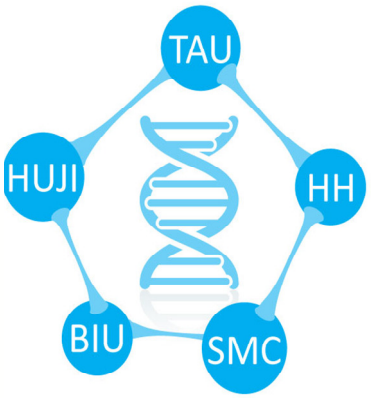
Chromatin States:
A quantitative genetics perspective

Irit Gat-Viks
Department of Cell research and Immunology
Life Sciences, Tel-Aviv University



The genetic basis of disease

Finding genes influencing a disease

Experimental Perturbations

Knockdown mice

↓

Phenotype

Natural Polymorphisms

... ATGATCGATACTCGATAACCCGA ...
... ATGATCGATACGCGATAACCCGA ...

SNP

Exploit natural polymorphisms to understand disease mechanisms

Genetic variant

Genome-wide association studies (GWAS)

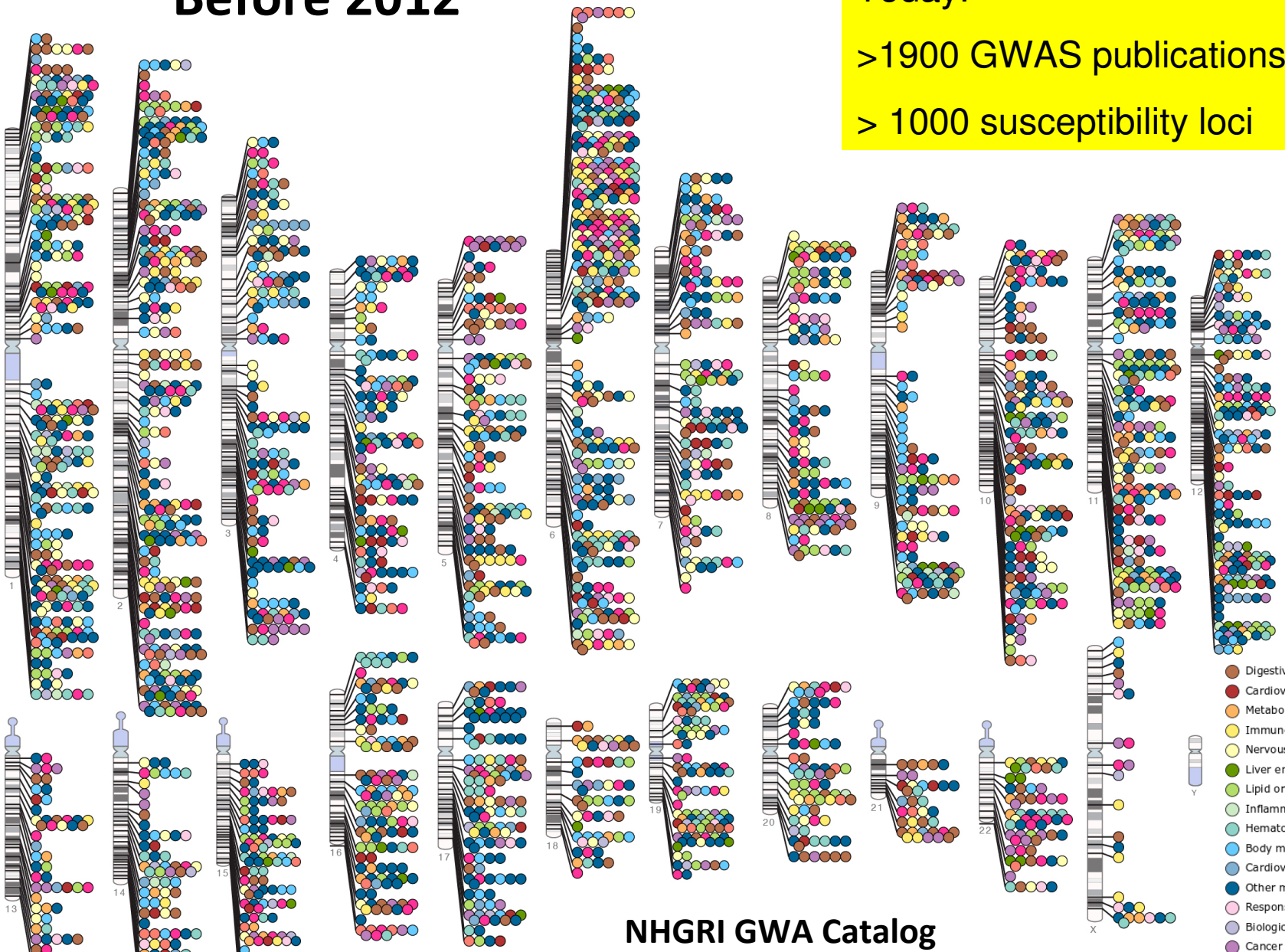
Complex phenotype

Before 2012

Today:

>1900 GWAS publications

> 1000 susceptibility loci



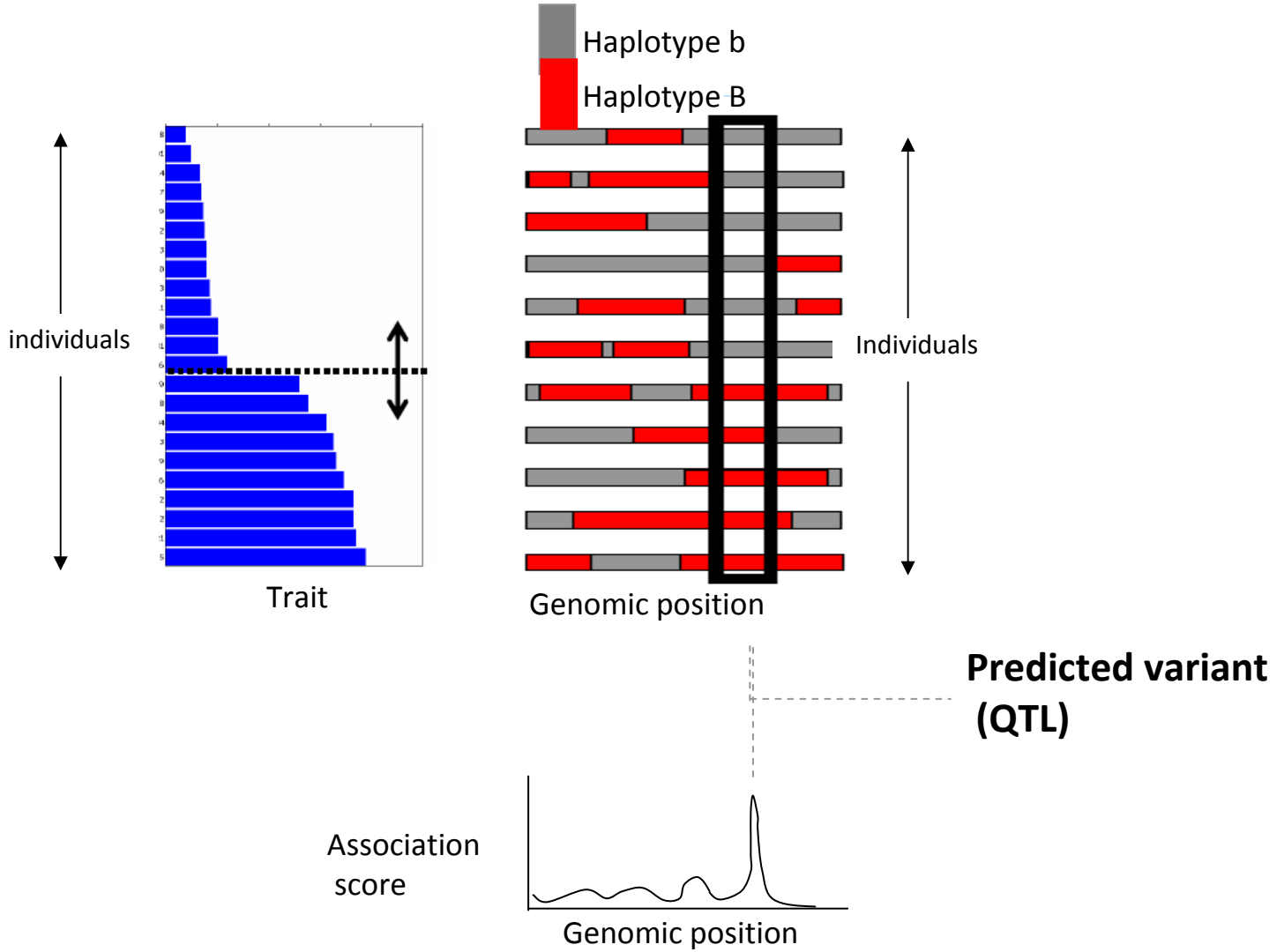
NHGRI GWA Catalog

www.genome.gov/GWAStu

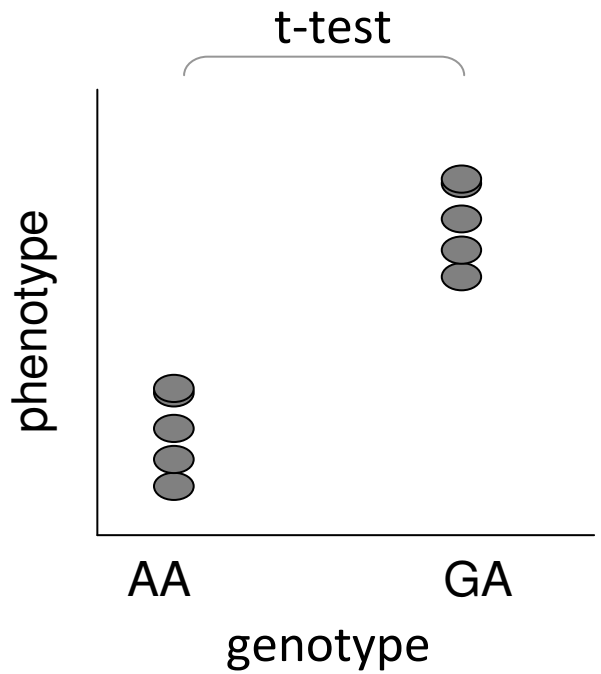
National Human Genome Research Institute

EMBL-EBI

Association study for a quantitative trait



QTL analysis: Standard linear models



QTL analysis: Standard linear models

1.ANOVA model

$$x_{ij} = \mu_i + e_{ij}$$

x_{ij} - Phenotype of j-th individual of marker genotype i

μ_i - Effect of marker i

e_{ij} - Residual error: the deviation of the jth individual from the expected value of the ith marker, E(eij)=0. Var(eij)=σ^2

The presence of a linked QTL is indicated by a significant between-marker variance

2. Multiple regression model $x_j = \mu + \sum_{i=1}^{n} b_i g_{ij} + e_j$

x_j - Phenotype of jth individual

g_{ij} - Indicator variables (one for each marker genotype)

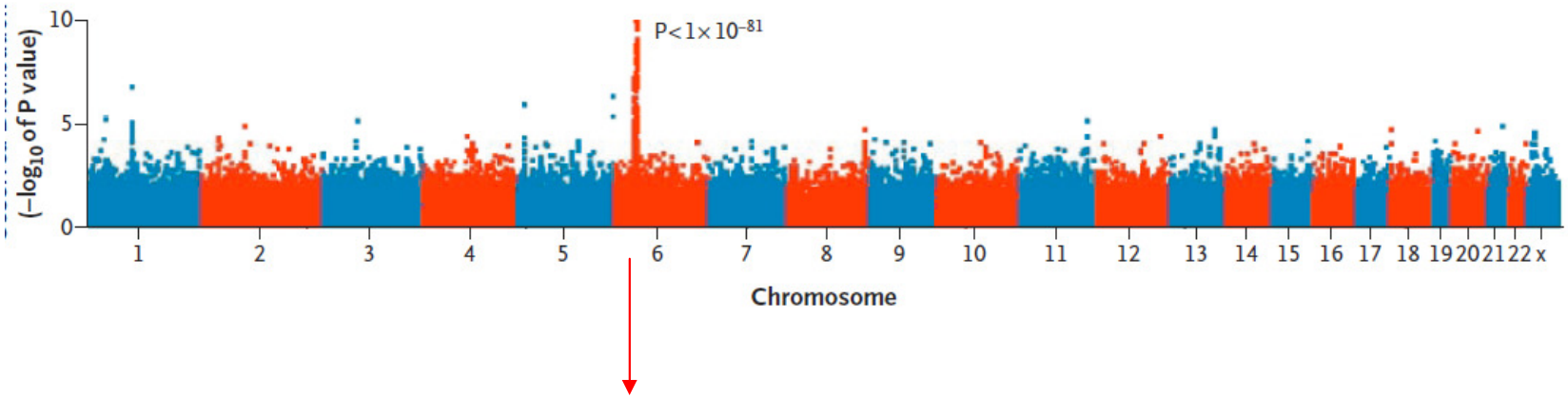
$$g_{ij} = \begin{cases} 1 & \text{if individual j has marker genotype i} \\ 0 & \text{otherwise} \end{cases}$$

e_j - Residual error

The presence of a linked QTL is indicated by a significant fraction of character variance accounted for by the marker genotype

Risk Alleles for Multiple Sclerosis

Manhatten plot



MHC region

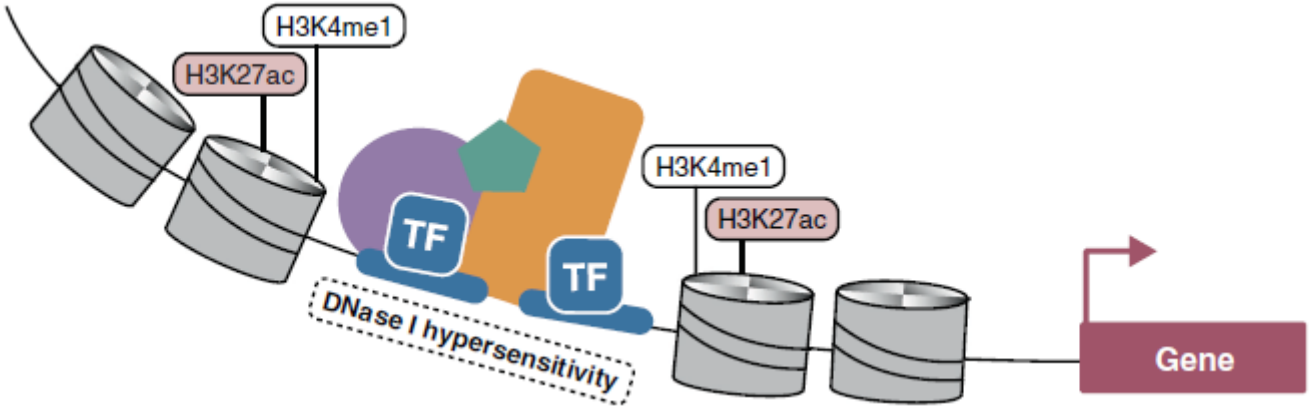
N Engl J Med 2007

A quantitative genetics perspective

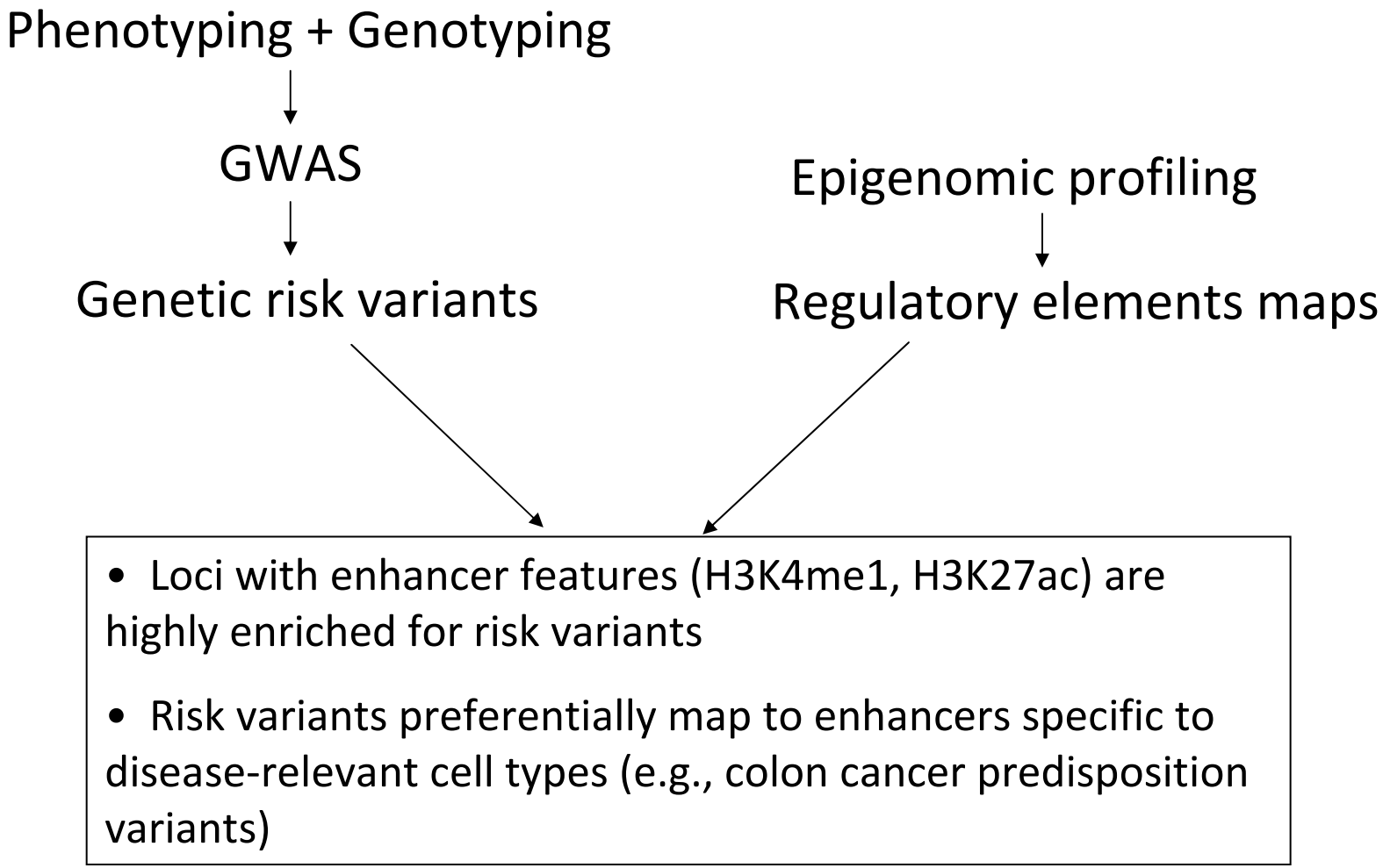
Constructing **regulatory elements maps** using epigenomic profiling

For example, revealing **enhancers**:

- The locations of enhancer elements coincide with DNase I hypersensitive regions of open chromatin flanked by nucleosomes marked with H3K4me1/2.
- H3K27ac and H4K16ac are associated with active chromatin.
- H3K27me3 and H3K9me3 are associated with repressed chromatin.

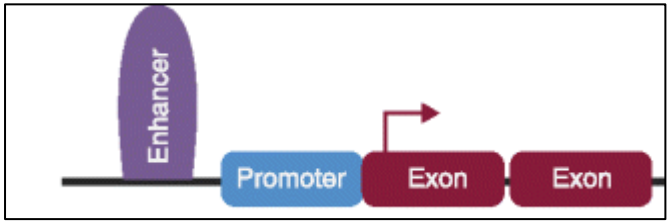


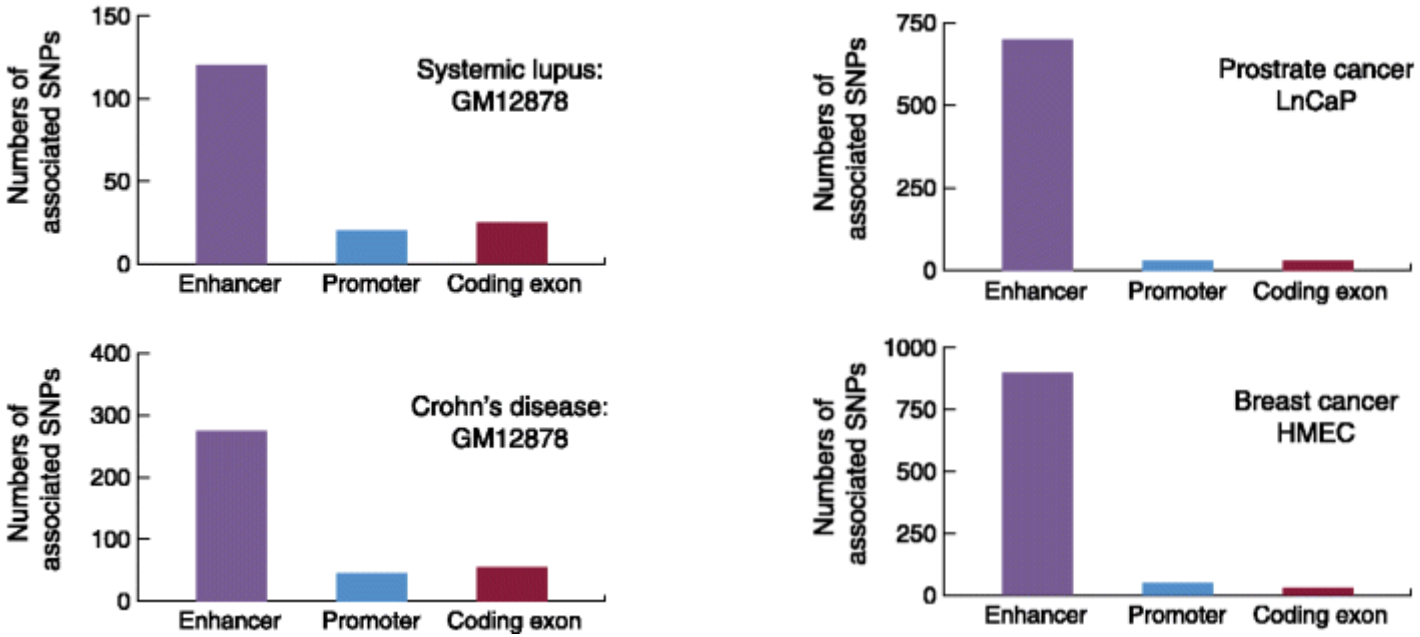
Interpreting susceptibility loci using epigenomic profiling



Corradin et al. genome medicine 2014

Enrichment of genome-wide association study variants in putative enhancer elements





The challenge: what is the **role** of non-coding variants in gene expression?

Corradin et al. genome medicine 2014

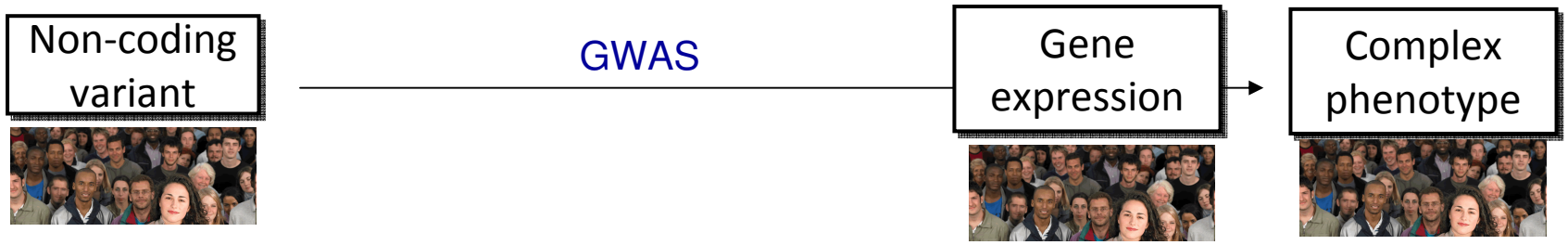
Understand the role of non-coding variants in gene expression

Non-coding variant

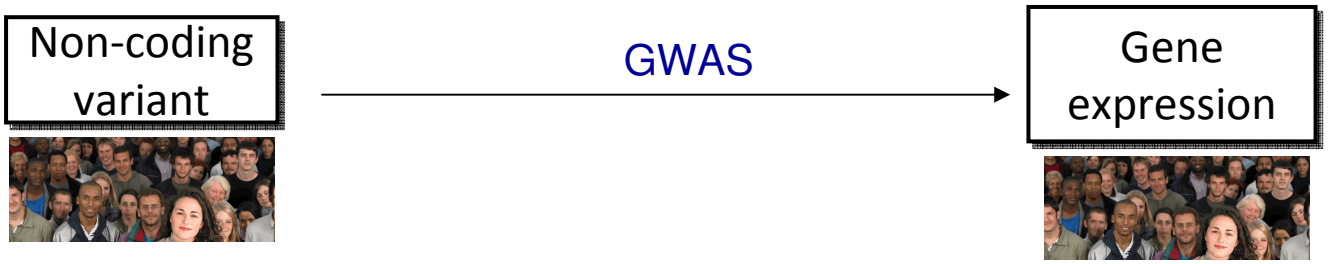
Genome-wide association studies (GWAS)

Complex phenotype

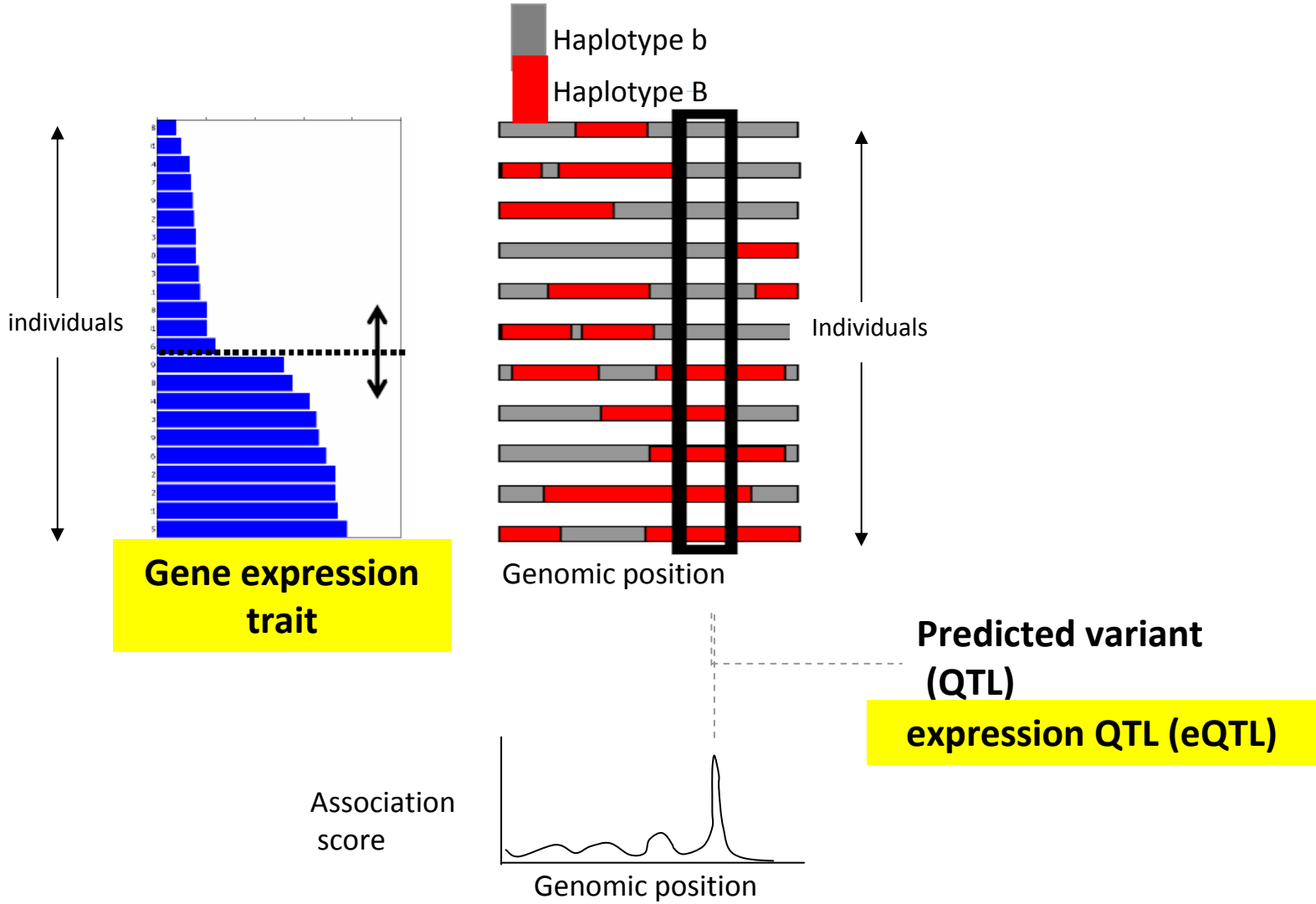
Understand the role of non-coding variants in gene expression



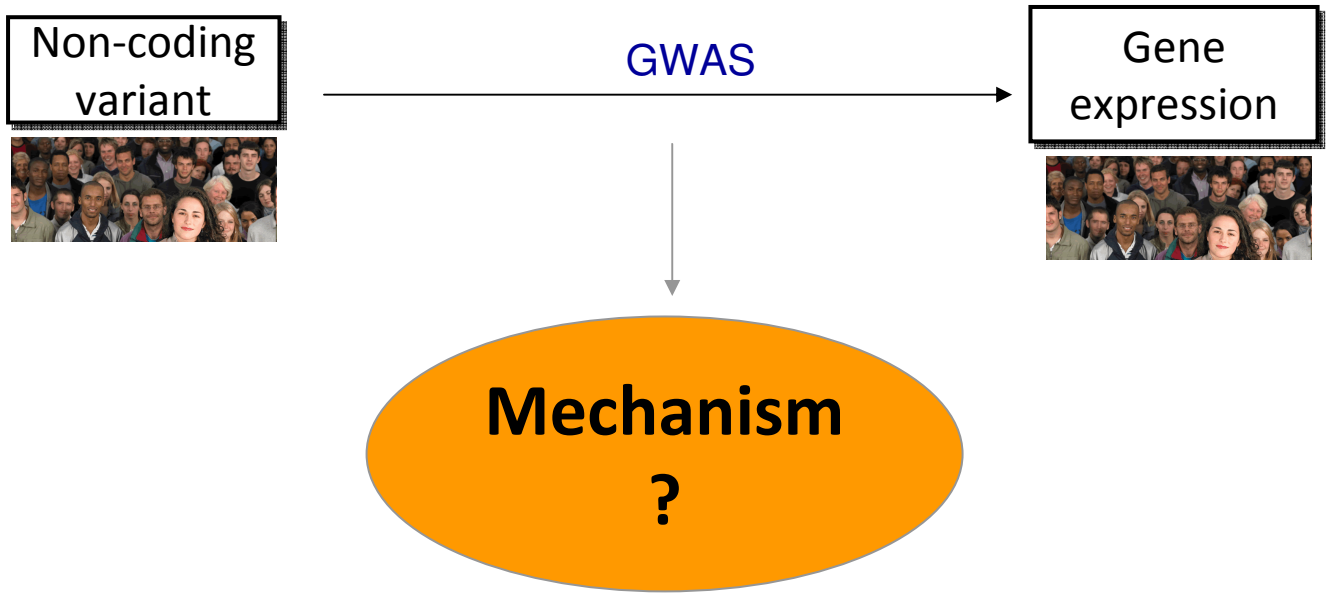
Understand the role of non-coding variants in gene expression



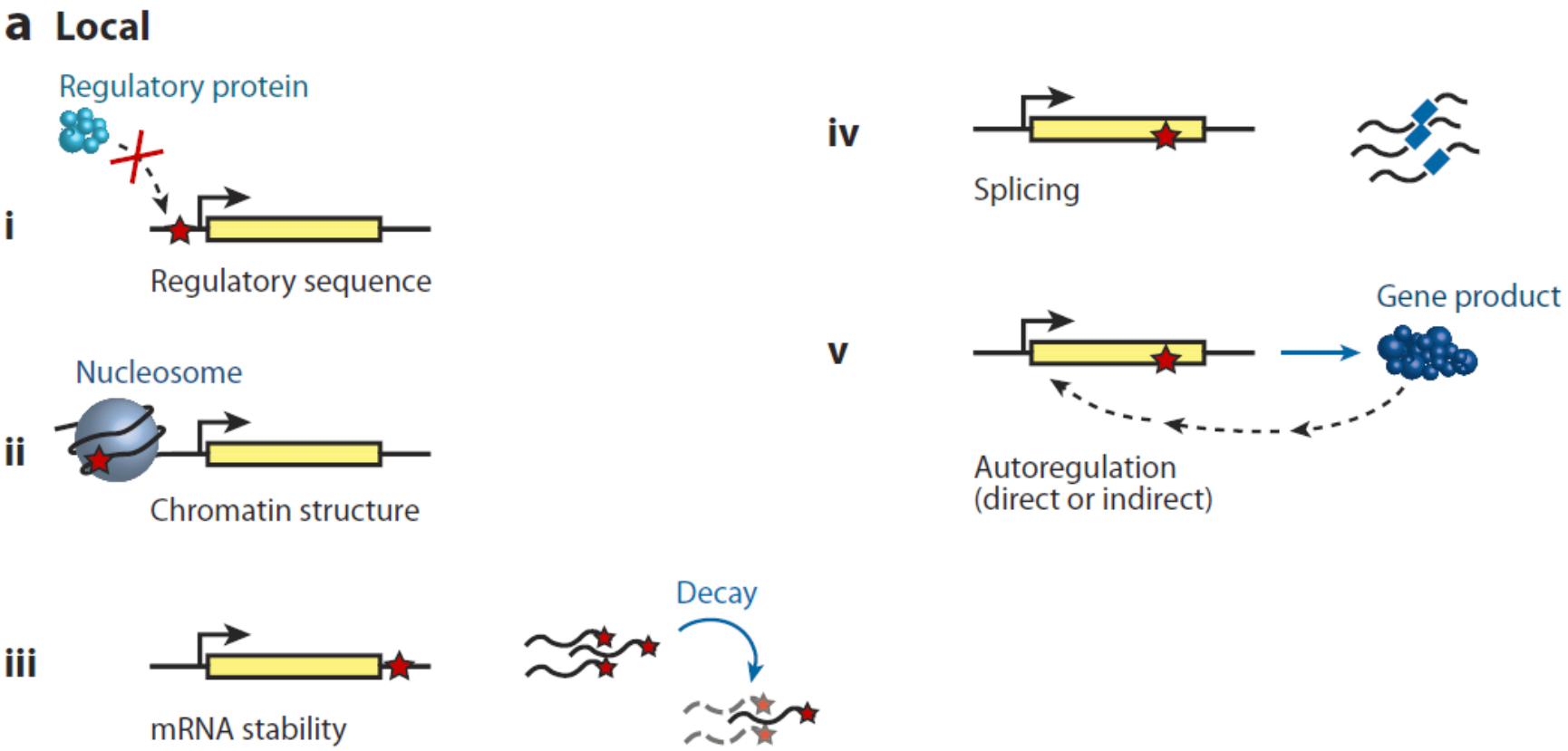
Understand the role of non-coding variants in gene expression



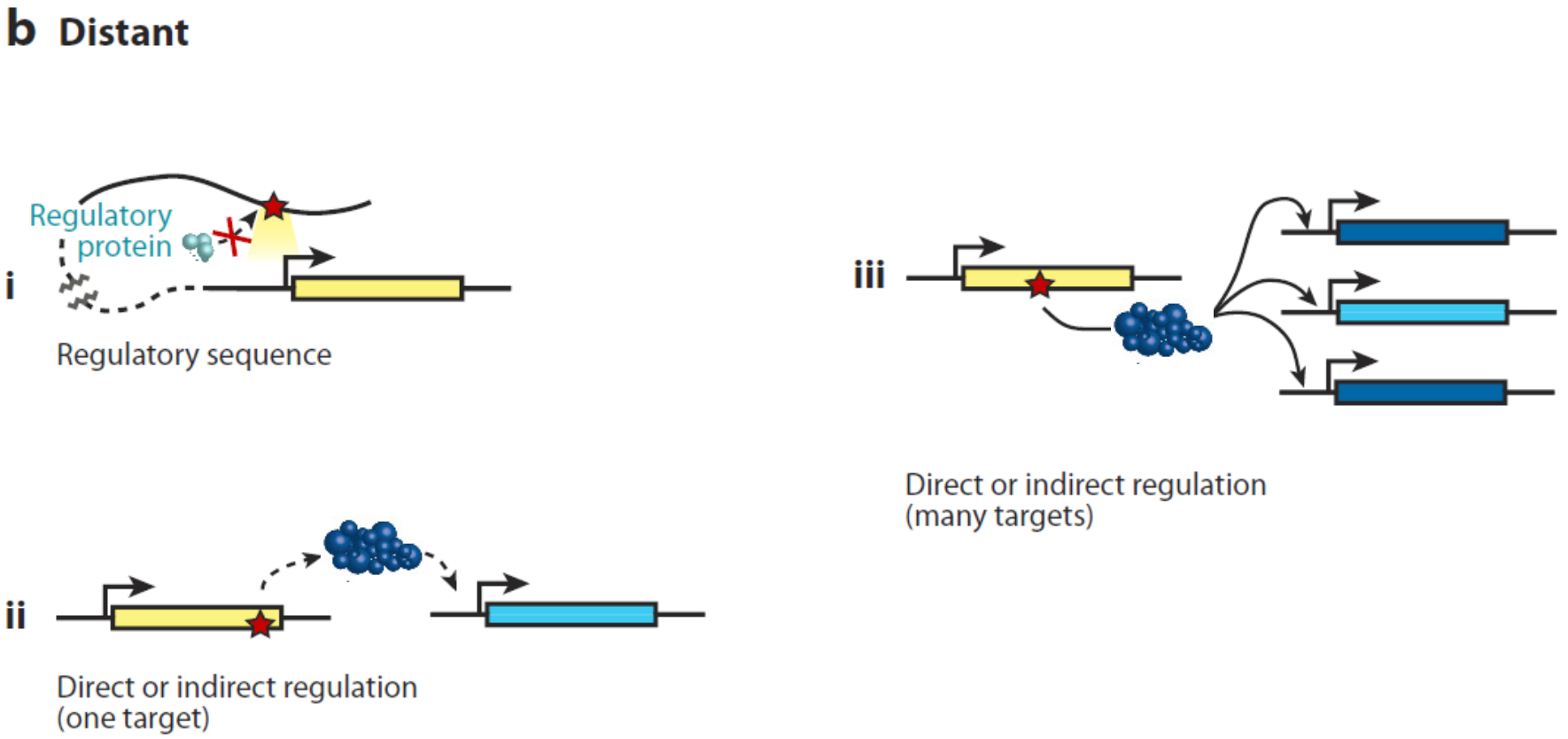
Understand the role of non-coding variants in gene expression



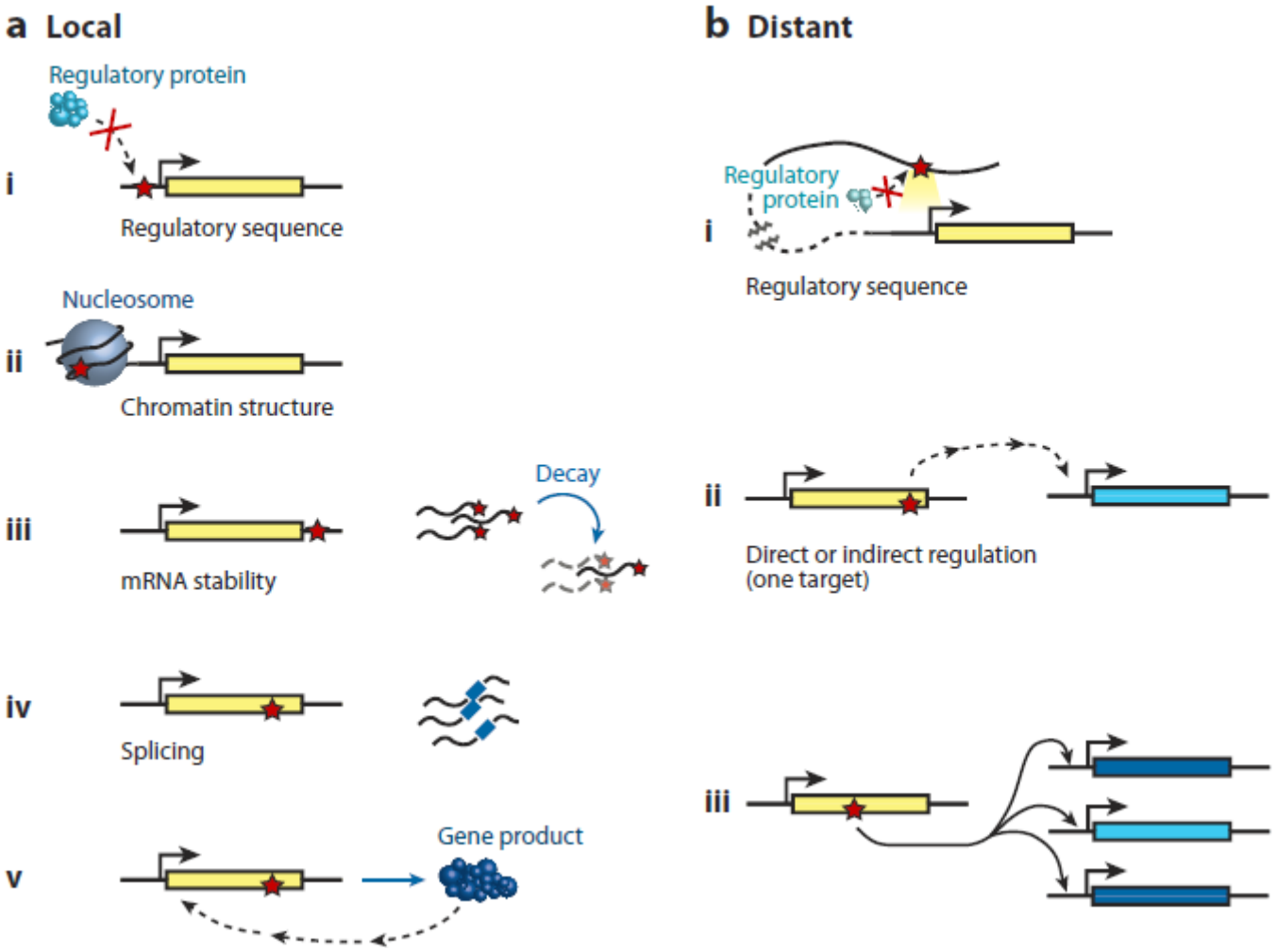
Understanding the role of variants in gene expression based on their genomic positions



Understanding the role of variants in gene expression based on their genomic positions



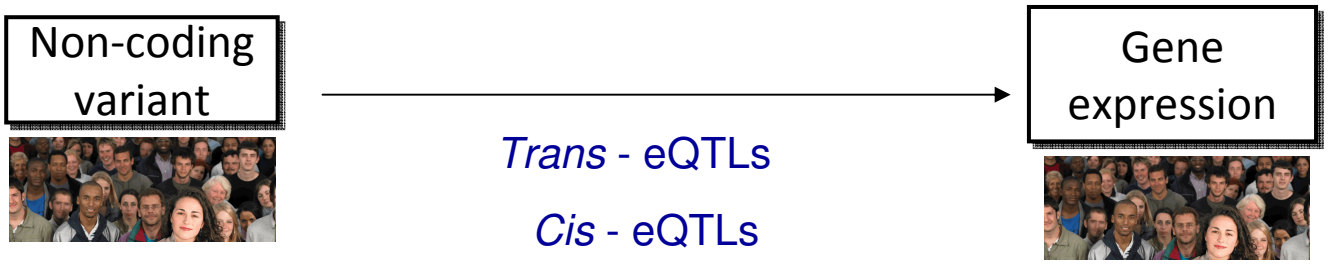
Understanding the role of variants in gene expression based on their genomic positions



Cis-acting eQTL

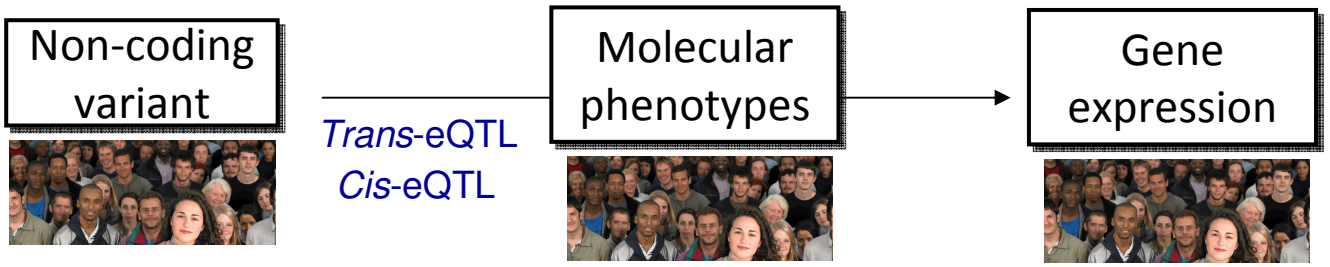
Trans-acting eQTL

Determine the mechanisms by which regulatory variants affect gene expression



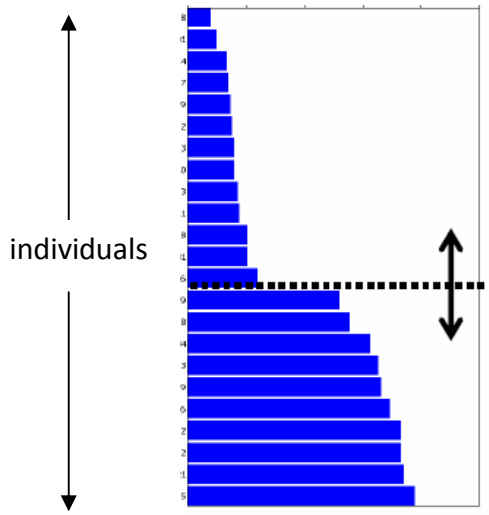
The challenge: eQTL analysis cannot reveal the complete functional mechanism by which non-coding variants influence gene expression

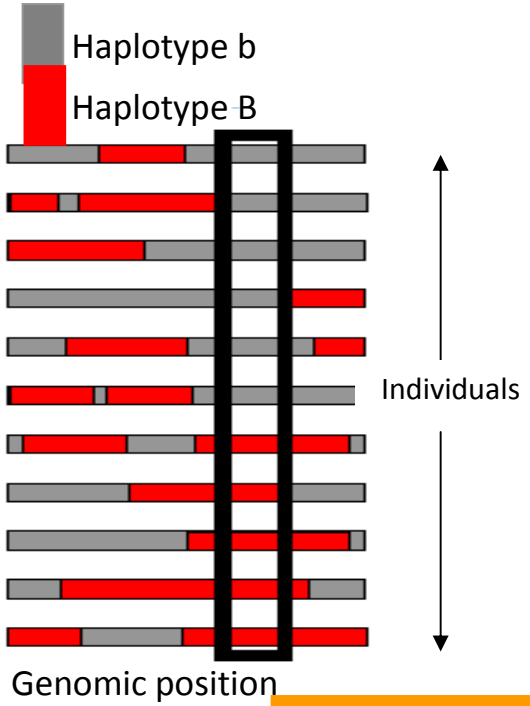
Determine the mechanisms by which regulatory variants affect gene expression



TF binding,
Histone marks,
DNase I sensitivity

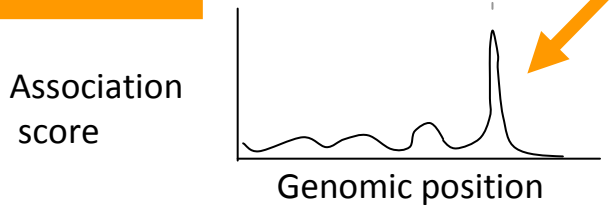
Association study for a quantitative trait





- Gene expression traits
- Open chromatin traits
- Histone modification traits

- *Cis*-association
- *Trans*-association



Example 1: Genetic landscape of open chromatin in yeast

Example 1: Genetic landscape of open chromatin in yeast

FAIRE-seq

Identify open chromatin peaks (OCRs)

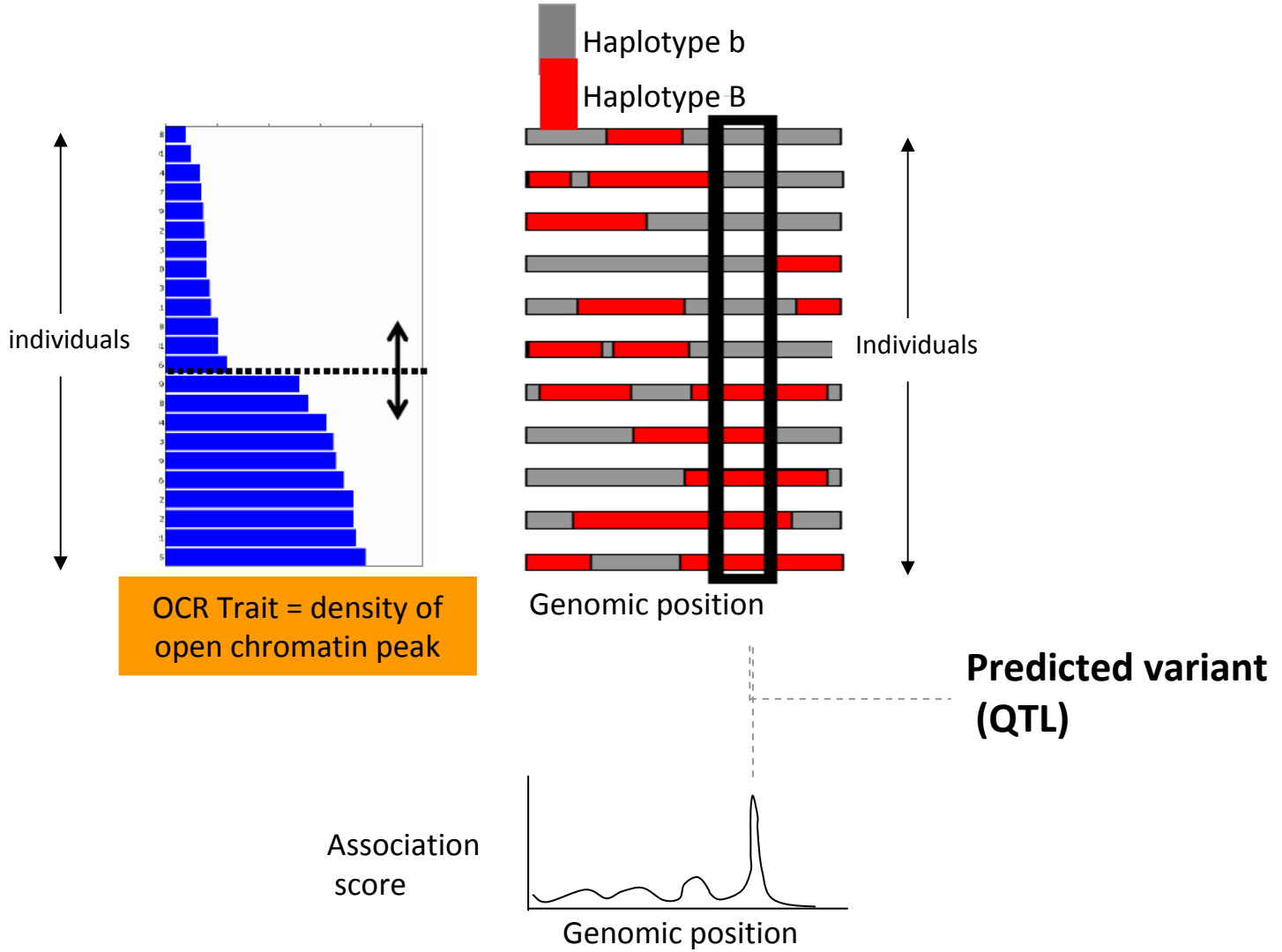
For each peak: QTL analysis of OCR traits

OCR Trait = density of peak

Broader peaks of naked DNA compared to DHS-seq and ATAC-seq

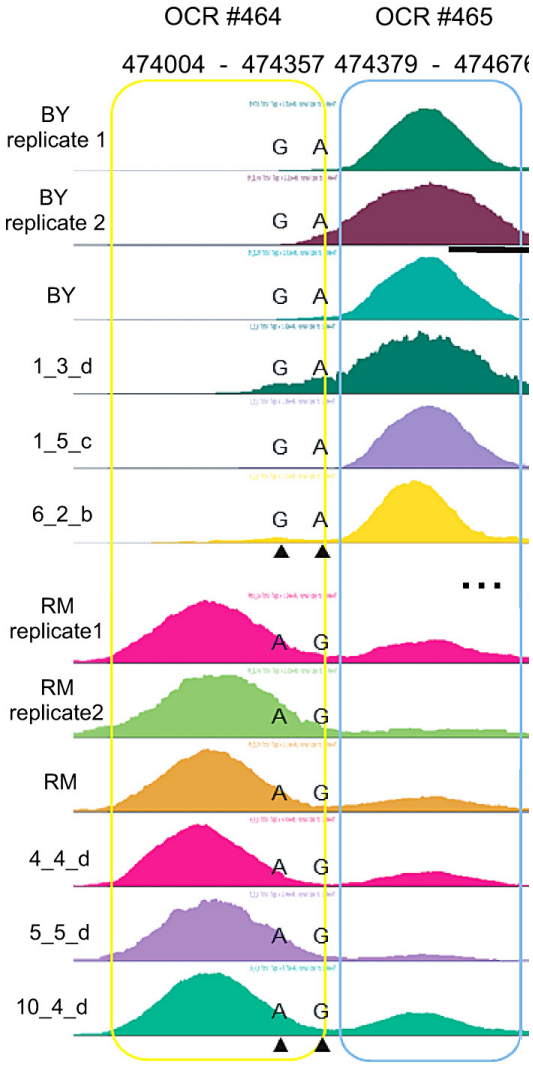
Lee et al. PLOS Genetics 2013

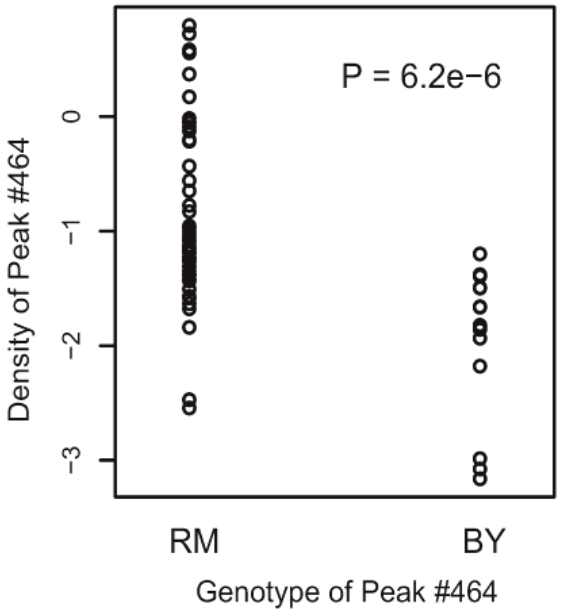
Association study for an open chromatin trait



Lee et al. PLOS Genetics 2013

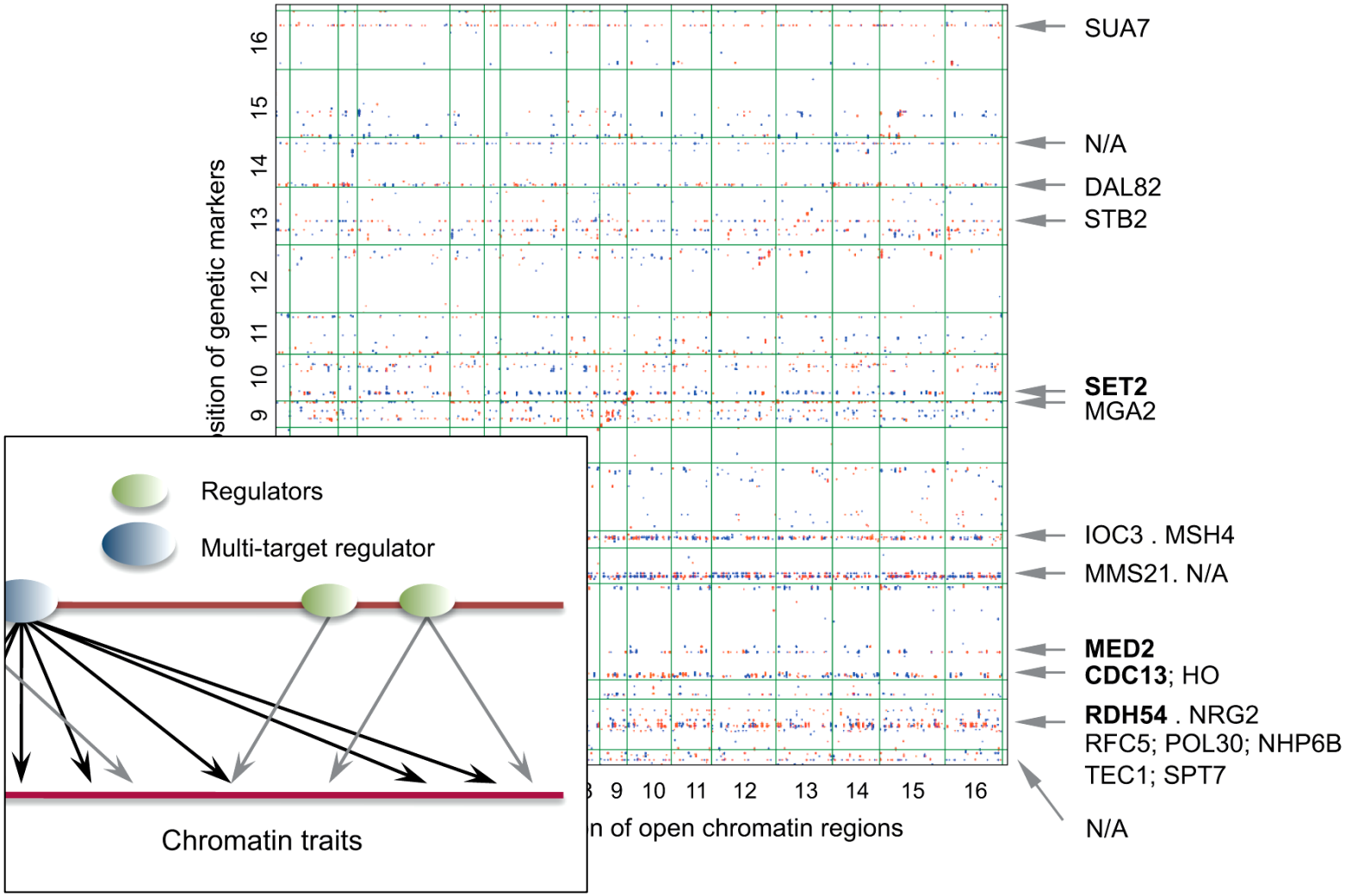
Characterization of *cis*-associations





Lee et al. PLOS Genetics 2013

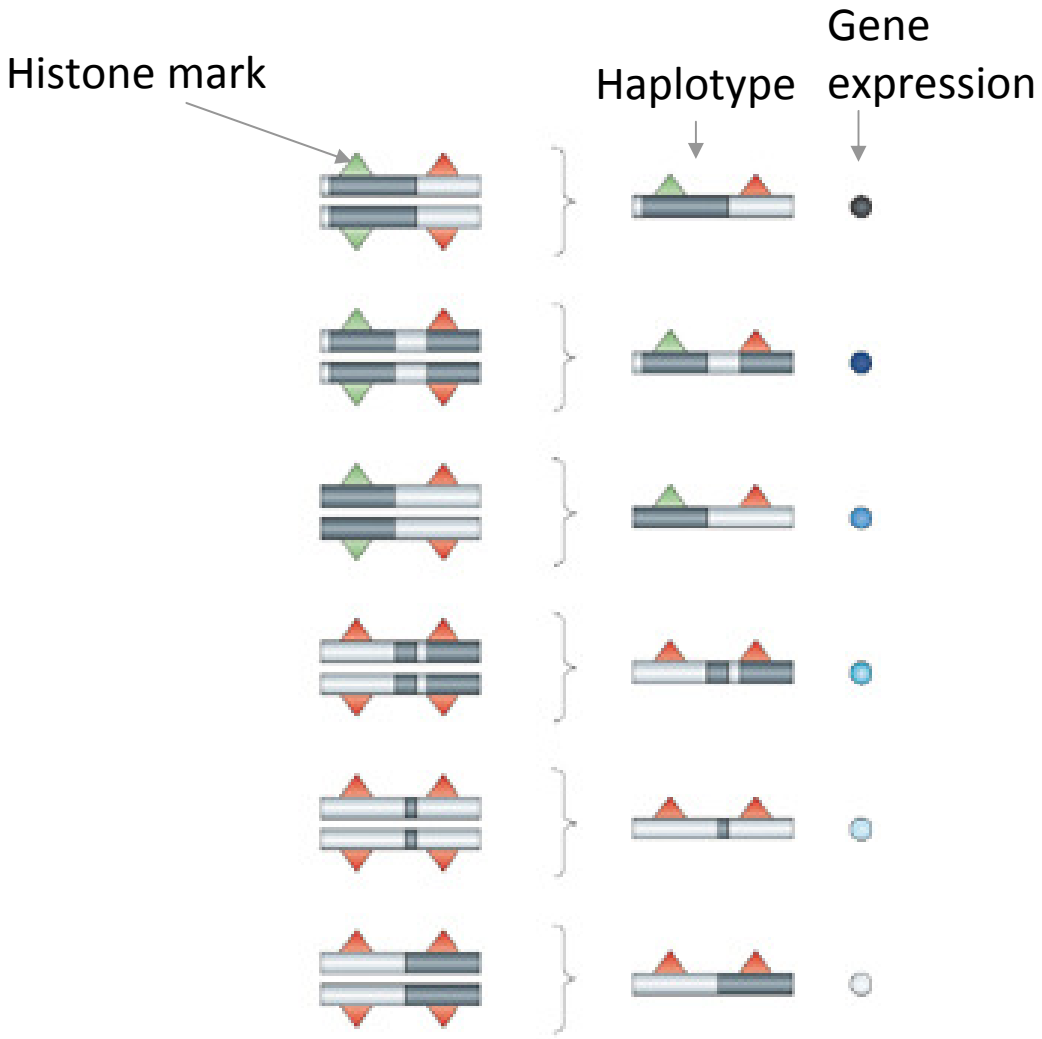
Characterization of *trans*-associations



Lee et al. PLOS Genetics 2013

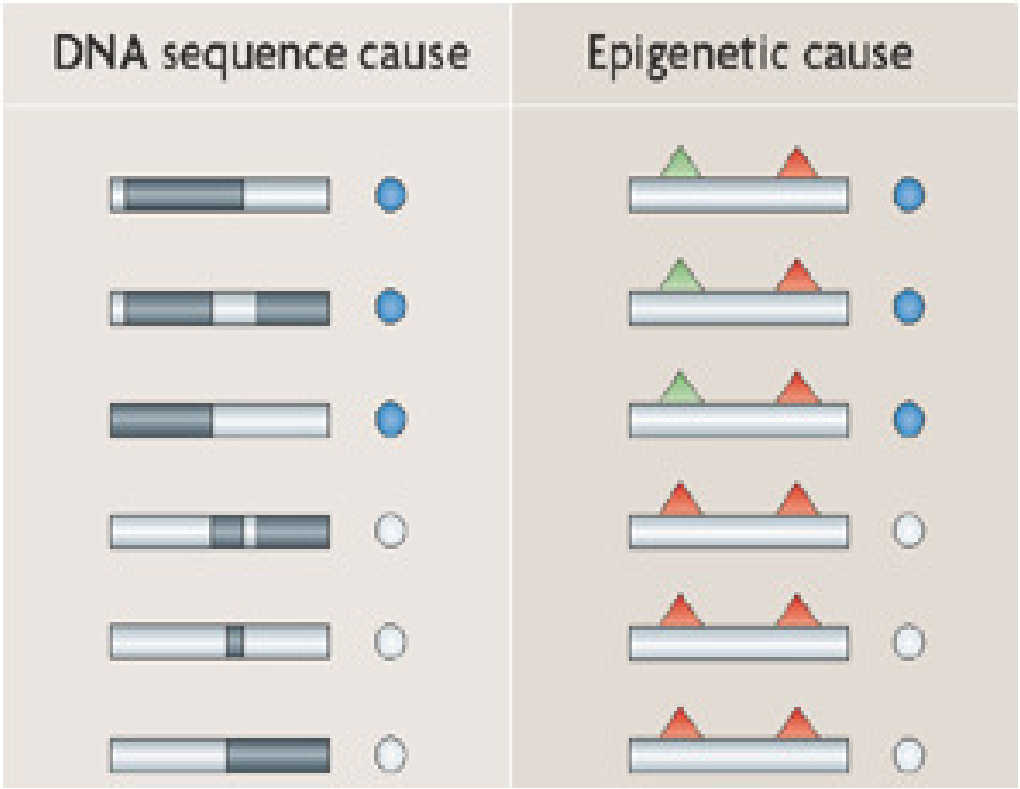
Example 2:
Integrate genetics + histone marks + expression traits

Integrate genetics + histone marks + expression traits



Johannes et al. Nature Reviews Genetics 2008

Integrate genetics + histone marks + expression traits

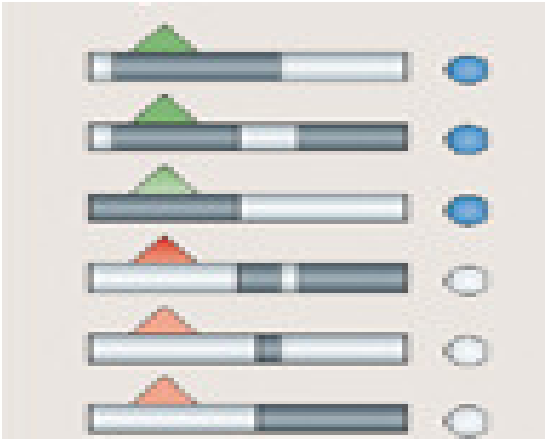


Johannes et al. Nature Reviews Genetics 2008

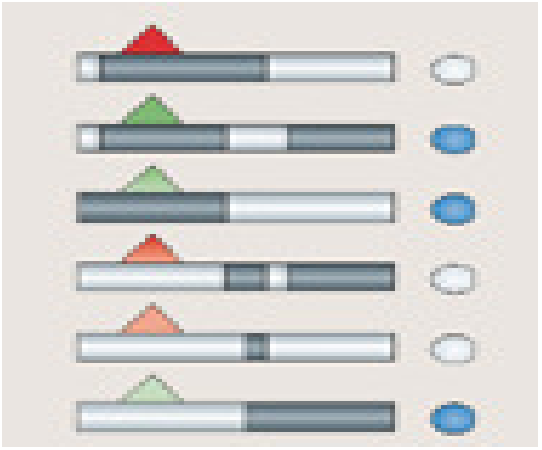
Integrate genetics + histone marks + expression traits

Relationships between histone marks and gene expression

Sequence dependent



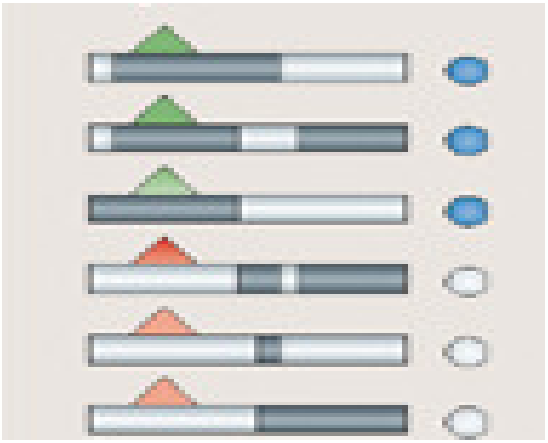
Sequence independent

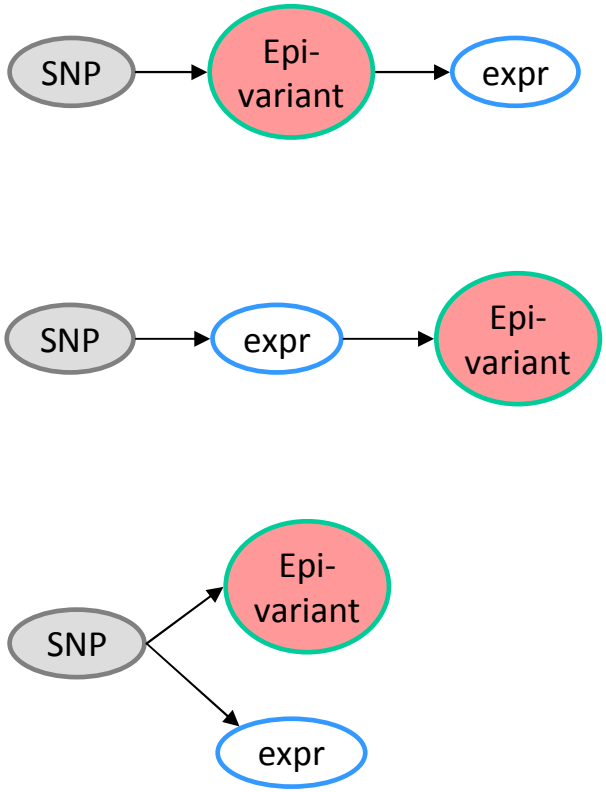


Johannes et al. Nature Reviews Genetics 2008

Integrate genetics + histone marks + expression traits

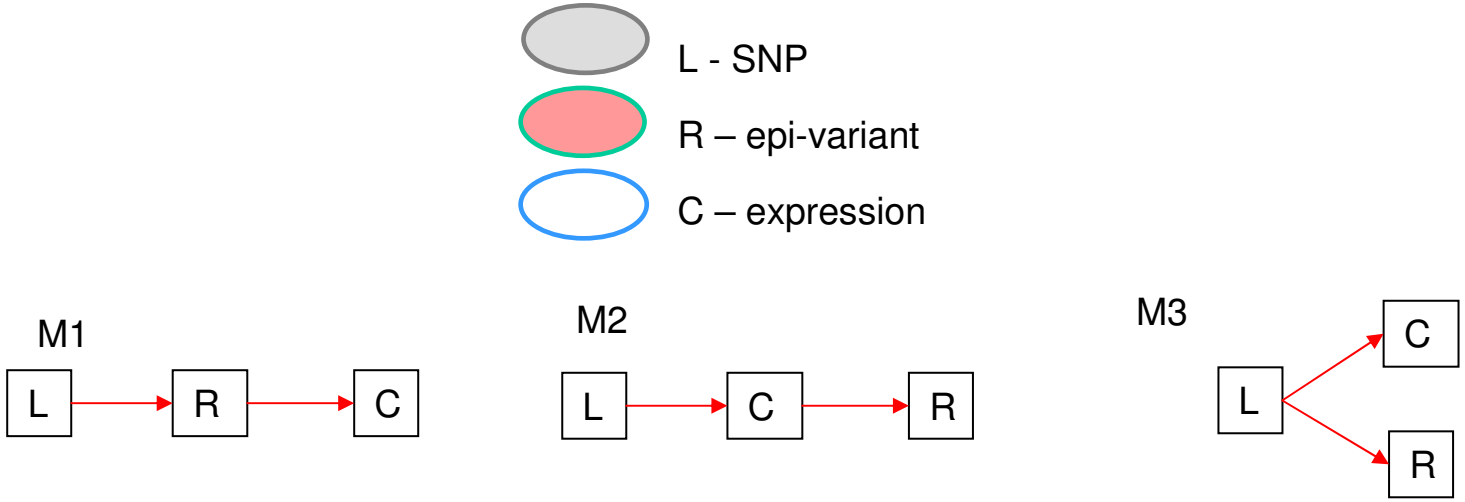
Sequence dependent





Johannes et al. Nature Reviews Genetics 2008

Inferring causal relations



$$\text{M1}.\ P(\text{L},\text{R},\text{C}) = P(\text{L})\,P(\text{R}|\text{L})\,P(\text{C}|\text{R})$$

$$\text{M2}.\ P(\text{L},\text{R},\text{C}) = P(\text{L})\,P(\text{C}|\text{L})\,P(\text{R}|\text{C})$$

$$\text{M3}.\ P(\text{L},\text{R},\text{C}) = P(\text{L})\,P(\text{C}|\text{L})\,P(\text{R}|\text{C},\text{L})$$

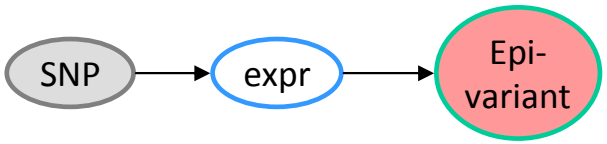
Schadt et al. Nature Genetics 2005

The likelihood for each model over all individuals in the population of interest are given by:

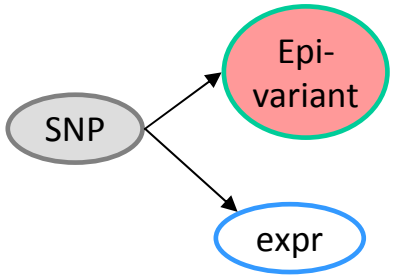
$$likelihood \quad function \quad = L(\theta; M) = p(data \mid \theta_M)$$

$$P(L,R,C \mid \theta_{M1}) = \prod_{i=1}^{n} P(L_i) \cdot P(R_i \mid L_i) \cdot P(C_i \mid R_i)$$

$$P(L,R,C \mid \theta_{M2}) = \prod_{i=1}^{n} P(L_i) \cdot P(C_i \mid L_i) \cdot P(R_i \mid C_i)$$



$$P(L,R,C \mid \theta_{M1}) = \prod_{i=1}^{n} P(L_i) \cdot P(R_i \mid L_i) \cdot P(C_i \mid L_i)$$



Schadt et al. Nature Genetics 2005

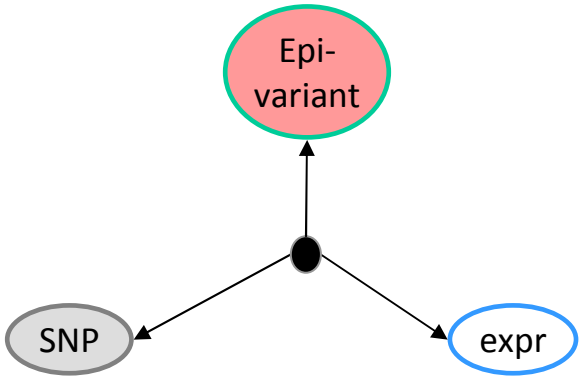
Integrate genetics + histone marks + expression traits

Rintish et al. 2014; 30 Rat BXH/BXB strains, liver and heart

- 18.1% and 14.5% of all H3K4me3 and H3K27me3 QTLs were also eQTL.
- 20% of all eQTL were also QTL for a histone mark.

Degner et al. (2012), 70 Yoruba lymphoblastoid cell lines

- 16% of DNase I sensitivity QTL (dsQTL) were also eQTL
- 23% of eQTL were also dsQTL



Genetic landscape of histone modifications in rat liver and heart

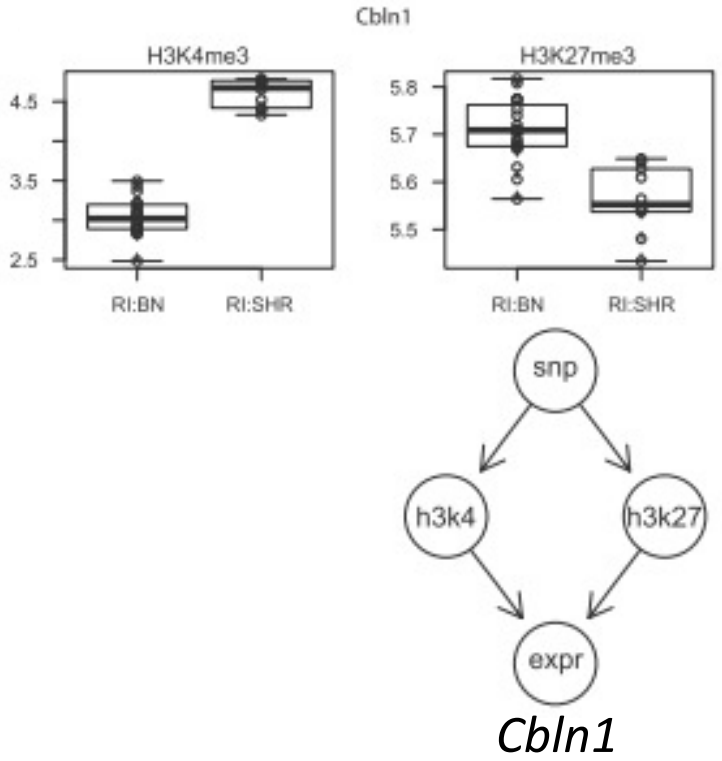
Table 2. HistoneQTL mapping results in heart and liver tissue

	Heart		Liver		Both tissues	
H3K4me3 (traits)	**(25,064)**		**(31,447)**		**(20,076)**	
FDR-cutoff[a]	*cis*	*trans*	*cis*	*trans*	*cis*	*trans*
0.05	2638	2504	2945	1464	931	82
0.01	2024	812	2235	243	698	31
0.001	1414	232	1454	59	467	13
1×10^{-4}	776	30	834	26	232	4
1×10^{-5}	0	0	460	12	0	0
	Heart		**Liver**		**Both tissues**	
H3K27me3 (traits)	**(4,214)**		**(3,776)**		**(2,688)**	
FDR-cutoff[a]	*cis*	*trans*	*cis*	*trans*	*cis*	*trans*
0.05	102	15	196	38	35	1
0.01	74	4	166	7	30	1
0.001	57	2	131	2	23	0
1×10^{-4}	50	1	97	0	18	0
1×10^{-5}	50	1	97	0	18	0

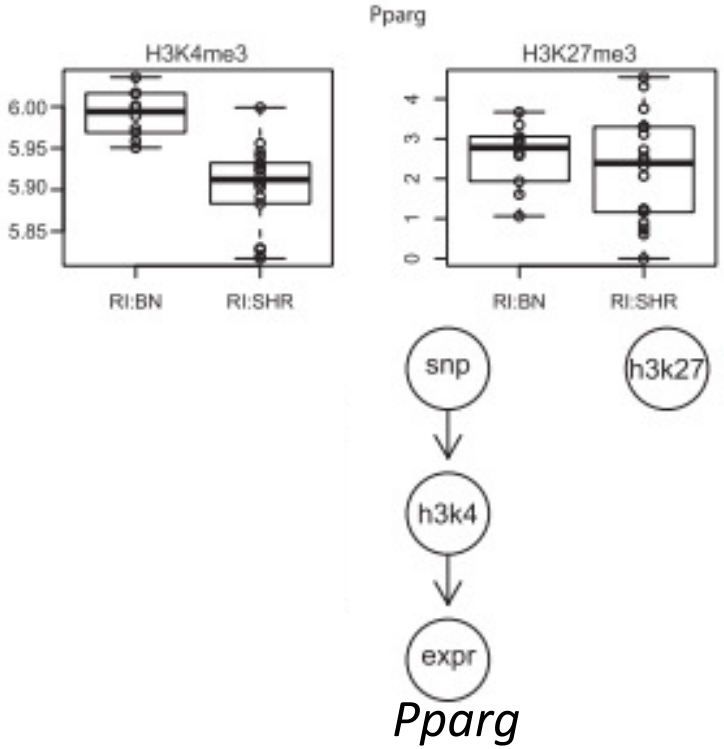
[a]Cutoff for limitation of false discoveries.

Rintisch et al. Genome research 2014

Integrate genetics + histone marks + expression traits



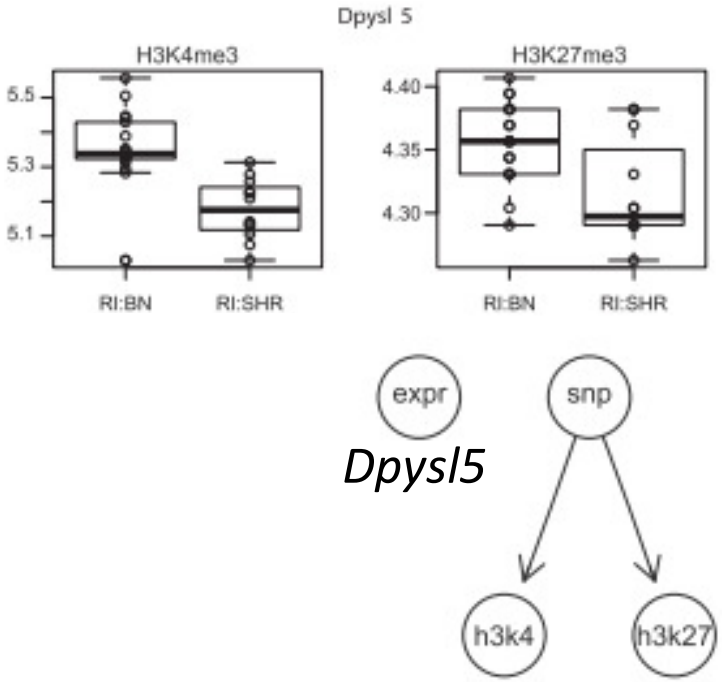
Cbln1



Pparg

Rintisch et al. Genome research 2014

Integrate genetics + histone marks + expression traits

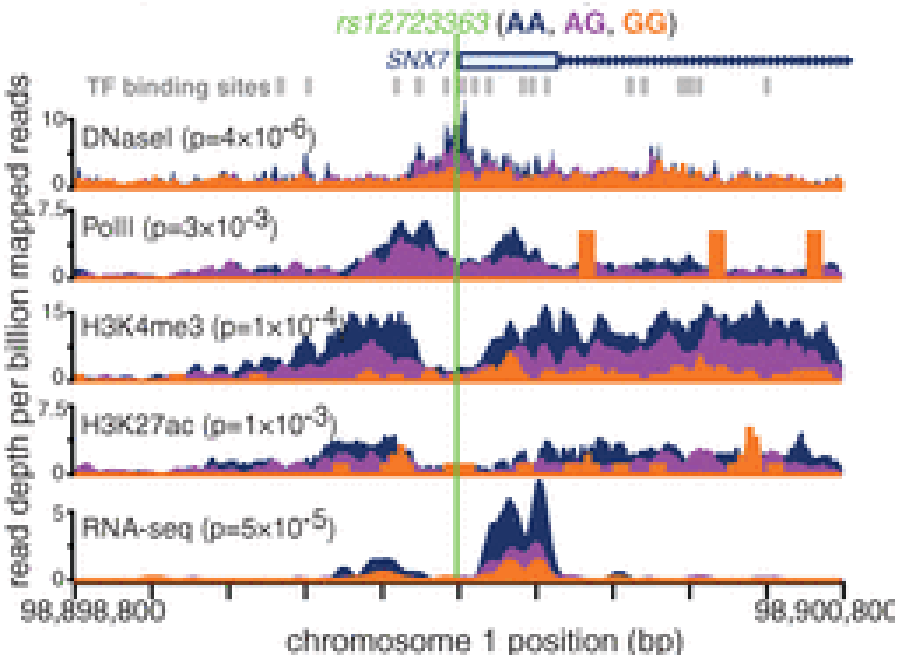


Rintisch et al. Genome research 2014

Example 3: Revealing functionally linked modifications

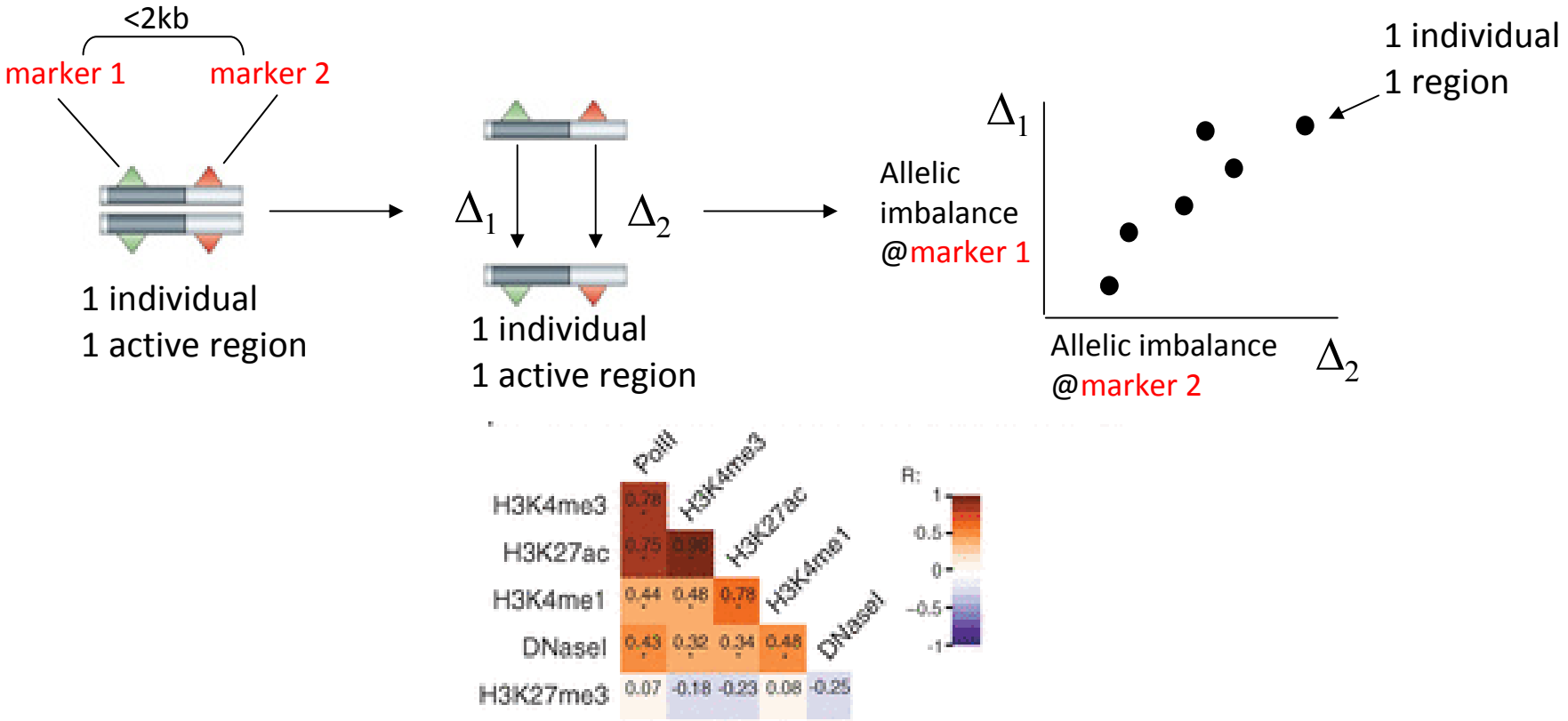
Coordinated change in histone marks along ~2kb regions

Lymphoblastoid cell lines (LCLs) from 70 Yoruba (Nigeria) individuals



McVicker et al. (Pritchard & Gilad labs) Science 2013

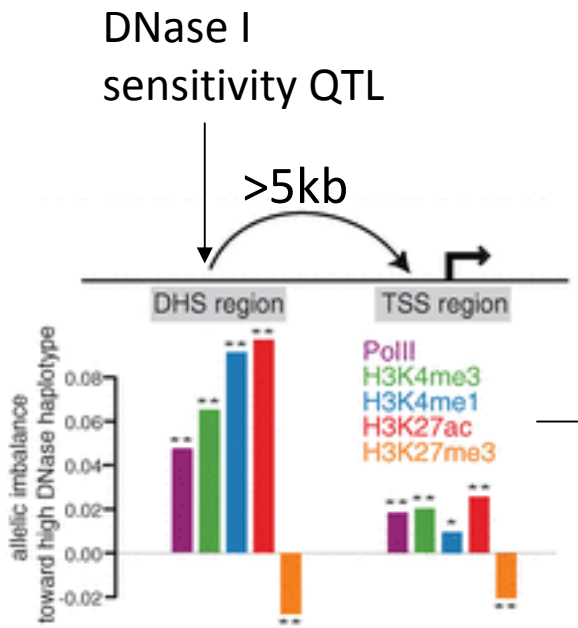
Revealing functionally linked modifications, depending on the same genetic element



Correlation in **allelic imbalance** between histone marks at DNase I sensitive QTL sites (dsQTLs)

McVicker et al. Science 2013

Coordinated change in histone marks between distal (>5kb) regions



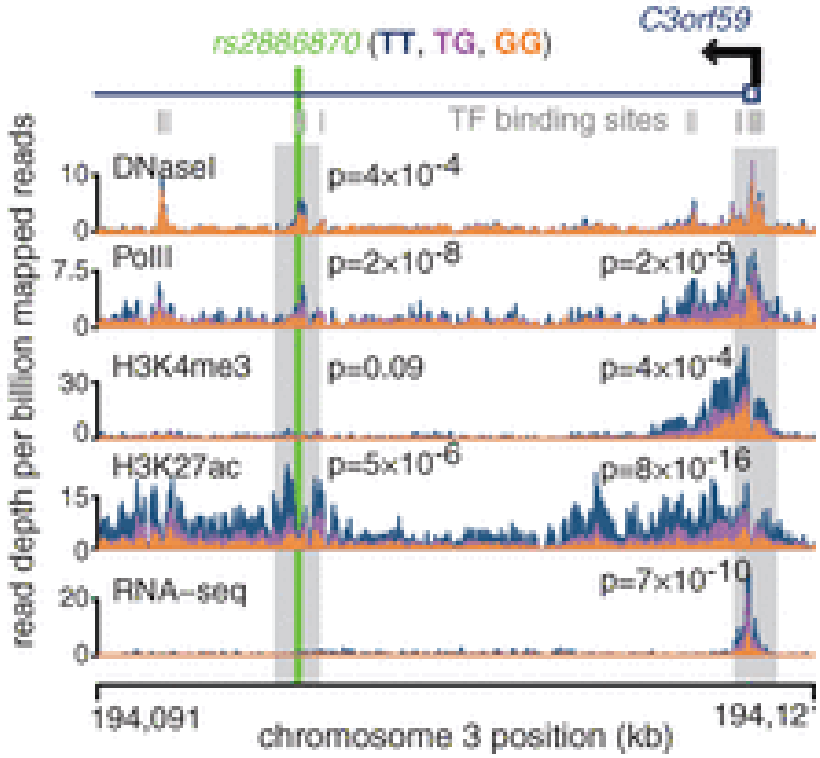
Possibly reflecting the 3D organization of the chromatin

Allelic imbalance across DNase I sensitivity QTLs and eQTLs

ChIP-Seq mark	Functional association
H3K4me3	Active promoters
H3K4me1	Active enhancers
H3K27ac	Active promoters and enhancers
H3K27me3	Inactive chromatin
RNA Pol II	Transcription

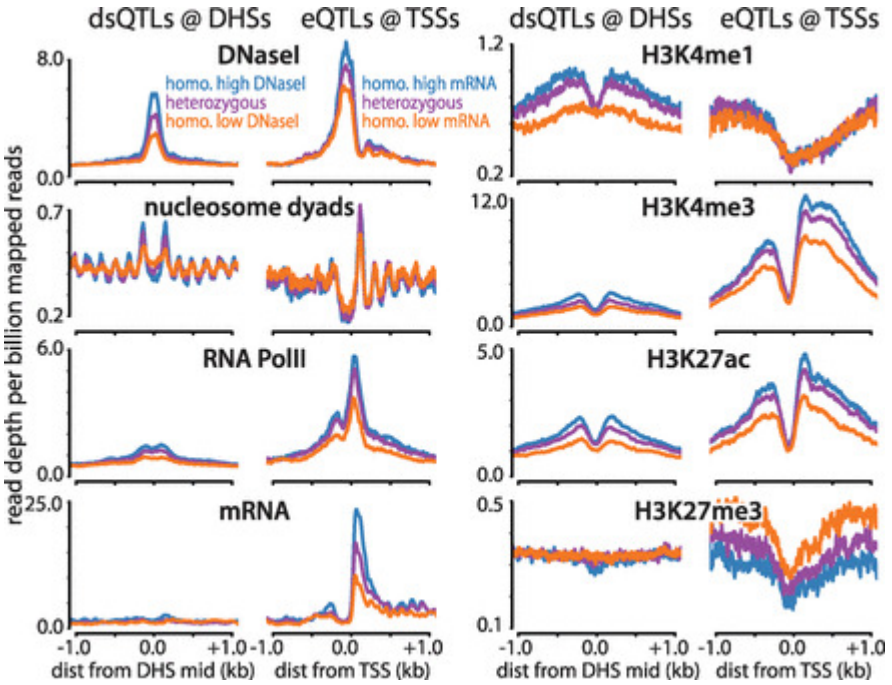
McVicker et al. Science 2013

Coordinated change in histone marks between distal (>5kb) regions



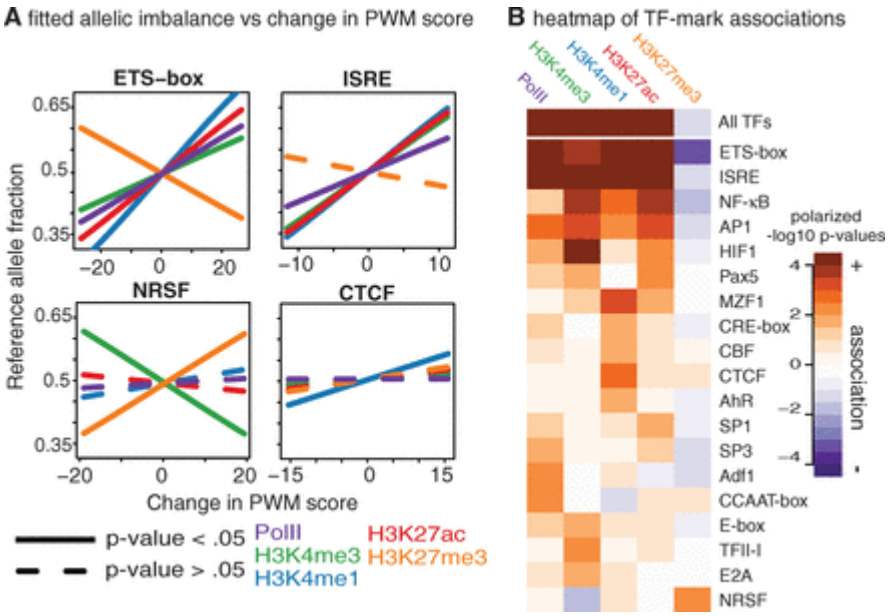
McVicker et al. Science 2013

Coordinated change in histone marks between distal regions



McVicker et al. Science 2013

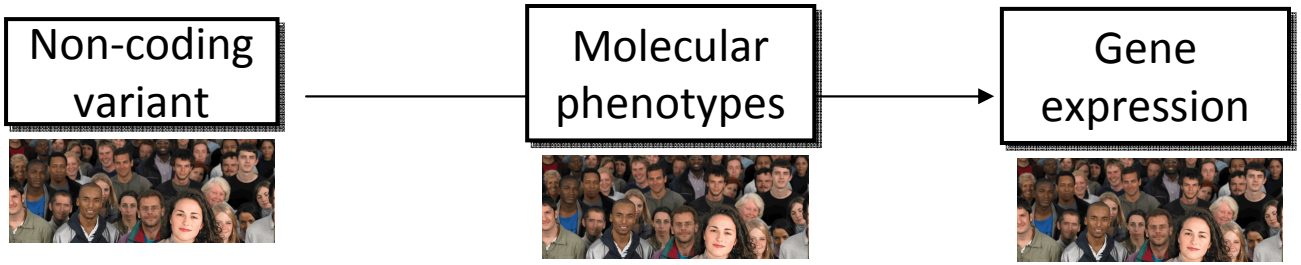
Coordinated change in histone marks between distal regions



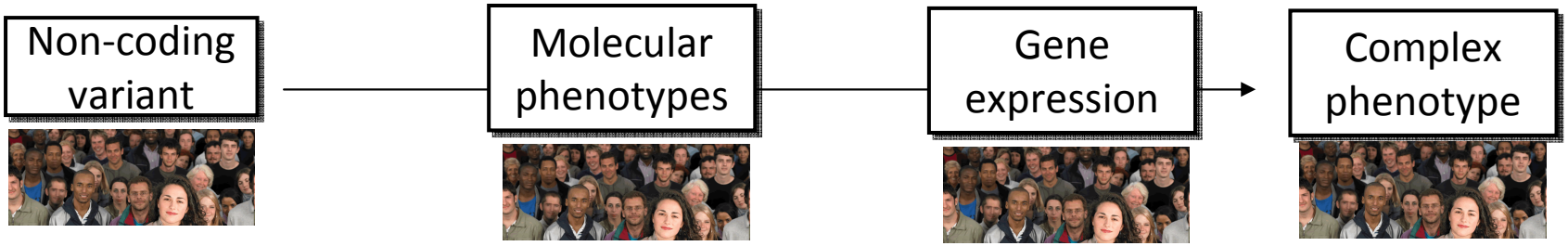
McVicker et al. Science 2013

Example 4: Reveal the function of chromatin states in complex physiological traits

Example 4: Reveal the function of chromatin states in complex physiological traits



Example 4: Reveal the function of chromatin states in complex physiological traits



A molecular basis for classic blond hair color in Europeans

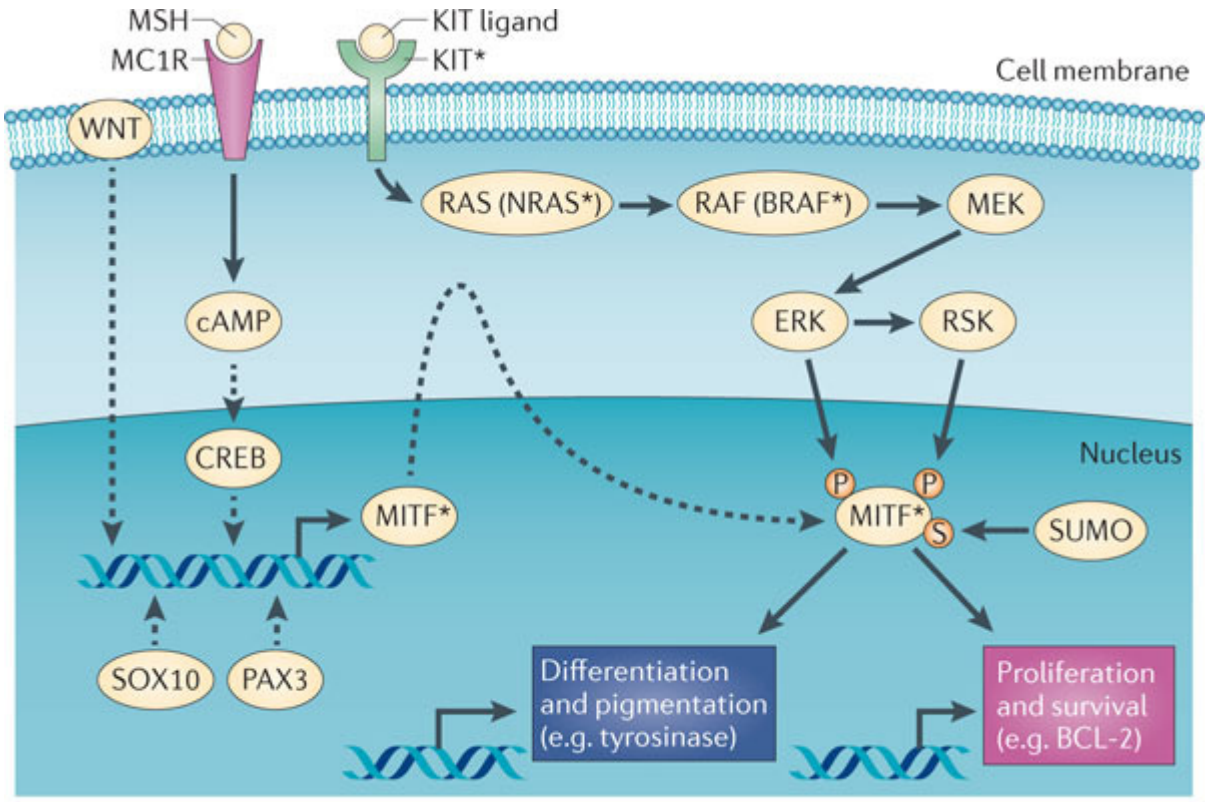
- Genetic variants linked to eight genes in humans are significantly associated with blond hair color in Europeans.

- Some variants alter the coding regions of genes known to be involved in pigmentation.

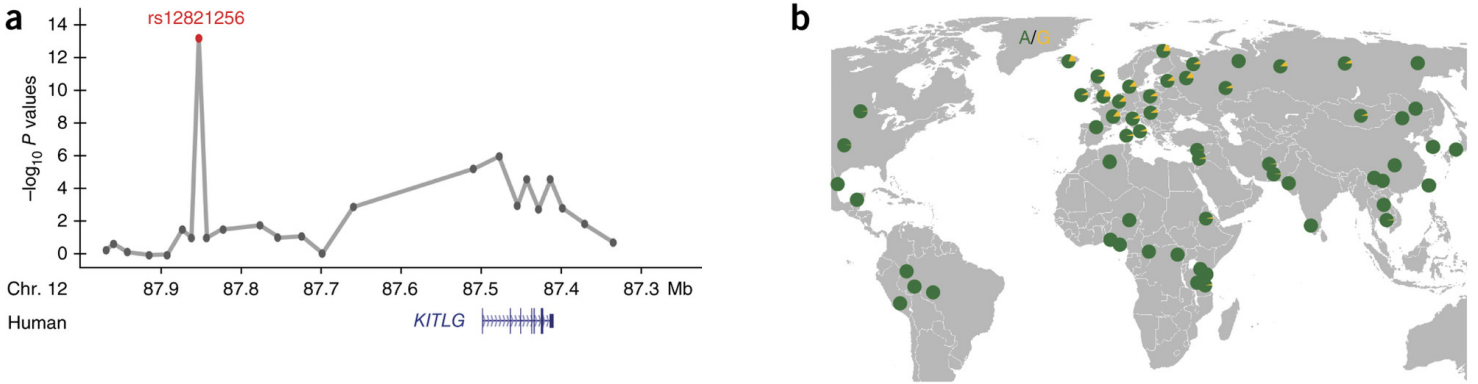
- Some variants map outside the protein-coding regions of pigmentation genes.

The Human *KITLG* gene (mouse *Kitl*)

• Encodes a secreted ligand for the KIT receptor tyrosine kinase and has an essential role in the development, differentiation and pigmentation.



The Human *KITLG* gene (mouse *Kitl*)

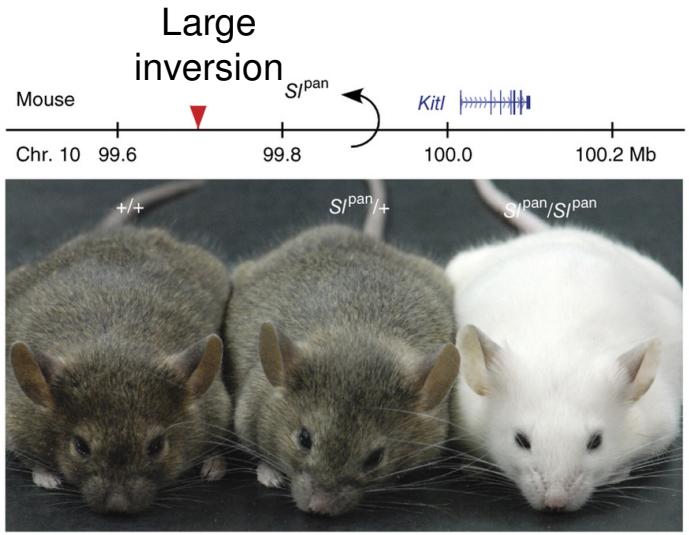


• A noncoding SNP (rs12821256) located over 350 kb upstream of KITLG is significantly associated with blond hair color in Iceland and The Netherlands.

• The blond-associated A>G substitution at this position is prevalent in northern European populations but virtually absent in African and Asian populations

Guenther et al. Nature Genetics 2014

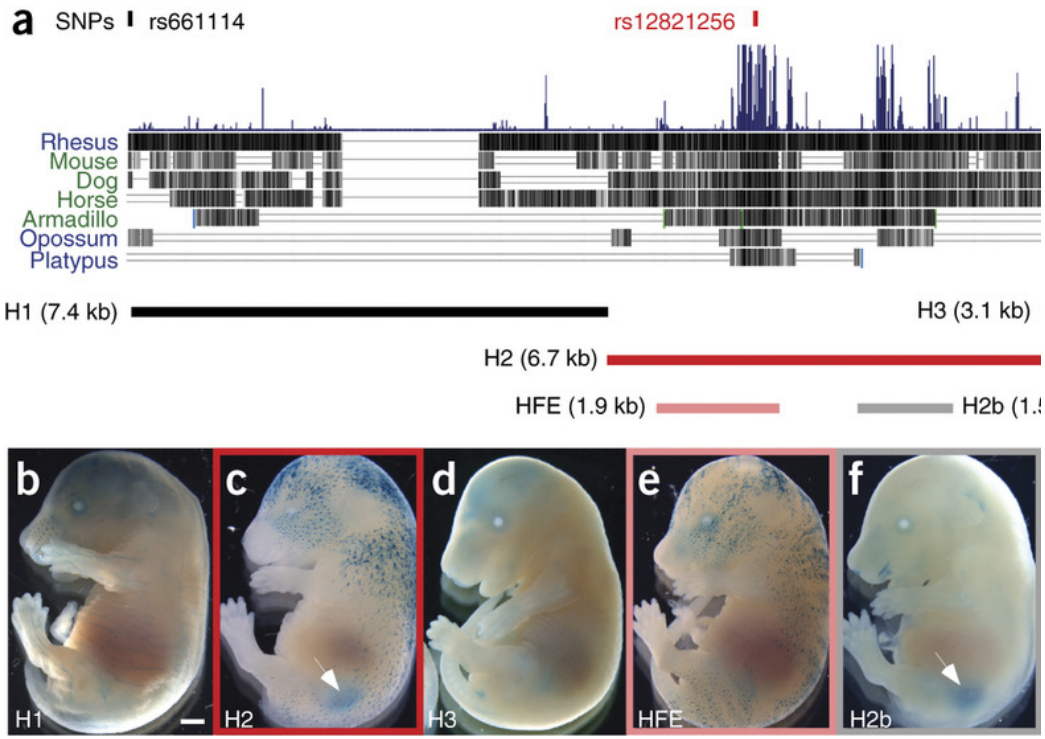
An inversion spanning the noncoding SNP rs12821256



Displacement of a single copy of the distant upstream regulatory sequences for *Kitl* is sufficient to reduce *Kitl* expression and lighten hair color.

Guenther et al. Nature Genetics 2014

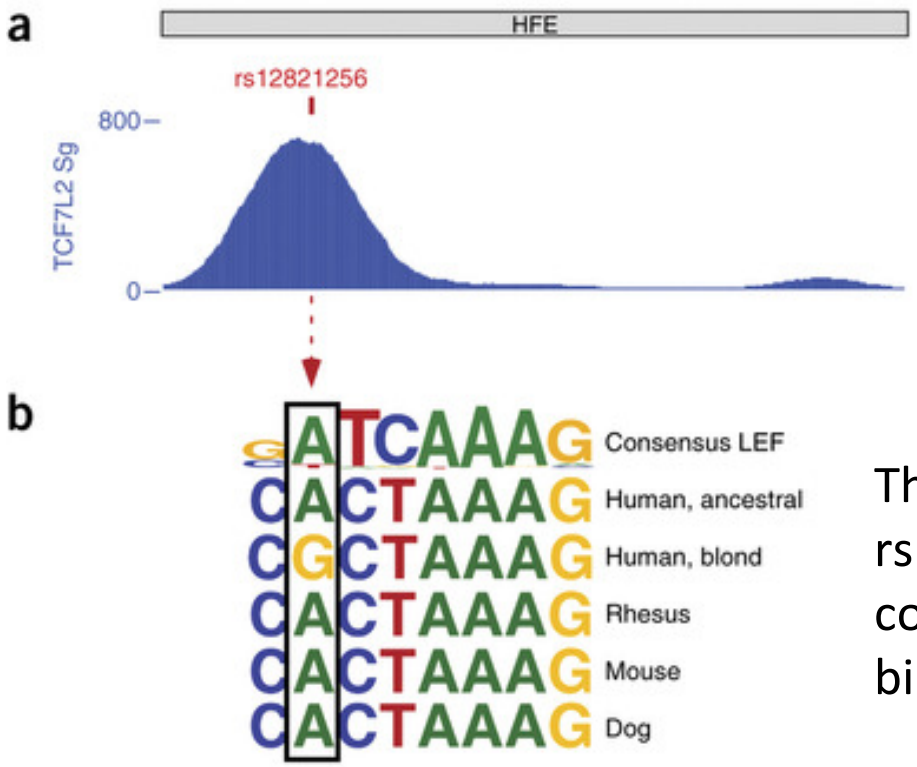
Searching for the functional enhancer



Five human fragments were cloned upstream of a *lacZ* reporter gene and tested for *in vivo* enhancer activity in transgenic mice.

Guenther et al. Nature Genetics 2014

rs12821256 alters a TCF/LEF binding site and reduces LEF responsiveness in keratinocytes

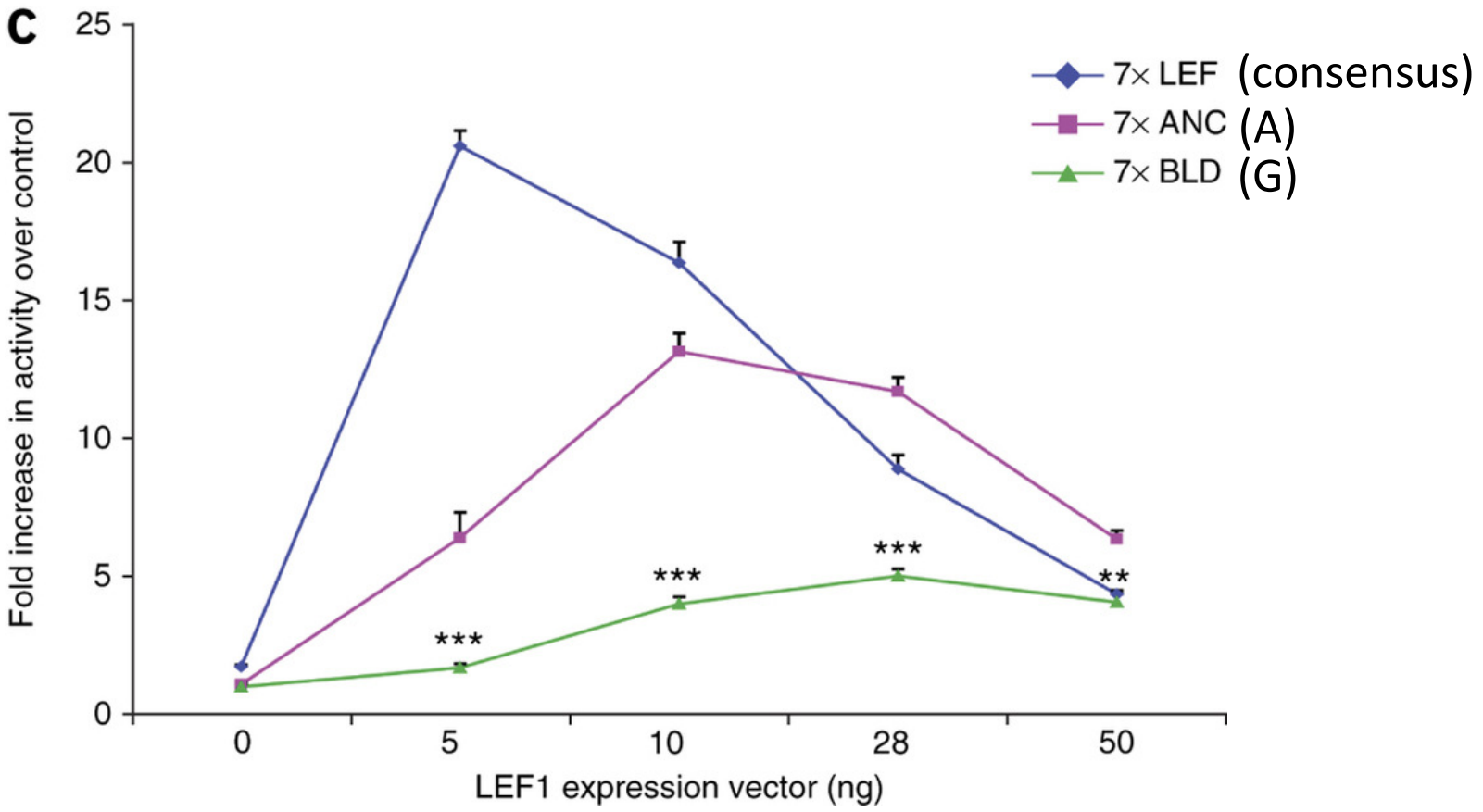


TCF ChIP-Seq reads from the ENCODE project

The sequence surrounding rs12821256 resembles a consensus TCF/LEF binding motif

Guenther et al. Nature Genetics 2014

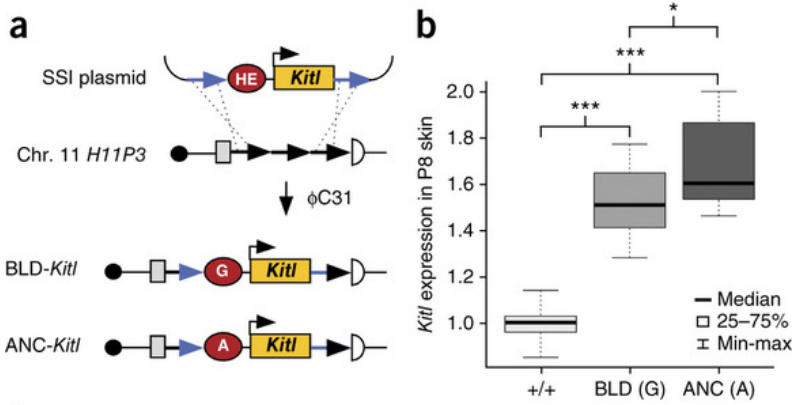
rs12821256 alters a TCF/LEF binding site and reduces LEF responsiveness in keratinocytes



Guenther et al. Nature Genetics 2014

Mouse lines differing at a single base-pair position in the *KITLG* enhancer show differences in hair color

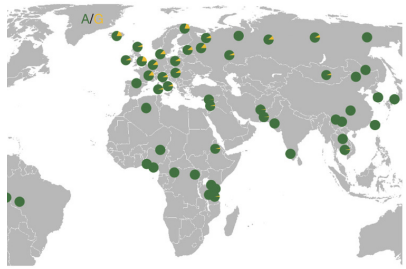
Matched lines of site-specific integration in transgenic mice



no enhancer

blond-variant

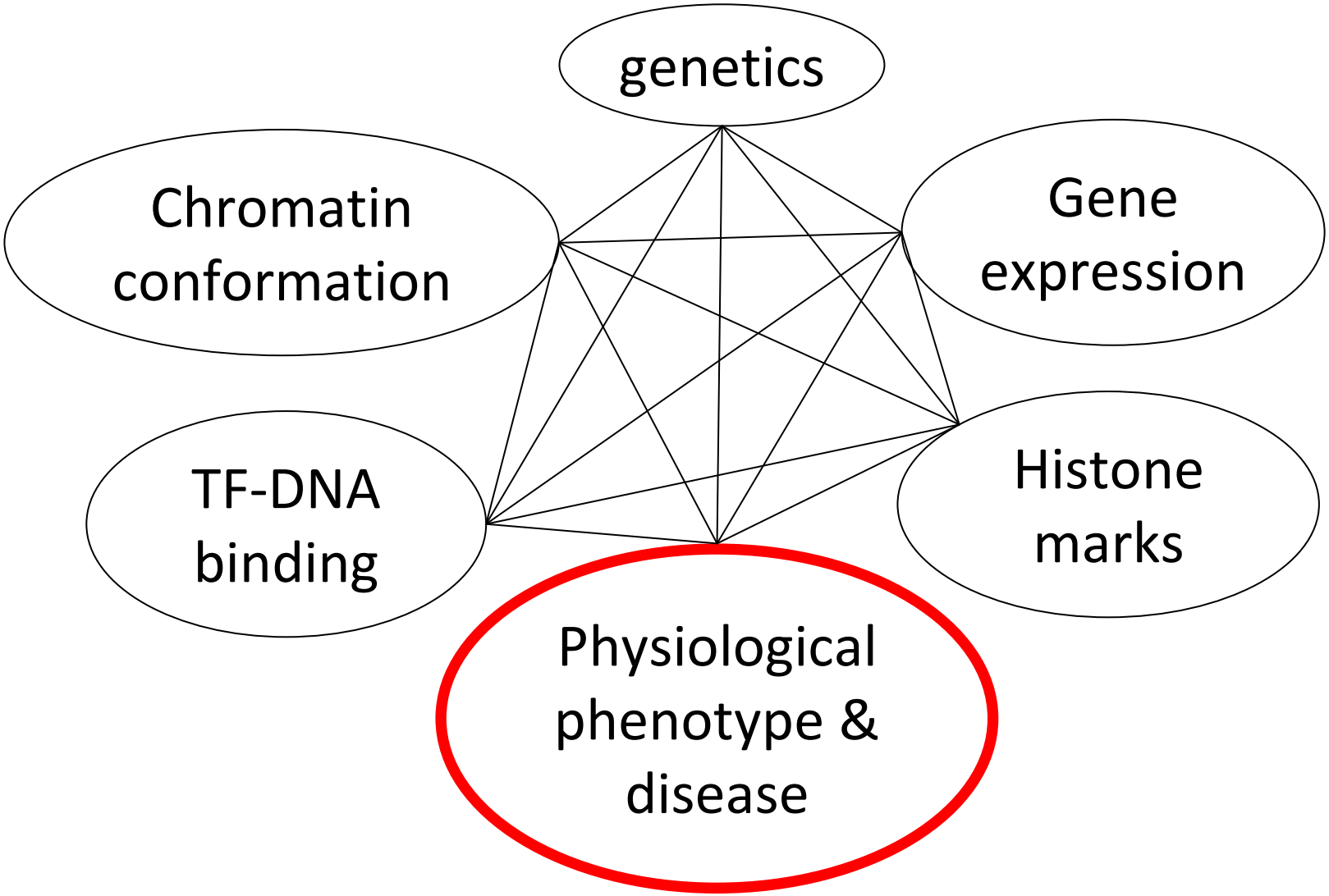
ancestral - variant



Small (20%) quantitative changes in enhancer activity were sufficient to alter hair color *in vivo*

Guenther et al. Nature Genetics 2014

The molecular basis of disease

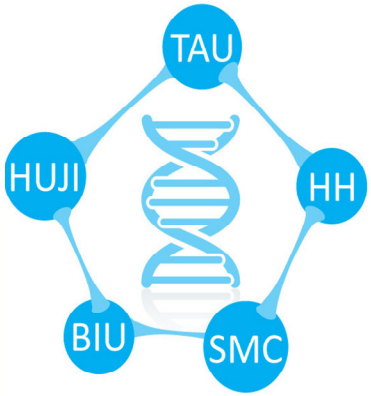


Utilizing genetics to understand transcriptional circuitry and the regulatory conformation of the genome

- Interpreting susceptibility loci using epigenomic profiling
- Revealing susceptibility loci that impact the chromatin landscape: Which genetic variants determine histone marks, open chromatin and TF binding?
- Reveal functionally linked histone marks between nearby or distal regions
- Reveal chromatin regulators
- Reveal the function of chromatin states in common disease

Thank you!


i-CORE
Israeli Centers of Research Excellence


TAU
HUJI
HH
BIU
SMC
Gene Regulation in
Complex Human Disease