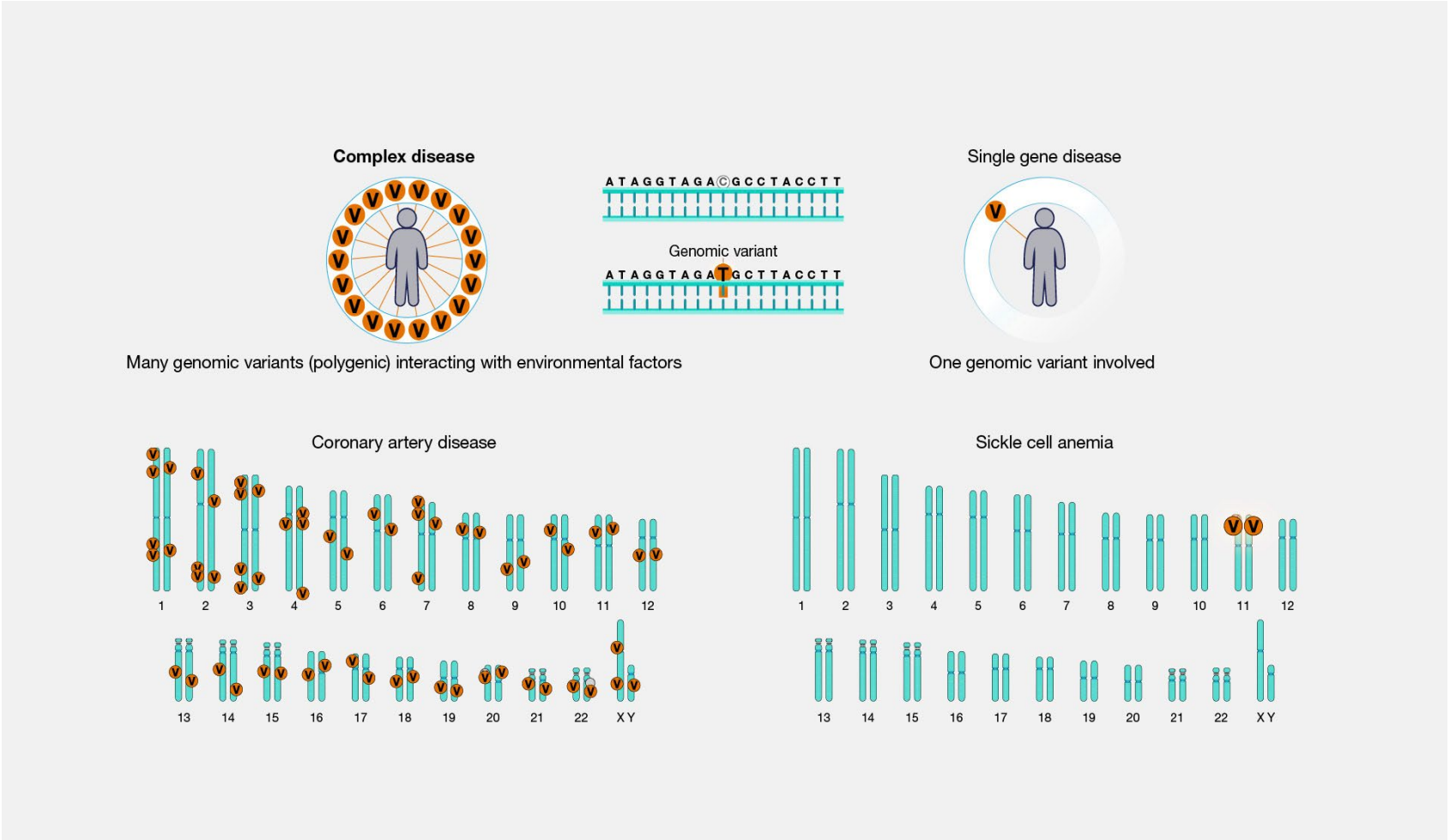
Post-GWAS

One gene ⇨ one mutation ⇨ one outcome

- Beadle and Tatum in 1941
  - Out dated

- Many of the most common and burdensome diseases, such as cardiovascular disease, cancer, Alzheimer's disease, Parkin-son's disease, and type 2 diabetes, are typically ***not (or never)*** caused by single mutations.
- Complex traits, by definition, are influenced by many genes

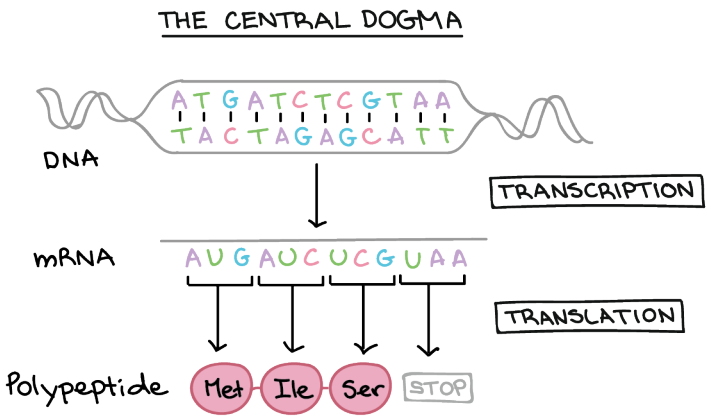


https://www.genome.gov/genetics-glossary/Complex-Disease

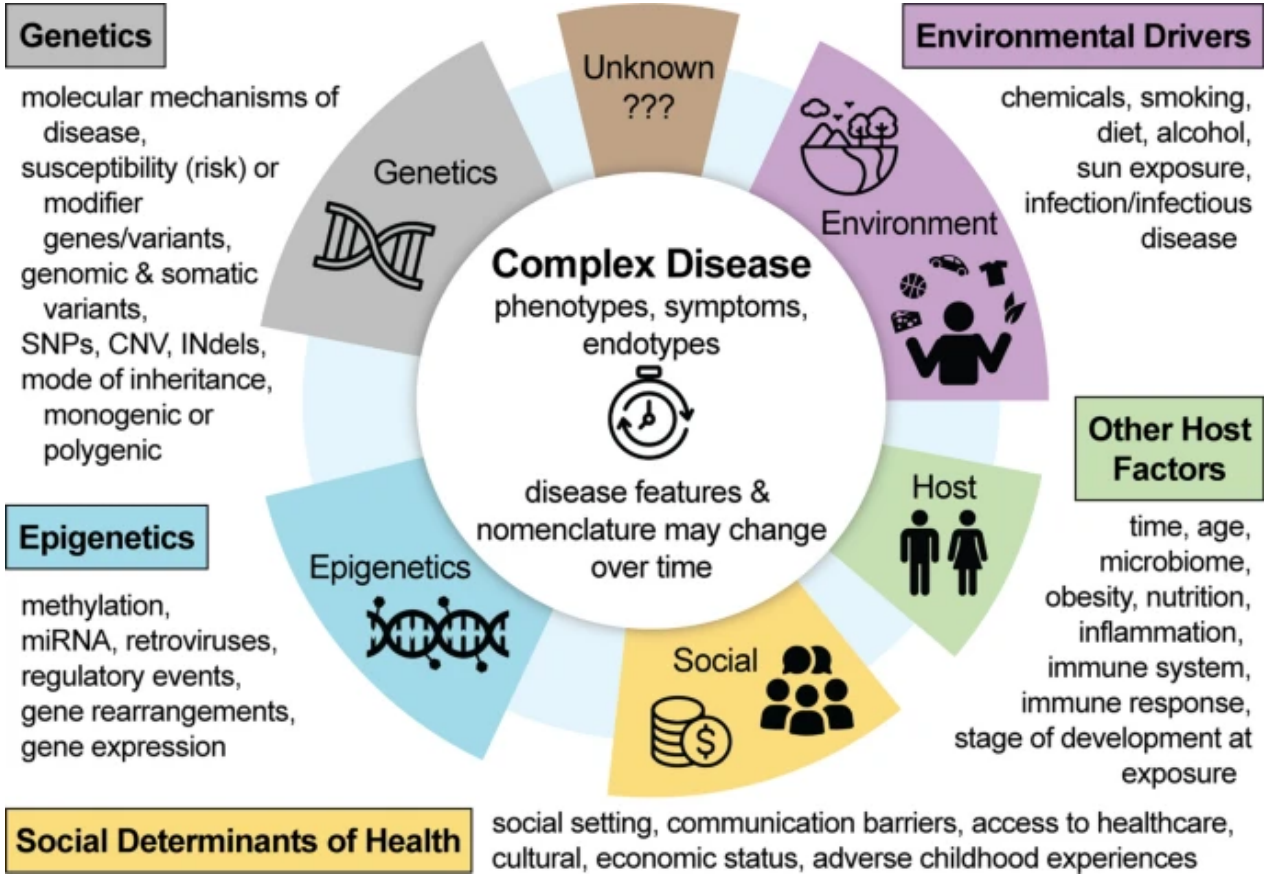
Complex Disease (multifactorial diseases)

- Not a singleular mechanism
- Rather multi-modal, including epigentices, biome, envirome, etc

- Challenges the central dogma



https://www.khanacademy.org/



Schriml LM, Lichenstein R, Bisordi K, et al: Modeling the enigma of complex disease etiology. J Transl Med 21:148, 2023

Beyond Locus Identification

- Gap between locus identification and functional understanding

- The association of a locus with disease does not specify which variant (or variants) at that locus is actually causing the association (the "causal variant"), nor which gene (or genes) is affected by the causal variant (the "target gene"
- There are often many co-inherited variants in strong linkage disequilibrium (LD) with the most significant (or "sentinel") disease-associated variant, comprising a haplotype
  - within the haplotype, genetic variants in strong LD often have statistically indistinguishable associations with disease risk

Disease associated variants

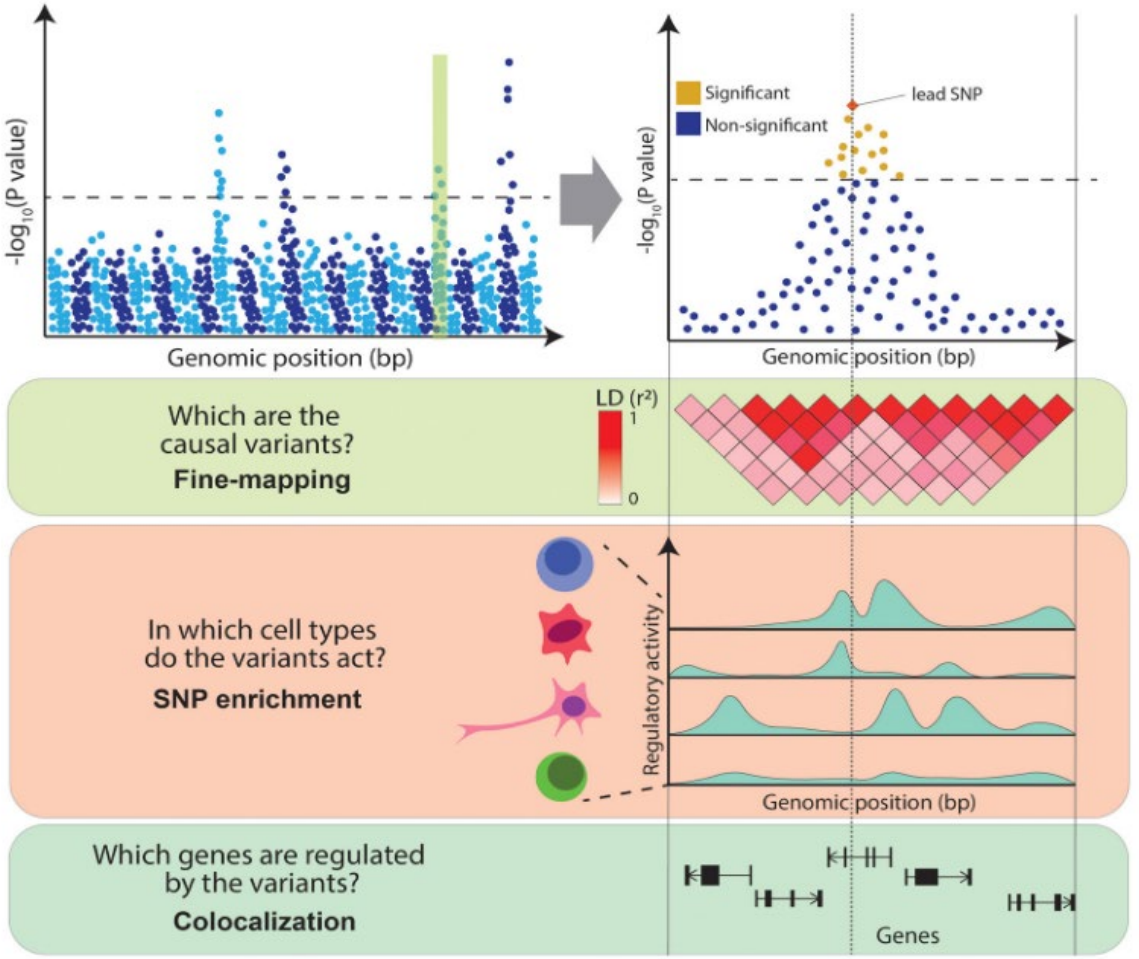
- Disease associated variants, as well as variants in strong LD with them, are enriched in predicted transcriptional regulatory regions, called ‘‘cis-regulatory elements’’ (CREs)
  - > 90% of disease-associated variants are located in non-protein-coding regions of the genome, and many are far away from the nearest known gene.
  - regulatory elements (RE) such as enhancers, promoters, various non-coding RNAs, and CTCF binding sites; and may also alter other elements including splice junctions
- Disease associated variants with expression quantitative trait loci (eQTLs) specifically, disease associated variants are more likely to be associated with the expression (mRNA) levels of one or more genes than would be expected by chance

Three Traditional Strategies

1) Fine-mapping
   - Causal inference

2) SNP Enrichment
   - Meta analysis

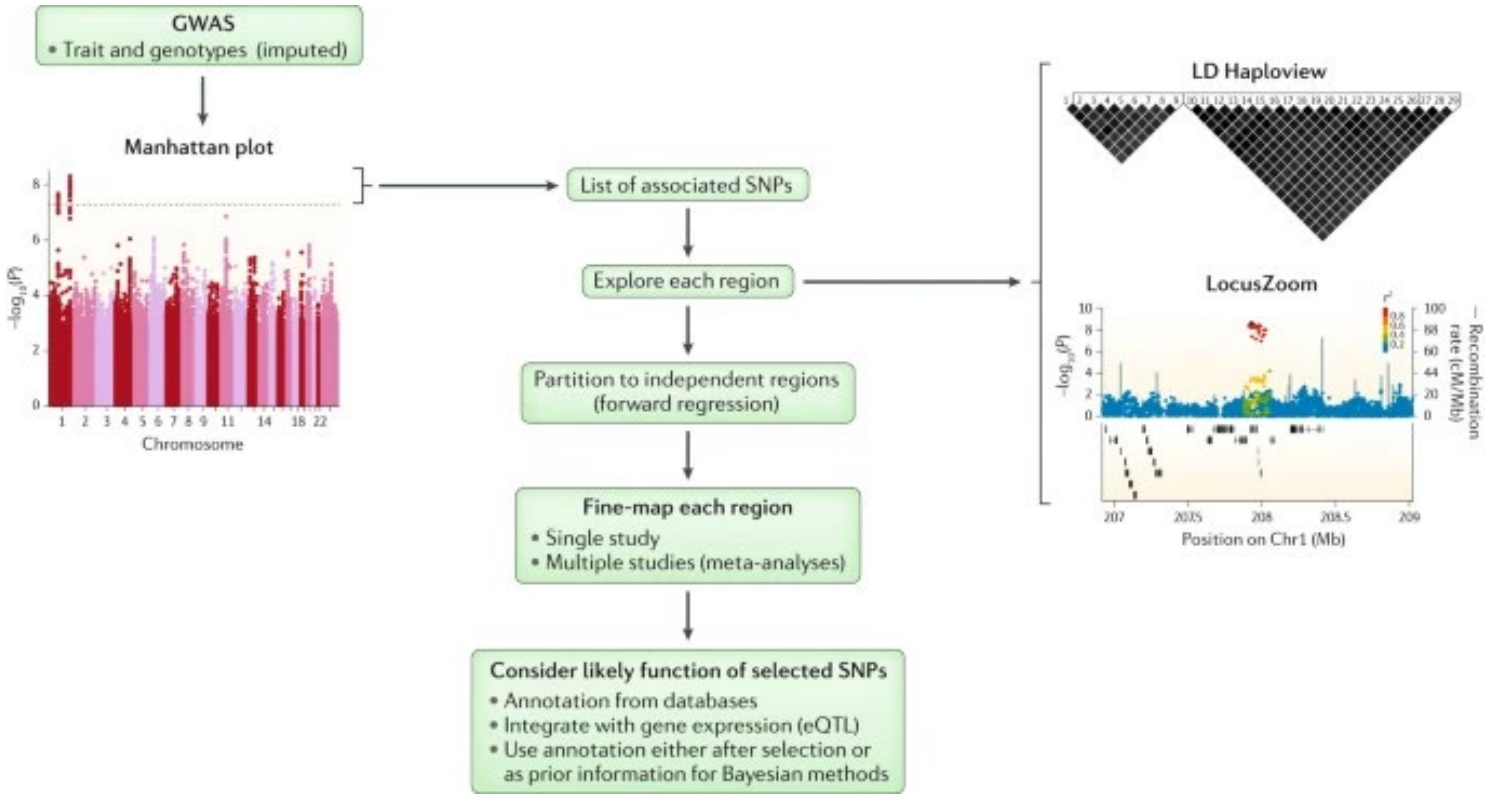
3) Colocalization
   - Down stream genes



Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

Fine Mapping

- Heuristic, Bayesian, Posterior Probabilities, Penalized regression approaches
- Influenced by SNP density, LD structure, effect size, sampling size



Schaid DJ, Chen W, Larson NB: From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 19:491-504, 2018

Fine Mapping Tools

Software	Input	Output	Reference
FINEMAP	sumstat: beta, se, LD, n causal	PIP and Bayesian	Web Page [50]
CaviarBF	sumstat: z value, LD, eQTL, fixed causal	PIP and Bayesian	Web Page [47]
PAINTOR	sumstat: z value, LD, eQTL, fixed causal, multi LD	PIP and Bayesian	Web Page [48]
CAVIAR - eCAVIAR	sumstat, LD, eQTL fixed causal	probability and confidence set	Web Page [51, 52]

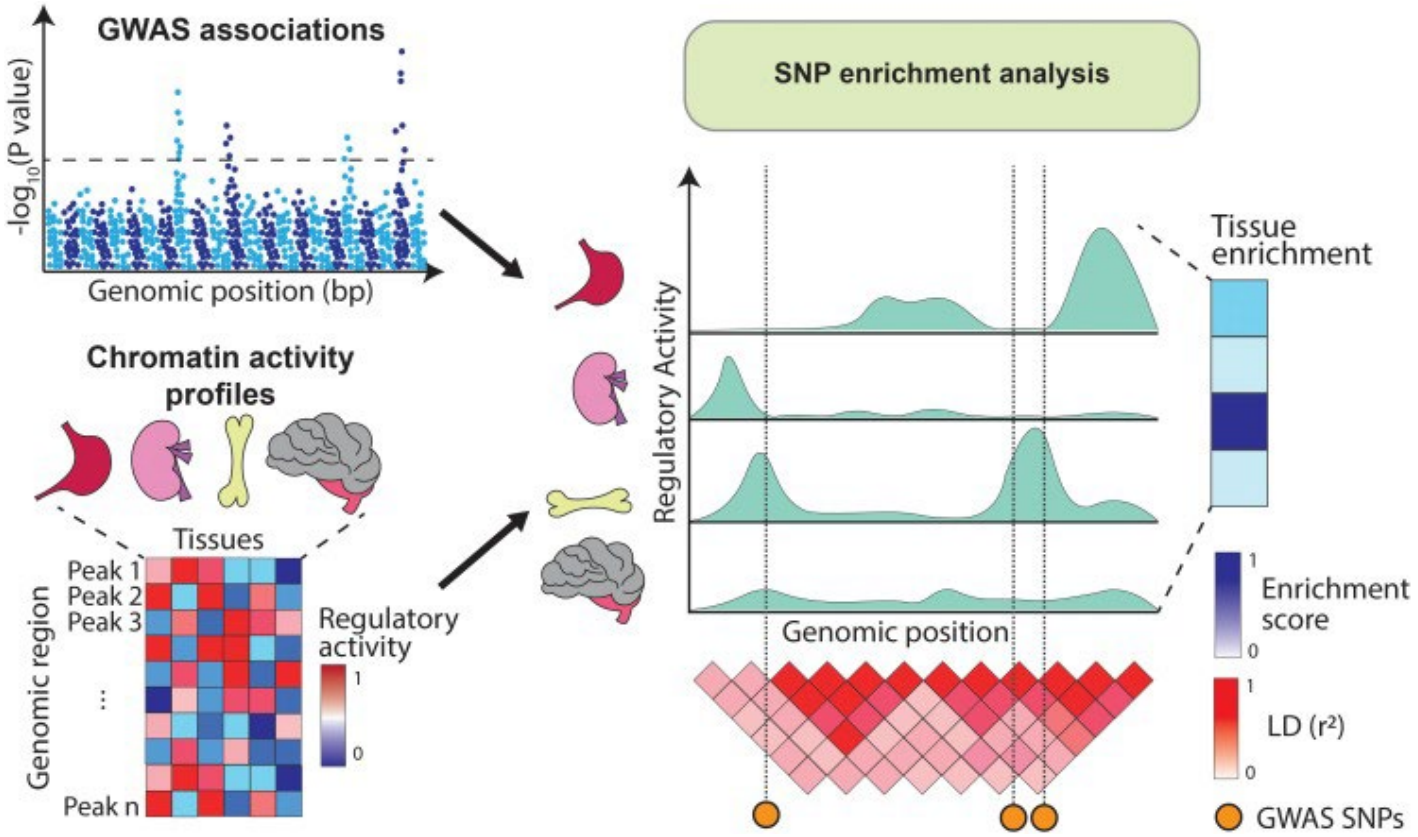
Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

https://catalog.gwaslab.org/Tools_Fine_mapping_README/ (more complete)

SNP Enrichment

- GWAS SNPs are overlapped with regulatory elements
- Assigned an enrichment score to tissues

- Alternatively, gene and pathway based



Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

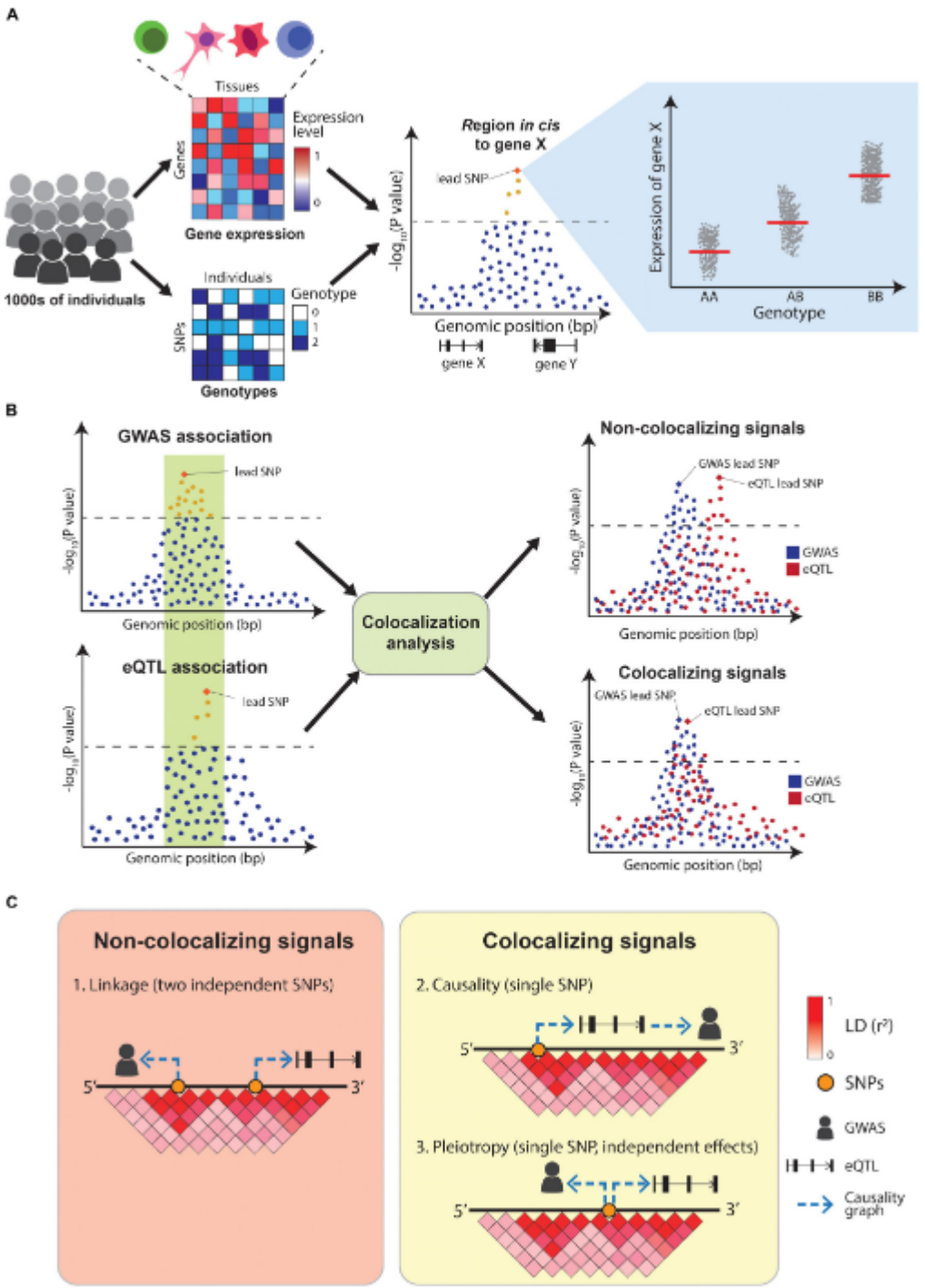
SNP Enrichment Tools

Method	Publications	Hypothesis tested	Input data
SNPsea	Hu et al., 2011; Slowikowski et al., 2014	Accumulation of GWAS variants near genes with high tissue specificity	Gene expression, GWAS index variants
EpiGWAS	Trynka et al., 2013	Accumulation of GWAS variants near highly active regulatory elements	Chromatin marks, GWAS index variants
GREGOR	Schmidt et al., 2015	Accumulation of GWAS variants in regulatory elements	Chromatin marks, GWAS index variants
GoShifter	Trynka et al., 2015	Intersection of GWAS variants with regulatory annotations (based on local-shifting of annotations)	Functional annotations, GWAS index variants
fGWAS	Pickrell, 2014	Higher GWAS effect sizes observed if a loci and a SNP overlap a functional annotation	Functional annotations, GWAS summary statistics
CHEERS	Soskic et al., 2019	Accumulation of GWAS variants in regulatory elements with high tissue specificity	Chromatin marks (quantitative), GWAS index variants
GARFIELD	Iotchkova et al., 2019	Higher GWAS effect sizes observed in variants that overlap regulatory annotations	Chromatin annotations, full GWAS summary statistics
RolyPoly	Calderon et al., 2017	Higher GWAS effect sizes observed near highly expressed genes	Gene expression, full GWAS summary statistics
LDSC	Finucane et al., 2015	Accumulation of heritability in variants overlapping a functional annotation	Chromatin annotations, full GWAS summary statistics
LDSC-SEG	Finucane et al., 2018	Accumulation of heritability near tissue specific genes	Gene expression, full GWAS summary statistics

Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

Colocalizations

- Establish linkage between SNP and QTL-mapping studies:
  - gene expression (eQTLs)
  - protein expression (pQTLs)
  - exon splicing (sQTLs)
  - DNA methylation (mQTLs)
  - chromatin acetylation (acQTLs)
  - chromatin accessibility (caQTLs)



Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

Colocalization Tools

Method	Publication	Approach	Input data
Regulatory trait concordance (RTC)	Nica et al., 2010	Conditional regression	Individual genotypes
Proportionality test	Wallace et al., 2012	Test for concordance of effects	Individual genotypes
Sherlock	He et al., 2013	Genome-wide comparison of association "signatures"	Summary statistics
COLOC	Giambartolomei et al., 2014	Bayesian test	Summary statistics
gwas-pw	Pickrell et al., 2016	Bayesian test	Summary statistics
eCAVIAR	Hormozdiari et al., 2016	Bayesian fine-mapping and colocalization	Summary statistics
enloc	Wen et al., 2017	Bayesian test for enrichment, fine-mapping and colocalization	Summary statistics
MOLOC	Giambartolomei et al., 2018	Bayesian test for multiple traits	Summary statistics

Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

Machine Learning Approaches

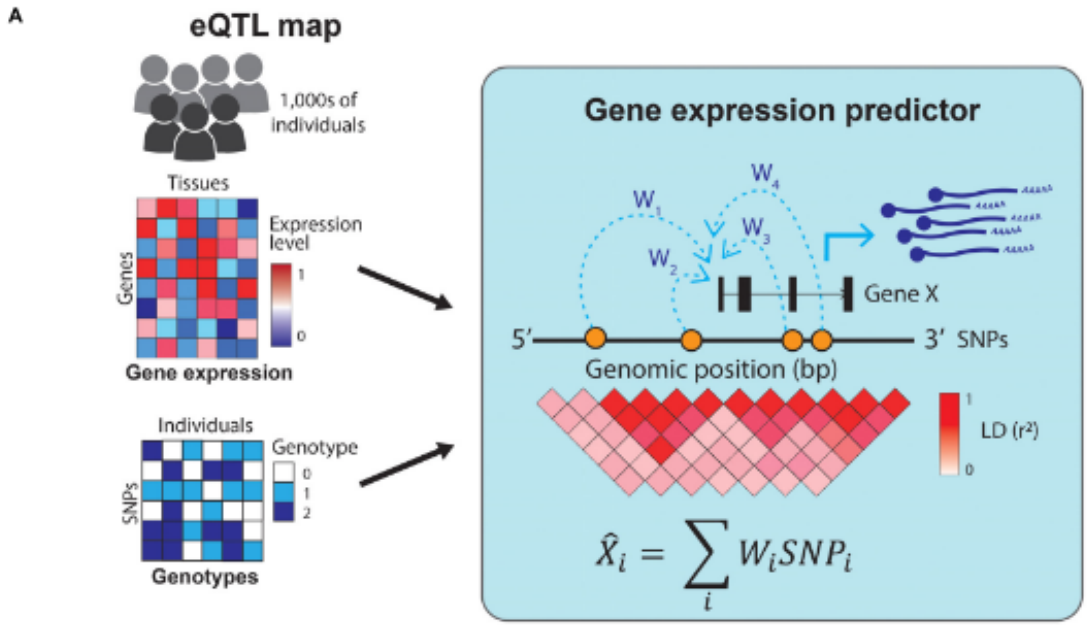
- Transcriptome-Wide Association Studies (TWAS)
  - Integrate GWAS with expression databases to directly infer genes & traits

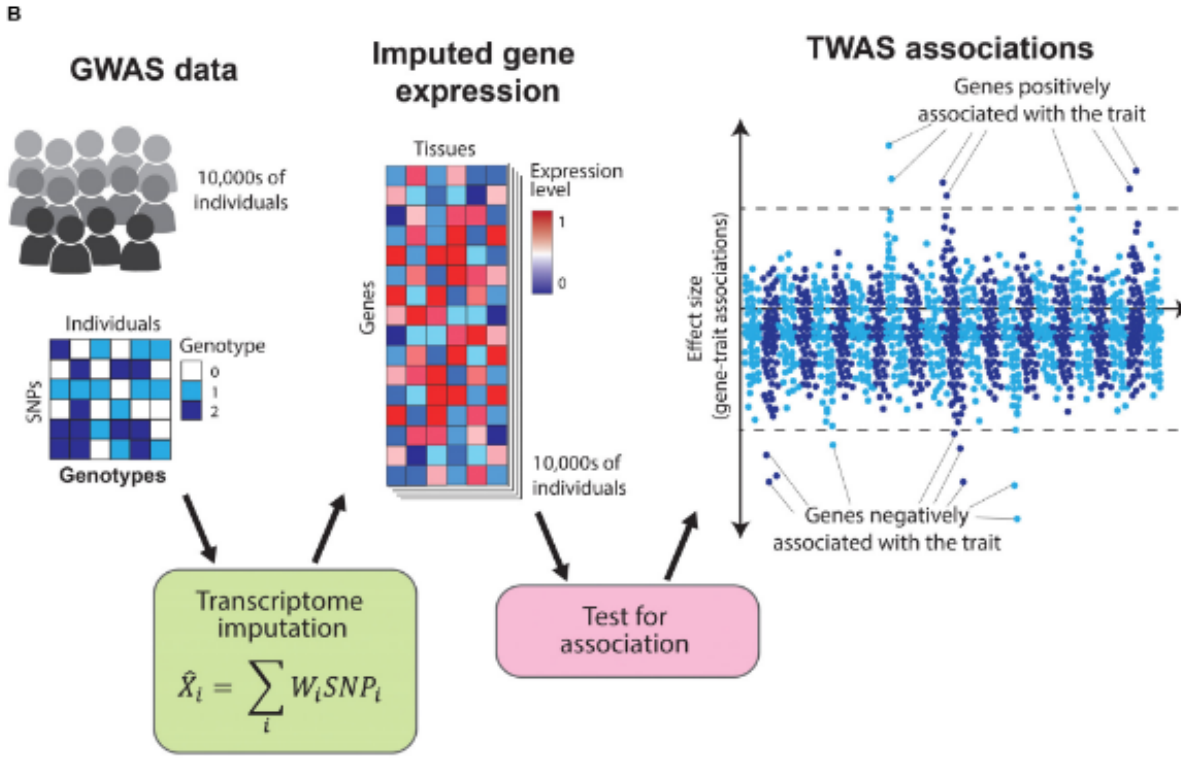
- Large Language Models (LLMs)
  - Model Context Protocol

Transcriptome-wide Association Studies

- Train on tissue-specifice eQTL maps (thousands)
- Build gene expression predictors
  - For cis SNPs, estimate expression levels

- Impute gene expression for GWAS study
- Derive disease associated genes





Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

TWAS Tools

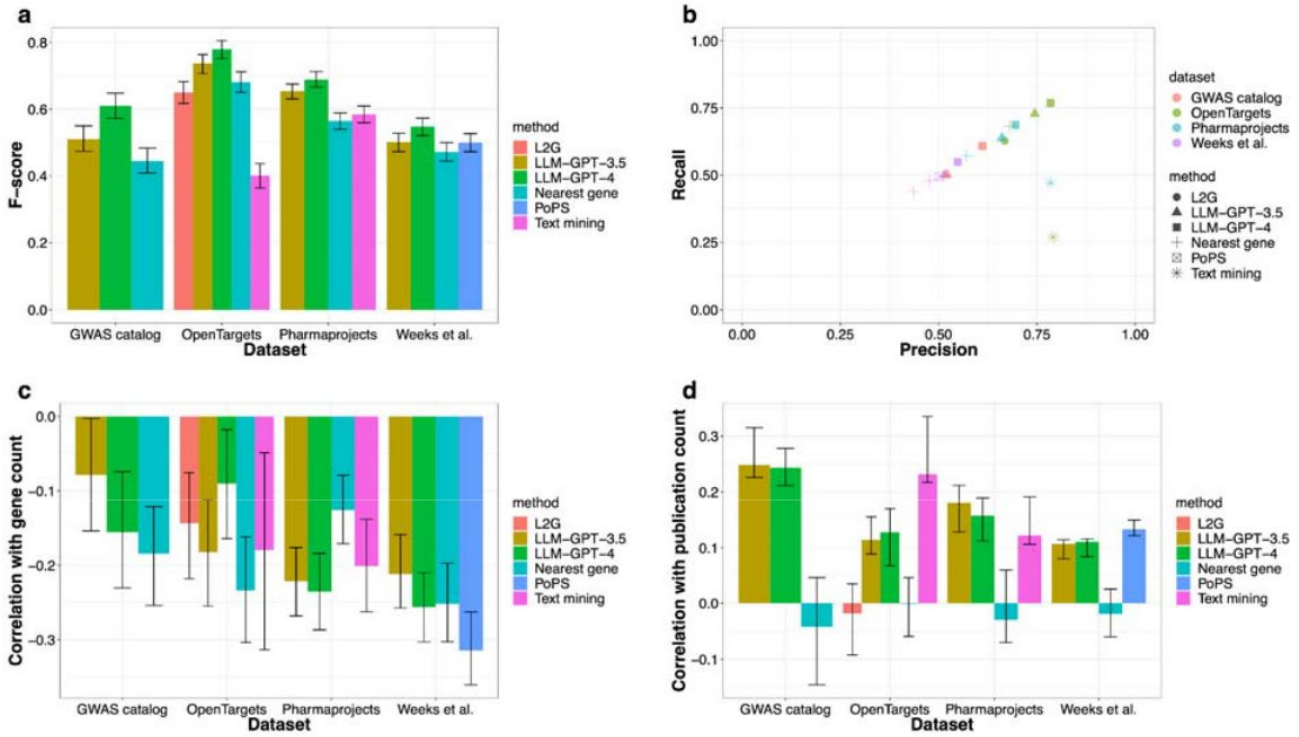
Software	specificity	remarks	link and publication
GWAMA	FE, RE, GC	short manual	Web Page [50]
Meta-Soft	FE, RE, RE2, FE2 and BE, Q, I^2 GC	R script to plot effect of study	Web Page [70]
MR-MEGA	FE, RE, Q, I^2	manual limited	Web Page [76]
METAL	FE, RE, Q, I^2 , SOC, p-value, GC		Web Page [65]
PLINK (1.9)	FE, RE	Few options described	Web Page [15]

Cano-Gamez E, Trynka G: From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet 11:424, 2020

Large Language Models

- Comparing the performance of different LLMs
- Extracting from different online DBs

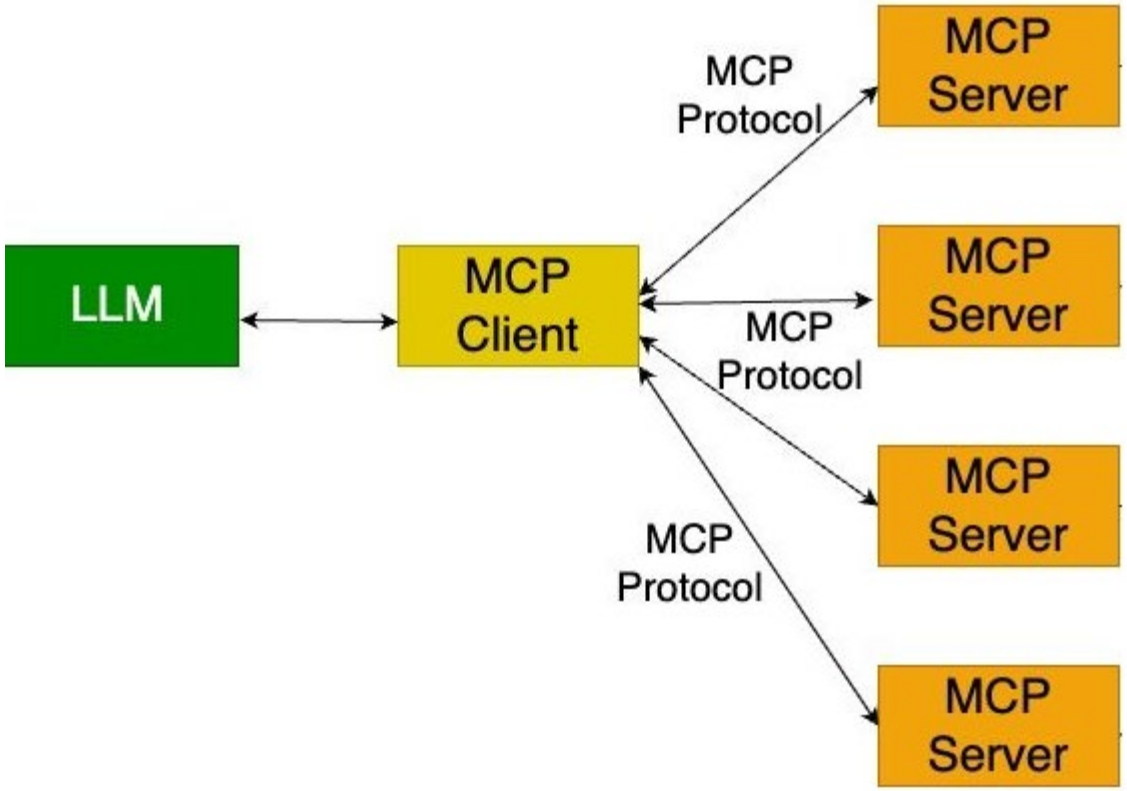
- GPT-3.5 and GPT-4 outperform state-of-the-art methods in identifying putative causal gene



https://www.medrxiv.org/content/10.1101/2024.05.30.24308179v1.full.pdf

Model Context Protocol

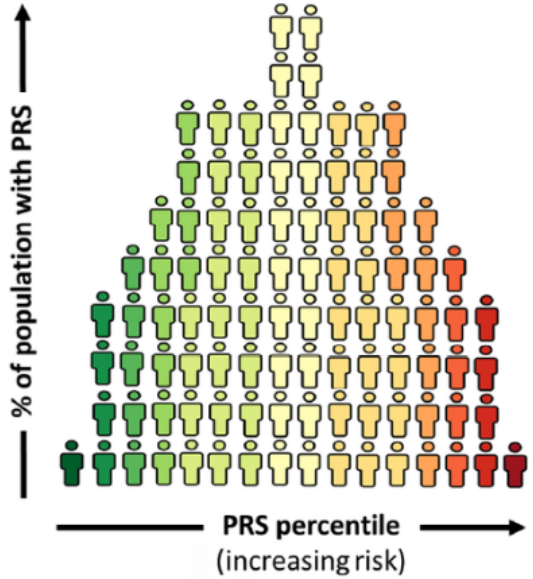
- LLMs
  - Easy to implement
  - Rather stupid (hallucinations)
  - Intelligence is from proper tuning
- MCP adds a management layer
  - Instructions specifically for formatting and interacting via REST API's
  - Providing a feedback loop for improving LLM performance



Polygenic Risk Scores

Polygenic Risk Scores (PRS)

PRS percentile	Risk of disease vs. reference group
0-1	Lowest
1-5	
5-10	
10-20	
20-40	
40-60 (reference)	1
60-80	
80-90	
90-95	
95-99	
99-100	Highest



https://www.rgare.com/knowledge-center/article/polygenic-risk-scores-combining-thousands-of-genetic-variants-to-predict-disease

SysBiolPGWAS

- A pipeline tools for Post-GWAS