

Module 2: Understanding the Genetic Basis of Diseases Using Public Resources

Database	Purpose
UCSC Genome Browser	Provides genomic and epigenomic data, including visualization tools for genome annotations.
OMIM	A database of human genes and genetic disorders. Use this to investigate gene-disease relationships and find information on genetic disorders.
GeneCards	Integrates information about genes, their functions, expression, and relevance to diseases.
GeneReviews	Expert-authored summaries on genetic conditions, including diagnostic and treatment guidelines.

What is a genome browser?¹

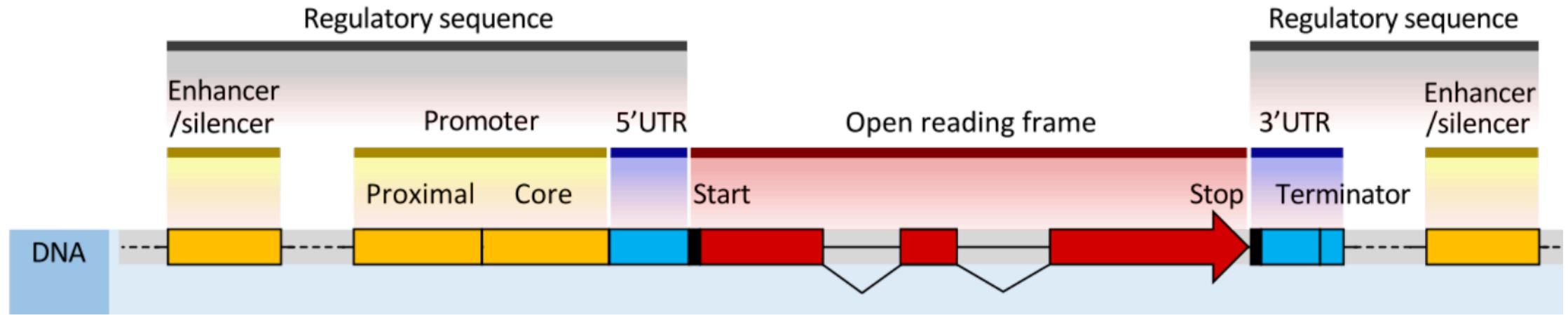
Genome browsers *integrate genomic sequence and annotation data* from different sources and provide a platform to search, browse, retrieve, and analyze genomic data.

They *differ from ordinary biological databases* in that they display data in a graphical format, with genome coordinates on one axis with annotations or space-filling graphics to show analyses of the genes, such as the frequency of the genes, their expression profiles, etc.

What is meant by annotation of data?

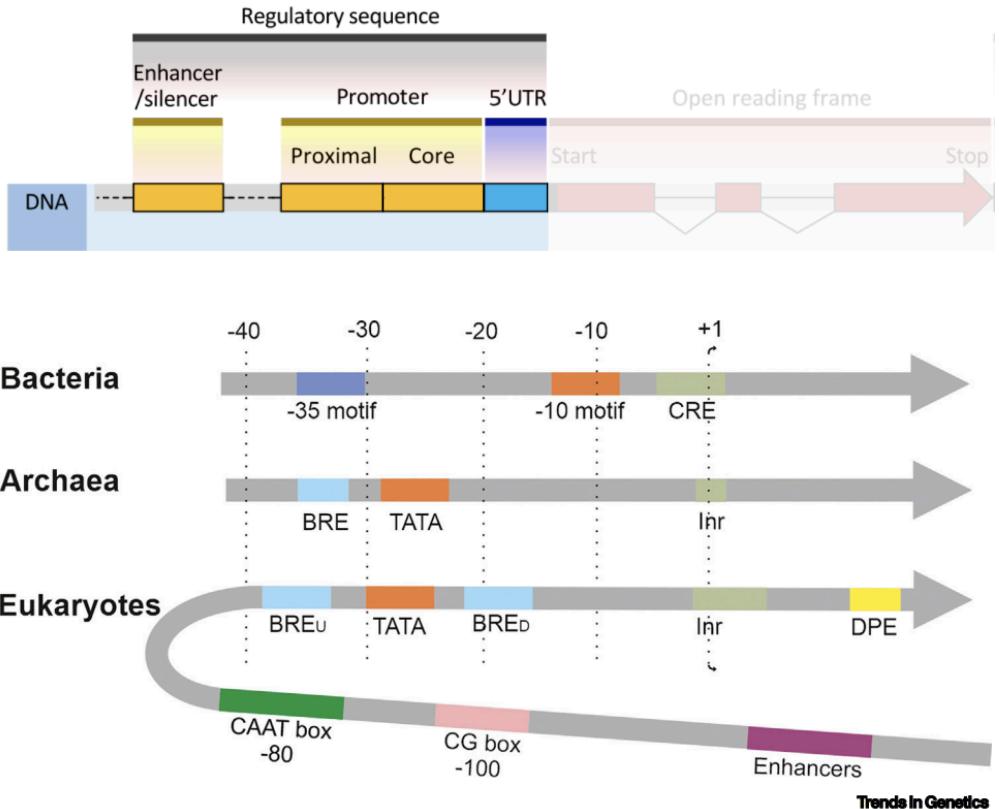
- Annotation means attaching biological information to sequences.

Gene Transcription



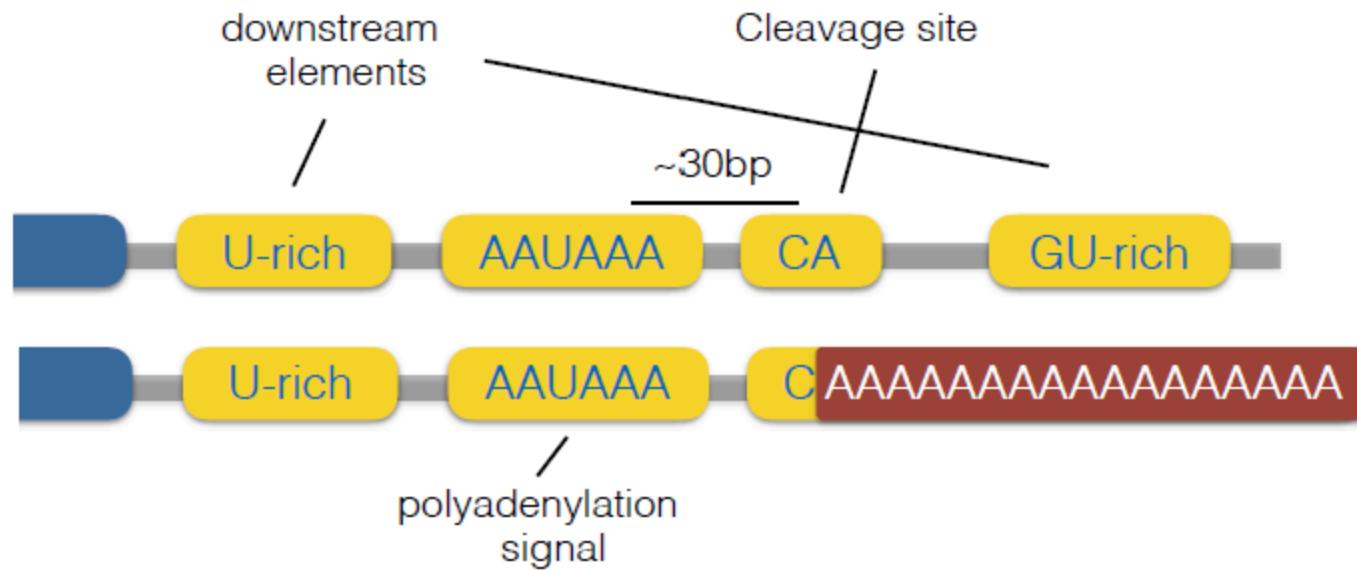
- Transcription starts at transcription start site (TSS) upstream of translation start site.
- Must end downstream of translation stop site (stop codon)
- 5' UTR is transcribed but not translated
- Enhancers are transcription factor binding sites that helps promoter to bind

Gene Transcription



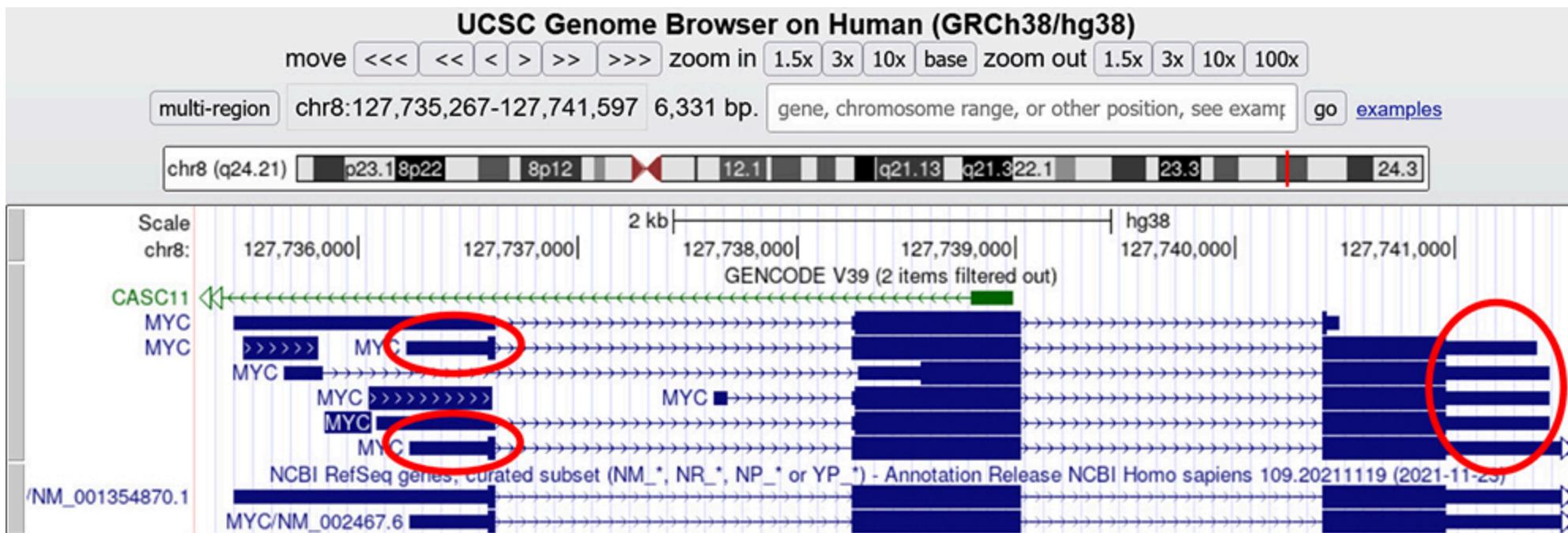
- Genes have promoters immediately upstream of TSS (+1).
- Most promoters range from 100 to 1000 bp in length.
- Transcription can also be affected by **enhancers** and **repressors**. Both enhancers and repressors bind remotely from the 'core promoter' (up to 2–3 Mbp in some genomes)

3' UTR and Transcription Termination

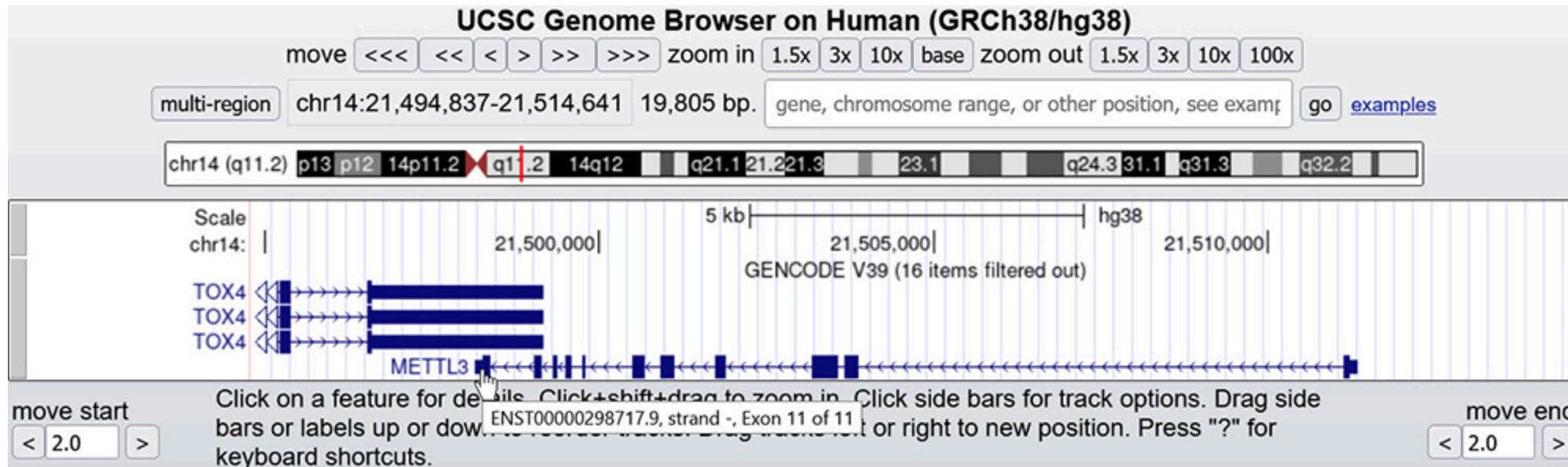


- Bacteria don't have 3' UTR. They use hairpins and U-rich sequence to terminate transcription.
- Eukaryotes remove part of the 3' end and add poly(A) tail to transport transcripts out of the nucleus.

Find the UTRs ²



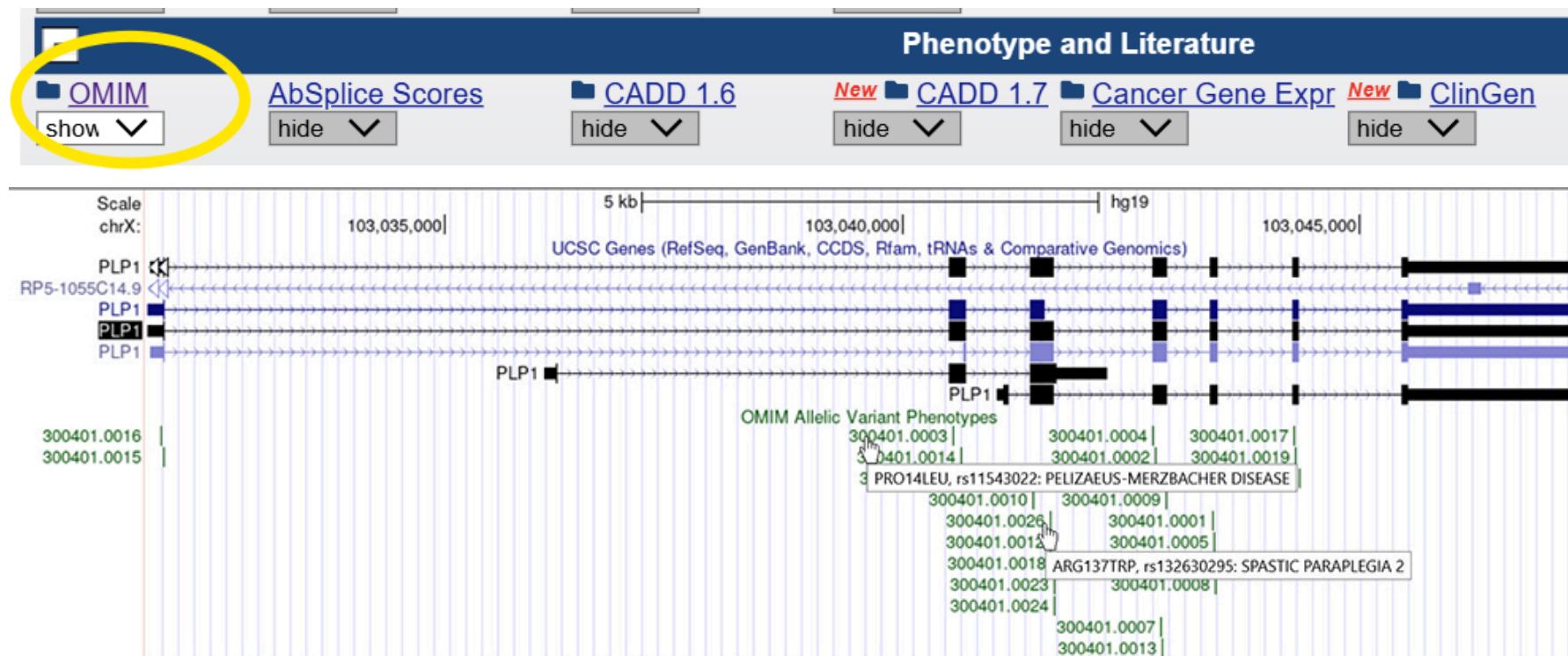
Identify strand direction and exon²



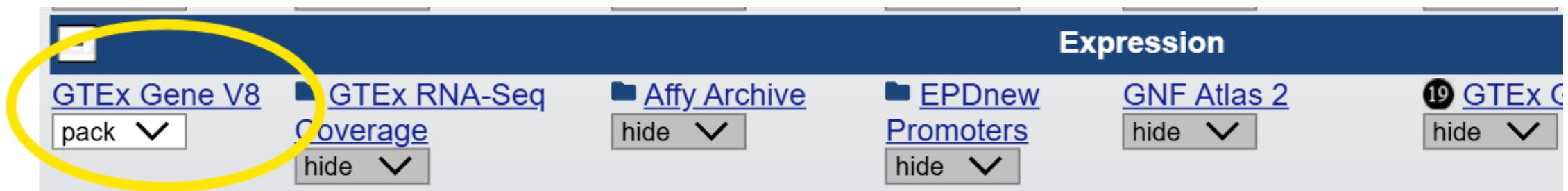
- Direction of arrows: 5' > 3' of a gene
- Exons: solid blue bars

Find disease variants using OMIM track

Different variants of PLP1 are associated with different disease states ³



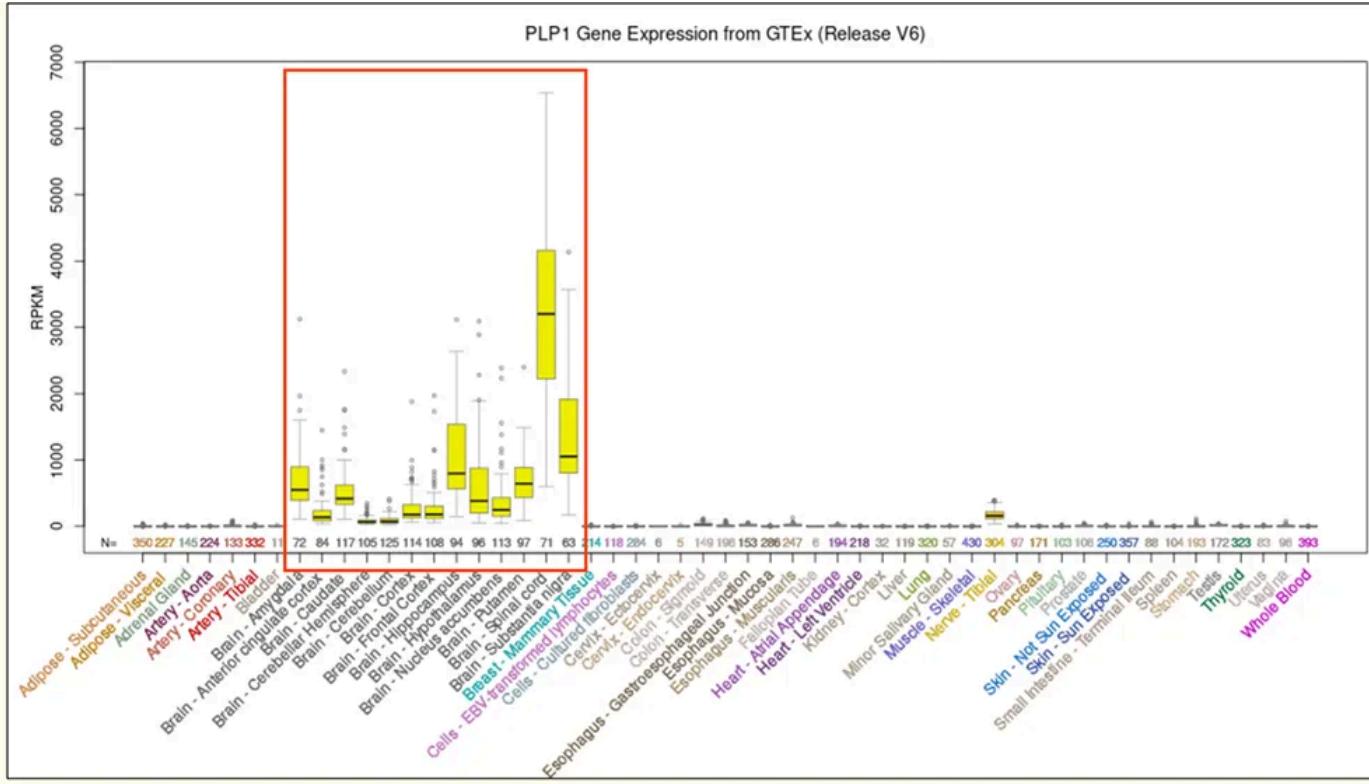
Find tissue-specific gene expression using GTEx V8 track



Tissue Specificity of Human Disease Module ⁴

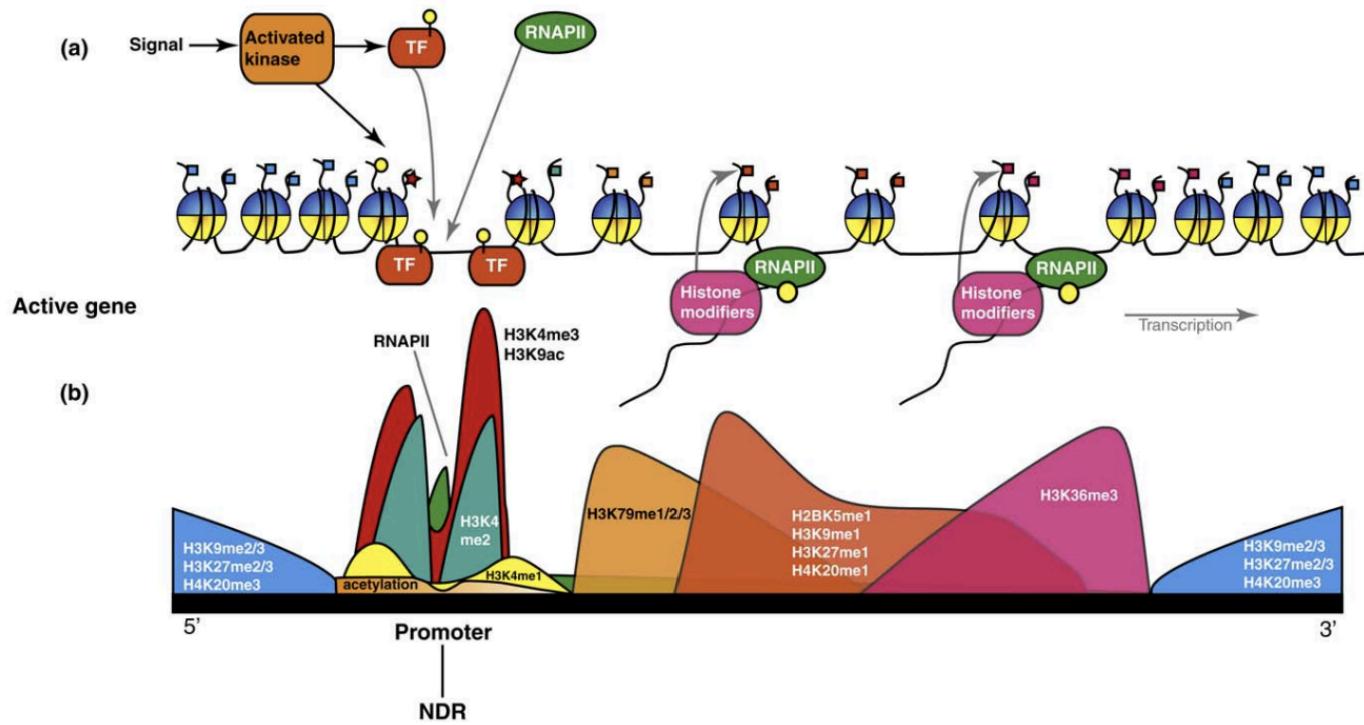
- Disease-associated genes are more likely to exhibit tissue-specific expression than non-disease-associated genes
- The integration of gene expression, disease manifestation, molecular network connectivity, and tissue specificity data leads to better predictions of novel disease-gene candidates than any of these elements alone

Find tissue-specific gene expression using GTEx V8 track



- Genes can be expressed ubiquitously or only in a specific cell type
- How do gene expression at specific tissues relate to the clinically expressed symptoms of this disease?

Distribution of histone modifications on active genes⁵

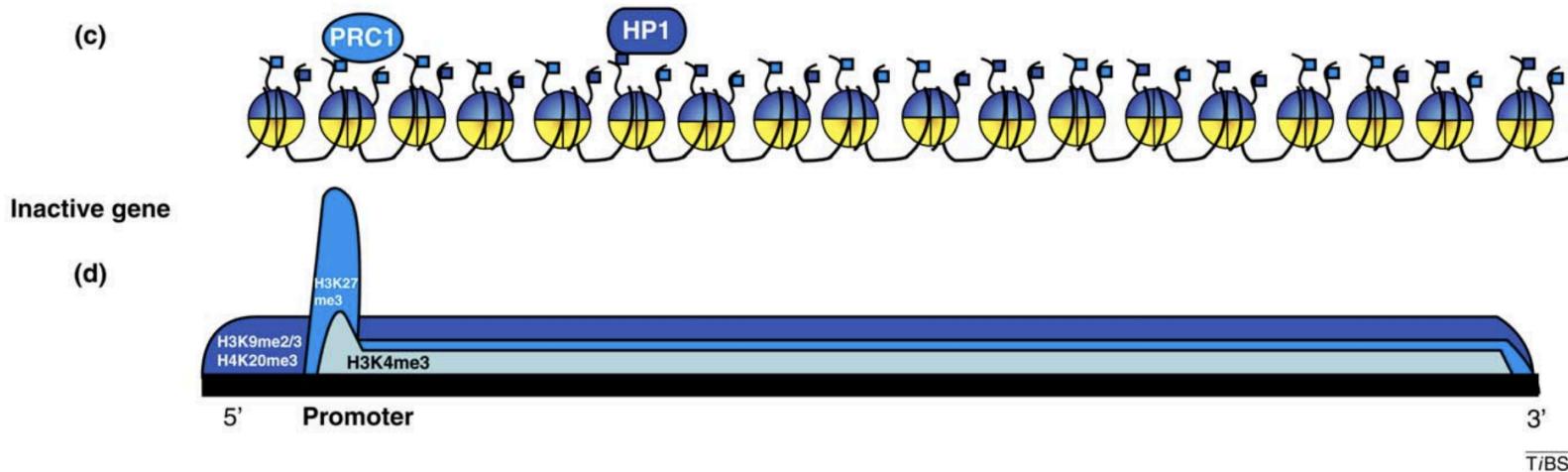


- Acetylated histones are found at or around most actively transcribed regions.
- High levels of active modifications such as acetylations and methylation of H3K4 indicate an actively transcribed gene.

Distribution of histone modifications on active genes⁵

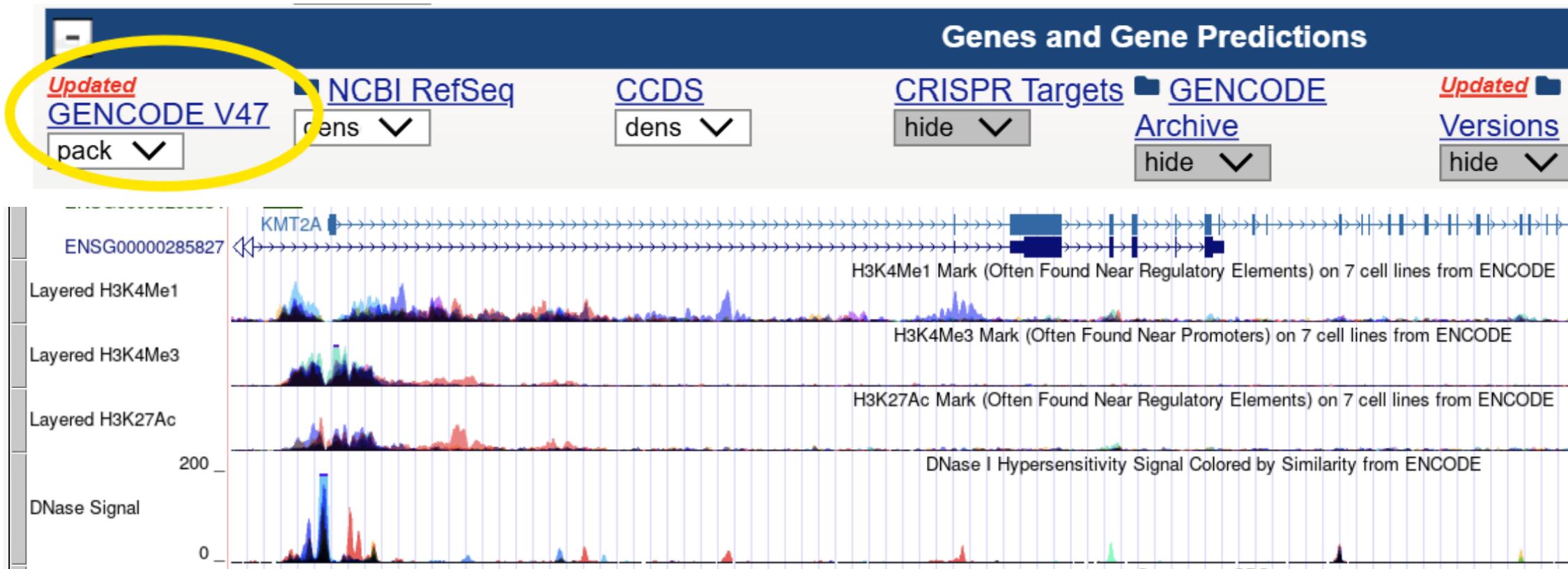
- The acetylation or the phosphorylation of histones serve to modulate gene expression in response to external signals rather than by establishing a stable gene expression pattern.
- The methylation of histones plays a key role in the stabilization of epigenetic traits.
 - Genetics mutations affecting histone lysine methyltransferases or proteins that bind to methylated lysines frequently lead to major developmental defect.
(e.g. KMT2A)

Distribution of histone modifications on silenced genes⁵



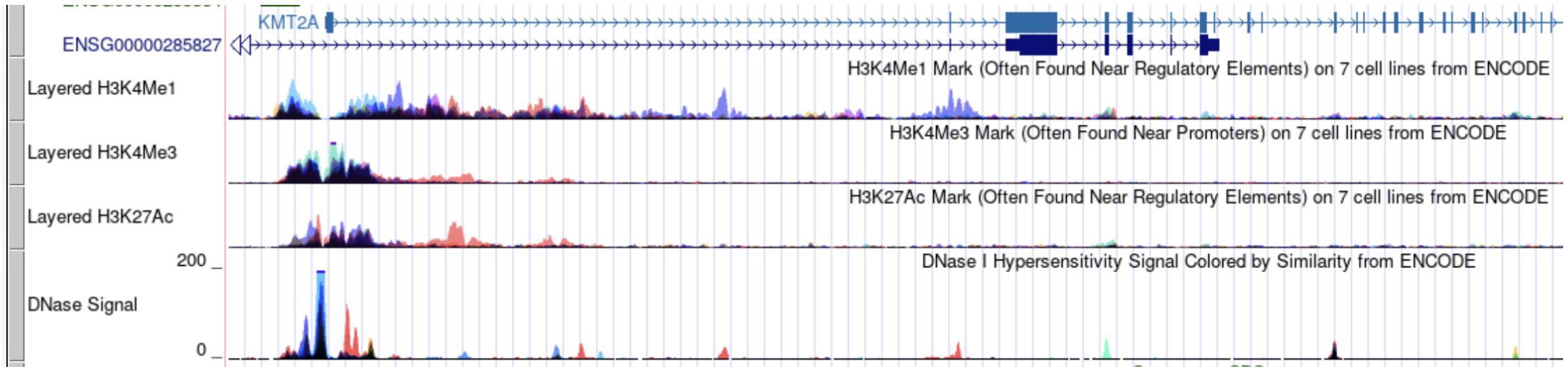
- H3K9 methylation and H4K20 methylation are silencing modifications that are evenly distributed across inactive genes.
- H3K27 methylation is enriched in the promoter.

Locate regulatory elements using ENCODE Tracks⁶



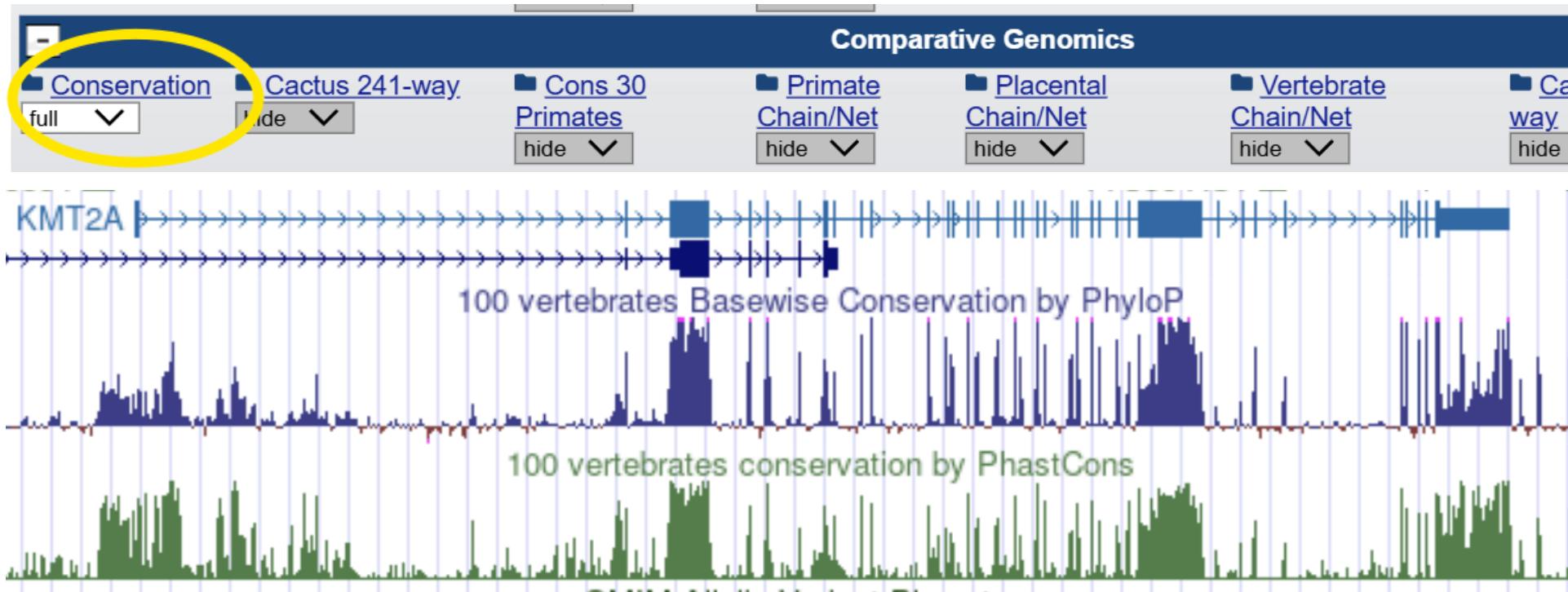
- Data is an overlay from seven different cell lines.

Locate regulatory elements using ENCODE Tracks⁵



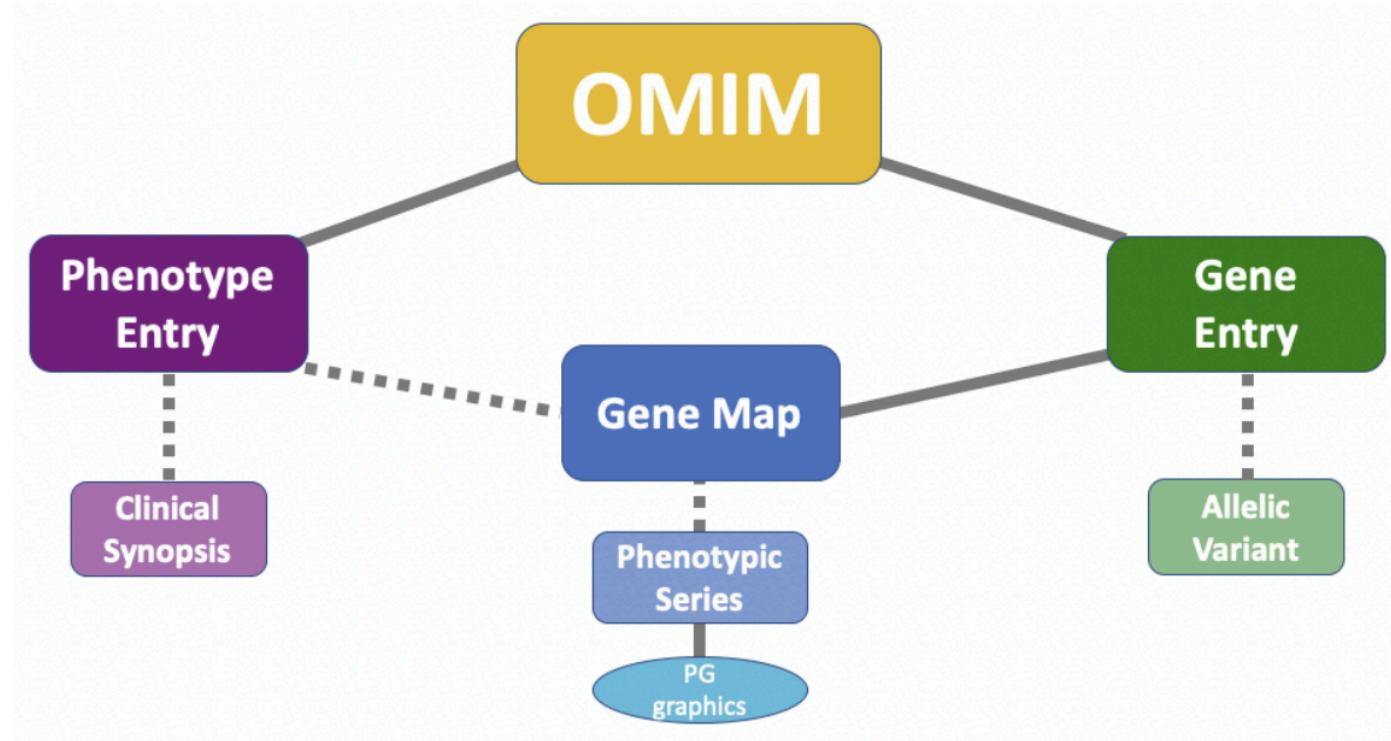
- **H3K Tracks:** Height represents intensity of acetylation of lysine on the H3 histone.
 - Higher peaks = more acetylation
 - High histone acetylation often indicates regulatory elements.
- **DNase I Hypersensitivity Track:** peaks indicate potential regulatory or promoter regions

Find conserved regions using Conservation Tracks⁵



- Highly conserved sequences often have important biological functions
 - Exons are highly conserved across species.
- Most pathogenic variants are located at conserved positions (high positive phyloP scores)⁶

OMIM



- OMIM focuses on the relationships between phenotypes and genes.
- Not all genes have allelic variants; not all phenotypes are mapped; not all phenotypes have Clinical Synopses; and mapped phenotypes are not necessarily part of a Phenotypic Series.

OMIM

1. Search by name of the gene or genetic disorder.

OMIM®

An Online Catalog of Human Genes and Genetic Disorders

Updated November 15th, 2024

2. There are a few ways to view the search results. This is the default view.

KMT2A Search Options ▾

View Results as: Gene Map Table Clinical Synopsis ?

Display: Highlights

Search: 'KMT2A '

Results: 15 entries.

Show 100 | Download As ▾ | « First | < Previous | [Next](#) | Last »

1: * [159555. LYSINE-SPECIFIC METHYLTRANSFERASE 2A; KMT2A](#)
MLL/AF4 FUSION GENE, INCLUDED
Cytogenetic location: [11q23.3](#), Genomic coordinates (GRCh38): [11:118,436,492-118,526,832](#)
Matching terms: [kmt2a](#)
► Gene-Phenotype Relationships ► ICD+ ► Links

2: # [605130. WIEDEMANN-STEINER SYNDROME; WDSTS](#)
Cytogenetic location: [11q23.3](#)
Matching terms: [kmt2a](#)
► Phenotype-Gene Relationships ► ICD+ ► Links

3. Select "Gene Map Table"

KMT2A

View Results as: **Gene Map Table** Clinical Synopsis

Display: Highlights

Search: 'KMT2A'

Results: 15 entries.

Show 100 | Download As ▾ | « First | < Previous | [Next](#) | Last »

Now, you can see more information (e.g. phenotype and inheritance pattern)

Search: 'KMT2A (Search in: Entries with: Genemap; Retrieve: gene map)'

Results: 13 entries.

Show 100 | Download As ▾ | « First | < Previous | [Next](#) | Last »

[Phenotype Only Entries](#) [All Entries](#)

Genomic context table	Location (from NCBI, GRCh38)	Gene/Locus	Gene/Locus name	Gene/Locus MIM number	Phenotype	Phenotype MIM number	Inheritance	Pheno map key	Comments	Mouse symbol (from MGI)
1:	2:218,710,835 2q35	TTLL4	Tubulin tyrosine ligase-like 4	618738						Ttl4
2:	4:48,497,357 4p11	FRYL, KIAA0826	FRY-like transcription coactivator	620798						Fryl
3:	8:41,929,479 8p11.21	KAT6A, MYST3, MOZ, ZNF220, ARTHS	K(lysine) acetyltransferase 6A	601408	Arboleda-Tham syndrome	616268	<input type="checkbox"/>	AD	3	Kat6a
4:	10:806,914 10p15.3	LARP4B, LARP5, KIAA0217	La ribonucleoprotein 4B	616513						Larp4b
5:	11:118,436,492 11q23.3	KMT2A, MLL, HRX, WDSTS	Lysine-specific methyltransferase 2E	159555	Wiedemann-Steiner syndrome	605130	<input type="checkbox"/>	AD	3	fuses with ENL, AF4, AF9, GMPS

OMIM: Disease Variants

604580 Download As ▾

FIBULIN 5; FBLN5

Allelic Variants (16 Selected Examples) :

Number	Phenotype	Mutation	SNP	gnomAD	ClinVar
.0001	CUTIS LAXA, AUTOSOMAL RECESSIVE, TYPE IA	FBLN5, SER227PRO	rs28939370	-	RCV000005809...
.0002	CUTIS LAXA, AUTOSOMAL DOMINANT 2 (1 patient)	FBLN5, 483-BP DUP	-	-	RCV000005810
.0003	MACULAR DEGENERATION, AGE-RELATED, 3	FBLN5, VAL60LEU	rs121434299	rs121434299	RCV000005811...
.0004	MACULAR DEGENERATION, AGE-RELATED, 3	FBLN5, ARG71GLN	rs121434300	rs121434300	RCV000005812...
.0005	MACULAR DEGENERATION, AGE-RELATED, 3	FBLN5, PRO87SER	rs121434301	rs121434301	RCV000005813...
.0006	▼ ALLELIC VARIANTS (16 Selected Examples):				
.0007	Table View ClinVar				
.0008	.0001 CUTIS LAXA, AUTOSOMAL RECESSIVE, TYPE IA				
.0009	FBLN5, SER227PRO rs28939370 RCV000005809...				
.0010	<p>Loeys et al. (2002) studied a large consanguineous Turkish family, originally described by Van Maldergem et al. (1988), in which 4 patients were affected by autosomal recessive cutis laxa type I (ARCL1A; 219100) and demonstrated homozygosity for a 998T-C transition in the FBLN5 gene. The mutation was predicted to result in a ser227-to-pro (S227P) substitution in the fourth cbEGF-like domain of fibulin-5 protein. Because serine is found in this position in mouse and rat fibulin-5 as well as in human fibulin-3, substitution for this amino acid may have important structural and functional consequences for normal elastogenesis. </p>				
.0011	<p>Hu et al. (2006) showed that S227P mutant fibulin-5 is synthesized and secreted at a reduced rate compared to wildtype. The mutant also failed to be incorporated into elastic fibers by transfected rat lung fibroblasts. Purified recombinant S227P fibulin-5 showed reduced affinity for tropoelastin in solid-phase binding assays as well as impaired association with fibrillin-1 microfibrils; in addition, the mutant protein triggered an endoplasmic reticulum (ER) stress response, as indicated by strong colocalization of the mutant with folding chaperones in the ER and by increased rates of apoptosis in</p>				

All ClinVar Variants

External Links

- ▶ Genome
- ▶ DNA
- ▶ Protein
- ▶ Gene Info
- ▶ Clinical Resources
- ▶ Variation
 - ClinVar
 - gnomAD
 - GWAS Catalog
 - GWAS Central
 - HGMD
 - NHLBI EVS
 - PharmGKB
- ▶ Animal Models
- ▶ Cellular Pathways

- You can find disease-causing variants in the Allelic Variants section of gene entries.

- Only select mutations are included. Selection criteria include:
 - the first mutation to be discovered,
 - high population frequency,
 - distinctive phenotype,
 - historic significance,
 - unusual mechanism of mutation,
 - unusual pathogenetic mechanism,
 - and distinctive inheritance.
- To see more variants in a gene, follow links in OMIM to ClinVar, gnomAD, and many other variant resources.

Lab Assignment 1 Instructions

- You are assigned to a group. Each group is responsible for one genetic disease from Module 14.
- Use the gene of interest for your group's genetic disease to complete Lab Assignment 1.