GENETICS · STUDY GUIDE

# Exam 5 Study Guide

This guide covers chapters 9 and 10. Chapter 9 asks how cells decide which genes to use, from bacterial operons to chromatin marks and small RNAs. Chapter 10 hands you the tools geneticists use to read, copy and change DNA, and the genomic thinking those tools made possible. Both chapters reward careful bookkeeping, so the worked examples matter. Work one chapter at a time, copy the tables by hand, cover the steps and try each example. Then take the ten item self-test. The cram sheet at the end holds every rule and formula.

WHAT THIS GUIDE COVERS

- Chapter 9 · cis and trans, positive and negative control, synthesis and decay, the lac and trp operons, partial diploids, attenuation, activators and enhancers, chromatin marks, DNA methylation, imprinting, X inactivation, bisulfite sequencing, siRNA and miRNA.
- Chapter 10 · restriction enzymes and cloning, PCR and qPCR, Sanger and modern sequencing, coverage and assembly, CRISPR and its relatives, genome size and annotation, RNA-seq, orthologs and paralogs, and dN/dS.

CHAPTER 9

## Gene Regulation and Epigenetics

### What to know

Regulation saves resources and builds cell identity. Expression can be controlled at chromatin, transcription initiation, RNA processing, export, mRNA stability, translation and protein turnover. Transcription initiation is the most common and cheapest control point.

A **cis-acting element** is a sequence, such as a promoter, operator or enhancer, that controls only the gene on the same molecule. A **trans-acting factor** is a diffusible product, such as a repressor,

activator or microRNA, that can act anywhere. In a partial diploid, a wild type trans factor rescues a defective one, but a defective cis site is never rescued.

Two separate questions classify a system. Is the bound protein a repressor (negative control) or an activator (positive control)? Does the small molecule turn the pathway on (inducible) or off (repressible)? Catabolic operons are usually inducible and biosynthetic ones repressible. Levels reflect synthesis and decay: $k = \ln 2 \div t_{1/2}$ and steady state $N = s \div k$, and a new level is about 97 percent reached in five half lives.

The **lac operon** (*lacZ*, *lacY*, *lacA*) is negative and inducible. Without lactose the repressor sits on the operator. Allolactose pulls it off. It is also under positive control: when glucose is low, cAMP rises and CAP-cAMP helps polymerase bind. Glucose also blocks permease (inducer exclusion). High expression needs lactose present and glucose absent. In partial diploids, $lacI^-$ is recessive, $lacI^S$ is dominant, and $lacO^c$ is cis dominant.

The **trp operon** is negative and repressible: tryptophan is the corepressor that lets TrpR bind the operator. **Attenuation** fine tunes it. A leader with two Trp codons decides which hairpin forms. With plenty of charged $tRNA^{Trp}$, the ribosome covers region 2 and the 3:4 terminator forms. With little, the ribosome stalls on region 1 and the 2:3 antiterminator forms. With no translation, 1:2 forms and 3:4 terminates. Repression and attenuation multiply to about 600 fold.

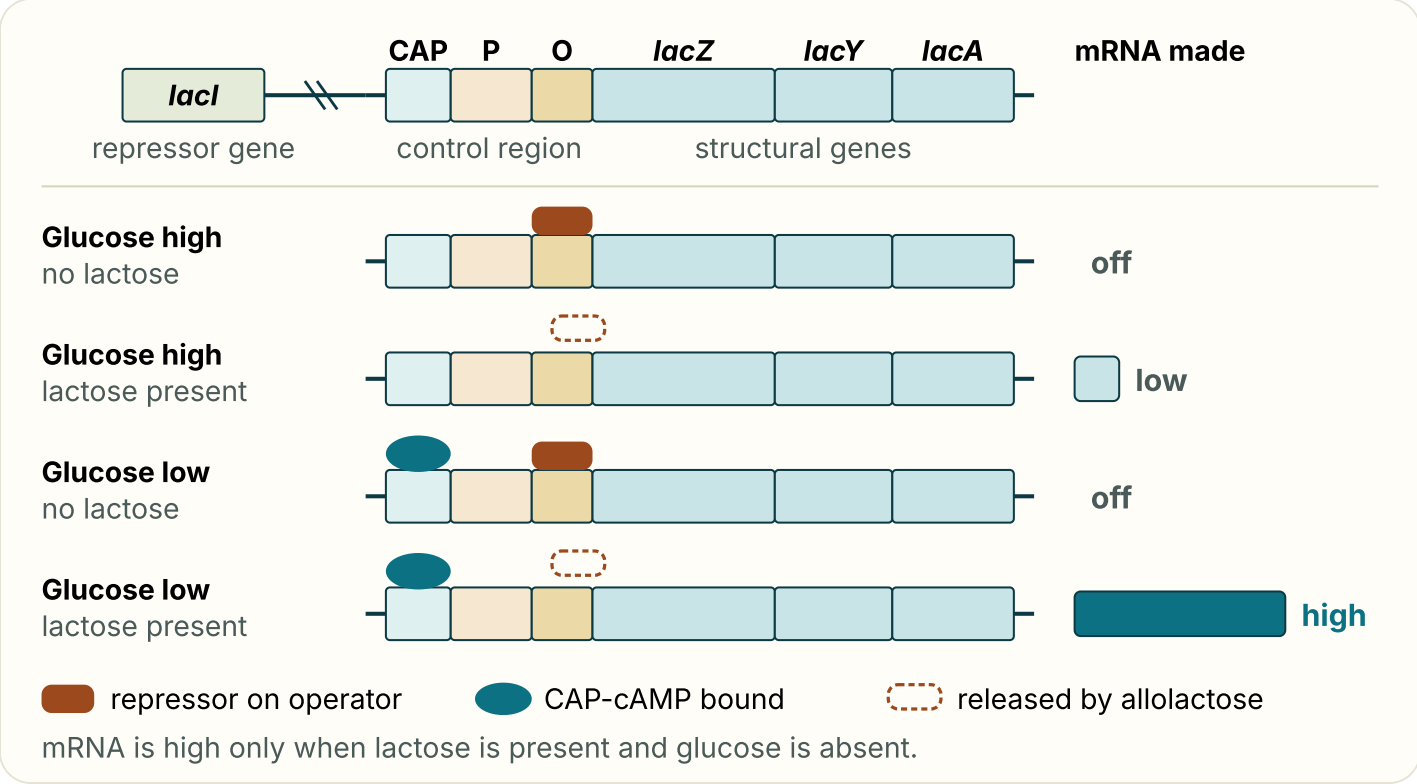
Eukaryotic **activators** have separable DNA binding and activation domains, as GAL4 domain swaps showed. Binding motifs include helix-turn-helix, zinc fingers, leucine zippers and helix-loop-helix. **Enhancers** work at a distance, in either orientation and from any side, by looping to the promoter through **Mediator**. Insulators bound by CTCF fence them into TADs. Reporter genes with deletions map enhancers.

Chromatin gates access. Acetylation (H3K27ac) and H3K4me3 mark active genes. H3K27me3 (Polycomb) and H3K9me3 (HP1) mark silent ones. **CpG methylation** silences promoters. DNMT3A and DNMT3B set it, DNMT1 copies it at hemimethylated sites after replication, and TET enzymes remove it. That copying is true epigenetic memory. **Imprinting** uses parent specific methylation, as at *Igf2* and *H19* with CTCF, and X inactivation uses Xist RNA. **Bisulfite** turns unmethylated C into U (read as T) while 5-methylcytosine stays C.

Small RNAs guide Argonaute. **siRNAs** from long dsRNA pair perfectly and direct slicing. **miRNAs** come from the cell's own hairpins: Drosha-DGCR8 crops them in the nucleus, Exportin-5 moves them out and Dicer cuts the loop. Animal miRNAs pair through the **seed** (nucleotides 2 to 8) in 3′ UTRs and cause repression and decay. A target site is the reverse complement of the seed. piRNAs with PIWI proteins guard the germ line.

**LOCK THESE IN**

- Cis affects only its own molecule. Trans diffuses.
- Negative or positive asks what the protein does. Inducible or repressible asks what the signal does.
- $N = s \div k$, $k = \ln 2 \div t_{1/2}$. Doubling half life doubles level.
- lac is high only with lactose present and glucose absent.
- *lacI*$^{-}$ recessive; *lacI*$^{S}$ dominant; $O^{c}$ cis dominant.
- Trp plenty: 3:4 terminator. Trp scarce: 2:3 antiterminator. No translation: 1:2, then 3:4.
- DNMT1 copies methylation at hemimethylated CpGs.
- miRNA seed = nucleotides 2 to 8. The mRNA site is its reverse complement.




**Two switches on one operon.** The repressor is the negative switch and CAP-cAMP is the positive one. Transcription is high only when lactose is present and glucose is absent.

## Rules and formulas for regulation

| Item | Rule or formula | Note |
|---|---|---|
| **Decay constant** | $k = \ln 2 \div t_{1/2}$ | per unit time |
| **Steady state** | $N = s \div k$ | s = synthesis rate |
| **Time to new level** | 50% in 1 $t_{1/2}$; 75% in 2; about 97% in 5 | set by the new half life |
| **lac logic** | ON only if lactose present AND glucose absent | repressor off, CAP-cAMP on |
| **Partial diploid, repressor** | any functional repressor reaches every $O^+$ | trans |
| **Partial diploid, operator** | $O^c$ makes only its own linked genes constitutive | cis |
| **Superrepressor** | $I^S$ shuts every $O^+$ for good; $O^c$ ignores it | dominant in trans |
| **trp total range** | repression about 70 × attenuation about 8 to 10 ≈ 600 fold | they multiply |
| **Reporter fold activation** | construct activity ÷ minimal promoter activity | necessary vs sufficient |
| **Bisulfite reading** | C stays C: methylated; C reads T: unmethylated | non CpG C must read T |
| **Percent methylation** | methylated CpG calls ÷ (CpGs × clones) | average over the region |
| **miRNA site** | 7mer = reverse complement of seed; 8mer adds A across from nt 1 | seed = nt 2 to 8 |

## Worked examples

**WORKED EXAMPLE 1 · SYNTHESIS, DECAY AND STEADY STATE**

An mRNA is made at 20 molecules per minute and has a half life of 3.0 minutes. (a) Find its steady state level. (b) An RNA binding protein doubles the half life to 6.0 minutes. Find the new level and about how long it takes to get there.

#### STEP 1 · DECAY CONSTANT

k = 0.693 ÷ 3.0 = 0.231 per minute.

#### STEP 2 · STEADY STATE

N = 20 ÷ 0.231 = **about 87 molecules**.

#### STEP 3 · NEW CONSTANT

k = 0.693 ÷ 6.0 = 0.116 per minute. N = 20 ÷ 0.116 = **about 173 molecules**, twice the old level.

#### STEP 4 · TIME

About five new half lives: 5 × 6.0 = **30 minutes** to get about 97 percent of the way.

### WORKED EXAMPLE 2 · A LAC PARTIAL DIPLOID

A strain has the chromosome $I^+ O^+ Z^- Y^+$ and an F′ with $I^S O^c Z^+ Y^-$. Glucose is absent. Predict β-galactosidase and permease with and without lactose.

#### STEP 1 · THE REPRESSORS

Both copies make repressor. The superrepressor from $I^S$ cannot bind inducer, so it sits on every $O^+$ operator and never leaves.

#### STEP 2 · B-GALACTOSIDASE

The only working *lacZ* is on the F′, next to $O^c$. No repressor binds $O^c$, so it is **constitutive**: made with or without lactose.

#### STEP 3 · PERMEASE

The only working *lacY* is on the chromosome, next to $O^+$. The superrepressor holds that operator even with lactose. So permease is **uninducible**: absent in both conditions.

#### THE HABIT

Work one structural gene at a time. Find its working copy, then ask what its own operator will allow.

### WORKED EXAMPLE 3 · READING A BISULFITE CLONE

A reference fragment is 5′-GCGTACGATCCGT-3′. One bisulfite clone reads 5′-GCGTATGATTCGT-3′. (a) Which CpGs were methylated? (b) Did conversion work? (c) Across 8 clones, CpG 1 is methylated in 8, CpG 2 in 1 and CpG 3 in 5. What is the average methylation?

#### STEP 1 · FIND THE CPGS

Number the bases: G1 C2 G3 T4 A5 C6 G7 A8 T9 C10 C11 G12 T13. CpG 1 is C2-G3, CpG 2 is C6-G7 and CpG 3 is C11-G12. C10 is followed by C, so it is not a CpG.

#### STEP 2 · COMPARE

Position 2 is still C: methylated. Position 6 reads T: unmethylated. Position 11 is still C: methylated. So **CpGs 1 and 3**.

#### STEP 3 · THE CONTROL

Position 10, a non CpG C, reads T. **Conversion worked**.

#### STEP 4 · AVERAGE

(8 + 1 + 5) ÷ (3 × 8) = 14 ÷ 24 = **58 percent**. CpG 2 is mostly unmethylated while CpG 1 is fully methylated.

### WORKED EXAMPLE 4 · A MIRNA SEED AND ITS TARGET

The mature miR-21 is 5′-UAGCUUAUCAGACUGAUGUUGA-3′. Find the seed and write the 7mer and 8mer sites as they would appear in an mRNA, 5′ to 3′.

#### STEP 1 · THE SEED

Number from the 5′ end: U1 A2 G3 C4 U5 U6 A7 U8. The seed (2 to 8) is 5′-AGCUUAU-3′.

#### STEP 2 · COMPLEMENT

Complement of each seed base, in order 2 to 8: U C G A A U A.

#### STEP 3 · REVERSE IT

The mRNA pairs antiparallel, so read that backward: **5′-AUAAGCU-3′** is the 7mer.

#### STEP 4 · THE 8MER

Add an A on the 3′ side, across from nucleotide 1: **5′-AUAAGCUA-3′**.

#### COMMON SLIP

A UTR that contains the seed itself (AGCUUAU) cannot pair. You need the reverse complement.

### TRAPS PEOPLE FALL INTO

- **Calling the trp system positive because it is repressible.** TrpR is a repressor, so it is negative control. Repressible is the other axis.
- **Expecting a wild type operator elsewhere to rescue $O^c$.** Operators act only in cis.
- **Expecting a wild type *lacI* to rescue $lacI^S$.** The superrepressor is dominant. It binds and never leaves.
- **Forgetting the ribosome in attenuation.** Ask where the ribosome sits, then which pairings are left.
- **Thinking enhancers must sit upstream.** They work upstream, downstream, in introns and in either orientation.
- **Calling acetylation a long term memory mark.** Most acetylation turns over in minutes. DNA methylation is copied by DNMT1.
- **Copying the seed as the target site.** Write the reverse complement.

- **Treating miRNA and siRNA outcomes the same.** Perfect pairing slices. Seed pairing represses and causes decay.

## Chapter 9 self-test · ten items

### 1. A mutation in a promoter affects only the gene on the same DNA molecule, even with a wild type copy elsewhere. What does that tell you?

A. The promoter makes a trans-acting product

B. The mutation must be in a repressor gene

C. The gene must be part of a regulon

D. The promoter is a cis-acting element

**D. The promoter is a cis-acting element.** Cis sites are binding sites, not products. They cannot be rescued by another copy of the region.

### 2. How is the trp operon classified?

A. Positive and repressible

B. Negative and repressible

C. Positive and inducible

D. Negative and inducible

**B. Negative and repressible.** TrpR is a repressor (negative), and tryptophan turns the pathway off (repressible).

### 3. Lactose and glucose are both present. What is lac transcription like?

A. High, since CAP-cAMP is bound at the promoter

B. High, since the repressor has been released

C. Low, since cAMP is low and permease is blocked

D. Off, since the repressor stays on the operator

**C. Low, since cAMP is low and permease is blocked.** Lactose frees the operator, but glucose keeps cAMP low and causes inducer exclusion. High expression needs glucose absent too.

### 4. A strain is $I^- O^+ Z^+$ / F′ $I^+ O^+ Z^-$. How is β-galactosidase made?

A. Inducible

B. Constitutive

C. Uninducible

D. Absent entirely

**A. Inducible.** Repressor from the F′ diffuses to the chromosomal operator, so $Z^+$ there is repressed until inducer arrives. *lacI*$^-$ is recessive.

### 5. In the trp leader, the start codon is destroyed. Tryptophan is scarce and *trpR* is mutant. What happens?

A. High expression, since the repressor is missing

B. High expression, since the antiterminator forms

C. Low expression, since 1:2 forms and then 3:4 terminates

D. Low expression, since TrpR binds the operator

**C. Low expression, since 1:2 forms and then 3:4 terminates.** With no ribosome on the leader, region 1 pairs with region 2. That leaves 3 free to pair with 4, the terminator.

### 6. A DNA fragment raises reporter output when moved 3 kb downstream and when flipped. What is it?

A. An enhancer
B. A TATA box
C. A core promoter
D. An insulator

**A. An enhancer.** Working at a distance and in either orientation are defining features of an enhancer.

### 7. Which histone mark is written by Polycomb PRC2 and keeps developmental genes silent?

A. H3K4me3
B. H3K36me3
C. H3K27ac
D. H3K27me3

**D. H3K27me3.** PRC2 (EZH2) deposits H3K27me3. H3K4me3 marks active promoters, H3K27ac active enhancers, and H3K36me3 transcribed gene bodies.

### 8. Why are DNA methylation patterns copied through cell division?

A. Histone acetylation guides methyl groups to DNA
B. DNMT1 methylates the new strand at hemimethylated CpGs
C. TET enzymes add methyl groups after replication
D. DNMT3A methylates every CpG in each new cell

**B. DNMT1 methylates the new strand at hemimethylated CpGs.** After replication each CpG is methylated only on the old strand. UHRF1 finds these sites and recruits DNMT1 to fill in the new strand.

**9. *Igf2* is expressed only from the paternal allele. A heterozygous null female is crossed to a wild type male. What fraction of pups are small?**

A. One quarter
B. None
C. One half
D. All of them

**B. None.** Every pup gets a working paternal allele from the father. The maternal copy is silent anyway.

**10. What decides whether Argonaute slices a target or represses it?**

A. How fully the guide pairs with the target
B. Whether the target has a 5′ cap
C. Whether the guide is 21 or 22 nt long
D. Whether Drosha or Dicer cut the guide

**A. How fully the guide pairs with the target.** Full pairing lets AGO2 slice, as with siRNAs. Seed only pairing, typical of animal miRNAs, leads to repression and decay.

Read the full chapter: Chapter 9 · Gene Regulation and Epigenetics

CHAPTER 10

# Molecular Genetic Tools and Genomics

## What to know

Type II **restriction enzymes** cut palindromic sites, usually 4 to 8 bp, leaving sticky or blunt ends. EcoRI cuts G^AATTC and leaves a 4 nt 5′ overhang. Ends from the same enzyme anneal, and **DNA ligase** seals them. Two different enzymes give directional cloning. In random sequence an n base site appears once per $4^n$ bp. Gels sort fragments by size, with smaller ones running farther.

A cloning **vector** needs an origin, a selectable marker and a multiple cloning site. Blue-white screening flags inserts that disrupt *lacZα*. Plasmids hold about 10 to 15 kb, BACs 100 to 300 kb and YACs a megabase or more. To express a human protein in bacteria, clone **cDNA** made by reverse transcriptase from mRNA, because bacteria cannot remove introns.

**PCR** cycles through denaturation (94 to 98 °C), annealing (50 to 65 °C) and extension (72 °C) with two primers and Taq polymerase. Copies ideally double each cycle: $N = N_0 \times 2^n$. In **qPCR** the threshold cycle $C_t$ falls one cycle for each doubling of template. Relative expression uses $\Delta\Delta C_t$: fold change = $2^{-\Delta\Delta Ct}$, normalized to a reference gene.

**Sanger sequencing** uses ddNTPs that lack a 3′-OH, so each one ends a chain. Fragments read from shortest to longest give the new strand 5′ to 3′, and the template is its reverse complement. Short read sequencing by synthesis gives hundreds of millions of 100 to 300 base reads. Long read methods span repeats and can detect modified bases.

**Coverage** is $C = L \times N \div G$, and about $e^{-C}$ of bases get no read. Assembly builds contigs and scaffolds, and repeats longer than reads break it. The first fully complete human sequence arrived only in 2022, with long reads.

**CRISPR** began as bacterial adaptive immunity. SpCas9 scans for an NGG **PAM**, checks the 20 nt just 5′ of it against the guide, and makes a blunt cut 3 bp upstream of the PAM. NHEJ leaves indels, often knocking out the gene. HDR copies a donor to make a precise knock-in. Base editors change C•G to T•A or A•T to G•C without a double strand break, prime editors write small edits, and dCas9 fusions turn genes down or up.

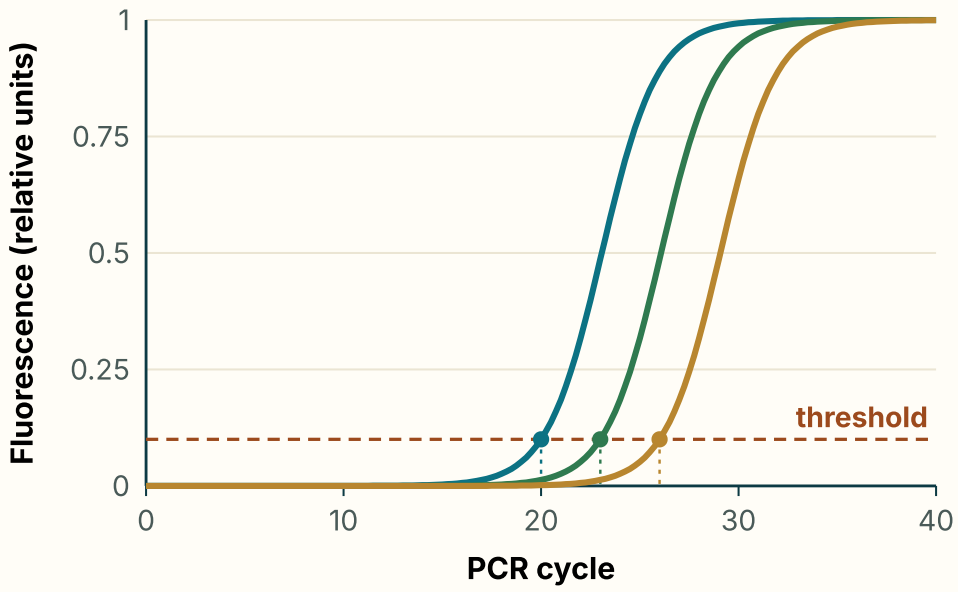
Genome size tracks repeats, not complexity: this is the **C-value paradox**. About 1.5 percent of the human genome codes for protein and about half is transposable element sequence (L1 about 17 percent, Alu about 10 percent). Annotation combines long ORFs, splice signals, RNA-seq and conservation. In random sequence a stop appears about every 21 codons, so long ORFs flag real genes.

RNA-seq counts are normalized, for example as counts per million, and changes are reported as $\log_2$ fold change. ChIP-seq maps protein binding and ATAC-seq maps open chromatin. **Orthologs**

were split by speciation and **paralogs** by duplication. The ratio ω = dN/dS shows selection: below 1 is purifying, near 1 is neutral, above 1 is positive.

**LOCK THESE IN**

- A random n base site appears once per $4^n$ bp: 256, 4,096, 65,536.
- PCR: $N = N_0 \times 2^n$.
- qPCR fold change = $2^{-\Delta\Delta Ct}$. One cycle earlier = twice the template.
- Sanger, shortest to longest, gives the new strand 5′ to 3′.
- Coverage C = LN/G; missed fraction about $e^{-C}$.
- SpCas9: 20 nt protospacer, then NGG; cut 3 bp upstream of the PAM. The guide never includes the PAM.
- About two thirds of random indels shift the frame.
- Orthologs by speciation, paralogs by duplication. ω < 1 means purifying selection.




8× template, $C_t$ 20
1× template, $C_t$ 23
⅛ template, $C_t$ 26

**$8 = 2^3$, so 8× more template crosses the threshold 3 cycles earlier.**

One cycle earlier means twice the template.

$N = N_0 \times 2^n$
fold change = $2^{-\Delta\Delta Ct}$

**Reading $C_t$ in real time PCR.** More starting template crosses the threshold sooner. Each doubling moves $C_t$ one cycle earlier, which is why fold change is a power of 2.

## Formulas for genomic tools

| Quantity | Formula | Note |
|---|---|---|
| Site frequency | once per $4^n$ bp | n = site length |
| Expected sites | genome length ÷ $4^n$ | real genomes deviate |
| PCR yield | $N = N_0 \times 2^n$ | ideal doubling |
| $\Delta C_t$ | $C_t$(target) − $C_t$(reference) | per sample |
| $\Delta\Delta C_t$ | $\Delta C_t$(treated) − $\Delta C_t$(control) | |
| qPCR fold change | $2^{-\Delta\Delta Ct}$ | above 1 is up |
| Coverage | C = L × N ÷ G | L read length, N reads, G genome size |
| Uncovered fraction | $e^{-C}$ | Lander-Waterman |
| Stop free run of k codons | $(61/64)^k$ | stops every 21 codons on average |
| Counts per million | reads ÷ total reads in millions | corrects for depth |
| $\log_2$ fold change | $\log_2$(treated CPM ÷ control CPM) | +1 = doubling |
| dN, dS and ω | differences ÷ sites of that kind; ω = dN/dS | < 1 purifying; > 1 positive |

## Worked examples

WORKED EXAMPLE 1 · MAPPING TWO RESTRICTION SITES

A 12 kb linear fragment cut with EcoRI gives 5 and 7 kb. HindIII gives 2 and 10 kb. Both together give 2, 3 and 7 kb. Map the sites. Then estimate how often an 8 bp site appears in a 3.1 Gb genome.

STEP 1 · PLACE ECORI

Put the EcoRI site 5 kb from the left end.

STEP 2 · TRY HINDIII AT 2 KB

Pieces would be 0 to 2, 2 to 5 and 5 to 12: 2, 3 and 7 kb. That matches.

#### STEP 3 · TRY HINDIII AT 10 KB

Pieces would be 5, 5 and 2 kb. That does not match.

#### STEP 4 · ANSWER

HindIII at 2 kb and EcoRI at 5 kb from the same end, **3 kb apart**. Check: 2 + 3 + 7 = 12.

#### STEP 5 · AN 8 BP SITE

$3.1 \times 10^9 \div 4^8 = 3.1 \times 10^9 \div 65{,}536 =$ **about 47,000 sites**.

### WORKED EXAMPLE 2 · QPCR WITH A REFERENCE GENE

A target gene has $C_t$ 26.0 in control cells and 28.0 in treated cells. The reference gene has $C_t$ 17.0 in control and 16.5 in treated. Find the fold change.

#### STEP 1 · $\Delta C_T$ CONTROL

26.0 − 17.0 = 9.0.

#### STEP 2 · $\Delta C_T$ TREATED

28.0 − 16.5 = 11.5.

#### STEP 3 · $\Delta\Delta C_T$

11.5 − 9.0 = 2.5.

#### STEP 4 · FOLD CHANGE

$2^{-2.5}$ = **0.177**, a drop of about 5.7 fold.

#### WHY NORMALIZE

The raw gap is only 2 cycles, which would suggest a 4 fold drop. The treated sample had a little more RNA loaded, so the real drop is larger.

### WORKED EXAMPLE 3 · COVERAGE AND GAPS

A 3.1 Gb genome is sequenced with 900 million reads of 100 bases. (a) What is the coverage? (b) At only 3x, how many bases would get no read?

#### STEP 1 · COVERAGE

$C = 100 \times 900{,}000{,}000 \div 3{,}100{,}000{,}000 = 90{,}000{,}000{,}000 \div 3{,}100{,}000{,}000 =$ **about 29x**.

#### STEP 2 · MISSED FRACTION AT 3X

$e^{-3} = 0.0498$.

#### STEP 3 · MISSED BASES

$0.0498 \times 3.1 \times 10^9$ = **about 154 million bases**, about 5 percent of the genome.

### WORKED EXAMPLE 4 · DESIGNING A CAS9 GUIDE

Design an SpCas9 guide for this strand and find the cut: 5′-TCAGGTACCATGGACTTCGAAGTCCTGCATCGACTTAGCTAGGTCTAA-3′ (48 nt).

#### STEP 1 · FIND NGG

NGG appears at positions 3, 11 and 41. Only the one at 41 to 43 (AGG) has 20 bases upstream.

#### STEP 2 · THE PROTOSPACER

Positions 21 to 40: AGTCCTGCATCGACTTAGCT.

#### STEP 3 · THE GUIDE

Same sequence in RNA, without the PAM: **5′-AGUCCUGCAUCGACUUAGCU-3′**.

#### STEP 4 · THE CUT

3 bp upstream of the PAM, between positions 37 and 38: ...CGACTTA^GCT AGG.

#### WHAT FOLLOWS

NHEJ leaves small indels. Any indel not divisible by 3 shifts the frame, so most cuts in a coding exon knock the gene out.

### TRAPS PEOPLE FALL INTO

- **Reading the Sanger ladder as the template.** The ladder gives the new strand. The template is its reverse complement.
- **Skipping the reference gene in qPCR.** Raw $C_t$ gaps mislead when loading differs.
- **Comparing raw RNA-seq counts.** Normalize to counts per million first.
- **Putting the PAM in the guide.** The guide is the 20 nt protospacer only.
- **Thinking Cas9 makes the edit.** Cas9 only cuts. The cell's NHEJ or HDR decides the outcome.
- **Expecting genome size to track complexity.** Repeats, not genes, drive size.
- **Calling human and mouse β-globin paralogs.** They split at speciation, so they are orthologs.
- **Using raw difference counts for dN/dS.** Divide by the number of sites of each kind first.

## Chapter 10 self-test · ten items

### 1. About how often does a 4 bp restriction site appear in random sequence?

A. Once every 16 bp

B. Once every 4,096 bp

C. Once every 1,024 bp

D. Once every 256 bp

**D. Once every 256 bp.** $4^4 = 256$. A 6 bp site gives 4,096 and an 8 bp site 65,536.

### 2. Why clone cDNA rather than genomic DNA to make a human protein in bacteria?

A. cDNA carries its own bacterial promoter and origin

B. cDNA is always shorter than any bacterial gene

C. Bacteria cannot remove introns, and cDNA has none

D. Genomic DNA cannot be cut by restriction enzymes

**C. Bacteria cannot remove introns, and cDNA has none.** cDNA is copied from mature mRNA, so the introns are already gone.

### 3. A PCR starts with 50 copies. How many copies are there after 20 perfect cycles?

A. About $1.0 \times 10^3$

B. About $5.2 \times 10^8$

C. About $5.2 \times 10^7$

D. About $2.6 \times 10^6$

**C. About $5.2 \times 10^7$.** $50 \times 2^{20} = 50 \times 1{,}048{,}576 = 52{,}428{,}800$, about $5.2 \times 10^7$.

**4. In qPCR, sample A crosses threshold three cycles before sample B, with equal reference values. How much more template did A start with?**

A. About 3 times as much

B. About 6 times as much

C. About 9 times as much

D. About 8 times as much

**D. About 8 times as much.** Each cycle earlier means twice the template: $2^3 = 8$.

**5. Why does a ddNTP end chain growth in Sanger sequencing?**

A. It lacks the base needed to pair with the template

B. It lacks the 3′-OH needed for the next bond

C. It pairs only with the primer, not the template

D. It carries a dye that blocks the polymerase site

**B. It lacks the 3′-OH needed for the next bond.** Polymerase needs a 3′-OH to attach the next nucleotide. Without it, the chain stops.

**6. A genome is sequenced to an average of 4x. About what fraction of bases gets no read at all?**

A. About 1.8 percent

B. About 25 percent

C. About 4 percent

D. About 0.1 percent

**A. About 1.8 percent.** $e^{-4} = 0.018$. Real coverage is less even than Poisson, so the true gap is larger.

### 7. Where does SpCas9 cut relative to its PAM?

A. 3 bp upstream of the NGG, on both strands

B. Within the NGG, on the target strand only

C. 20 bp downstream of the NGG, on both strands

D. At the 5′ end of the guide, on one strand

**A. 3 bp upstream of the NGG, on both strands.** The HNH and RuvC domains each cut one strand, making a blunt break 3 bp 5′ of the PAM.

### 8. You need to change one A•T pair to G•C with no double strand break. Which tool fits?

A. A cytosine base editor

B. dCas9 fused to a repressor

C. Cas9 nuclease with NHEJ

D. An adenine base editor

**D. An adenine base editor.** An adenine base editor converts A•T to G•C. The cytosine version does C•G to T•A, and dCas9 fusions change expression, not sequence.

### 9. Gene X has 300 reads in a 25 million read control library and 1,200 reads in a 40 million read treated library. What is the $\log_2$ fold change?

A. About 0.50

B. About 1.32

C. About 2.00

D. About 4.00

**B. About 1.32.** Control: 300 ÷ 25 = 12 CPM. Treated: 1,200 ÷ 40 = 30 CPM. 30 ÷ 12 = 2.5, and $\log_2$ 2.5 = 1.32. Raw counts would suggest 4 fold.

**10. Two orthologs show 27 nonsynonymous differences at 450 sites and 12 synonymous differences at 150 sites. What does ω suggest?**

A. 2.25: positive selection on the protein

B. 1.00: neutral evolution with no constraint

C. 0.75: mostly purifying selection

D. 0.44: complete purifying selection

**C. 0.75: mostly purifying selection.** dN = 27 ÷ 450 = 0.06. dS = 12 ÷ 150 = 0.08. ω = 0.06 ÷ 0.08 = 0.75, below 1.

Read the full chapter: Chapter 10 · Molecular Genetic Tools and Genomics

ONE PAGE

# Cram sheet

Everything from chapters 9 and 10 in one place. Cover the right column and see if you can rebuild each row.

## Chapter 9 at a glance

| Topic | Key facts |
|---|---|
| **Cis and trans** | cis = sequences on the same molecule; trans = diffusible products |
| **Classes** | negative (repressor) or positive (activator); inducible or repressible |
| **Steady state** | $k = \ln 2 \div t_{1/2}$; $N = s \div k$; about 97% of the way in 5 half lives |
| **lac** | allolactose releases repressor; CAP-cAMP when glucose low; inducer exclusion; IPTG gratuitous |
| **lac mutants** | $I^-$ recessive constitutive; $I^S$ dominant uninducible; $O^c$ cis dominant constitutive |
| **trp** | TrpR + Trp represses; attenuation: 3:4 terminator vs 2:3 antiterminator; about 600 fold range |
| **Activators** | DNA binding + activation domains; GAL4/Gal80/Gal3; HTH, zinc finger, bZIP, bHLH |
| **Enhancers** | distance, either orientation, any side; Mediator; CTCF insulators and TADs |
| **Histone marks** | active: H3K27ac, H3K4me3; silent: H3K27me3 (Polycomb), H3K9me3 (HP1) |
| **Methylation** | CpG; DNMT3A/B set, DNMT1 copies, TET removes; bisulfite C to T if unmethylated |
| **Imprinting and X** | *Igf2* paternal, *H19* maternal, CTCF at ICR; Xist coats the inactive X |
| **Small RNAs** | siRNA perfect, slices; miRNA seed 2 to 8, represses; Drosha-DGCR8, Exportin-5, Dicer, AGO; piRNA guards germ line |

## Chapter 10 formulas and facts

| Name | Formula or fact | Note |
|---|---|---|
| **Site frequency** | $4^n$ bp | 256; 4,096; 65,536 |
| **PCR** | $N = N_0 \times 2^n$; 94 to 98, 50 to 65, 72 °C | Taq lacks proofreading |
| **qPCR** | $2^{-\Delta\Delta Ct}$ | normalize to a reference gene |
| **Sanger** | ddNTP lacks 3′-OH; read short to long = new strand 5′ to 3′ | 700 to 1,000 bases |
| **Coverage** | $C = LN/G$; missed about $e^{-C}$ | aim for 30x |
| **SpCas9** | 20 nt protospacer + NGG; cut 3 bp upstream | NHEJ indels; HDR knock-in |
| **Editors** | CBE: C•G to T•A; ABE: A•T to G•C; prime editing; CRISPRi and CRISPRa | no double strand break |
| **Human genome** | 3.1 Gb; about 20,000 protein genes; about 1.5% coding; about half transposons | C-value paradox |
| **ORFs** | stop every 21 codons; P(no stop in k codons) = $(61/64)^k$ | long ORF = likely gene |
| **RNA-seq** | CPM; $\log_2$ fold change | normalize before comparing |
| **Homologs** | orthologs by speciation; paralogs by duplication | synteny helps |
| **Selection** | $\omega$ = dN/dS: < 1 purifying, ≈ 1 neutral, > 1 positive | divide by sites |


Dr. Garg Learning · CC BY-SA 4.0