All Questions

Vince Reuter

July 11, 2020

1 Evolutionary Classifications

Q1. What are the 3 clades of extant mammals?

Answer:

- Monotremes
- *Metatheria* (marsupials)
- *Eutheria* (*placentalia*)

Q2. What anatomical feature allows particularly long gestation in one of the clades? How?

Answer: The lack of an epipubic bone in Eutherians allows abdomenal expansion during pregnancy, in turn allowing longer gestation

Q3. What are the extant species of monotremes?

Answer:

- Platypus
- 4 different kinds of Echidna

Q4. Where are the monotremes found?

Answer: All 5 extant monotremes are in Australia.

2 Telomeres

Q1. **Telomere binding proteins (TBPs):** which protein is primarily responsible for *preventing end-to-end fusions* of chromosomes? How so?

Answer: ***TRF2*** prevents end-to-end fusions (NHEJ) by preventing tetramerization (?) of *Ku* proteins (spec., *Ku70* and *Ku80*) that would normally lead to NHEJ.

3 Nucleosomes

Q1. What fundamental properties of histones that set them apart from most proteins account for excellent performance of function as DNA packaging proteins?

Answer:

- **Size:** each of the histone proteins is quite small, useful for a functional role in a tight space like the nucleus, and for *finer resolution* positioning
- **Charge/pH:** DNA is negatively charged and generally acidic (yielding protons in solution); histones have many *basic* AA side chains (among highest combined K and R proportion of all proteins), strongly electrostatically attracting DNA.

Q2. What property of histones across species suggests their centrality of function to eukaryotic life?

Answer: The histones are among the most highly conserved proteins, typically differing by just $2 - 5$ amino acids.

Q3. What structural feature of histones is both functionally essential and conserved in their evolutionary history?

Answer: The *histone fold* domain is essential for the "handshake" by which the octamer forms, and it's the domain that's shared by the nucleosomal histones.

4 Balancer Chromosomes

Q1. What's a ***balancer* chromosome?**

Answer: A balancer chromosome is an artificial/synthetic/genetically engineered construct that features a number of structural variants designed to *inhibit recombination.*

Q2. What are 3 key properties of a balancer chromosome?

Answer:

- **Prevention of crossing over,** allowing a line with known mutations to be maintained without needing to constantly screen for the desired mutations
- **Homozygous lethality:** a balancer carries some recessive lethal allele, such that homozygosity is lethal (i.e., any balancer carrier is heterozygous.)
- **Dominant *marker(s):*** a good balancer carries marker gene(s) that facilitate visual identification and/or selection, e.g. via poison

Q3. Why do balancer chromosomes accumulate so many variants, particularly deletions and large structural variants o/w likely to be harmful?

Answer: Since they're *only found as heterozygotes,* balancer chromosomes are subject to *much* ***weaker negative selection*** against harmful variants.

Q4. How does a balancer chromosome prevent recombination (crossover suppression, i.e. preclude meiotic synapsis)?

Answer: The inversions tend to prevent recognition of the homologous chromosome relationship.

Q5. What if, despite the inversions' tendency to suppress recombination, a balancer chromosome recombines with a homolog?

Answer: The structural aberrations that will result from recombination between a balancer and its homolog should *confer* ***lethality,*** preventing a population that should be heterozygous from coming to include homozyogtes.

5 Experimental Methods General

Q1. What are the two main wins of *in situ* Hi-C compared to standard (dilution) Hi-C?

Answer:

Q2. What's the meaning of the *in situ* portion of *in situ* Hi-C? How's it differ from dilution Hi-C?

Answer: The *in situ* potion of *in situ* Hi-C means that

Q3. How does *in situ* Hi-C attain its main wins relative to dilution Hi-C?

Answer:

- For greater resolution, *in situ* Hi-C (at least the Rao/Lieberman-Aiden original from 2014) deployed a 4-cutter rather than a 6-cutter restriction enzyme.
- The *in situ* aspect of *in situ* Hi-C actually appears to reduce artifactual ligation events that can occur in dilution Hi-C.

Q4. Why/how does *hydroxyurea* stall DNA replication

Answer:

- **Ribonucleotide reductase (RNR)** is an enzyme that reduces ribose (specifically, the 2' carbon) to deoxyribose by removing the hydroxyl group from 2' carbon.
- Hydroxyurea inhibits the catalytic activity of RNR, limiting the pool of dNTPs (DNA replication raw materials) in the cell.

Q5. Why may PCR on extract from cell culture or from a genetically engineered animal, after conditionally deleting an allele, still have *some* minor signal from the deletion target?

Answer: These are *populations* of cells (bulk) from which the material for PCR is being taken, and *recombination will likely be incomplete/nonuniversal.*

Q6. Why use a method like 3/4/5-C over Hi-C? What's an additional requirement, especially for 3/4C?

Answer:

- The "lower" C-based methods provide *greater granularity/resolution* relative to Hi-C.
- There's an inherent cost savings that only having a restricted regional interest (i.e., not genome-wide) allows to leverage, since we need not sequence so much.
- The methods require specific regional primer design, and to have restricted interest/focus to one or a few loci.

Q7. ***Capture-C* vs. *traditional 3C*** What improvement does Capture-C make relative to traditional 3C?

Answer: Capture-C adds *biotinylation-based pulldown* for better purification of the library (?)

6 Body Sections Axes Planes

Q1. What are the 3 main "slices," or "sections," of tissue that may be taken?

Answer:

- *Coronal/frontal:* an instance of a longitudinal plane, the coronal plane runs parallel to the long body axis, perpendicular to the transverse plane, dividing dorsal from ventral
- *:*
- *Sagittal/longitudinal/median:*

Q2. Which body plane does the coronal plane intersect as a vertical line?

Answer: The coronal plane intersects the sagittal plane as a vertical line.

Q3. Which body plane does the coronal plane intersect as a horizontal line?

Answer: The coronal plane intersects the transverse plane as a horizontal line.

Q4. Which two body poles does the coronal plane separate?

Answer: The coronal plane separates ventral from dorsal.

Q5. Which body plane does the sagittal plane intersect as a vertical line?

Answer: The sagittal plane intersects the coronal plane as a vertical line.

Q6. Which body plane does the sagittal plane intersect as a horizontal line?

Answer: The sagittal plane intersects the transverse plane as a horizontal line.

Q7. Which body poles does the sagittal plane divide/define?

Answer: The sagittal plane defines/divides the left and right. Any non-median sagittal plane implies a lateral vs. medial division.

Q8. To which two body planes is the transverse plane perpendicular

Answer: A transverse plane is perpendicular to both a sagittal plane and a coronal plane

Q9. Which two body poles does a transverse plane define/divide?

Answer: A transverse plane divides anterior from posterior.

7 Analytical Methods

Q1. In which kind of graph/plot does a **power law** reveal itself as a **linear** relationship? Why?

Answer:

- A **power law** relation between variables is revealed as a *linear* visualization in a **log-log plot**.
- $y = bx^a \implies \log(y) = a\log(x) + b$, so taking logs gives a linear equation relating log-response to log-input, with power as slope.

8 DNA repair

Q1. What is *DNA **resection?***

Answer: DNA resection is the *exonucleolytic "chewback"* of bases, typically from a double-stranded break

Q2. In what processes is *DNA **resection*** a key step/subprocess?

Answer: ***DNA repair***, via *homologous recombination (HR):* DNA resection is critical for the homologous recombination (HR), typically as part of the *double-stranded break repair **(DSBR)*** response.

Q3. How does *DNA **resection*** perform its function (facilitation of initiation of HR)?

Answer: DNA resection *creates **3' overhang,*** forming the ssDNA that will perform the **strand *invasion*** step of HR, invading a nearby duplex region.

Q4. ***Double* strand break repair (DSBR):** The balance between which 2 proteins is a primary determinant of the procession and pathway chosen for DSBR? How So?

Answer: The balance between *BRCA1* and *TP53BP1* is a major determinant of the balance between candidate DBSR pathways, exerting great influence over which pathway is chosen.

Q5. What process does *TP53BP1* disfavor, and how does that influence a cell's response to / pathway chosen for double-stranded break repair (DSBR)?

Answer: *TP53BP1* disfavors end resection. Since end *resection promotes HR,* TP53BP1 tilts the balance of power away from HR and in favor of more error-prone NHEJ, which is worse for the cell.

Q6. Which process does *BRCA1/2* favor, and how does that influence a cell's response to / pathway chosen for DSBR?

Answer: *BRCA1/2* associate with *Rad51* to help initiate/facilitate/orchestrate **homologous recombination.** As such, cells deficient in those proteins will tend to more frequently use NHEJ to repair double-strand DNA break damage, leading to more costly and dangerous repair errors.

9 General

Q1. Why care about interactions between nucleic acids and proteins? Why are they important?

Answer:

- Events' ***timing*** **specificity** (think TF concentration in **developmental *stage***)
- Events' ***spatial*** **specificity** (think TF concentration in **subsellular *compartment***)

Q2. Describe/characterize the *relative merits* / "tradeoff" of virtue between EMSA and footprinting methods.

Answer: ***Sensitivity*** **vs.** ***specificity:*** EMSA is more sensitive (looser robustness requirement for probe binding by protein), but it provides a coarser picture / less resolution than footprinting for *where* a protein of interest binds.

Q3. What's a key piece of information likely to be of interest that *neither EMSA nor a footprinting method* provides? What class of method does?

Answer:

- **Protein identity:** neither EMSA nor footprinting provides much information about the identity of protein binding.
- ChIP (chromatin immunoprecipitation) methods aim provide binding protein identity information.

10 ChIP

Q1. What are the 2 main ChIP kinds/strategies and how do they differ?

Answer:

- ***Primer*-based ChIP** targets a specific, *individual* region of interest.
- ***Chip*- or *Sequencing*-based CHIP** (ChIP-Chip and ChIP-seq, respectively) can examine an *entire genome* for evidence of protein binding.

Q2. Briefly describe **ChIP-PCR.** How is it used relative to other ChIP methods? What're strengths/weaknesses relative to other ChIP methods?

Answer:

- As with any other chromatin immunoprecipitation method, the availability of an antibody specific for a protein, a complex, or a modification of interest is critical.
- DNA abundance is quantified and compared between an experimental and control sample group; experimental/treatment receives specific antibody while control receives nothing or a nonspecific antibody.
- ***Scope* difference:** a key difference with other ChIP methods is in the scope of investigation; ChIP-PCR examines one or few loci while others can examine many.
- The *scope limitation* arises from the *need to design primers.* Primers targeting a control (expected unbound) locus and the region(s) of interest must be designed.
- Owing to limited scope, ChIP-PCR is typically deployed where the line of inquiry is either ***locally specific** to a region,* or as a **cost-effective *control*** for a method with broader scope.
- Strength : cost :: weakness : scope

Q3. What are the *main **validation** questions* to address in a **ChIP-PCR** experiment?

Answer:

- Interest is to compare control amplicon vs. target amplicon in ***input** DNA* (no/nonspecific antibody) for ***primer bias,*** and to compare control amplicon vs. target in ***treated** DNA* to assess ***antibody efficacy.***
- ***Antibody* efficacy:** If the antibody effectively binds the protein/complex/modification of interest, the treatment samples' target region should be enriched (lower cycle threshold) relative to those samples' control region. This suggests that the antibody's sensitive and specific.
- ***Primer* efficacy/bias:** targeting a ***control* locus/amplicon** helps to *establish a **baseline*** for amplification of the respective regions. Then we can *compare the **change in CT*** for a given sample. This facilitates comparison of the target to control in the treated sample, which is ultimately what we're after.
- Example:
  (a) In input DNA, both the control amplicon and the target amplicon show CT of 27.
  (b) In the IP sample(s), the control amplicon has CT of 30 while treated (specific Ab) samples have CT of 27. This implies $8x = 2^3x$ enrichment of the target.
  (c) We can use the raw difference in the CT between the regions in the treated sample because the baseline difference is $0 = 27 - 27$ between the amplicons in the input DNA.

Q4. What's ***cycle threshold?*** What information does it provide?

Answer:

- The **cycle threshold** is the *number of PCR cycles* needed to sufficiently amplify some region of interest for detection by whatever machine (spectrophotometer?)
- The CT is a *measure of **abundance*** (of molecules). Each PCR cycle amplifies (in theory) each molecule 2x, so abundance is inversely exponentially proportional to the CT.

11 EMSA

Q1. What's EMSA stand for?

Answer: Electophoretic mobility shift assay

Q2. What's EMSA measure directly?

Answer: EMSA measures the **distance traveled** by a biomolecule or biomolecular process, though a polyacrylamide gel

Q3. Broadly speaking, how does PAGE/gel electophoresis work? How does it provide a readout of intermolecular binding?

Answer:

- 1. Gel impedes movement as negatively charged molecules migrate away from negative source
- 2. Larger molecules experience stronger resistance
- 3. More binding means larger molecular complexes.
- 4. $d = rt \wedge\ r \propto$ molecular (complex) size

Q4. What biomolecular phenomenon does EMSA aim to measure?

Answer: EMSA aims to measure binding between proteins and nucleic acids (and proteins-to-proteins). Most generally, EMSA/PAGE/shift assays measure molecular (complex) size

Q5. What two main kinds of labeling are used to detect complexes in a shift assay?

Answer:

- Fluorescent (fluorphore conjugation)
- Autoradiography (make one end of nucleic acid radioactive)

Q6. What main precaution must be taken when using a fluorophore to label molecules for detection in a shift assay?

Answer: The fluorophore must not interfere with binding between the molecules of interest

Q7. In simplest form, EMSA provides information, e.g., **that *a*** protein bound DNA, but **not *which specific* protein(s)** bound DNA. How can that be assessed?

Answer: **Antibody *supershift.*** Antibody binding can be protein-specific, and thus will specifically decrease mobility of compatible protein(s), allowing identity inference.

Q8. How may the presence of multiple proteins simultaneously binding to DNA be examined, via shift assay?

Answer: Each binding event will increase size and weight of the molecular complex, further reducing its mobility

Q9. EMSA says nothing about ***specific* sequence** bound by a protein, just about a larger fragment. How to *hone in?*

Answer:

- **Directed mutagenesis:** specific subsequence(s) within bound fragment(s) may be targeted for mutation
- **Excess oligomer:** relative to the initial fragment(s), smaller hypothesized oligomers for binding can be added, and will "soak up" protein binding

Q10. In what 2 main ways may the search space for sequence specificity me reduced?

Answer:

- Prior empirical findings
- Established sequence conservation

Q11. What are *2 main* ***weaknesses*** of EMSA? How may each be addressed?

Answer:

- ***Identity* of binding partner(s):** EMSA doesn't directly provide information about *which specific* protein(s) are binding, just the added size.
- ***Identity* of binding partner(s)** may be addressed by **specific antibody binding;** an antibody will *selectively* ***decrease mobility,*** specifically when its complementary protein is bound. This requires **antibody availability,** though.
- ***Sequence specificity:*** EMSA says binding occurs to a fragment, but not *which part(s)* (subsequences/"motif(s)") of the fragment.
- ***Sequence specificity*** may be addressed by either **directed *mutagenesis*** or addition of **excess putative oligomer**

12 Footprinting Methods

Q1. Why do footprinting methods *require more robust binding* between DNA probe and protein of interest?

Answer: Each in this class of methods leverages ***lack** of signal* as its source of information, so to generate enough "background" signal, binding must be robust

Q2. What's are **2 main *weaknesses*** of footprinting methods? How may each be addressed?

Answer:

- **Robust requirement:** the requirement for more robust (relative to EMSA) protein binding to probe cannot really be ameliorated.
- **Protein identity lack:** as with EMSA, the identity of the protein isn't apparent from footprinting data. Unlike with EMSA, there's not a great way to discern this.

Q3. What's a major strength of footprinting methods relative to EMSA?

Answer: **Finer *resolution:*** Each footprinting method provides much more fine-grained detail about ***where*** *the protein of interest binds.*

Q4. Name and briefly describe each of the **3 main footprinting methods.**

Answer:

- ***Nuclease*** **footprinting:**
- ***Chemical*** **footprinting:**
- ***Chemical interference*** **footprinting:**

Q5. What are the main steps/principles in nuclease and chemical footprinting?

Answer:

- Label DNA with a radioactivity or with a fluorophore
- Incubation to allow protein binding
- Introduction of fragment breaks (nuclease or chemical)
- Run gel elecrophoresis
- Infer location of protection from "hole" in the fragment size distribution (where presumably the protein of interest bound)

Q6. Why should titration be done to aim for about one fragment break per probe?

Answer: Since the label is at just one end of the DNA molecule, only the length of the first fragment in the linear sequence will be observed.

Q7. How does **chemical *interference*** footprinting work?

Answer: Chemically induce DNA probe breakoints that distort the molecule and see which do or don't impair binding. Effectful alterations are inferred to be where the protein binds.

Q8. The order of which 2 main footprinting steps is reversed for chemical interference footprinting?

Answer: Induction of fragment breaks comes before incubation for protein binding, since the assay leverages impairment of protein binding by the breaks, rather than the binding's protection from breaks.

Q9. Unlike EMSA, with a footprinting method why is there not a straightforward way to tackle protein binding identity?

Answer: Footprinting leverage signal absence. There's no way to specifically target a void. There's no way to get more information from a null and still have it be a null.

13 SELEX

Q1. What's **SELEX** stand for?

Answer: **SELEX** is the systematic evolution of ligands by exponential enrichment.

Q2. Briefly describe the SELEX steps.

Answer:

- 1. Create *combinatorial library* of DNA molecules, each with primers flanking a randomized sequence region
- 2.

Q3. Why does SELEX necessarily rely on PCR (amplification)?

Answer: Each unique molecule in the combinatorial DNA library is very unlikely to bind strongly to the protein of interest, and each molecule that does will be few in number.

Q4. Typically, about how large is the combinatorial region of each DNA molecule in the library?

Answer: Similar to TF motifs (analogous *in vivo* biological application), each combinatorial region is usually about $10 - 12$ bp.

Q5. What are the main components of each DNA molecule in the combinatorial library?

Answer:

- Each molecule has a $10 - 12$ bp randomized region in the middle (synthetic binding site to test).
- Each randomized test region is flanked by primers to facilitate amplification via PCR.

Q6. What are the **two main methods for purifying/selecting** protein-bound molecules after incubation of DNA with the protein of interest?

Answer:

- Generally, **EMSA** (electrophoretic mobility shift assay) may be used, as the protein-bound DNA molecules will be larger and more unusually shaped.
- Alternatively, if an antibody against the protein of interest is available, purification/selection may be done via **immunoprecipitation.**

14 DNA Replication General

Q1. **Replication *safety/security:*** Name and describe the replication machine/enzyme that's tightly cell cycle regulated to assure *exactly one* replication per cycle?

Answer:

- The **replicative *helicase*** is tightly coupled to the cell cycle to assure loading at necessary locations and only one period of activation per cell cycle.
- In G1 helicases may be *loaded but not active*; during the remainder of the cell cycle, *loading's prohibited but activation is license.*

15 classic experiments

Q1. ***Transforming* principle:** which classical experiment demonstrated that DNA is the "transforming principle?" That is, that it's the kind of molecule that transformed bacteria from benign to virulent/pathogenic.

Answer: The **Avery-McCarty-MacCleod** experiment showed that DNA is the transforming princple.

Q2. Briefly describe the **Avery-McCarty-MacCleod** experiment and its significance.

Answer:

- 1. Smooth bacteria are virulent, rough ones aren't; heat-killed smooth bacteria aren't virulent, nor are heat-killed rough ones.
- 2. Rough bacteria incubated with heat-killed smooth bacteria become virulent.
- 3. When incubating with parts (cell isolates) of heat-killed smooth bacteria, DNA is the only kind of molecule that confers virulence upon the rough bacteria.
- 4. We may infer that *DNA is "taken up" by bacteria (as plasmid)* and confers functional properties that other molecules don't.
- 5. This *clarified* ***Griffith's 1928 experiment*** that observed the transforming phenomenon when enjecting pairs of rough/smooth dead/live bacteria.

Q3. Which fundamental biomolecular genetics question did the **Avery-McCarty-MacCleod** experiment address?

Answer: The **Avery-McCarty-MacCleod** addressed the *transforming principle,* a sort of *functional heredity,* showing that DNA is the only molecule so empowered.

Q4. Which fundamental question did the ***Meselson-Stahl* experiment** address?

Answer: The **Meselson-Stahl experiment** addressed the ***mode** of DNA replication,* demonstrating that it's **semiconservative.**

Q5. *How* did the Meselson-Stahl experiment address the question of mode of DNA replication?

Answer: **Isotope labeling:** the experimental design used an alternate *nitrogen isotope*, showing that it's proportional contribution to each DNA molecule diminished over time.

Q6. What was the experimental/observational readout from the Meselson-Stahl experiment? That is, how was the proportion of heavier nitrogen assessed?

Answer: When run on a gel, a DNA molecule with more of the heavier nitrogen isotope moves more slowly and therefore travels a lesser distance.

Q7. How did the **Avery-McCarty-MacCleod** experiment address the alternative hypotheses that protein and/or RNA were responsible for bacterial transformation?

Answer:

- The experiment used various *digestive enzymes* that would break down members of alternative molecular classes like protein or RNA
- Things like ribonucleases (RNA) and trypsin (protein) that break down some a non-DNA target did nothing to the transformational capability of cellular extract, but so-called deoxyribonucleopolymerase destroyed it.

Q8. Describe the **Hershey-Chase** experiment.

Answer:

- 1. Label both S-35 and P-32, and incubate phages with those labeled isotopes.
- 2. Allow phages to infect cells.
- 3. See which labels go where during pathogenesis/invasion of host cell.
- 4. Sulfur outside $\implies$ protein doesn't really enter.
- 5. Phosphorus inside $\implies$ DNA does enter the cell.

Q9. What was the significance of the **Hershey-Chase** experiment?

Answer: The **Hershey-Chase** experiment further reinforced the idea that DNA is the transforming principle / encoding of heredity.

Q10. What biochemical difference between DNA and protein did the **Hershey-Chase** experiment leverage?

Answer: DNA contains much more phosphorus than protein, while DNA contains no sulfur (but some amino acids and therefore protein do, plus disulfide bridges). Therefore, labeling phosphorus and sulfur distinguishes protein from DNA.

16 DNA repair HR

Q1. What's a major risk of mitotic HR that must be avoided? What danger does it typically pose? Molecularly, how is the solution strategy implemented?

Answer:

- *Loss of heterozygosity (**LOH**)* is a major risk that a mitotic cell must avoid. That's when homologous recombination randomly selects the alternate allele rather than the sister chromatid for recombination.
- Having 2 copies of each gene safeguards against a defective copy. Losing one copy increases the likelihood of no working copy, which is especially risky for cancer.
- ***Cohesin** wraps sister chromatids,* keeping them quite close together. The spatial proximity of sister chromatids relative to homologous chromosomes favors usage of the sister chromatid template.

Q2. If ***HR*** *requires a close match* between template and strand to repair, how can **LOH** occur?

Answer: Because so much DNA is conserved between individuals, at a small scale with high probability *maternal and paternal alleles will closely match.* The match may be close enough to satisfy HR.

17 DNA Structure

Q1. What are 4 main functions/traits that DNA must provide?

Answer:

- **Replicability** with high *fidelity* and virtually without upper bound on number of replications
- **Packaging/Compression:** spatial constraints of the cell and the nucleus means that an otherwise large linear genome must become small.
- ***Developmental*** **regulatory control:** DNA must provide the instructions with which organismal develompent may be orchestrated
- **Day-to-day grind:** DNA must provide the instructions that a cell needs to make the proteins that it needs to function each day

Q2. What main problem/question did the double helical structure solve?

Answer: Replicability/generational passage: the *complementary base pairing* and concomitant redundancy of information addressed the replicability mechanism question.

Q3. Which 2 main goals/objectives/purposes of DNA must its packaging balance?

Answer:

- **Accessibility:** the DNA must be packed in such a way that access for transcription and replication is easy
- **Compression:** the DNA must be packed small enough to fit inside the tiny nucleus.

Q4. Characterize the *bond kinds* involved in the DNA double helix.

Answer:

- In the sugar-phosphate **backbone** covalent bonds stabilize the double helical structure while the bases bond noncavently, with hydrogen bonds.
- External : stronger : covalent :: internal : weaker : noncovalent
- C-G : 3 :: A-T : 2

Q5. Why are A-T bonds "weaker" than C-G bonds? What empirical property is often used as a metric for the strength?

Answer: A-T pair with 2 hydrogen bonds rather than 3. "Melting" temperature (Tm) measures this. A-T rich DNA has a lower Tm than G-C rich DNA.

Q6. Why are the strands of the DNA double helix said to be *antiparallel?* What gives **DNA *directionality?*1**

Answer: *Asymmetry* in the chemical bonding linking together deoxybiose subunits of the *backbone* is what gives DNA directionality and the "antiparallel" moniker.

Q7. What are some benefits of having antiparallel strands?

Answer:

- ***Proofreading*** **mechanism:** the 5'-to-3' *exnucleolytic proofreading* that allows DNA polymerase to boost its fidelity is enabled by DNA's directionality.
- **Coding space:** while most relevant for prokaryotes and viruses with simple genomes, DNA's directionality means that *genes may overlap*

Q8. About what percentage of cell volume is the nucleus?

Answer: About 10 percent.

Q9. What's a major logistical benefit of having a nucleus and nuclear envelope (i.e., walling off DNA from the rest of the cell)?

Answer: The nucleus is an instance of the more general strategy of *compartmentalization,* making desirable *molecular collisions more likely.* It shifts the *probability distribution* of molecular collisions in a productive way.

Q10. Characterize the percentage of the human genome devoted to protein coding vs. repetitive elements

Answer:

- Protein coding genome sequence is just about 1.5 percent of the total.
- Repetitive, transposable elements account for nearly half of the human genome.

Q11. How large is the average human gene, in total nucleotides base pairs and in amino acids (how main coding nucleotides?)

Answer:

- The average human gene spans abut 27 Kb.
- The average protein is only about 430 amino acids long, so about 1300 coding nucleotides.

Q12. Briefly, what are the 3 main specialized regions of each chromosome that regulate its structure?

Answer:

- **Replication origins:** many in eukaryotes to expedite replication of huge genomes; typically (always?) just one in prokaryotes
- **Centromere:** critically, there's *one and* ***only one.*** Mitotic segregation won't occur if there's no centromere, chromosomes will fail to properly segregate during mitosis.
- **Telomere:** distiguishing chromosome ends from breaks so that normal, healthy DNA tips aren't mistaken for DNA damage and then erroneously repaired

Q13. DNA replication creates a structural problem for chromosomes: how to know when to stop and how to finish? How do prokaryotic and eukaryotic cells differ in their solution strategies?

Answer:

- Prokaryotes use a *circular genome.*
- Eukaryotes use linear chromosomes and specialized end structures (telomeres).

Q14. Describe the central structure of a nucleosome?

Answer: The nucleosome core is a **histone *octamer,*** composed of a H3-H4 tetramer and a pair of H2A-H2B dimers.

Q15. Approximately what fold compression does packing into nucleosomes provide for DNA packing?

Answer: Wrapping around nucleosomes

Q16. About how long is linker DNA?

Answer: Linker DNA ranges from just a few base pairs to up to about eight base pairs, so the distance between nucleosomes is about 200 bp.

Q17. Describe size and conservation of histones relative to other proteins.

Answer: Histones are smaller and more conserved than most other proteins. They're only about $102-135$ amino acids, and differ by only a few amino acids, suggesting critical structural importance under intense stabilizing selection

Q18. What features do the histone proteins share?

Answer: Histones share the *histone* ***fold,*** a domain composed of three α helices connected by short loops

Q19. Account for the simultaneous strength and nonspecificity of affinity between DNA and the histone octamer. Why does virtually any sequence bind so strongly?

Answer:

- **Charge** and *amino acid composition:* DNA is negatively charged, while about 1/5 of the histone's amino acids are either *lysine* or *arginine.*

- Basic amino acids are positively charged at neutral pH (because they donate hydroxide ions to solution), so they modest positive charge of the histone core attracts negatively charged DNA.

Q20. Describe the extent of the interface between the histone octamer and the DNA that wraps around it.

Answer:

- In just 147 bp, or about one-and-a-half "wraps" around the octamer, DNA makes a whopping 142 hydrogen bonds with the octamer.
- Roughly half of the hydrogen bonds between DNA and the octamer are between the backbone and the histone core (not tails).

Q21. Characterize the nucleosome's sequence binding preference.

Answer: Certain dinucleotide pairs bend with relative ease; the DNA is more sharply bent on the nucleosome-adjacent side, so on that side, it's preferable to have A-T enrichment in the octamer-proximal minor groove.

Q22. Why must the sequence binding preference of nucleosomes be relatively weak?

Answer: Nucleosomal sequence binding preference may be viewed as an assist with the chromatin architectural plan, but greater functional exigency must be given preference for nucleosome positioning and DNA access, so nucleosome preference must not overwhelm TFs' ability to reposition them.

18 DNA Replication

Q1. What's it mean that *DNA replication is* ***semiconservative?***

Answer: ***Semiconservative*** *replication* means that in each "daughter duplex" that's generated, one strand is newly synthesized, and one strand is inherited from the parent cell.

Q2. Which *classic experiment* demonstrated that *DNA replication is* ***semiconservative?***

Answer: The **Meselson-Stahl experiment** demonstrated the DNA replication is semiconservative.

19 Chromatin Histones Nucleosomes

Q1. If DNA so frequently "breathes" by unwrapping briefly from the nucleosome (about 4 times per second, with each exposure $10 - 50$ms), why do eukaryotic cells need ATP-dependent chromatin remodeling complexes at all?

Answer: While contact between a TF and the DNA may be able to happen randomly in such frequent but brief windows, likelihood improves greatly with the remodeling factors, as they maintain accessibility much longer.

Q2. How do nucleosome remodeling complexes achieve *nucleosome replacement?* What's the approximate average frequency of that turnover in cells?

Answer:

- Remodeling complexes feature a ***helicase-like*** *subunit* that uses **ATP hydrolysis** to achieve the "sliding" action.
- Remodeling complexes associated/collaborate with *histone* ***chaperone*** *proteins* that assist with histone substitution of histones within the octamer.
- On average, each nucleosome is replaced once every $1 - 2$ hours, or around 15 times each day.

Q3. What's the most important factor for determining nucleosome position? Why is this functionally important?

Answer:

- Binding of transcription factors and other proteins, much more than histone-to-sequence binding preference, influences nucleosome position
- For functional plasticity and response to signaling / changing conditions, it's critical that DNA-binding proteins' influence outweighs the (static) preference of histones for certain DNA sequence features.

Q4. Which histone has been less conserved through evolutionary history? How may this help explain interspecific chromatin structural differences?

Answer: The **"linker" histone (H1)** has been much less conserved than the others. Since it plays a more prominent role in higher-order chromatin structure, that lack of conservation may help account for interspecific differences in chromatin packing.

Q5. How does linker histone H1 facilitate chromatin condensation into higher-order nucleosomal arrays?

Answer: H1 *alters DNA exit path* from the nucleosome, in a way making the emergent DNA slightly more tethered/less flexible.

Q6. It's said that *heterochromatin is self-propagating.* What early experimental observation in flies supports this?

Answer:

- The so-called *position effect* from early fly genomics experiments supports the notion that heterochromatic state propagates
- Euchromatin translocated into a heterochromatic neighborhood has its genes *silenced.*

Q7. Histone PTMs and accessibility: which general histone tail modification is associated with chromatin accessibility? How does it foster access?

Answer:

- **Lysine *acetylation*** is strongly associated with chromatin accessibility, by *charge neutralization.*
- Specifically, the negative acetyl charge neutralizes the positive lysine charge, thereby weakening binding affinity between DNA and the the octamer.

Q8. Histone PTMs and chromatin: Which histone PTM is strongly associated with condensed heterochromatin? How so?

Answer: H3K9me3 is most associated with (esp. constitutive) heterochromatin. **Linker histone H1** is recruited by this PTM.

Q9. How do cells unevenly distribute histone protein synthesis? How does this relate to fulfillment of functional needs?

Answer:

- *Histone variants* (most) are synthesized continually *throughout interphase.*
- *Canonical histones* are synthesized primarily during *S phase.*
- Histone variants, by nature, typically replace a canonical histone to shift probability distributions of events in some functional way; as such, they're needed throughout much of the cell cycle.
- Synthesizing canonical histones primarily during S phase is both functionally efficient and tips the concentration of available histone protein for nucleosomes toward the canonical/constitutive rather than the specifically functionally facultative.

Q10. Describe how nucleosome remodeling complexes attain functional two kinds of specificity.

Answer:

- ***Spatial* specificity:** nucleosome remodeling complexes target specific genomic regions / motifs by incorporating *DNA-binding subunit(s).*
- ***Functional* specificity:** remodeling complexes favor binding to a particular histone variant through *protein-protein interactions;* specifically, complexes include subunit(s) for interaction with chaperone(s) associated with particular histone(s).

Q11. Broadly, what balance does a structure featuring a dense core but several looser, freer tails (histones) strike, and functionally what does this achieve?

Answer:

- The balance/tradeoff is struck between density and close spatial packing, and dynamism.
- Pairing a small, dense core with several short but flexible "tails" allows histones to both condense DNA into chromatin while leaving a "hook," or "platform" for functional alteration as cellular needs change.

Q12. Histone PTM *combinatorial specificity:* How do nucleosome remodeling complexes and other "readers" of chromatin implement combinatorial specificity?

Answer: Many/all individual PTMs have at least one "reader" motif/domain. Complexes or large proteins can achieve a sort of combinatorial mark specificity and stronger binding by *incorporating several domains/modules* with "reading" properties.

Q13. Chromatin *spreading:* what's the classic way in protein/enzymatic complexes "spread" chromatin marks?

Answer: A complex that *pairs reader with writer* is effective for "spreading" a particular mark. That is, a complex in which the writer leaves a mark that the reader binds.

20 connections between topics

Q1. ***Immunology*** *and genome* ***integrity/maintenance:*** which process is used by both *V(D)J recombination* and by *DNA repair?*

Answer: **NHEJ:** non-homologous end joining is used by both V(D)J recombination and by DNA repair.

Q2. ***Meiosis*** *and genome* ***integrity/maintenance:*** which process is used by both *meiosis* and by *DNA repair?*

Answer: *Homologous recombination (**HR**)* is used by both meiosis and by DNA repair.

Q3. How does the deployment of **HR** contextually differ?

Answer: sister chromatids : DNA repair :: homologous chromosomes : meiosis

Q4. Motivate the difference in ***HR*** *mechanics* that depends on usage context. That is, why do it each way in each context?

Answer:

- For **DNA *repair,*** a primary concern is *loss of heterozygosity (**LOH,**)* so a chromosome's sister chromatid is favored as the temlate.
- For **meiotic *recombination,*** a primary concern is *generation of genetic* ***diversity,*** so the homologous chromosome is favored as recombination partner.

21 math stat general

Q1. What's a good way to measure how quickly an object converges to 0 in some limit? That's is, what's a general strategy? Think **Blitzstein** and **Mark Low.**

Answer:

- ***Competition*** **between terms:** specifically, multiply by a *sufficiently large power* of n going to infinity.
- More generally, allow a term going to infinity to "compete with" a term going to 0, until the nonzero term is sufficiently powerful for nondegenracy.

Q2. How and under what conditions can we *define a* ***natural ordering of matrices?***

Answer:

- A natural ordering for matrices occurs when they're all squares of a common size and are positive definite
- The determinant (product of eigenvalues) then defines a natural ordering of the matrices.

22 Virology2020 L07 Transcription and RNA processing

Q1. Which **molecule type** (of nucleic acid) is required for transcription?

Answer: *Transcription* is a process linked by definition to **dsDNA**; dsDNA to mRNA.

Q2. Which viral classes involve transcription at all?

Answer: The classes that *include* ***dsDNA*** as part of the lifecycle use transcription; Class I, II, VI, and VII.

Q3. Which viruses enter the cell "ready to go" w.r.t. transcription? That is, which viruses *don't* need intermediate(s)?

Answer: **Class I, dsDNA**: The dsDNA viruses enter the cell ready for transcription, since their genome is the molecular input for transcription.

Q4. Which viruses are even more "ready-to-go" from the protein production perspective than the "ready-to-go" dsDNA viruses? That is, which viruse "skip" straight to translation?

Answer: **Class IV, sense-strand RNA** viruses enter the cell even more "ready" for protein production, as they have the property that $gRNA = mRNA$.