

Gene Regulatory Networks

02-710 Computational Genomics

Seyoung Kim

Transcription Factor Binding



Transcription Control

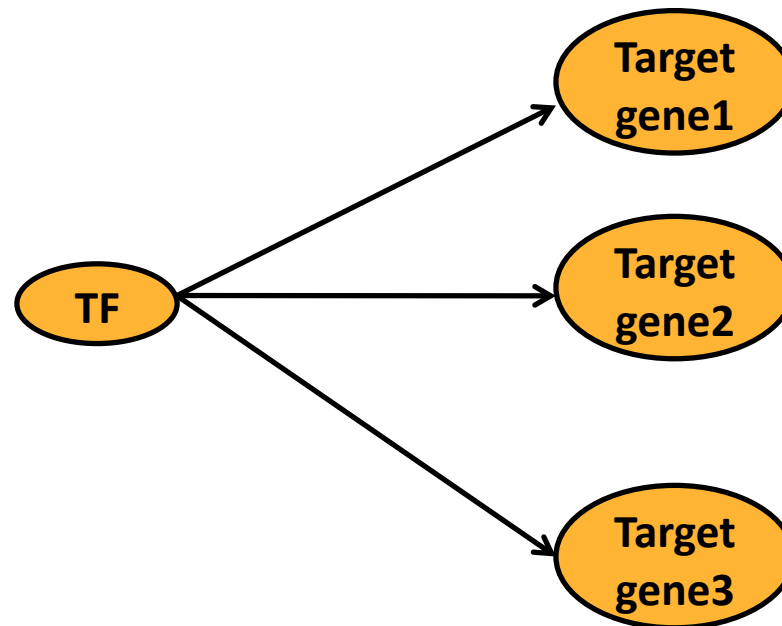
- Gene transcription is influenced by
 - Transcription factor binding affinity for the regulatory regions of target genes
 - Transcription factor concentration
 - Nucleosome positioning and chromatin states
 - Enhancer activity

Gene Transcriptional Regulatory Network

- The expression of a gene is controlled by cis and trans regulatory elements
 - **Cis regulatory elements:** DNA sequences in the regulatory region of the gene (e.g., TF binding sites)
 - **Trans regulatory elements:** RNAs and proteins that interact with the cis regulatory elements

Gene Transcriptional Regulatory Network

- Consider the following regulatory relationships:

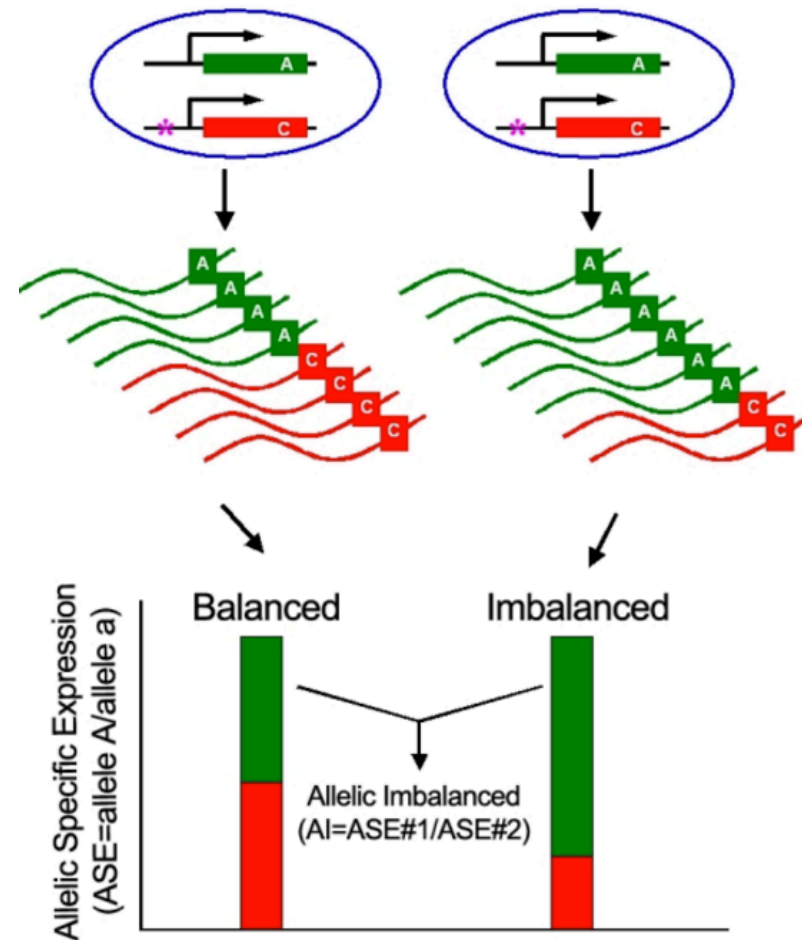


Gene Transcriptional Regulatory Network

- Cis and trans regulatory elements form a complex transcriptional regulatory network
 - Each trans regulatory element (proteins/RNAs) can regulate multiple target genes
 - Cis regulatory modules (CRMs)
 - Multiple different regulators need to be recruited to initiate the transcription of a gene
 - The DNA binding sites of those regulators are clustered in the regulatory region of a gene and form a CRM

How Can We Learn Transcriptional Networks?

- Leverage allele specific expressions
 - In diploid organisms, the transcript levels from the two copies of the genes may be different
 - RNA-seq can capture allele-specific transcript levels

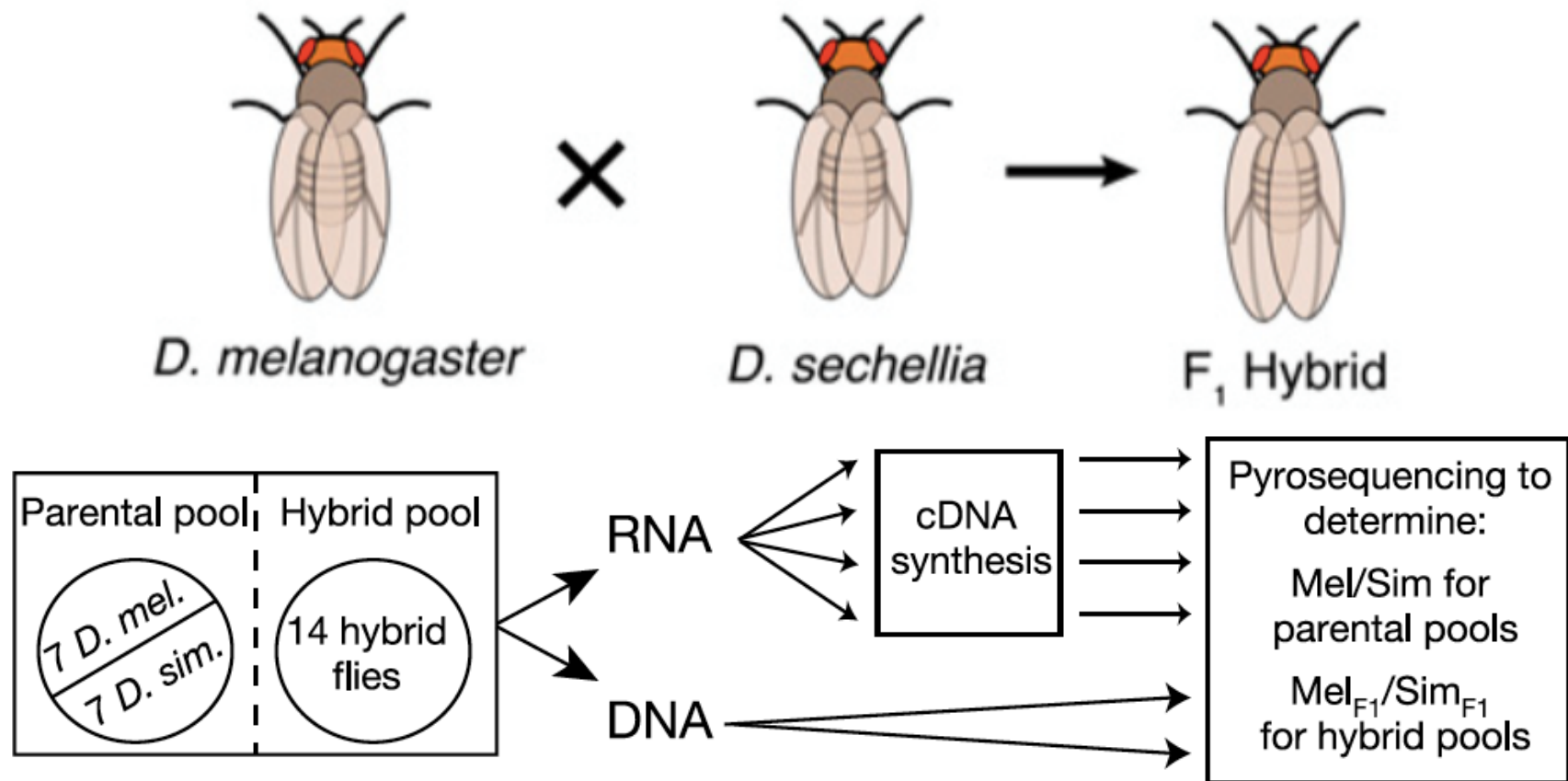


How Can We Learn Transcriptional Networks?

- Leverage allele specific gene expressions
 - Teasing out cis/trans regulatory divergence between two species (Wittkopp et al. Nature 2004)
 - First study that makes use of allele-specific gene expressions to learn about gene regulatory network
 - Examines genome/transcriptome data for two parent species and their F1 progeny
 - Allele-specific eQTL mapping
 - Extension of eQTL mapping
 - Examines population genome/transcriptome data

Cis/trans regulatory Divergence in Two Species

- Compare two parent species with their F1 progeny

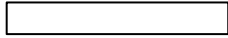





Only Cis-regulatory Divergence Between Two Species

D. melanogaster

TF	Target	Expr
		High
		High

D. sechellia


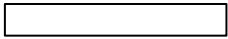

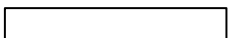
TF	Target	Expr
		Low
		Low

F1 Hybrid


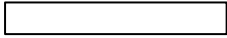


TF	Target	Expr
		High
		Low

Only Trans-regulatory Divergence Between Two Species


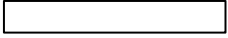

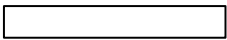
D. melanogaster

TF	Target	Expr
		High
		High

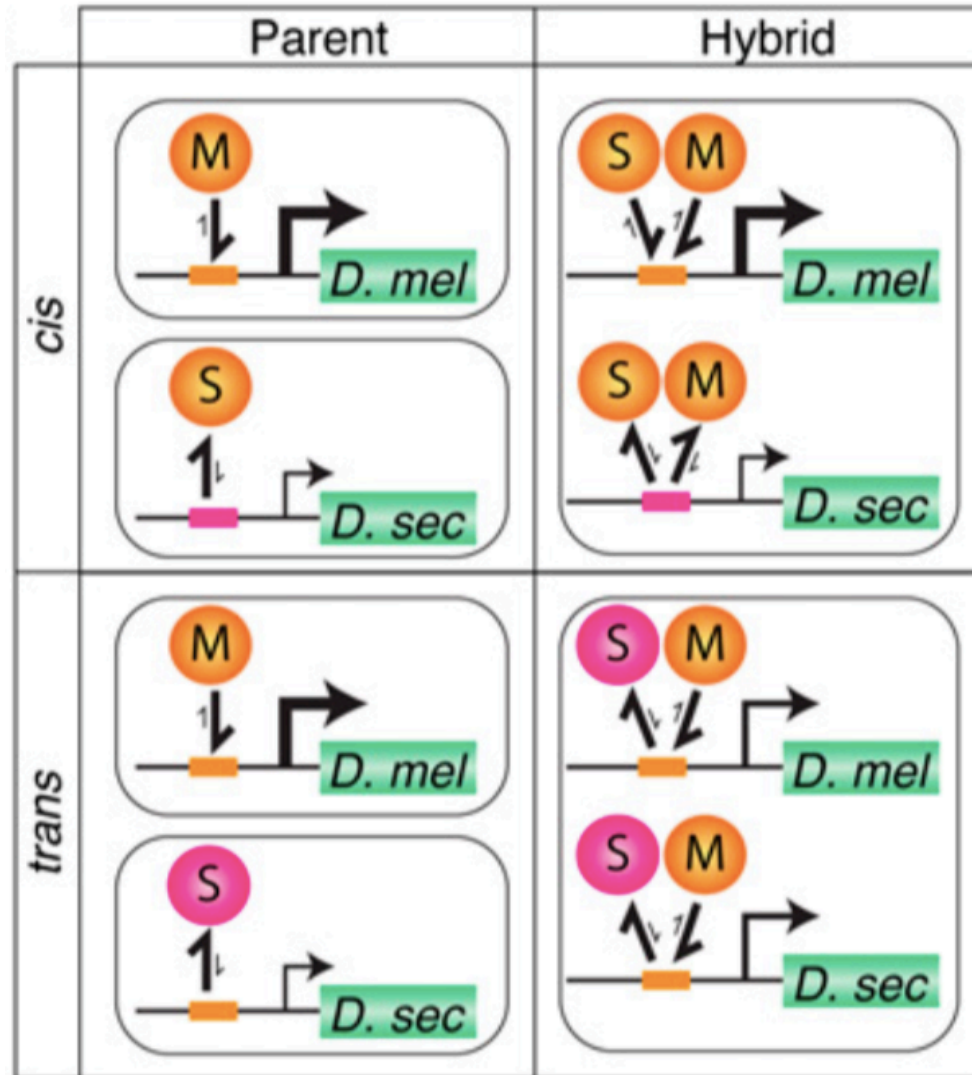
D. sechellia

TF	Target	Expr
		Low
		Low

F1 Hybrid

TF	Target	Expr
		Medium
		Medium

Cis/Trans Regulatory Divergence and Allele-Specific Expression



More Precisely...

- Only cis regulatory changes between the two species

$$\frac{A_p}{a_p} = \frac{A_{F1}}{a_{F1}}$$

- Only trans regulatory changes between the two species

$$\frac{A_{F1}}{a_{F1}} = 1$$

A, a: Two allele-specific transcript levels of target gene
p, F1: parent species and F1 hybrid

More Precisely...

- In order to test the presence of trans regulatory changes between the two species, test

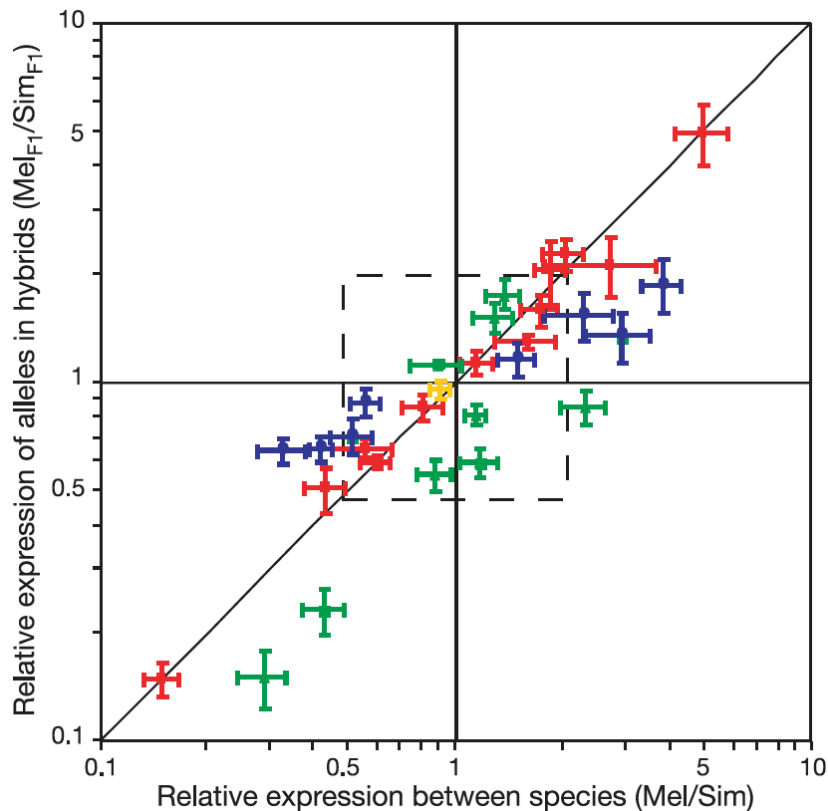
$$\frac{A_p}{a_p} \neq \frac{A_{F1}}{a_{F1}}$$

- In order to test the presence of cis regulatory changes between the two species, test

$$\frac{A_{F1}}{a_{F1}} \neq 1$$

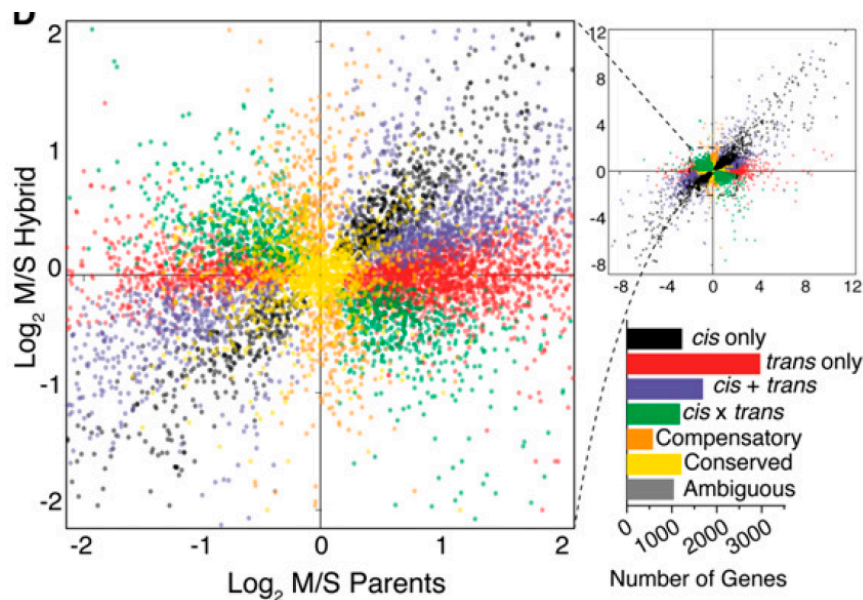
A, a: Two allele-specific transcript levels of target gene
p, F1: parent species and F1 hybrid

How Much Cis/Trans Regulatory Divergence Between *D. melanogaster* and *D. sechellia*



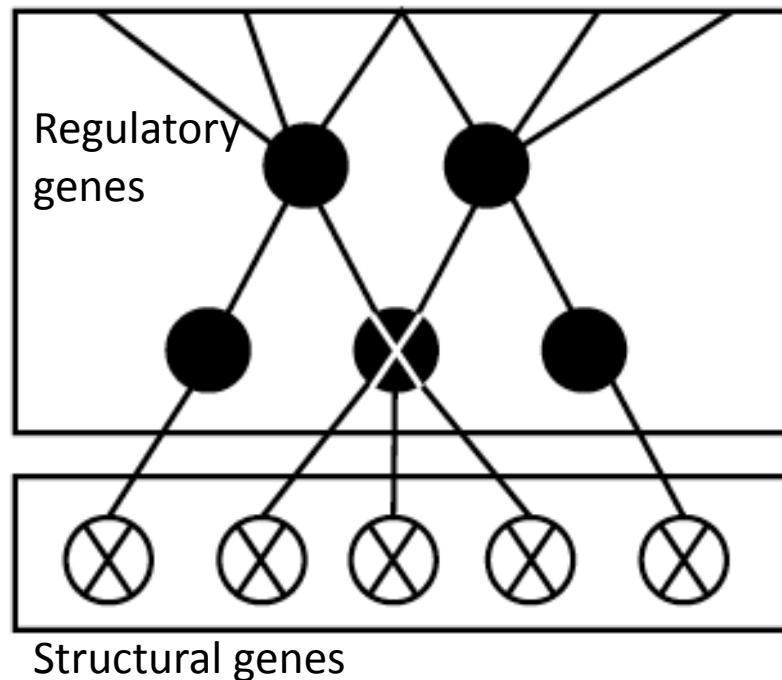
- Only cis-regulatory divergence: genes on the diagonal line ($y=x$)
- Only trans-regulatory divergence: genes on $y=1$
- Most genes are influenced by a combination of cis and trans regulatory divergence

How Much Cis/Trans Regulatory Divergence Between *D. melanogaster* and *D. sechellia*



- RNA-seq-based studies
 - Cis + trans
 - Allele specific ratios have the same signs in parents and hybrids
 - Cis/trans regulatory differences favored expression of the same allele during evolution
 - Cis x trans
 - Allele specific ratios have different signs in parents and hybrids
 - Cis/trans regulatory differences favored expression of the different allele during evolution
 - Conserved
 - no expression differences in the parents but differentially expressed in the hybrids.
 - cis and trans divergence compensated with each other

Overall Regulatory Networks



- Regulatory genes
 - encode TFs, signaling molecules
 - Expression changes lead to trans-acting effects on downstream genes
- Structural genes
 - Encode enzymes or cellular components
 - Usually terminal nodes with no trans-acting effects

Limitations

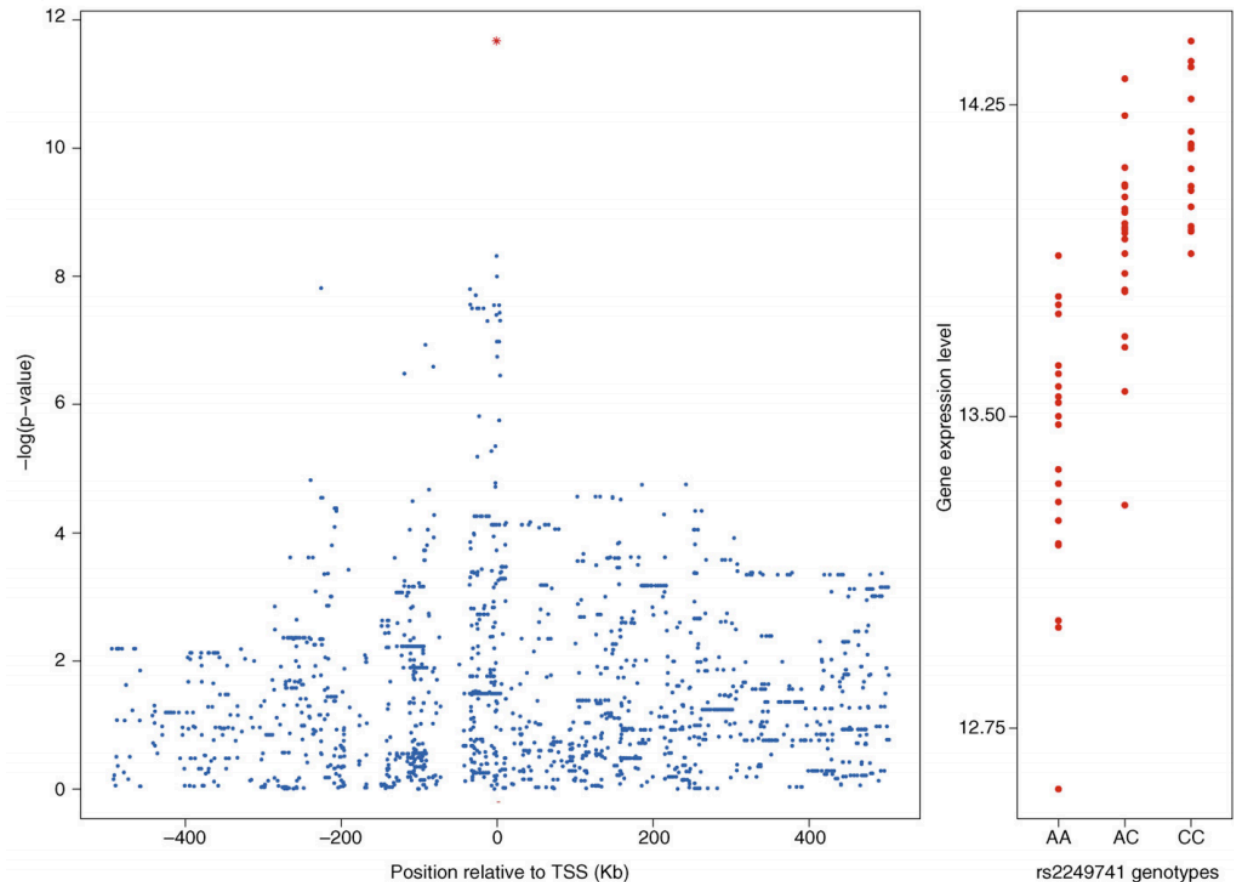
- Examined the data between **only two species and their progeny**
 - Limited statistical power
 - Captured only the overall contribution of cis/trans regulatory divergence
 - Unable to identify the precise cis/trans regulatory elements whose variation contribute to the gene expression divergence

How Can We Learn Transcriptional Networks?

- Leverage allele specific gene expressions
 - Teasing out cis/trans regulatory divergence between two species (Wittkopp et al. Nature 2004)
 - First study that makes use of allele-specific gene expressions to learn about gene regulatory network
 - Examines genome/transcriptome data for two parent species and their F1 progeny
 - Allele-specific eQTL mapping (Sun, Biometrics 2012)
 - Extension of eQTL mapping
 - Examines population genome/transcriptome data

Expression Quantitative Trait Locus (eQTL) Mapping

- Study of the genetic basis of gene expression using population gene expression and genotype data
- Essentially GWAS with gene expression as phenotype values



eQTLs for HLA-C gene expression in HapMap European samples

eQTL Mapping

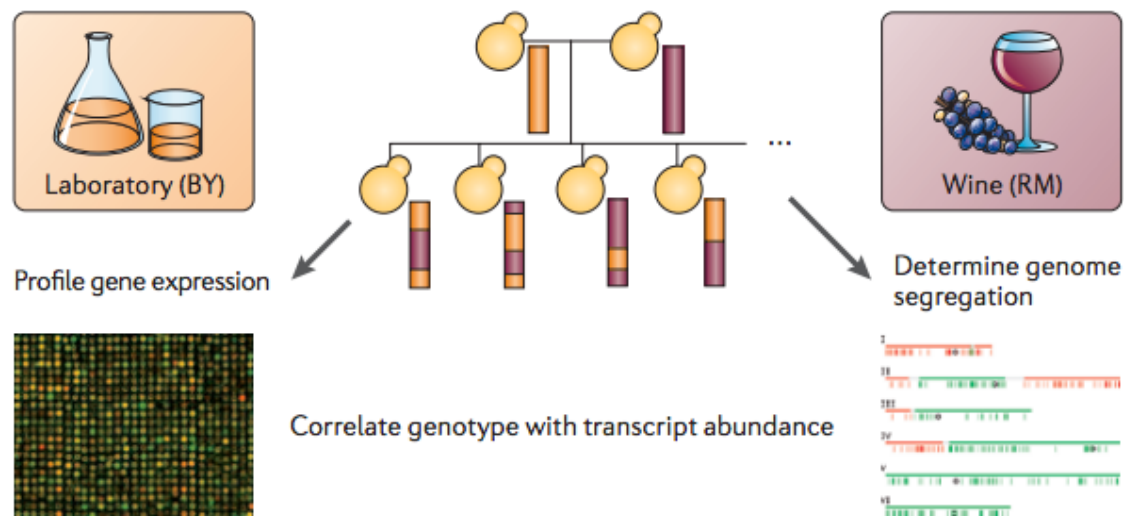
- Popular approach for studying the genetic basis of gene expressions for
 - Different tissue types
 - Different diseases

Year	Cell type or tissue studied	Design	Number of individuals	Disease or trait being compared to*	Refs
Blood					
2014	Whole blood	Twins	2,752 [†]	Many	63
	Whole blood	Unrelated	922	NA	47
2013	Whole blood	Unrelated	5,311 [§]	Type 1 diabetes and cholesterol metabolism	70 ^l
2012	Whole blood and LCLs	Families	862	NA	216
2011	Whole blood	Unrelated	1,469	Blood traits	71
Bone					
2011	Osteoblasts	Unrelated	113	Asthma	217
Brain					
2014	10 brain regions	Unrelated	134	Parkinson's disease and other brain disorders	218
2012	Cortex and cerebellum	Unrelated	400	Parkinson's disease and other brain disorders	219
2011	Developmental time series	Unrelated	269	NA	220
2010	4 brain regions	Unrelated	150	NA	221
Heart					
2014	Heart	Unrelated	129	Cardiac traits	222
Immune system					
2014	LCLs	Unrelated	869	Type 1 diabetes and ulcerative colitis	72
	Dendritic cells	Unrelated	534 [§]	Autoimmune and infectious disease	74
	Lymphocytes and monocytes	Unrelated	461	Autoimmune disease and neurodegenerative diseases	73
	T cells	Unrelated	348 [§]	Autoimmune disease	150
	Stimulated monocytes	Unrelated	432	Immunity-related (for example, bacterial infection, inflammation, multiple sclerosis and Crohn's disease)	67
2013	LCLs	Unrelated	462	Many	88
2012	Monocytes and B cells	Unrelated	283	Immunity-related (for example, ulcerative colitis and systemic lupus erythematosus)	66
2010	Monocytes	Unrelated	1,490	Many	148
2009	Fibroblasts, LCLs and T cells	Unrelated	75	NA	8
2007	Lymphocytes	Extended families	1,240	HDL-C	49
Liver					
2011	Liver tissue	Unrelated	266	Diabetes, drug response, lipid levels and prostate cancer	223
2008	Liver tissue	Unrelated	427	Type 1 diabetes, coronary artery disease and plasma LDL-C	224
Lung					
2012	Lung tissue	Unrelated	1,111	Asthma	225
Multiple tissue types					
2012	Adipose tissue	Twins	856	Triglyceride levels and birth weight	62
	Skin tissue			Melanoma	
	LCLs			Immunity-related	
2010	Liver tissue	Unrelated	960	Plasma LDL-C and myocardial infarction	100
	Subcutaneous fat		433		
	Omental fat		520		

Year	Cell type or tissue studied	Design	Number of individuals	Disease or trait being compared to*	Refs
Tumours					
2014	Colorectal tumours and normal tissue	Unrelated	103	Colorectal cancer	226
	5 tumour types [‡]	Unrelated	145–391	5 types of cancer	227
2013	Breast cancer	Unrelated	219	Breast cancer	228

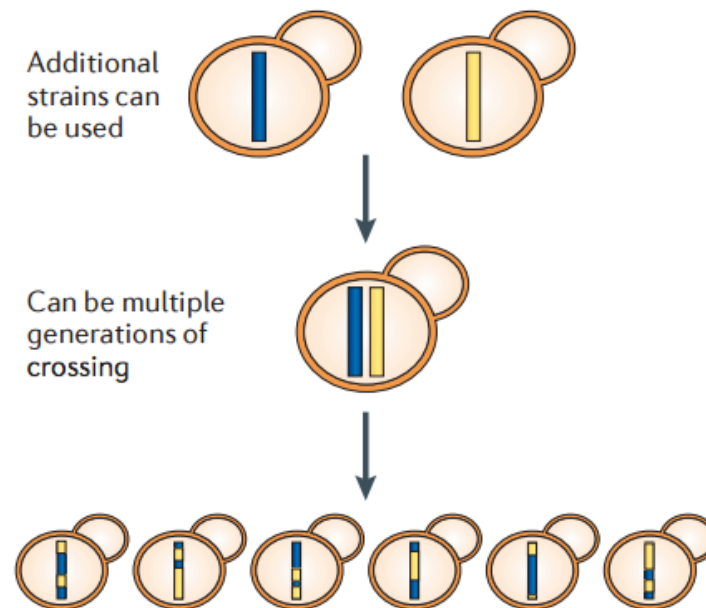
Expression Quantitative Trait Locus (eQTL) Mapping

- Recombinant inbred lines, instead of population of unrelated individuals
 - Has been applied to model organisms yeast, mouse, etc.



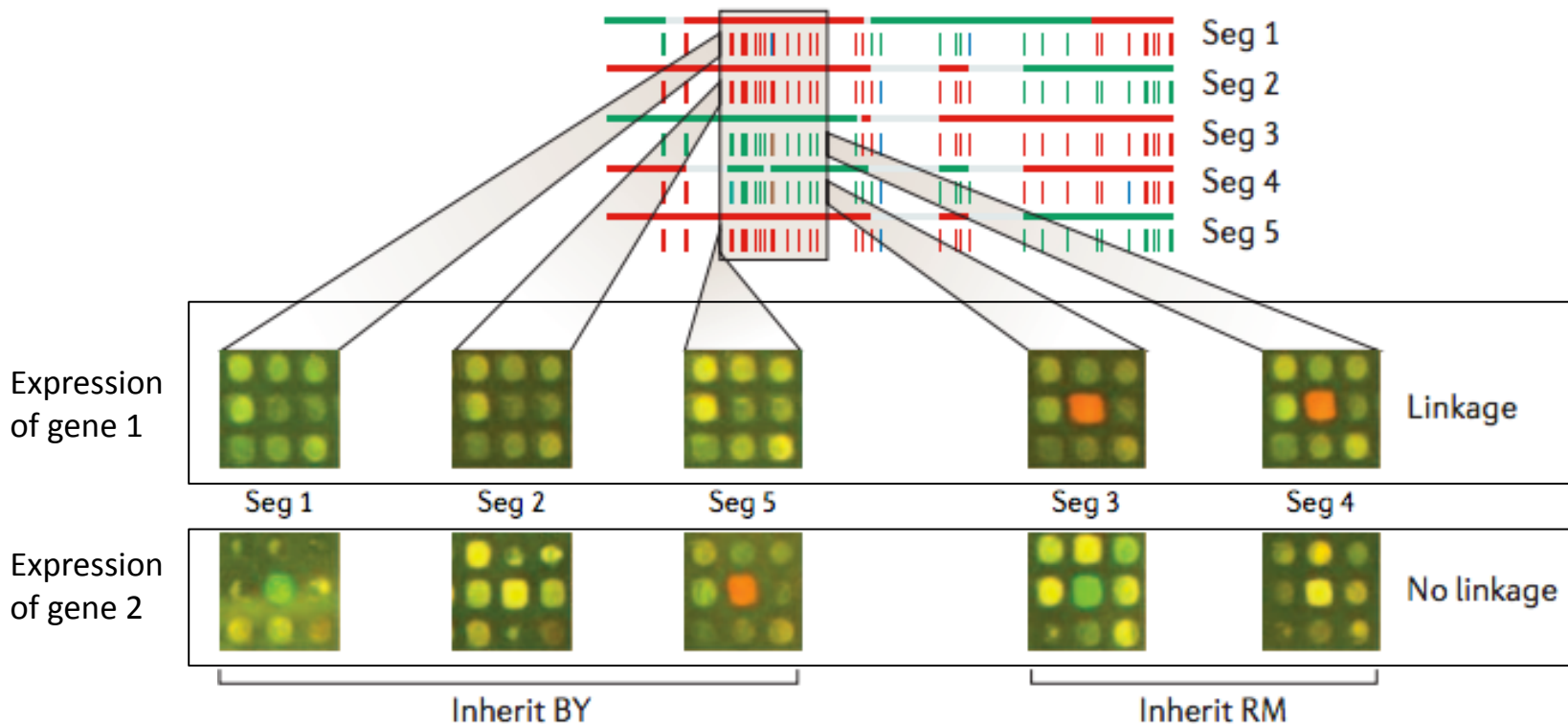
Expression Quantitative Trait Locus (eQTL) Mapping

- Recombinant inbred lines, instead of population of unrelated individuals
 - Key ideas: shuffle their genomes through mating, to understand the effects of genetic differences in parent species.
 - More generations of mating means more shuffling through recombinations and higher resolution for eQTLs



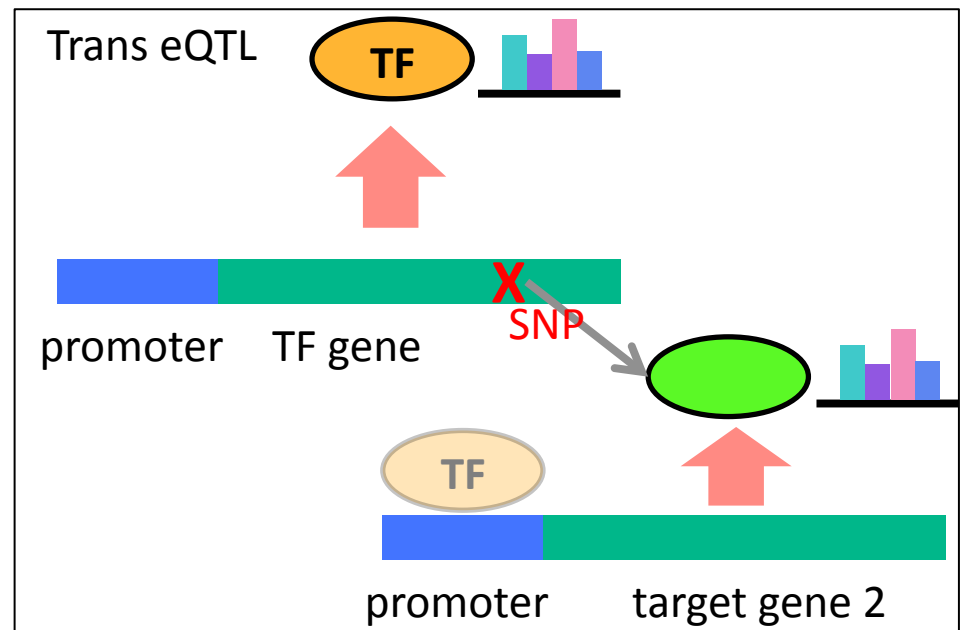
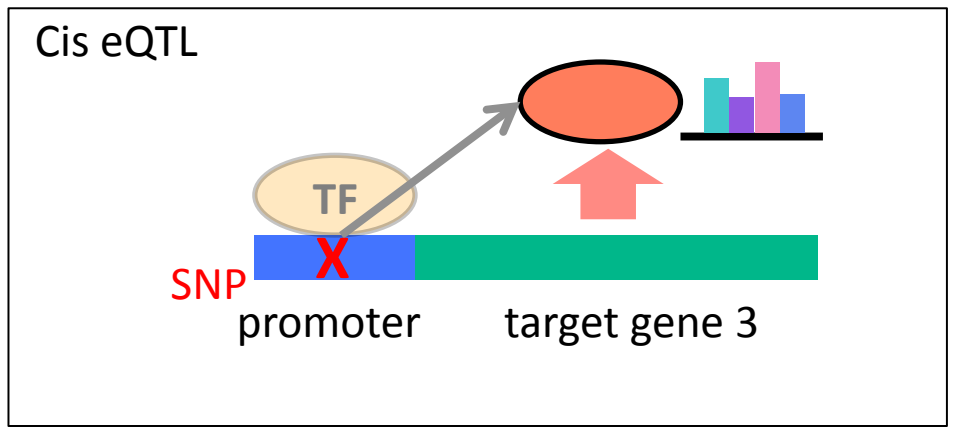
Expression Quantitative Trait Locus (eQTL) Mapping

- Recombinant inbred lines, instead of population of unrelated individuals
 - At a given locus, the samples are separated according to the inherited alleles



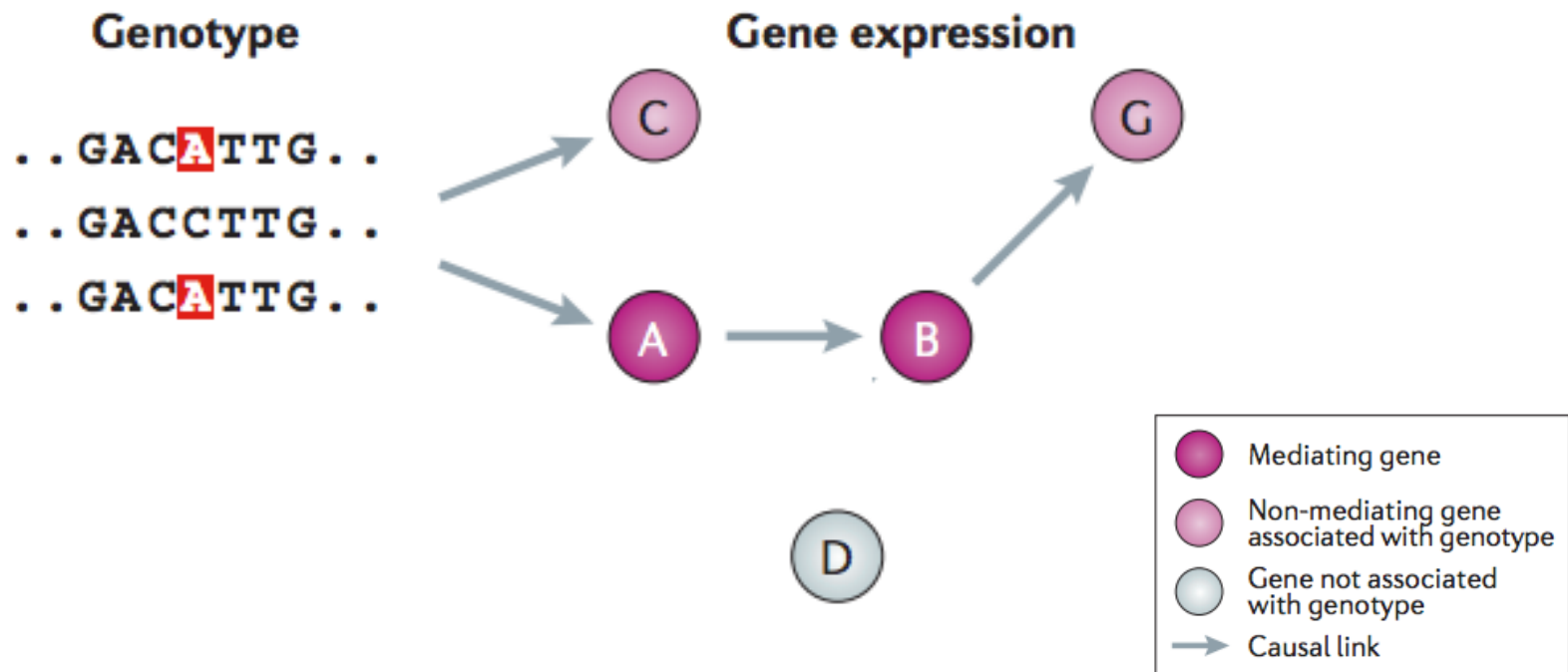
cis and *trans* eQTLs

- *cis* eQTL: in genomes, the eQTL is located near the gene whose expression is affected by the eQTL.
 - E.g., mutations in the upstream of a gene influences the expression level of the gene
- *trans* eQTL: in genomes, the eQTL is located far away (or on a different chromosome) from the gene whose expression is affected by the eQTL.
 - E.g., mutations in the transcription factor gene can influence the expression level of the TF target genes.



cis and *trans* eQTLs

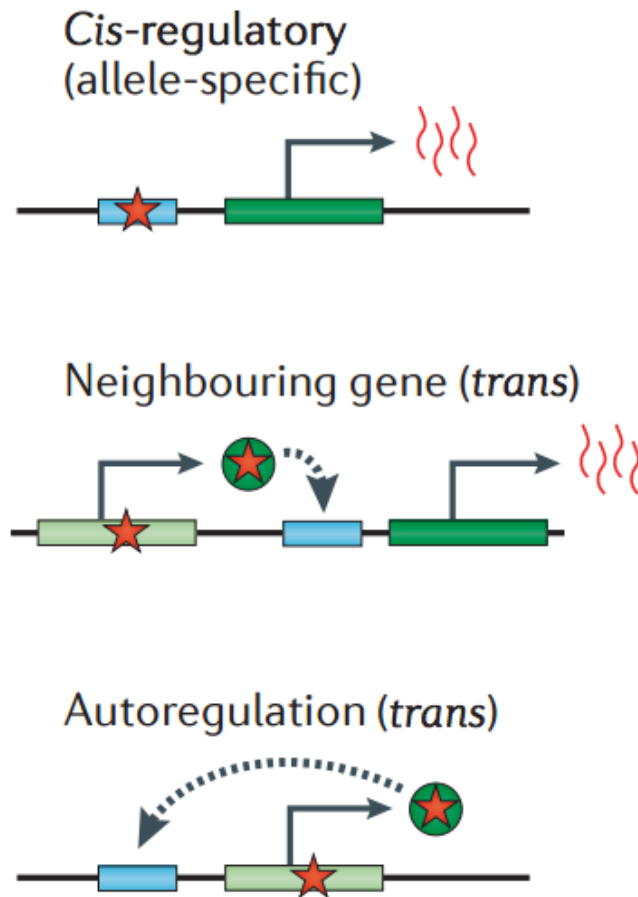
- Cis eQTLs and its downstream effects, leading to trans eQTLs



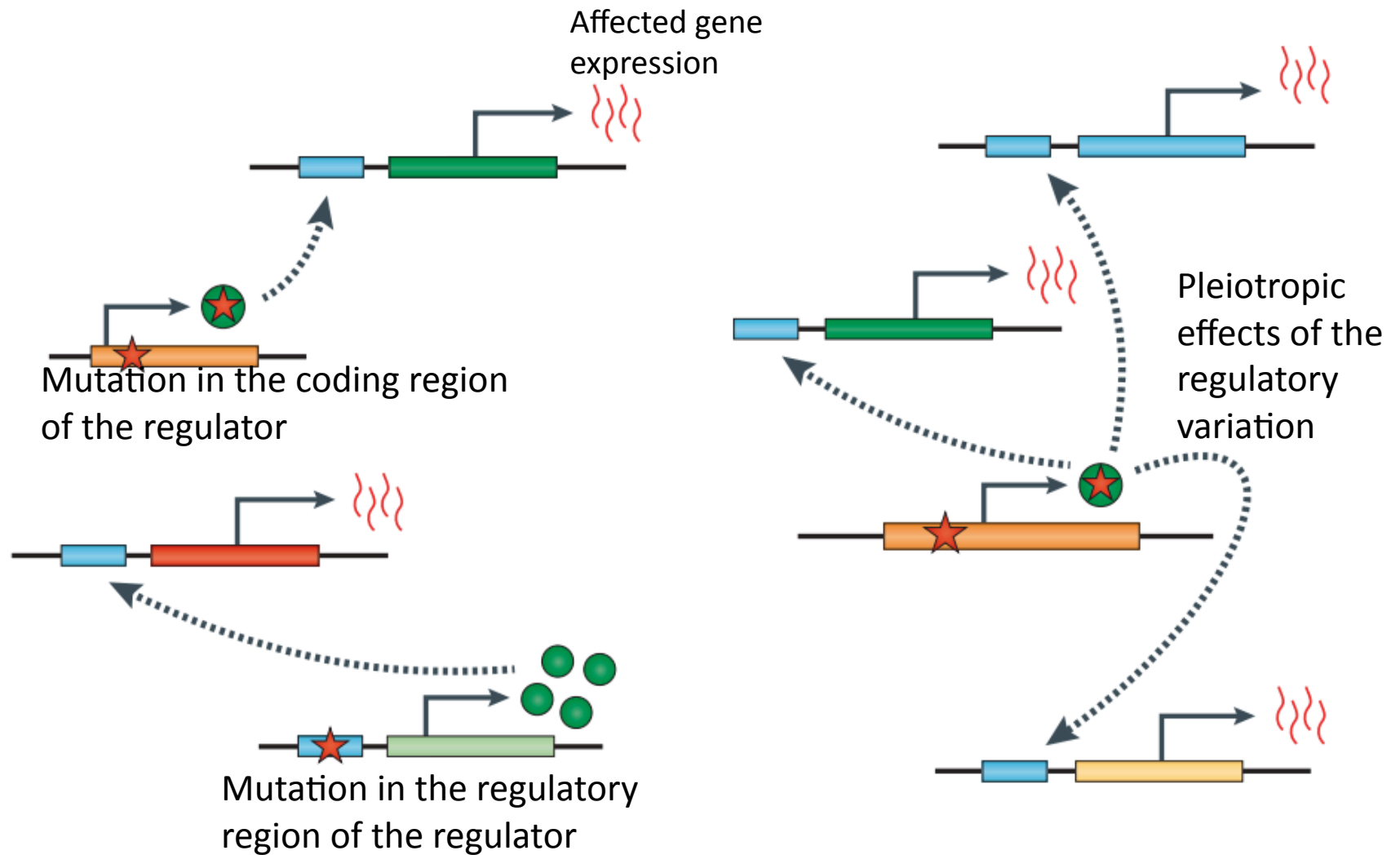
cis and *trans* eQTLs

- In traditional eQTL mapping, *cis* eQTLs as “local” eQTLs and *trans* eQTLs as “distant” eQTLs
- However, to be precise, not all *cis* eQTLs are “local” and not all *trans* eQTLs are “distant” (Rockman & Kruglyak, Nature Reviews Genetics 2006).
 - Precisely, *cis* eQTLs = *cis* regulatory variation
 - Precisely, *trans* eQTLs = *trans* regulatory variation

Not All Cis eQTLs are Cis-regulatory Variation



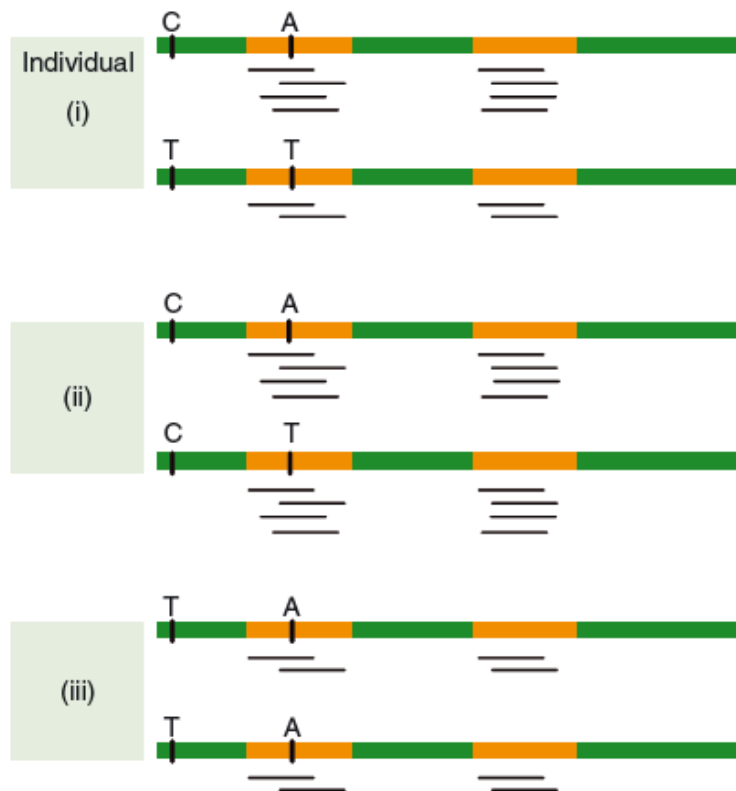
Distant Regulatory Variation



Data for Allele-Specific eQTL Mapping

- Data preparation

- Gene expression data: Allele-specific expression data can be measured only if the exons for the transcript contain heterozygous loci



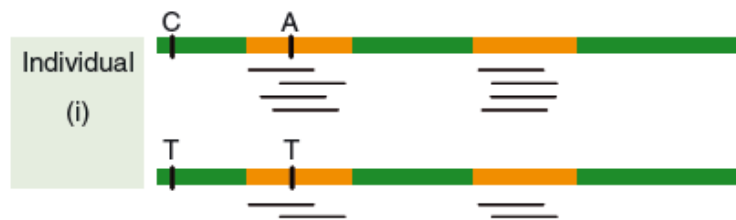
Allele-specific expression can be measured only for individuals 1 and 2

orange exon | SNP
green non-transcribed regions, e.g., intron etc.

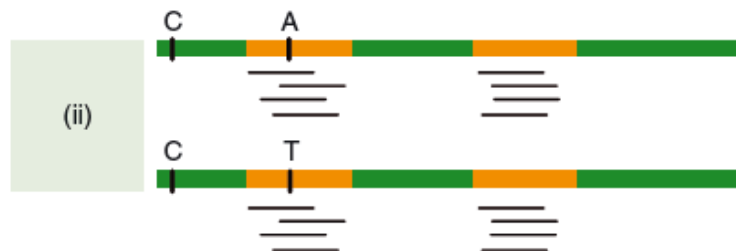
Data for Allele-Specific eQTL Mapping

- Data preparation

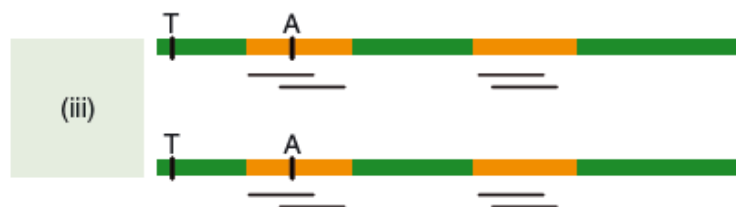
- Genotype data: cis-regulatory effects can be assessed only if the allele-specific expression is measured AND the locus is heterozygous



Cis-regulatory effects can be assessed only for individual 1



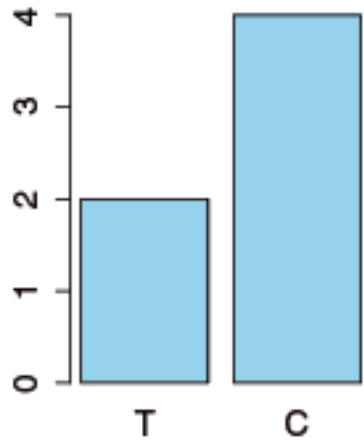
Haplotype information between SNPs and allele-specific transcript should be resolved



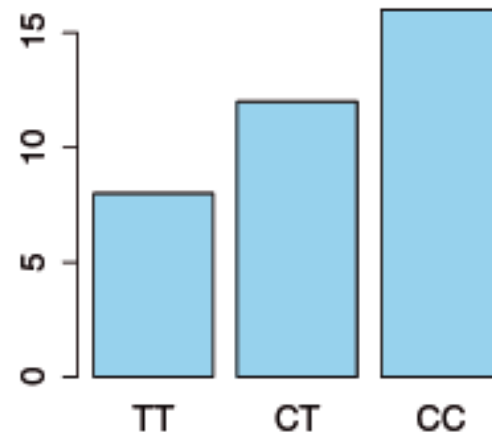
orange exon | SNP
green non-transcribed regions, e.g., intron etc.

Allele-specific eQTLs

- Cis eQTLs = cis-regulatory variation
 - Linear model with respect to the alleles



- Trans eQTLs = trans-regulatory variation
 - Linear model with respect to genotypes



Summary

- Cis/trans regulatory elements and the effects of their variation on gene expression
- How can we learn the underlying transcriptional network that consist of cis and trans regulatory elements?
 - Cis/trans regulatory divergence can be assessed based on allele-specific expressions of two parent species and their F1 hybrid. However, this approach has limited power
 - Cis/trans regulatory variation and their effects can be assessed in population genome sequence and allele-specific expression data