

Noncoding DNA

Molecular architecture, regulatory variation, and implications for pathology

Dr Abu Suraih Sakhri

MBBS, MD (Pathology)

www.abusuraihsakhri.com

Learning outcomes

- Classify the major functional non–protein-coding DNA elements.
- Explain promoter–enhancer regulation and the distinction between sequence and expression.
- Differentiate SNPs, linkage disequilibrium, and copy-number variants.
- Interpret how noncoding variation can contribute to phenotype and disease.

Molecular pathology begins before protein synthesis



Disease may arise from a changed regulatory program, even when the protein-coding sequence is intact.

Human genome: Approximate scale

- The human genome comprises approximately 3.2 billion DNA base pairs.
- Approximately 20,000–25,000 protein-encoding genes.
- Only about 1.5% of genomic DNA is protein-coding.

Coding sequence $\neq$ all biologically important sequence

Protein-coding sequence

- Specifies amino acid sequence
- Altered coding can change protein structure
- Occupies a small genomic fraction

Noncoding sequence

- Can regulate transcription or RNA fate
- Includes structural chromosomal regions
- May influence gene expression without altering protein sequence

Why gene number does not explain organismal complexity

- Organisms with similar gene counts can display major phenotypic differences.
- A gene can yield multiple transcripts and protein isoforms.
- Cellular identity also depends on when, where, and how much each gene is expressed.

“Genome dark matter”: A historical label

- Noncoding does not mean transcriptionally silent or biologically irrelevant.
- Pervasive transcription and gene-regulatory potential.
- Many transcribed or biochemically active segments remain incompletely characterized.

Biochemical activity versus proven biological function

Evidence of molecular activity

- RNA transcription detected
- Protein binding detected
- Accessible chromatin detected

Stronger functional inference

- Sequence perturbation changes output
- Phenotypic impact is reproducible
- Mechanism is supported in relevant cells

A biochemical signal alone is insufficient to establish essential organism-level function.

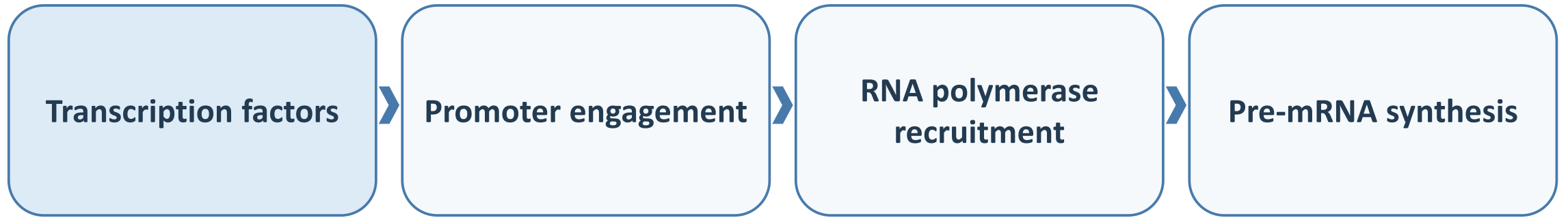
A practical map of noncoding genomic elements

- Regulatory DNA: Promoters and enhancers.
- Chromatin-organization sites: Architecture and access.
- Regulatory RNAs: Transcribed but not translated.
- Structural regions: Telomeres, centromeres, repeats.
- Mobile elements: Transposon-derived genomic sequence.

Promoters: Local initiation control

- Promoter regions recruit transcription machinery and associated regulatory proteins.
- Their relationship to the transcription start site helps define gene-specific initiation.
- A promoter can be present yet inactive in an inappropriate cellular context.

From promoter occupancy to primary transcript



Simplified mechanism; Chromatin accessibility and signaling state can modify every step.

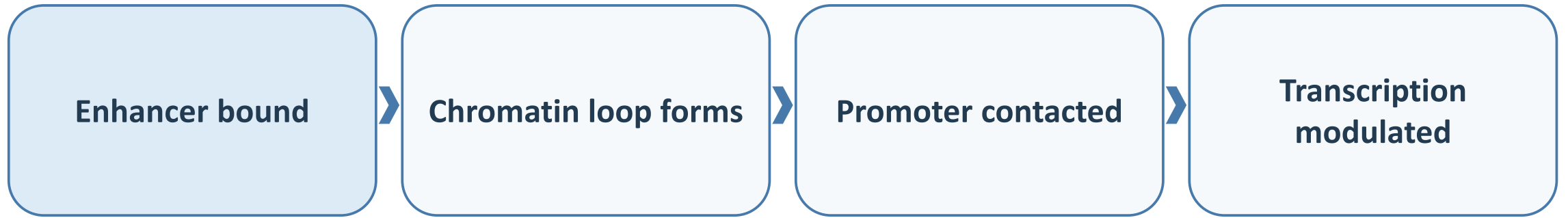
Promoter variants alter transcription, not necessarily protein sequence

- A noncoding substitution may modify a transcription-factor binding motif.
- The resulting effect can increase, decrease, or leave expression unchanged.
- Functional consequences depend on the relevant cellular environment.

Enhancers regulate transcription at a distance

- Enhancers recruit regulatory proteins and can influence transcription from remote genomic positions.
- DNA looping that allows an enhancer to interact with a promoter.
- Distance in the linear DNA sequence is not the same as distance in the folded nucleus.

Enhancer–promoter coupling



The same coding gene may have multiple enhancer inputs.

Cell-type specificity emerges from regulatory wiring

- A regulatory sequence may be active in one cell type and inactive in another.
- Cell lineage influences transcription-factor abundance and chromatin accessibility.
- Identical genomes can generate different expression programs and cellular identities.

Promoter versus enhancer

Promoter

- Typically near its target initiation region
- Central to assembly of transcription machinery
- Initiation-site oriented

Enhancer

- Can act across substantial genomic distances
- Modulates promoter output
- Often highly context-dependent

Cis-regulatory versus trans-acting factors

Cis-regulatory element

- DNA site located on the same DNA molecule
- Examples: Promoter or enhancer
- A local sequence variant may change binding

Trans-acting factor

- Diffusible RNA or protein regulator
- Often binds multiple DNA targets
- Its altered abundance can affect many genes

Silencers and insulators: Additional regulatory vocabulary

- Silencers reduce expression in a context-dependent manner.
- Insulator/boundary-associated elements can constrain inappropriate enhancer–promoter interactions.
- Their effects reflect the interplay of sequence, proteins, and three-dimensional chromatin.

One pathogenic regulatory route



A change in expression can be clinically important even without a coding mutation.

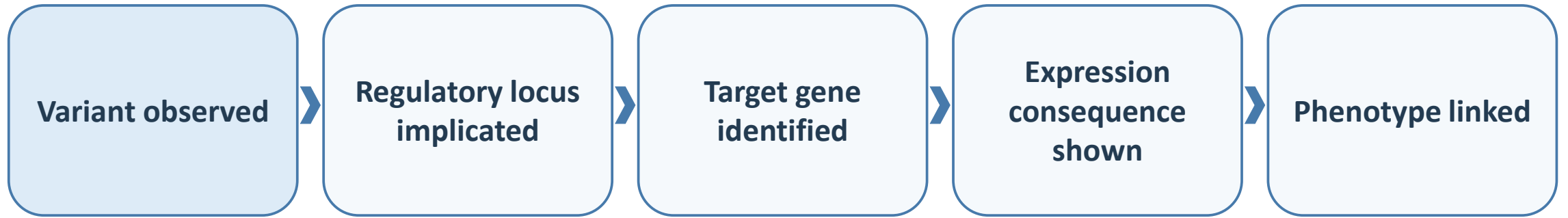
A regulatory variant can act by gain or loss

- Gain: Create or strengthen an activating motif.
- Loss: Disrupt an activating motif or a regulatory contact.
- Direction of gene-expression change is empirical; It cannot be inferred from variant location alone.

The gene nearest a variant is not always its target

- Long-range regulation can connect a distal element with a non-neighboring gene.
- Physical proximity in linear genome coordinates is an unreliable sole targeting rule.
- Experimental or multi-omic evidence helps connect the variant to its effector gene.

Regulatory variant interpretation: Causal logic



Association, regulatory localization, and functional validation are distinct evidentiary steps.

Chromatin organization is not merely packaging

- Genomic sites that organize and maintain higher-order chromatin.
- DNA folding can influence which regulatory elements contact one another.
- Architecture shapes access of regulatory proteins to genomic sequences.

Euchromatin and heterochromatin

Euchromatin

- Relatively dispersed on light microscopy
- Generally accessible and transcriptionally permissive
- Can support active gene expression

Heterochromatin

- Densely packed chromatin
- Often transcriptionally repressed
- Important for chromosome organization

Three-dimensional contacts complicate linear maps

- Enhancers and target promoters may be far apart along the DNA strand.
- Folding changes physical proximity within a nucleus.
- A regulatory variant may influence a gene outside its nearest annotated locus.

Chromatin state connects DNA to transcription



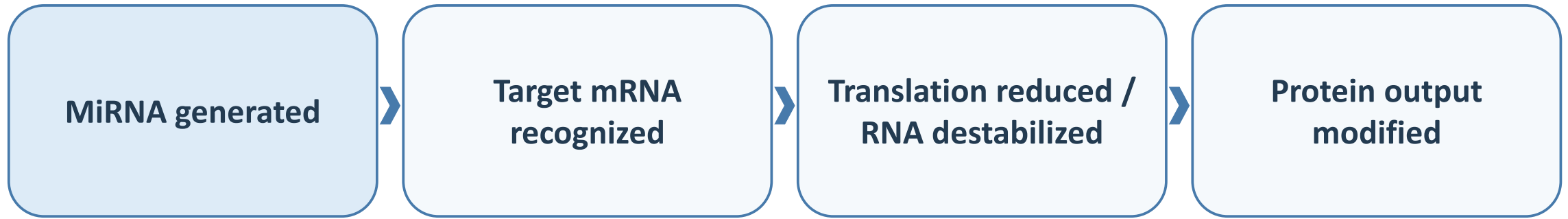
Interpretation pitfall: One regulatory map is not universal

- Chromatin and transcription-factor profiles vary across tissues and states.
- A candidate enhancer should be evaluated in the biologically relevant cell population.
- Absence of activity in one assay does not prove inactivity across all tissues.

Noncoding DNA can encode functional RNA

- Noncoding regulatory RNAs as a major class of noncoding elements.
- The source specifically highlights microRNAs and long noncoding RNAs.
- These RNAs regulate gene output without being translated into proteins.

MicroRNA: A post-transcriptional control route



A single miRNA can participate in a network of mRNA targets.

MicroRNA: Implications for cellular phenotype

- MicroRNAs frequently fine-tune rather than switch gene expression absolutely.
- One regulatory RNA can influence multiple components of a pathway.
- Dysregulation can contribute to changes in cell differentiation, survival, or proliferation.

Long noncoding RNA: Diverse mechanisms

- Long noncoding RNAs are a heterogeneous group, not one mechanistic category.
- Some participate in chromatin-associated regulatory complexes.
- Others act through RNA–protein or RNA–RNA interactions; Function must be established case by case.

Regulatory RNA versus untranslated region

Regulatory noncoding RNA

- Transcribed as its own RNA product
- Can regulate other transcripts or chromatin
- Examples: MiRNA, lncRNA

UTR of a coding mRNA

- Part of a protein-coding transcript
- Not itself translated into amino acids
- Can affect stability, localization, and translation

Why widespread transcription requires careful interpretation

- RNA detection reveals transcription, not automatically necessity.
- Noise, transient intermediates, and functional products must be distinguished.
- Perturbation plus appropriate phenotype provides stronger causal support.

Noncoding sequence can have structural functions

- Telomeres protect chromosome termini.
- Centromeres support chromosome segregation machinery.
- Satellite repeats contribute to centromeric and heterochromatic organization.

Telomeres: Protective chromosome termini

- Telomeres comprise repetitive DNA at the ends of linear chromosomes.
- They distinguish physiological chromosome ends from DNA breaks.
- Changes in telomere maintenance are relevant to replicative capacity and genome stability.

Telomere failure: Conceptual consequence



Centromeres: Chromosome-segregation sites

- Centromeric DNA is associated with kinetochore formation.
- Kinetochore machinery mediates interaction with the mitotic spindle.
- Centromere function is therefore central to accurate chromosome segregation.

Satellite DNA: Repeating structural sequence

- Arrays of short repeated units that can extend for large distances.
- Satellite repeats are prominent in centromeric regions.
- Their structural roles should not be confused with classic protein-coding functions.

Telomere versus centromere

Telomere

- Chromosome end
- Protects terminal DNA
- Relates to replication and end stability

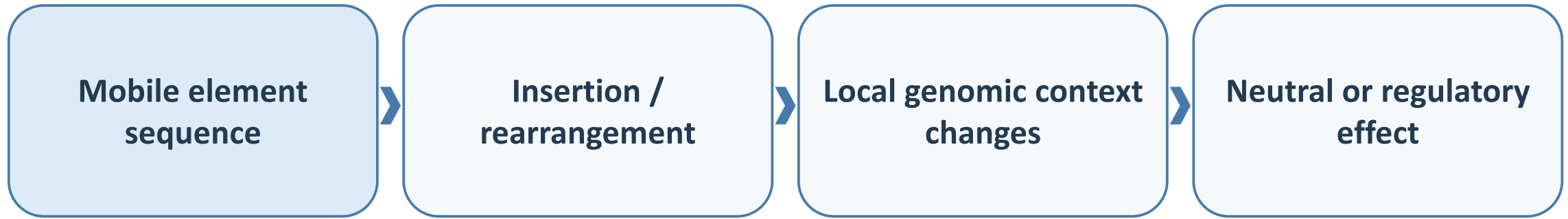
Centromere

- Chromosome segregation region
- Associated with kinetochore
- Relates to spindle attachment and partitioning

Mobile genetic elements occupy a major genomic fraction

- Mobile elements as accounting for more than one third of the human genome.
- Their historical movement contributed to genomic rearrangement and diversity.
- Many copies are no longer capable of autonomous mobility.

Transposon-derived change: Possible trajectories



The outcome depends on the element, its activity, and insertion site.

Possible biological consequences of mobile DNA

- Insertion may disrupt a coding or regulatory region.
- Mobile-element sequence may contribute regulatory motifs.
- Repeated elements can facilitate genomic rearrangements and structural diversity.

Repeat-rich DNA does not imply a single function

Structural or regulatory repeat

- Can support chromosome architecture
- May form regulatory binding environments
- Function can be context-specific

Mobile-element-derived repeat

- Originates from transposon activity
- Many copies are inactive remnants
- A subset affects gene regulation or rearrangement

Most inherited genomic differences are small—but consequential

- Human genomes are overwhelmingly sequence-identical.
- A relatively small differing fraction still represents many variable sites.
- Phenotypic differences reflect variants interacting with regulatory and environmental context.

Single-nucleotide polymorphism: Definition

- An SNP is variation at one nucleotide position among individuals.
- Common SNP loci as typically biallelic.
- SNPs occur in exons, introns, and intergenic sequence.

Coding SNP versus noncoding SNP

Coding-region SNP

- Can be synonymous or change an amino acid
- May influence protein structure or function
- Interpret through transcript and protein context

Noncoding SNP

- May influence transcription or RNA processing
- Often requires tissue-specific regulatory context
- Can be benign, causal, or a correlated marker

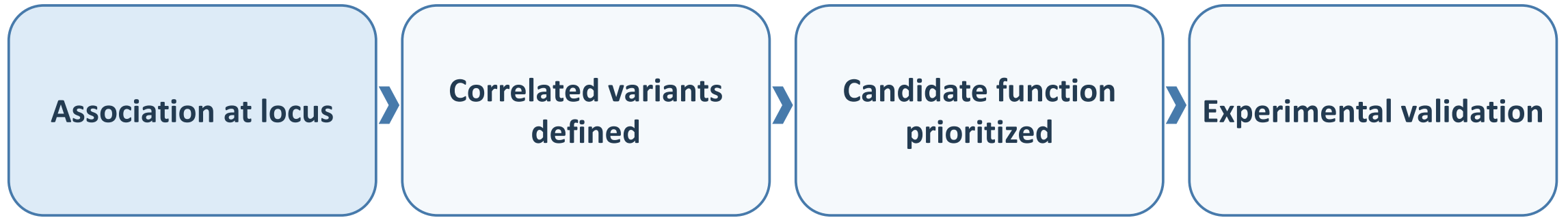
Not every noncoding SNP changes gene expression

- Many variants are functionally neutral in the tested context.
- Some alter a regulatory motif without measurable phenotypic effect.
- Variant classification requires evidence beyond genomic coordinates.

Linkage disequilibrium: Correlation between variants

- Variants physically near each other may be inherited together.
- A neutral SNP can tag a nearby functional variant.
- Thus, a strong association signal may not identify the causal nucleotide.

Genome-wide association: Avoid the causal leap



Linkage disequilibrium is central to the gap between association and mechanism.

A marker SNP versus a causal regulatory SNP

Marker / tag SNP

- Statistically associated with a phenotype
- May be inherited with the causal allele
- Not necessarily a functional DNA change

Causal regulatory SNP

- Sequence alters a functional element
- Changes target expression in relevant cells
- Mechanism connects to phenotype

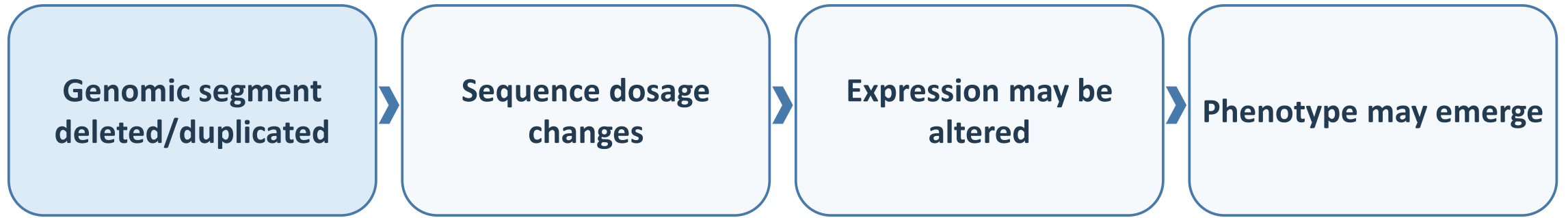
Population differences require careful interpretation

- SNP allele frequencies can vary across ancestral populations.
- Linkage patterns can also vary; A useful marker need not transfer identically between groups.
- Group-level variation cannot substitute for individual functional or clinical evidence.

Copy-number variation: Definition and scale

- CNVs involve deletion or duplication of contiguous genomic segments.
- A scale from roughly 1 kilobase to millions of bases.
- A CNV may include coding genes, regulatory DNA, or both.

Copy-number variation: A dosage framework



A CNV is not automatically pathogenic; Content, size, inheritance, and context matter.

SNP versus CNV

SNP

- Single nucleotide coordinate
- Usually described as alternate alleles
- Can change coding or regulatory sequence

CNV

- Larger DNA interval
- Different copy counts across individuals
- Can involve one or many genes / elements

Some CNVs act through regulatory elements

- A copy-number change can alter enhancer dosage or topology.
- The resulting effect can depend on which gene is controlled.
- A CNV can be significant even without deleting an entire coding gene.

From genomic finding to disease mechanism

- An observed variant is not equivalent to an established cause.
- Interpret the sequence, genomic context, inheritance, phenotype, and functional data together.
- Regulatory variants often demand target-gene and tissue-specific evidence.

Workup for a candidate noncoding variant



Clinical example: TERT promoter variants in cancer

- Somatic TERT promoter mutations occur in several cancer types.
- They can create or modify transcription-factor binding and increase TERT transcription.
- This is a direct example of a pathogenic effect from noncoding sequence.

TERT example: Causal chain



What the TERT example does not establish

- A TERT promoter mutation is not present in all malignancies.
- Detection alone does not establish a particular stage, grade, or prognosis.
- Clinical utility depends on tumor classification, assay, and validated evidence.

Case 1: A promoter variant

A tumor has a somatic nucleotide substitution upstream of a protein-coding gene. The gene's exons remain unchanged, but RNA output is increased. Which mechanism is most direct?

A. Altered promoter regulation

B. Missense substitution

C. Protein truncation

D. Chromosomal nondisjunction

Case 1: Interpretation

A. Altered promoter regulation

- Promoter mutations can change transcription-factor occupancy.
- Protein sequence may be completely unchanged.
- The effect requires gene-expression and context-specific validation.

Case 2: Association is not proof

A GWAS SNP is strongly associated with disease, but the nearest gene has no detectable expression change. What is the best next inference?

A. The SNP is necessarily causal

B. Search linked variants and distal targets

C. The GWAS signal is invalid

D. A coding mutation must explain it

Case 2: Interpretation

B. Search linked variants and distal targets

- Linkage disequilibrium can make a noncausal SNP a strong marker.
- Enhancers may act on distal genes; The nearest gene is not always the target.
- Fine-mapping and functional experiments can clarify mechanism.

Case 3: Interpreting a deletion

A microdeletion removes an intergenic regulatory element but leaves nearby coding exons intact. Which statement is defensible?

A. No effect is possible because no exon is deleted

B. A gene-expression effect is plausible

C. All deletions cause the same phenotype

D. Pathogenicity is proven from location alone

Case 3: Interpretation

B. A gene-expression effect is plausible

- Regulatory dosage and chromatin contacts may be disrupted.
- Target-gene assignment and phenotype must be evaluated.
- Genomic location alone cannot establish pathogenicity.

Common misconceptions in noncoding genomics

- “Noncoding” does not mean “nonfunctional.”
- “Transcribed” does not automatically mean “necessary.”
- “Associated” does not automatically mean “causal.”
- “Nearest gene” does not automatically mean “target gene.”

Integrated comparison: Sequence, regulation, phenotype

- Protein-coding variants may alter the properties of protein products.
- Regulatory variants can alter the abundance, timing, or cellular distribution of otherwise normal proteins.
- Copy-number changes can affect both coding content and regulatory architecture.

Take-home framework

- Recognize five broad noncoding element classes.
- Ask which molecular output a variant could change: Transcription, RNA fate, structure, or dosage.
- Connect molecular effect to a disease-relevant cell type.
- Demand causal evidence before converting genetic association into mechanistic certainty.

Selected references

Primary textbook	Robbins, Cotran & Kumar: Pathologic Basis of Disease, 11e (2026), Ch. 1, pp. 1–3
Supplementary genetics	MedlinePlus Genetics: Noncoding DNA; copy-number variation
Regulatory RNAs	Lee & Kim, Cell (2024); PMID 39547208
Function caveat	Doolittle, PNAS (2013); PMC3619371
Clinical example	Rachakonda et al., Int J Cancer (2021); PMID 34313327
Fine-mapping	Li & Zhou, Nat Rev Genet (2025); doi:10.1038/s41576-025-00869-4