selected genes and perturbations



Inspect differential expression for selected genes

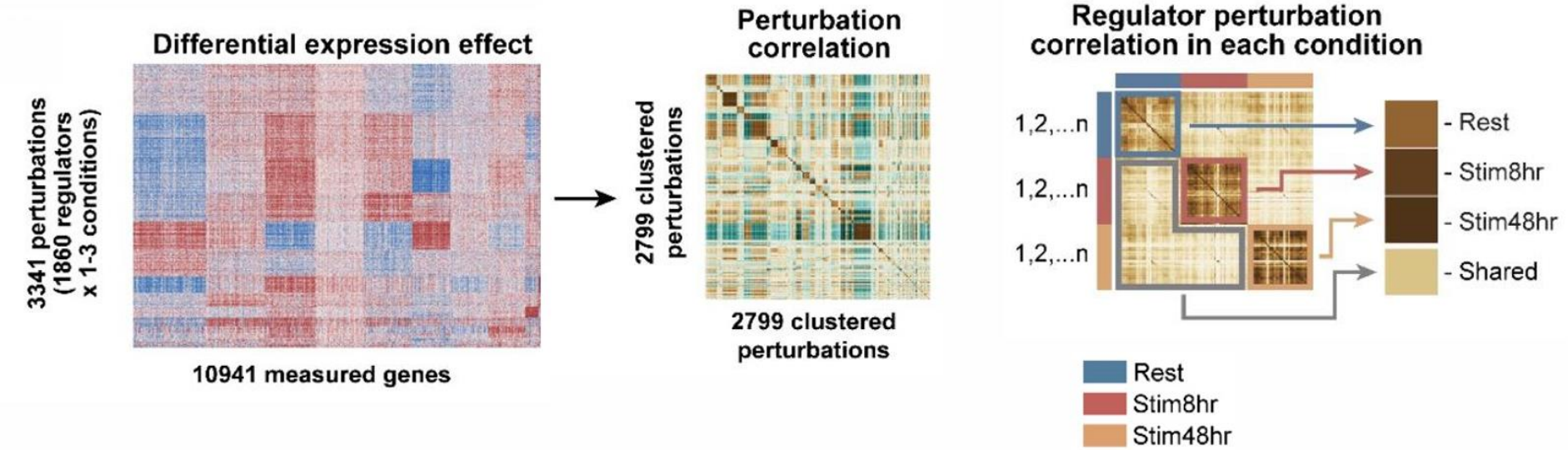




Zhu, Dann et al. BioRxiv (2025)

Which perturbations tend to illicit similar responses?

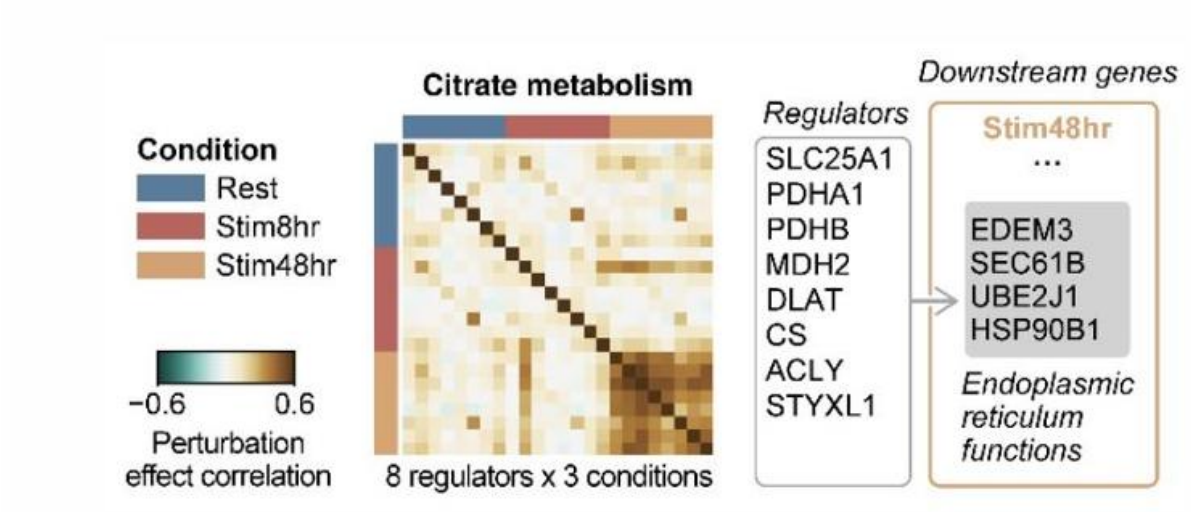
Observation: responses can vary a lot between stimulation conditions



Zhu, Dann et al. BioRxiv (2025)

Some perturbations have effects only in specific conditions

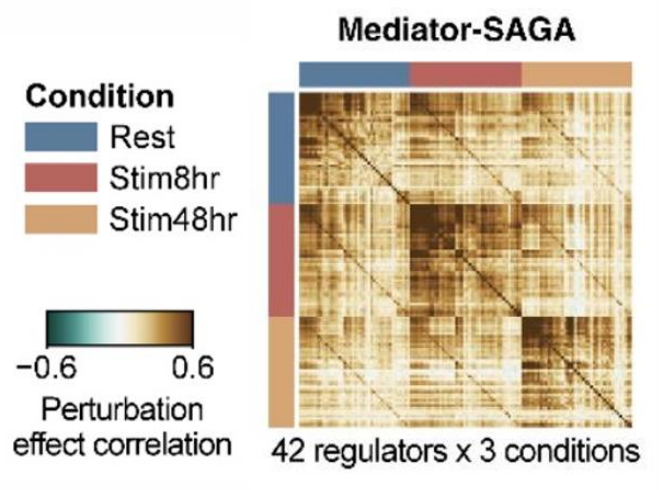
Example: Citrate metabolism is highly impactful late in T-cell activation

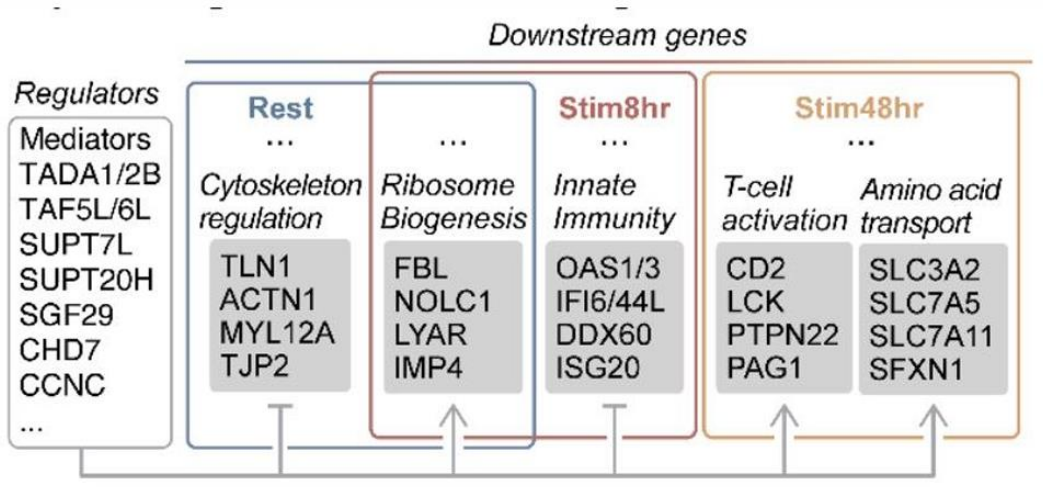


Zhu, Dann et al. BioRxiv (2025)

Some perturbations effect different functions in different conditions

Example: Mediator/SAGA complex targets shift over time through activation





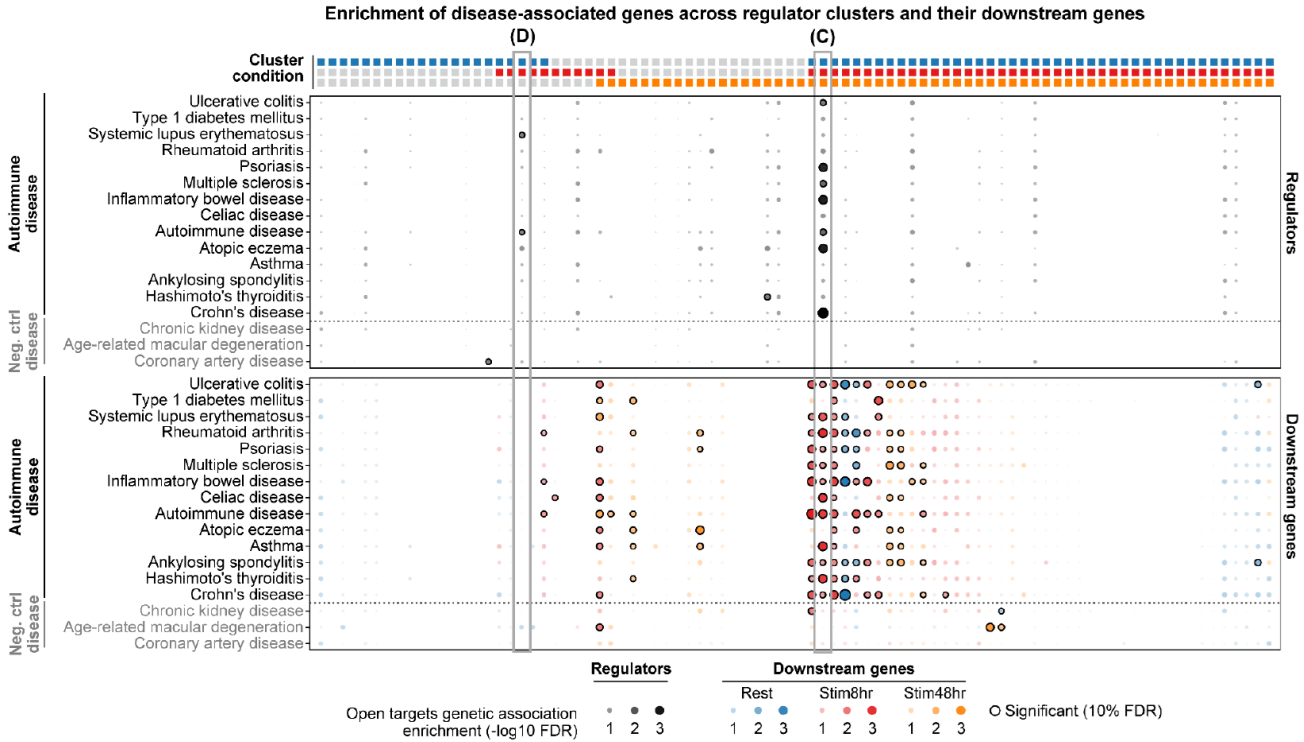
Zhu, Dann et al. BioRxiv (2025)

Interpreting GWAS through the Perturb-seq lens

Provides mechanistic insight and clarifies causal genes

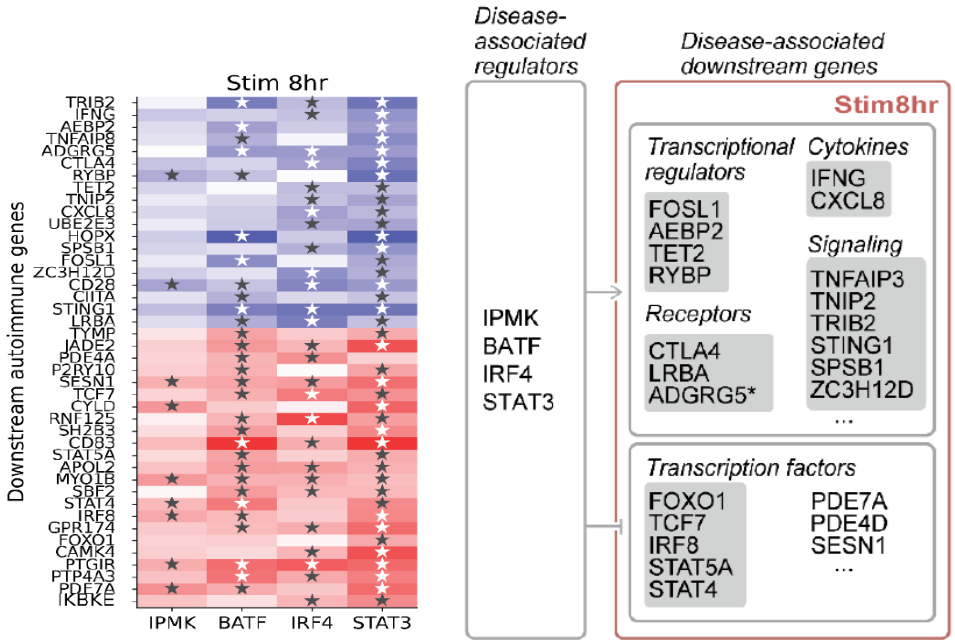
Regulators

Downstream genes



Zhu, Dann et al. BioRxiv (2025)

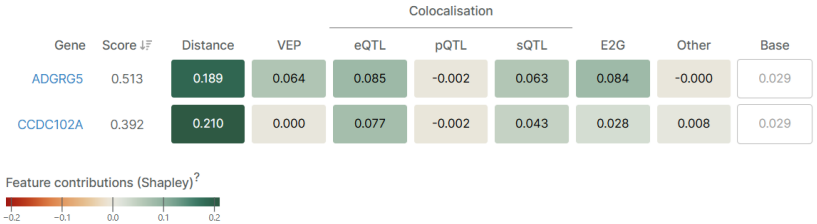
Co-enrichment of GWAS genes in Perturb-seq clusters may reveal causal mechanisms and help identify causal genes



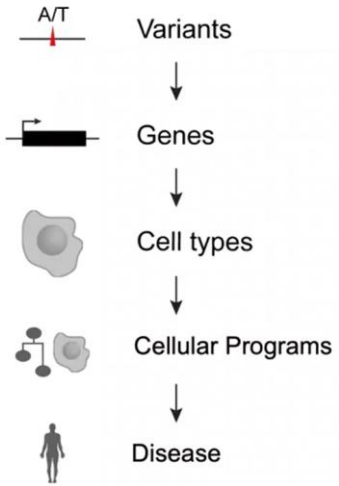
Perturb-seq suggests the hypothesis that ADGRG5 effects autoimmune disease downstream in T-cells downstream of STAT3

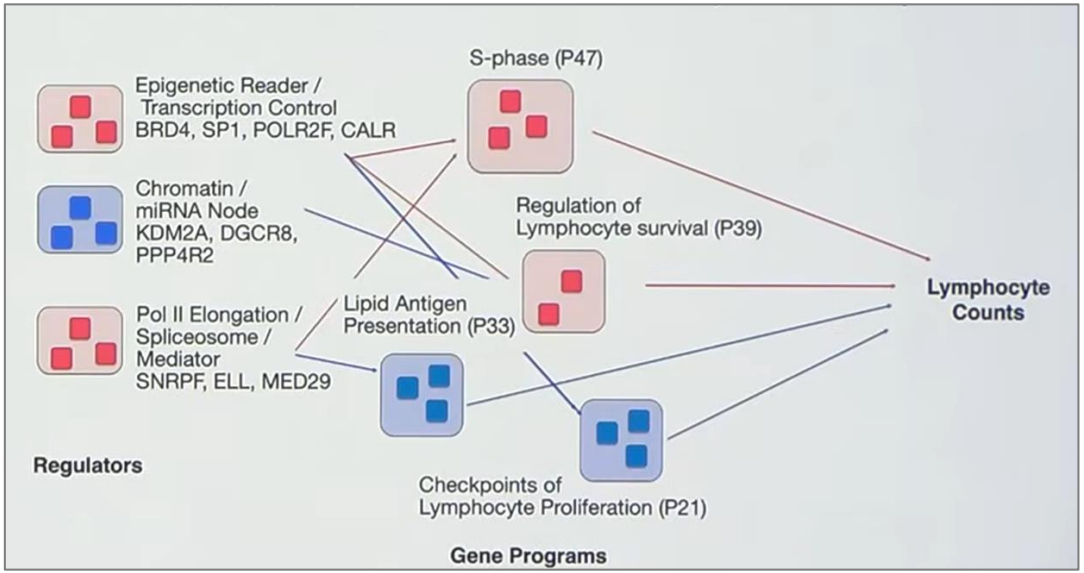
Regulatory genomics alone could not confidently prioritize ADGERG5 as the causal gene in its locus (Open Targets L2G)

Perturb-seq suggests 'guilt by association'



Vision: construct gene regulatory maps bridging regulators to gene programs to human traits





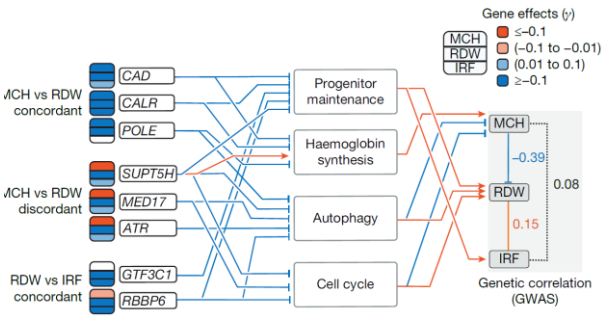
Prichard, Dann Lopez, Biology of Genomes, CSHL (2026)

Large scale Perturb-seq is scaling up functional genomics

Data are limited at the moment, but anticipate much more in next few years

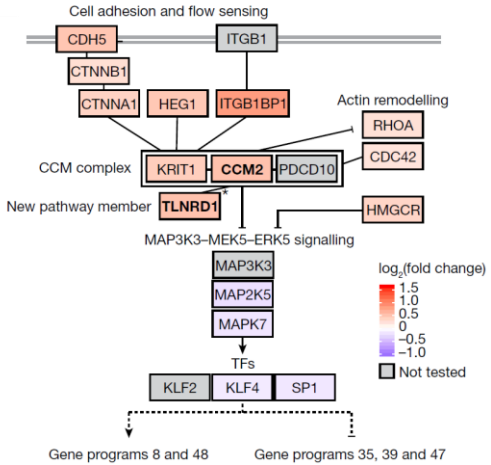
Maps of regulatory networks controlling erythroid traits

Genomewide Perturb-seq in K562s. Replogle, Saunders et al. Cell (2022); Ota et al. Nature (2025)



Prioritizing Coronary Artery Disease genes

Large-scale Perturb-seq in endothelial cells. Schnitzler, Kang et al. Nature (2024)

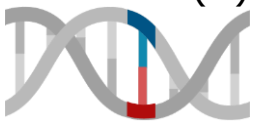


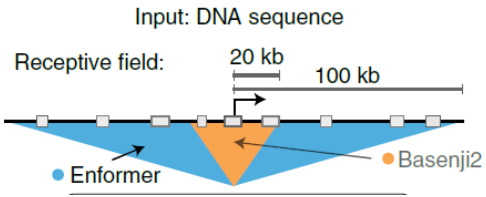
Many more large-scale Perturb-seq datasets to come!

Other existing data in iPSCs, ESCs, HCT116, HepG2

Scaling up further with AI (?)

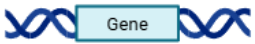
Human genome variations (V)

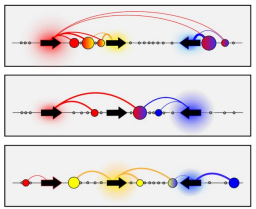




Variant Effect Prediction

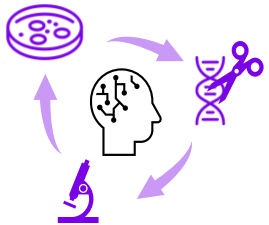
Gene (G)





Enhancer Target Prediction

Function (F)



Gene Function Prediction

Hope that AI models will learn from large scale perturbation (and other) data to predict perturbation responses:

- *Effect of genes in other cell types/states*
- *Combinations of genes*

Evaluating gene function prediction with AI

Models leveraging perturbation data in training are starting to show promise

bioRχiv

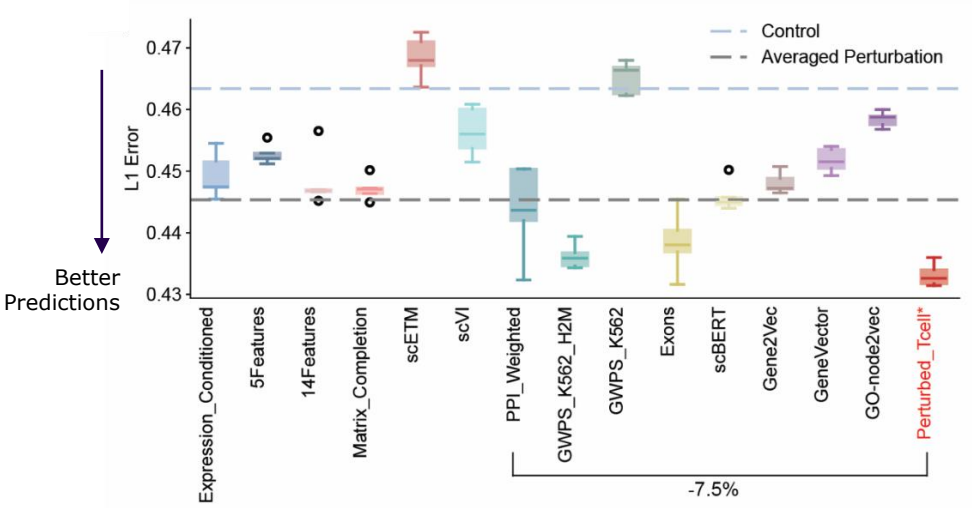
A community machine learning challenge to predict the effects of gene perturbations on T cell differentiation for cancer immunotherapy

Author: Jiaqi Zhang[1,2], Marc A Schwartz[3], Mohammed Mutaher[3], Oluwatomisin Olajide[3], Yuri Pritykin[4], Orr Ashenberg[1,5*], Nir Hacohen[3,6,*], Caroline Uhler[1,2,*]

1. Perturbation/Model: Perturb-seq of 73 T-cell regulators injected in a mouse melanoma model
2. Assay: Proportions of progenitor, effector, exhausted T-cells in tumors
3. Challenge: Predict cell type proportions
4. Challenge: nominate new targets that shift proportions

Lessons:

A. Most models cannot make useful predictions
B. 'winning models' leverage Perturbation data in similar cell types (e.g. K562, Primary T-cells)



Gene function prediction with AI

Application to 'in silico CRISPR screening'

bioRχiv

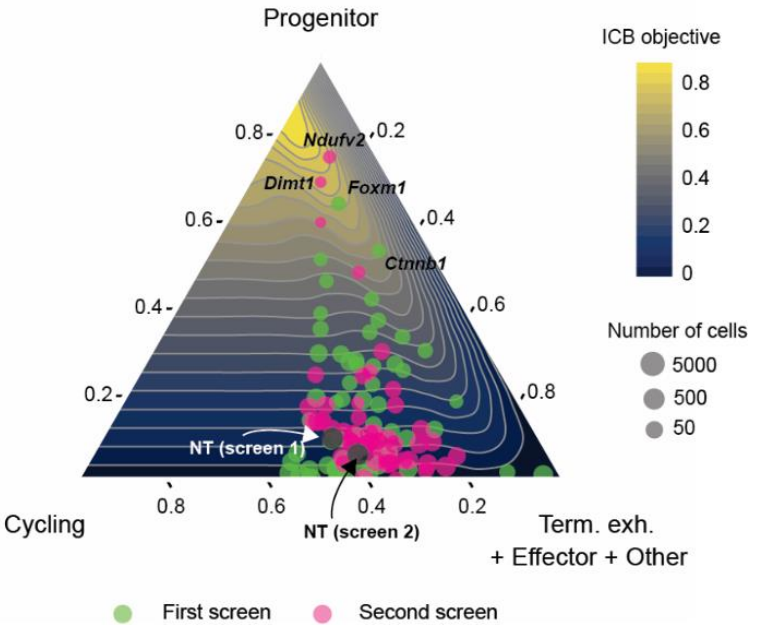
A community machine learning challenge to predict the effects of gene perturbations on T cell differentiation for cancer immunotherapy

Author: Jiaqi Zhang[1,2], Marc A Schwartz[3], Mohammed Mutaher[3], Oluwatomisin Olajide[3], Yuri Pritykin[4], Orr Ashenberg[1,5*], Nir Hacohen[3,6,*], Caroline Uhler[1,2,*]

1. Perturbation/Model: Perturb-seq of 73 T-cell regulators injected in a mouse melanoma model
2. Assay: Proportions of progenitor, effector, exhausted T-cells in tumors
3. Challenge: Predict cell type proportions
4. Challenge: nominate new targets that shift proportions

Lessons:

A. Most models cannot make useful predictions
B. 'winning models' leverage Perturbation data in similar cell types (e.g. K562, Primary T-cells)
C. Most AI-nominated targets are weaker than expert curated targets, but some novel targets emerge!



Part 2. Genetic Screening Conclusions

- Cas9 is a flexible tool enabling genetic screening at scale
- Perturb-seq is being adopted as a generalizable, scalable approach to study gene function
- Cell type / state is extremely important in perturbation experiments
- Combining human genetics variant to function “surveillance” analysis with genetic screening “interrogation” is an emerging approach to investigate disease biology.
- AI Perturbation prediction as a lot of promise, but requires well-matched perturbation data in relevant cell models

Thank You