MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

SNPs - GWAS - eQTLs

Sebastian Schmeier
s.schmeier@gmail.com
http://sschmeier.github.io/bioinf-workshop/
17.08.2015

THE ENGINE
OF THE NEW
NEW ZEALAND

Overview

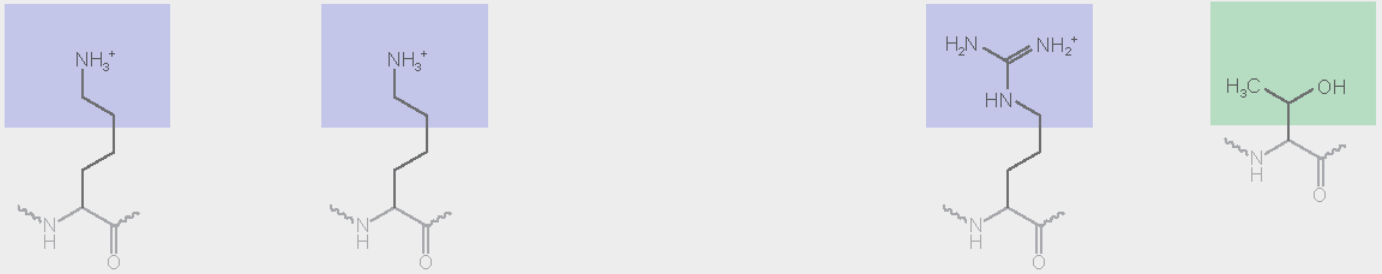
- Single nucleotide polymorphism (refresh)
- SNPs effect on genes (refresh)
- Genome-wide association studies (refresh)
- Quantitative trait loci (QTL)
- SNPs effect outside of the genic region
- Expression quantitative trait loci (eQTL)

Single nucleotide polymorphism (SNP)

- A SNP is defined as a single base change in a DNA sequence that occurs in a significant proportion (more than 1 percent) of a large population.
- SNPs result from replication errors and DNA damage

SNPs effects in coding regions

	No mutation	Point mutations			
		Silent	Nonsense	Missense	
				conservative	non-conservative
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	**Lys**	**Lys**	**STOP**	**Arg**	**Thr**

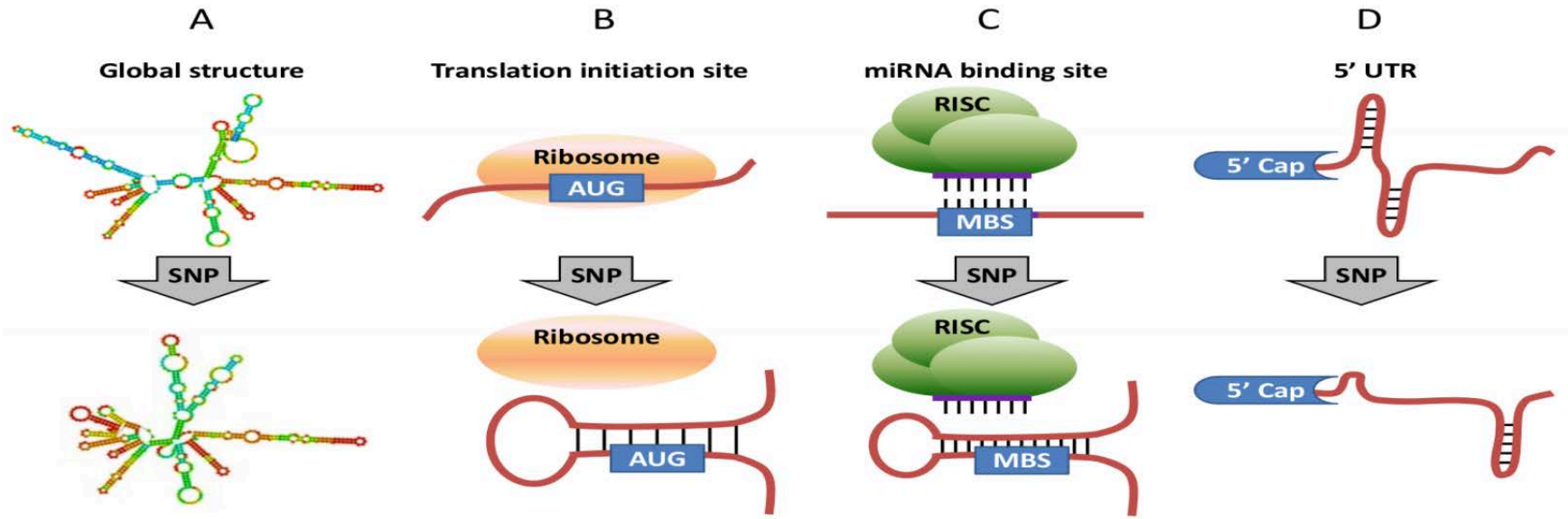


https://en.wikipedia.org/wiki/Point_mutation

basic
polar

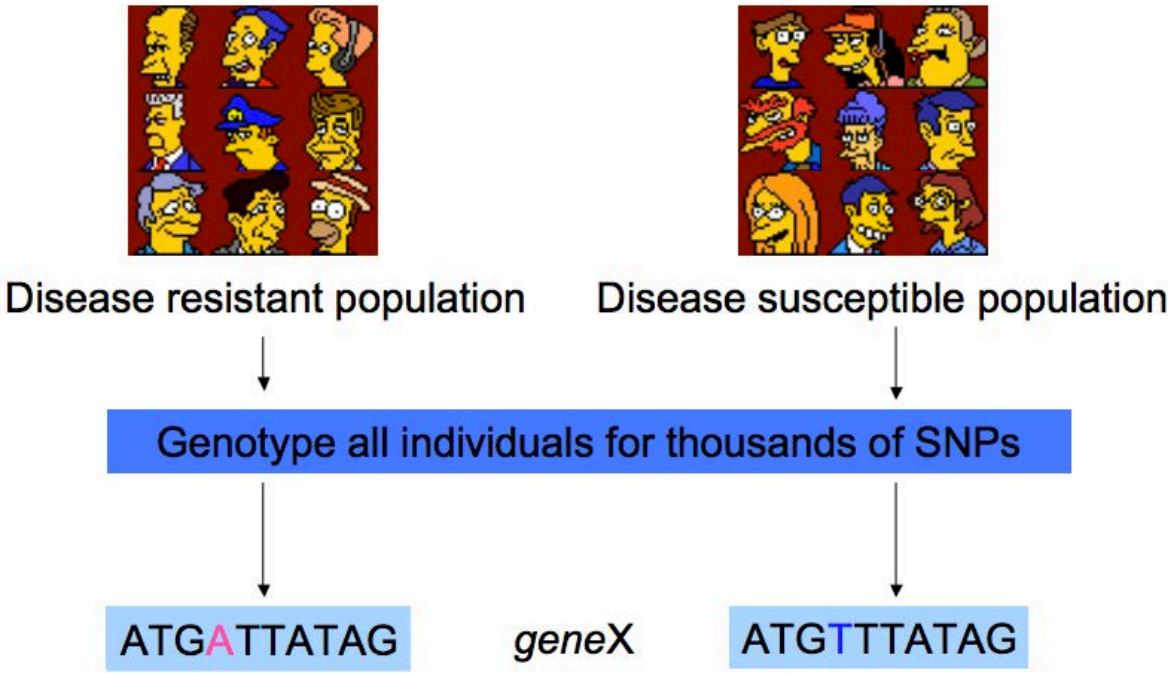
Structural effect caused by exonic SNPs

A. Exonic SNPs that cause substantial change in mRNA global structure and stability.
B. Exonic SNPs that change the mRNA local structure around translation initiation sites.
C. Exonic SNPs that change the structural accessibility of miRNA binding sites (MBSs).
D. Exonic SNPs in 5' UTR that may change mRNA local structure near 5' cap and thus affect miRNA-mediated translation inhibition.



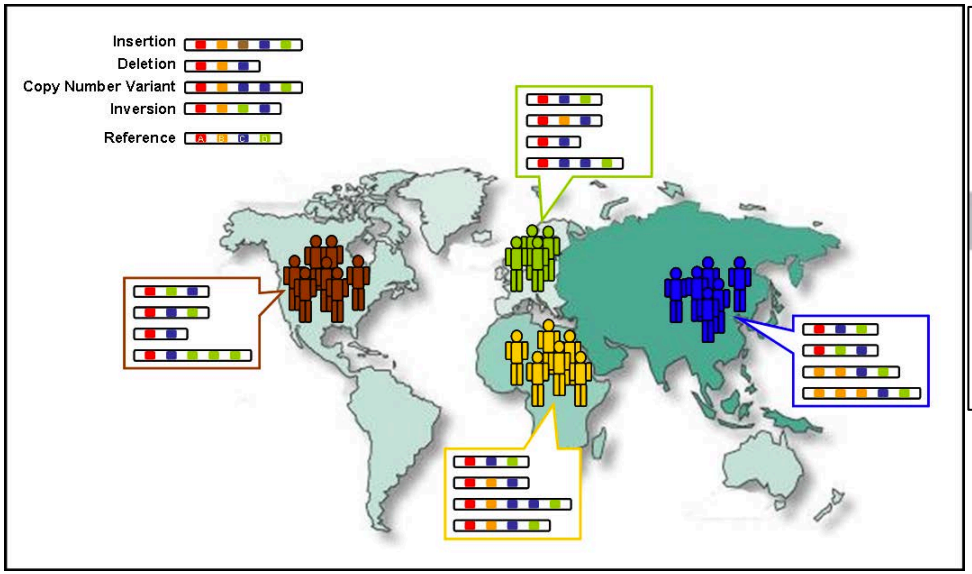
Why are SNPs useful?

- Example: Resistant people all have an 'A' at position 4 in gene X, while susceptible people have a 'T' (A/T are the SNP



1000 Genome project

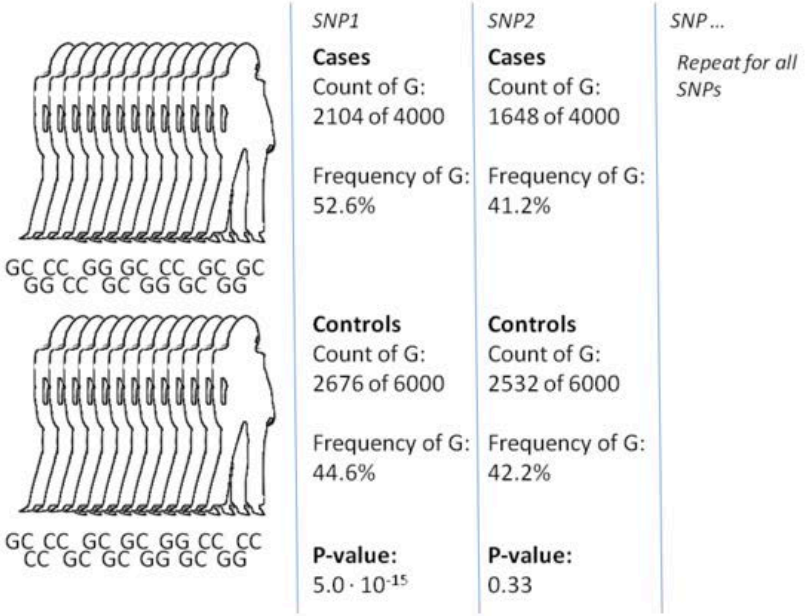
- First project to sequence the genomes of a large number of people, to provide a comprehensive resource on human genetic variation (2008-2010)
- Aim was to find most genetic variants that have frequencies of at least 1% in the populations



http://www.1000genomes.org/

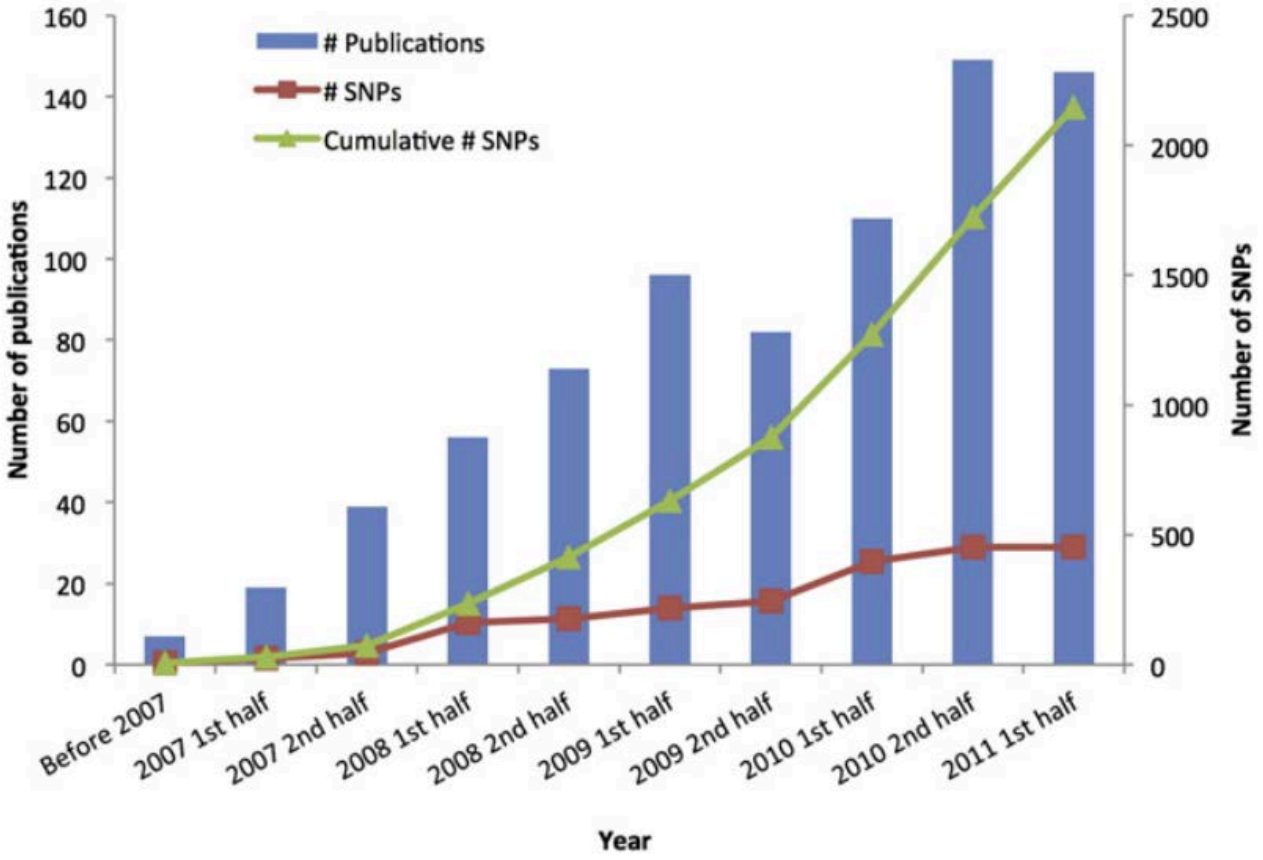
Genome-wide association study (GWAS)

- The aim is to identify genes involved in human disease/traits
- This method searches the genome for SNPs that occur more frequently in people with a particular trait/disease than in people without the trait/disease.
- GWAS scans many SNPs at the same time using SNP arrays or NGS-based methods



https://en.wikipedia.org/wiki/Genome-wide_association_study

Genome-wide association study (GWAS)



Five Years of GWAS Discovery. Visscher *et al.* Am J Hum Genet. 2012 Jan 13; 90(1): 7–24.

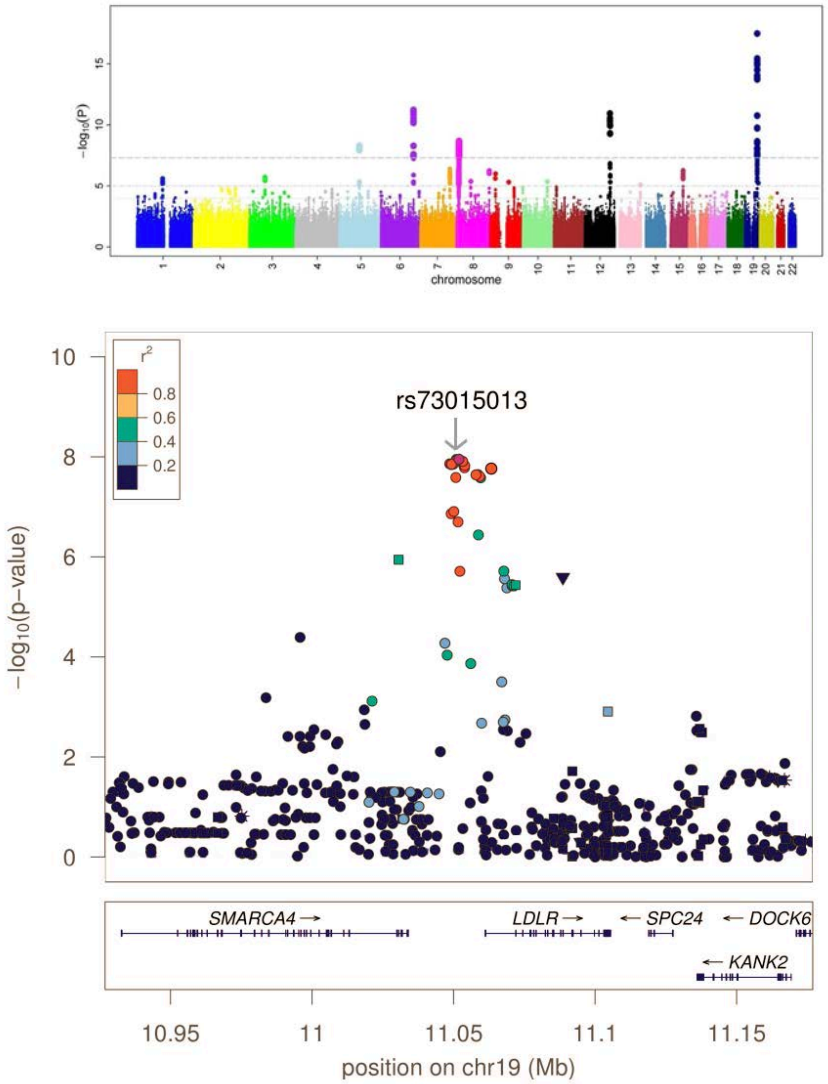
Genome-wide association study (GWAS)

"To date, genome-wide association studies (GWAS) have published hundreds of common variants whose allele frequencies are statistically correlated with various illnesses and traits. *However, the vast majority of such variants have no established biological relevance to disease or clinical utility for prognosis or treatment.*"

Genetic heterogeneity in human disease. McClellan J1, King MC. Cell. 2010 Apr 16;141(2):210-7. doi: 10.1016/j.cell.2010.03.032.

Genome-wide association study (GWAS)

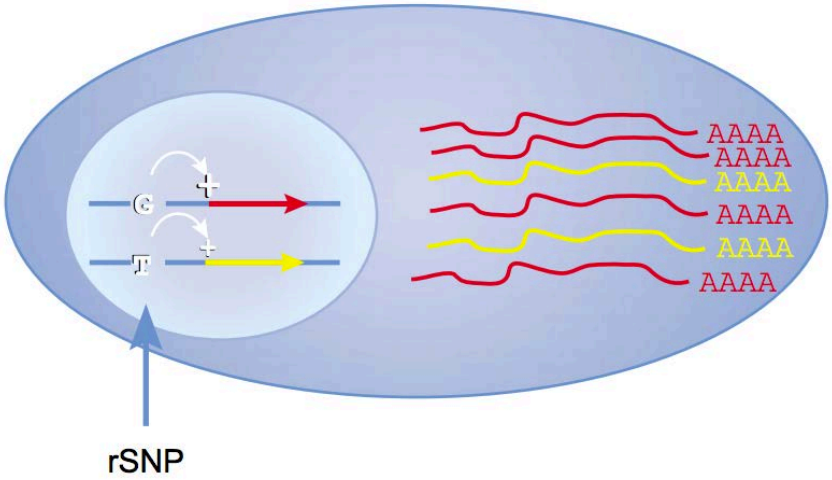
- GWAS identify SNPs and other variants in DNA which are associated with a disease or trait, but cannot on their own specify which genes are causal
- The molecular mechanisms by which genetic variation predisposes individuals to diseases are still poorly characterized
- The majority of GWAS SNPs are non-coding!!!!



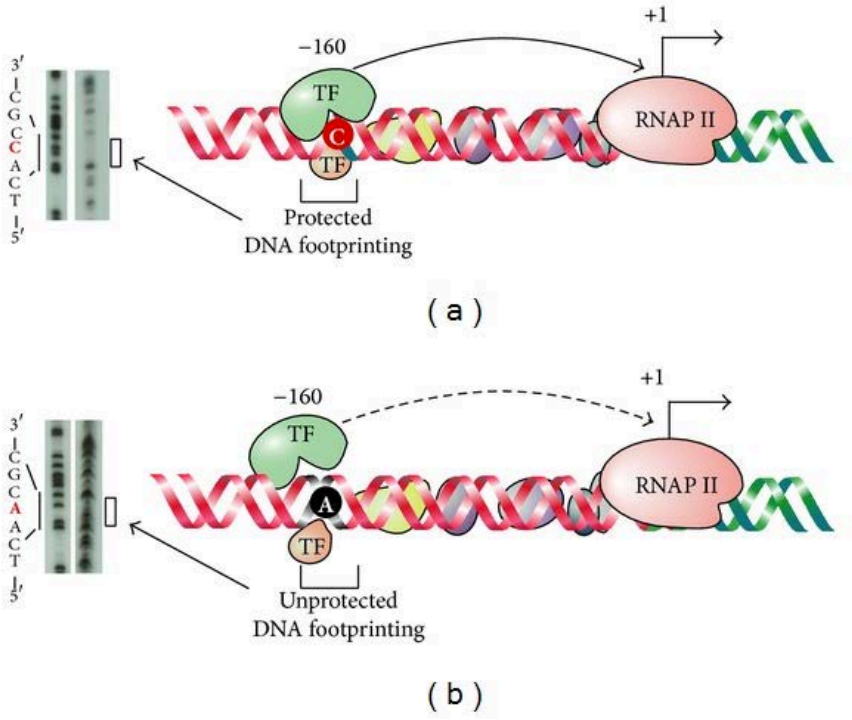
https://en.wikipedia.org/wiki/Genome-wide_association_study

SNP effect in non-coding regions

- Regulatory SNPs (rSNPs)
- SNPs regulate gene transcription by affecting transcription factor (TF) binding



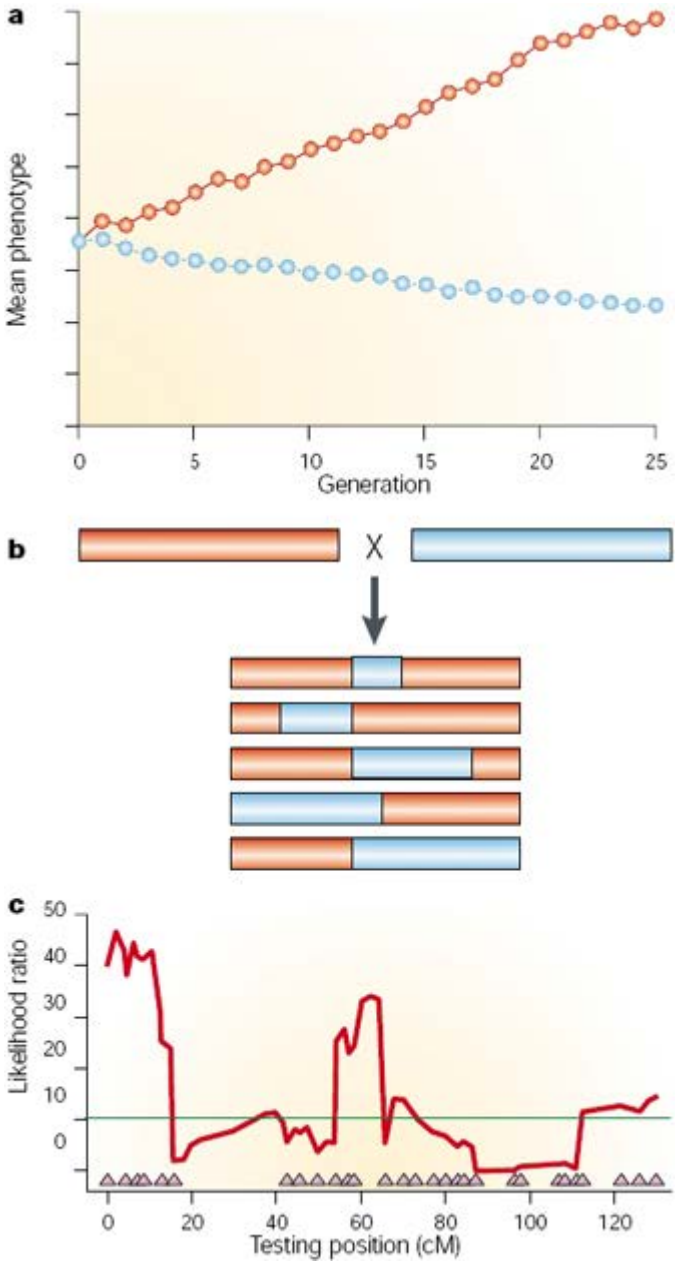
doi:10.1155/2014/967565



http://www.nature.com/ng/journal/v33/n4/full/ng0403-439.html

Quantitative trait locus (QTL)

- A QTL is a section of DNA (the **locus**) that correlates with variation in a phenotype (the quantitative **trait**)
- Requires parental strains (A) with genetic differences in the trait
- Parental lines get crossed to create F1 offspring, which are crossed among themselves to create F2 that contain diff. fractions of the parental genomes (B)
- Statistics to test regions association to trait (C), highest likelihood at 60cM



http://www.nature.com/scitable/topicpage/quantitative-trait-locus-qtl-analysis-53904

expression QTL (eQTL)

- A **genetic locus** where the genotype of a variant is significantly associated with gene expression levels of one or more genes.
- An eQTL usually contains multiple DNA variants, i.e. supposedly regulatory SNPs (rSNPs)
- Mapping eQTLs is done using standard QTL mapping methods that test the linkage between **variation in expression and genetic polymorphisms**.

expression QTL (eQTL)

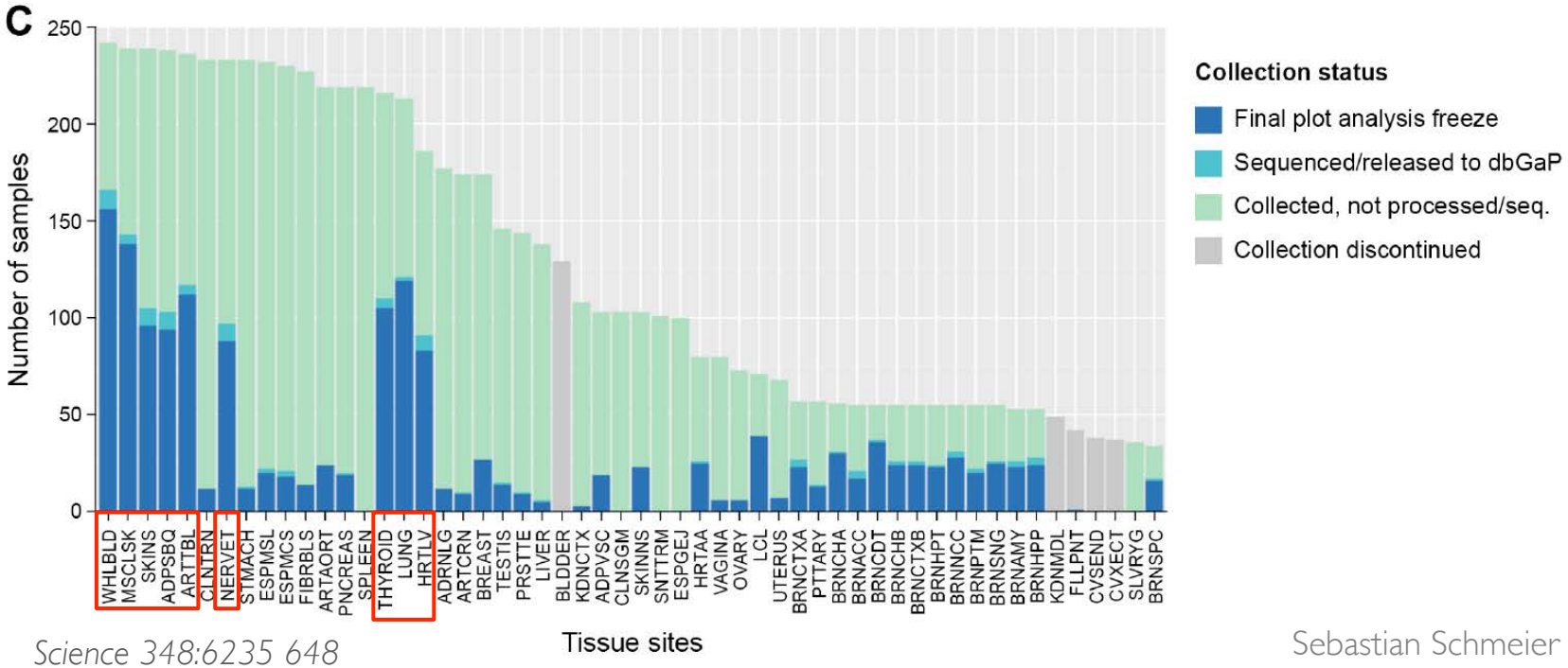
- **Cis-eQTL**: A genetic variant that influences the expression levels of a proximal gene on the same chromosome in an allele-specific manner.
- **Trans-eQTL**: A genetic variants that affects gene expression through an intermediate trans factor, such as a protein or RNA regulator. Trans-eQTLs usually lie far away from the target gene or on a separate chromosome.

Genotype-tissue expression (GTEx)

- Large-scale project to discover eQTLs in human tissues
- Pilot study data published in 2015 *(Science 348:6235 348)*
- The study uses SNPs as markers for eQTLs
- ~6.8 million SNPs with *minor allele frequency* (MAF) >= 5% were tested
- The ultimate goal is to provide a framework fir biological interpretation of disease-related variants (→what is missing from GWAS)

Genotype-tissue expression (GTEx)

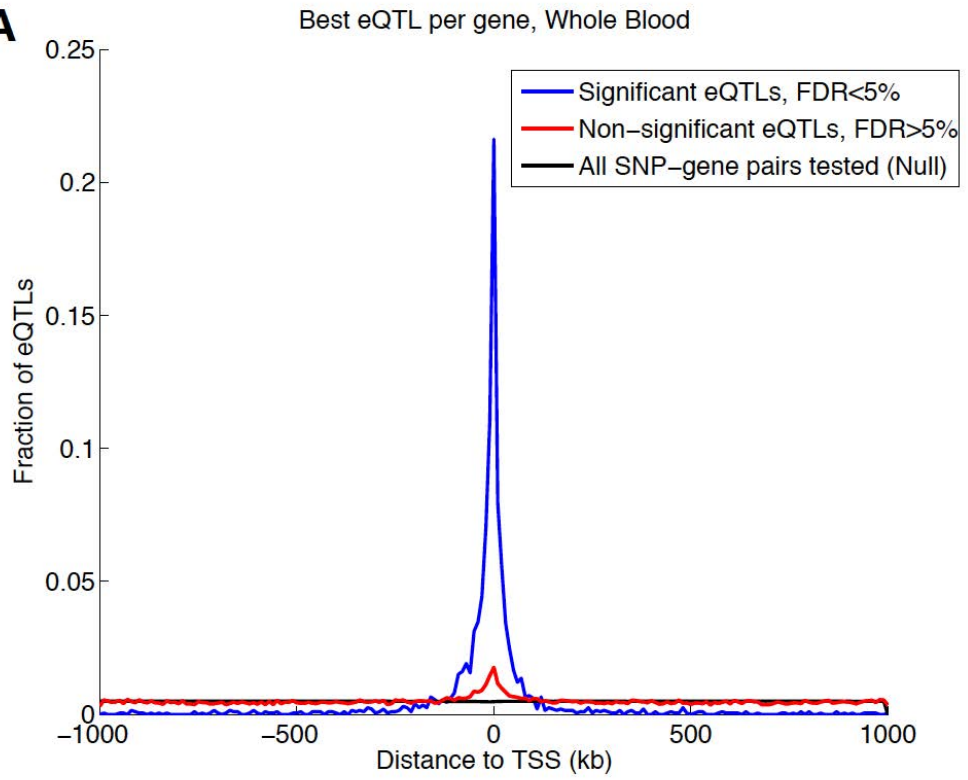
- 237 postmortem donors
- 28 tissues per donor
- 54 distinct body sites



Science 348:6235 648

Genotype-tissue expression (GTEx)

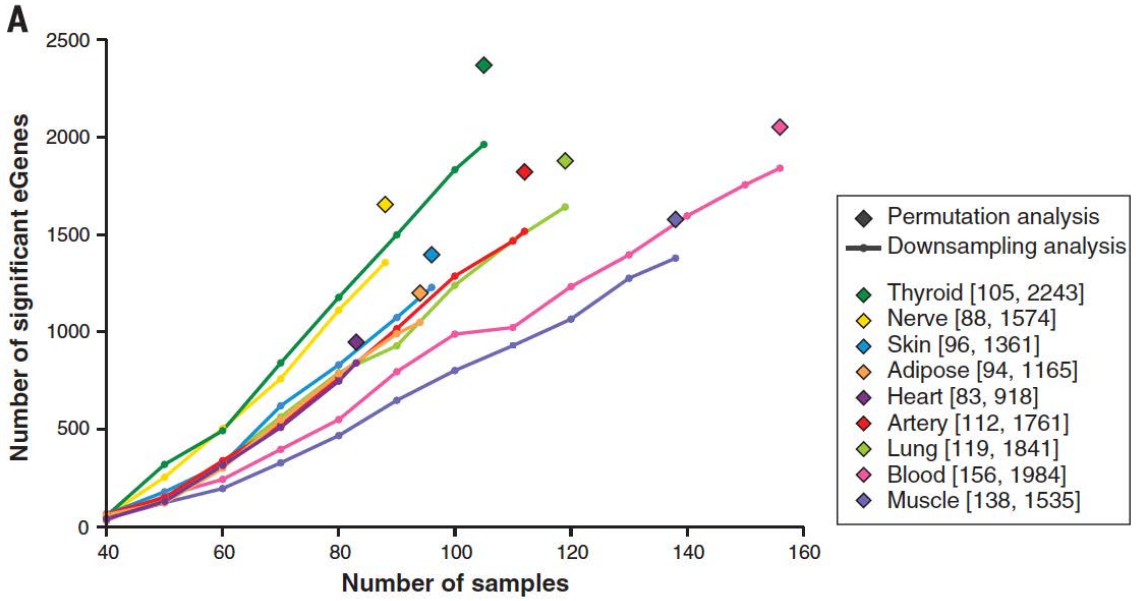
- Focuses on *cis-eQTLs*
  → here 1Mbp surrounding TSS of each gene
- The eQTL signal shows an upstream bias (~60% of all eQTL are upstream)

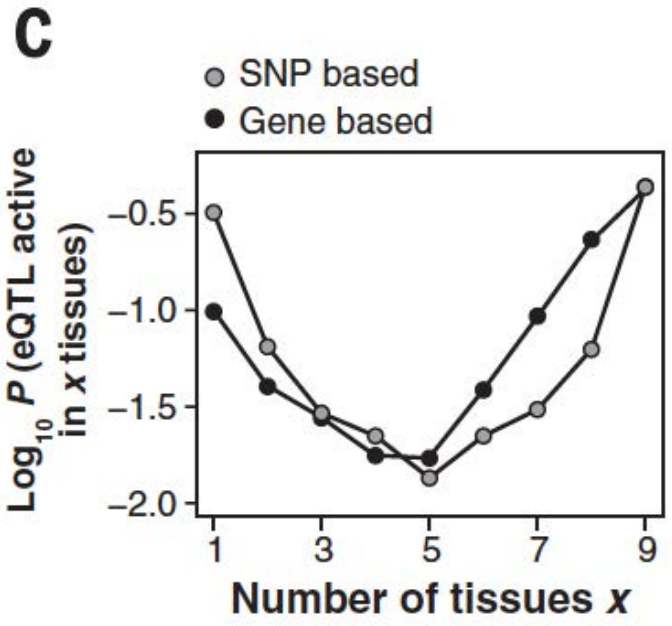


Science 348:6235 648

Genotype-tissue expression (GTEx)

- Single tissue analysis revealed different number of eQTL genes per tissue
- Multitissue analysis reveals the most likely configurations are for active eQTLs in few or many tissues

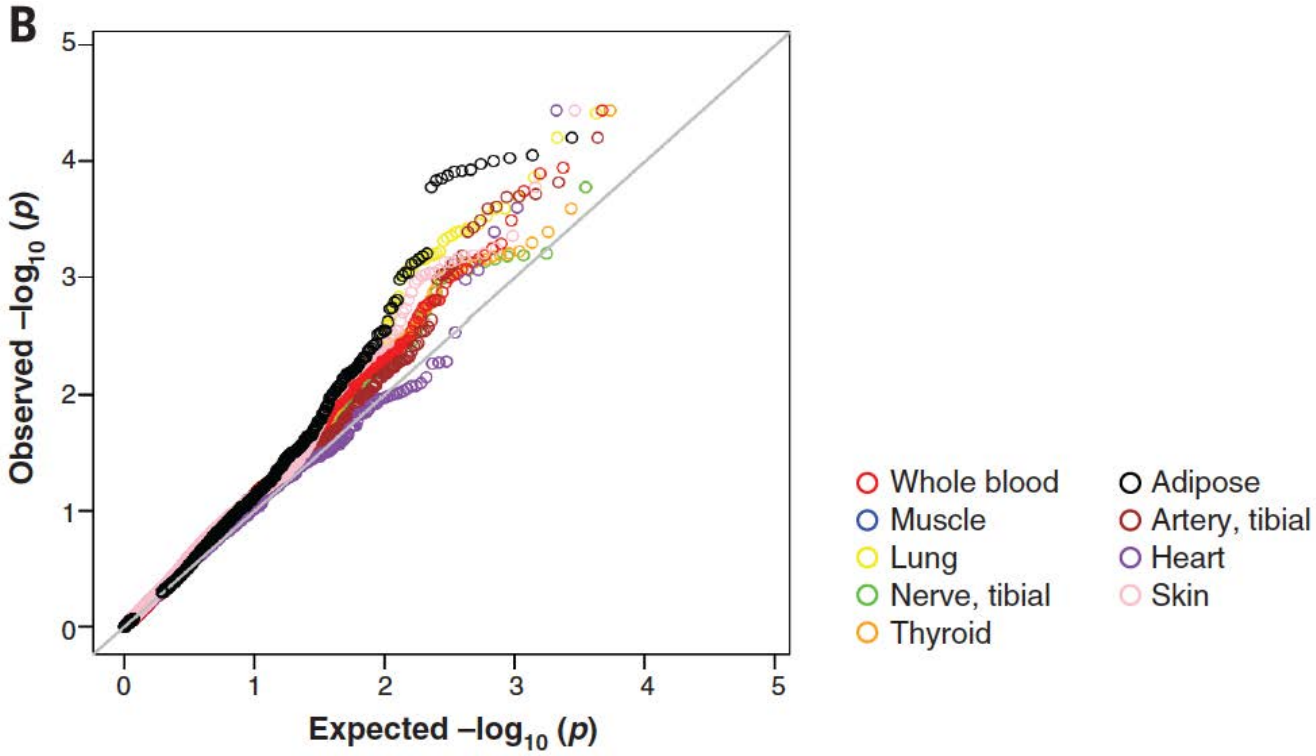




Science 348:6235 648

Genotype-tissue expression (GTEx)

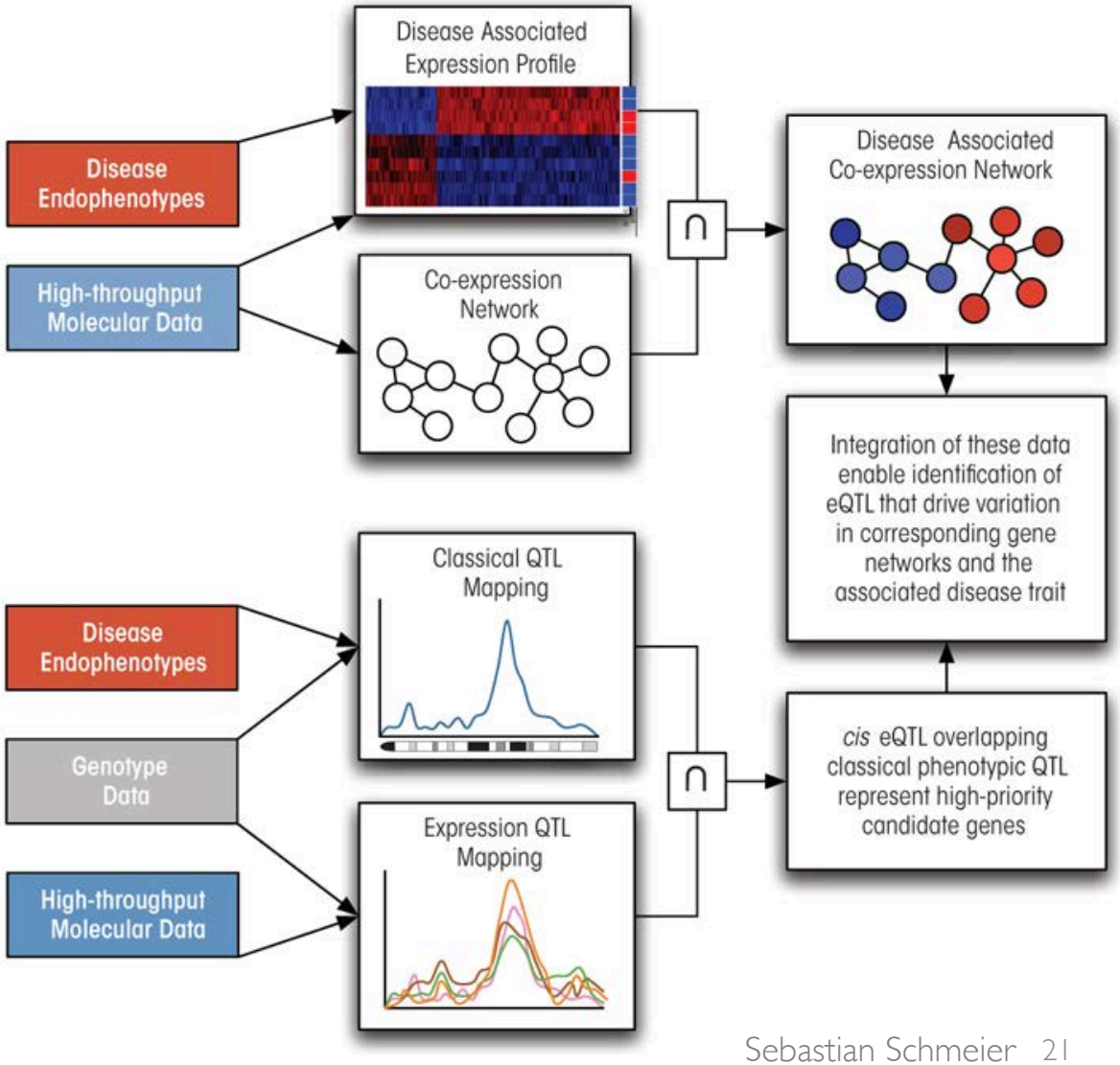
- Enrichment for eQTLs for disease associations is tissue-dependent →the example shows hypertension



Science 348:6235 648

Integration of expression and eQTLs to identify gene networks and candidate genes for complex traits

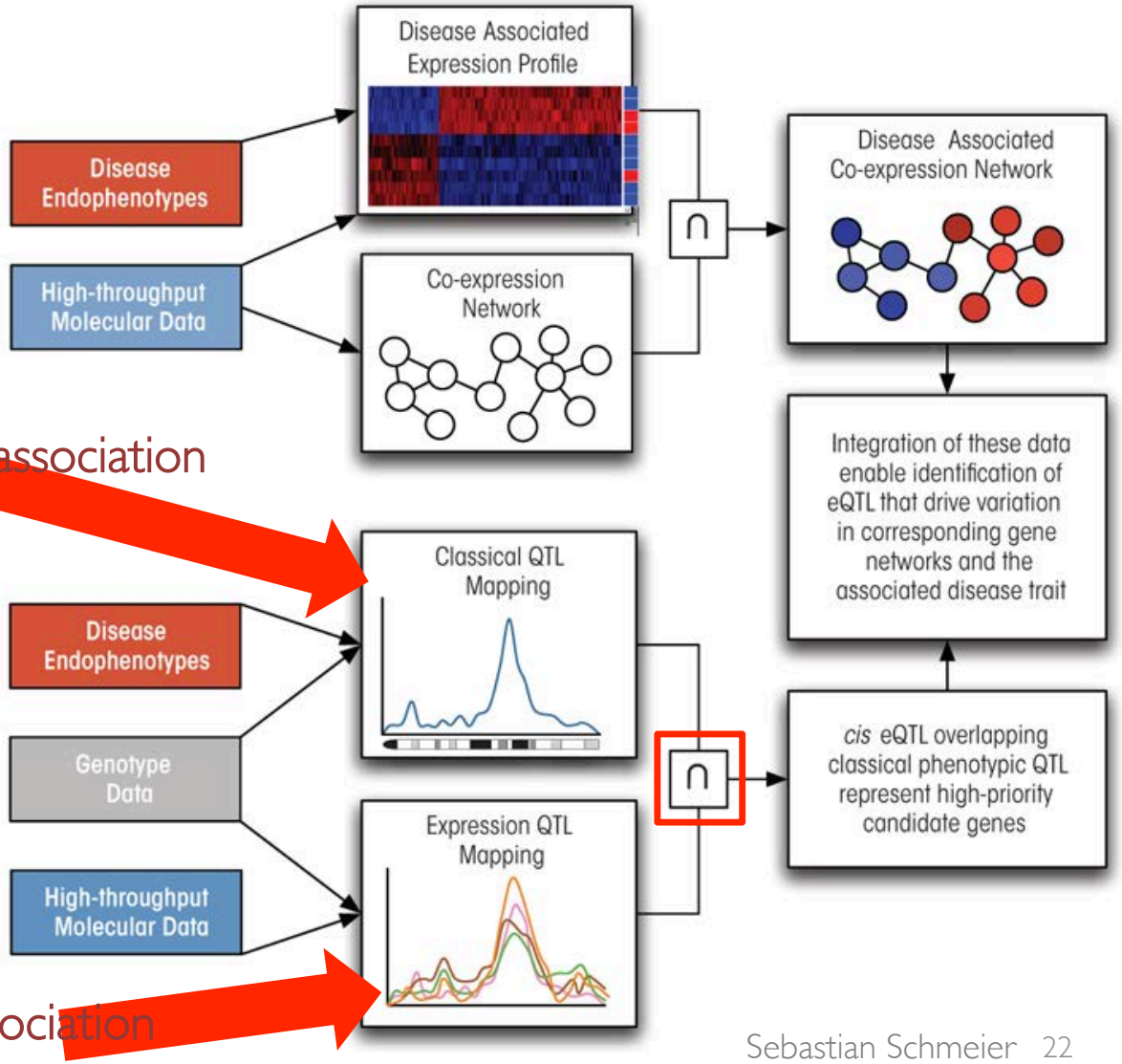
- Example of how such an approach might look like.



http://pubs.niaaa.nih.gov/publications/arcr343/306-317.htm

Integration of expression and eQTLs to identify gene networks and candidate genes for complex traits

- Example of how such an approach might look like.



Variation to trait association

Importantly, eQTLs allow for associating known disease-related SNPs to be associated to regulated genes

Variation to gene association

http://pubs.niaaa.nih.gov/publications/arcr343/306-317.htm

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Questions?

Miles, C. & Wayne, M. (2008) Quantitative trait locus (QTL) analysis. *Nature Education 1(1):208*

The GTex consortium. (2015). The Genotype-Tissue Expression (GTex) pilot analysis: Multitissue gene regulation in humans. *Science 348:6235 648*

Sebastian Schmeier
s.schmeier@gmail.com
http://sschmeier.github.io/bioinf-workshop/

THE ENGINE OF THE NEW NEW ZEALAND