All Questions

Vince Reuter

Last updated Friday, August 7, 2020

Contents

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 abdominal 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 Bioinformatics File Formats

Q1. Descibe the BED coordinate system.

Answer: BED uses 0-based inclusive starts and exclusive ends, so a half-open interval; the first nucleotide has 0 start and 1 end.

Q2. What justifies the (perhaps quirky) BED coordinate system choice?

Answer: Computation time justifies BED coordinates; specifically, length of a region/feature becomes simple subtraction of start from end.

8 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.

9 Nucleic Acid Protein Interactions

Q1. What biochemical property would be desirable for the 3D amino acid region of a protein that must bind a nucleic acid? In other words, what amino acid trait might we expect to be enriched in a nucleic acid binding region of a protein? Why?

Answer: Since nucleic acids are negatively charged, higher affinity binding between the protein and the nucleic acid will be achieved if the protein's amino acids in that region (again, perhaps discontinuous / not contiguous in linear amino acid sequence) are ***positively* charged.** Note, however, that while higher binding affinity may generally be preferred, that's very likely not *always* the case.

Q2. What properties together of a plasma membrane and amino acids in a protein make for good binding?

Answer:

- ***Negatively* charged cytoplasmic leaflet** of the plasma membrane, due to *flippases'* regulatory action
- ***Positively* charged amino acid** subsequence or folded region of a protein

10 Classic Experiments

Q1. Which experiment first to suggest that *genetic material may be* ***horizontally*** *transmitted*?

Answer: **Griffith's experiment** was the first to suggest the possibility of horizontal transfer of genetic material.

Q2. When was **Griffith's experiment**?

Answer: 1928

Q3. What did **Griffith's experiment** establish (i.e., what was the main finding)?

Answer: Griffith's experiment established that *bacteria can* ***transfer genetic material*** horizontally.

Q4. How was Griffith's experiment able to show what it did?

Answer: Mixing extract heat-killed, virulent bacteria with live, benign bacteria made the previously benign bacteria pathogenic.

Q5. What was the process of horizontal gene transfer initially called (in Griffith's experiment)?

Answer: The process now known as HGT was called *transformation*, and the material was called the *transforming principle*.

Q6. Which experiment, and when, *solidified evidence for* ***DNA as the transforming principle***?

Answer: Results published in **1944** by *Avery, MacLeod, and McCarty* supported DNA (not bacteria) as the transforming principle.

Q7. Why was it not thought possible for DNA to be the material of the genetic code? What was the alternative?

Answer: DNA's *biochemical boringness* was oft cited as reason that it couldn't encode life's complexity; *protein* was the alternative.

Q8. How did *Avery, MacLeod, and McCarty's* work elaborate Griffith's work to demonstrate *DNA as the* ***transforming principle***?

Answer: Avery, MacLeod, and McCarty subjected the extract to a *variety of* ***digestive enzymes***, and the only one that could break down the extract was a deoxyribonuclease (not various proteases and ribonucleases).

Q9. Which classic experiment (and when) *affirmed AMM*, demonstrating that **DNA is the transforming principle**?

Answer: The **Hershey-Chase experiment** of 1952 affirmed Avery, MacLeod, and McCarty's work, showing that DNA is the transforming principle.

Q10. How did the **Hershey-Chase** experiment show what it did?

Answer:

- The Hershey-Chase experiment demonstrated that DNA is the transforming principle by *radiolabeling specific elements*.
- Phosphorus is specific to DNA (not protein), and sulfur is specific, through cysteine, to proteins (not DNA).
- Phages incubated with either radiolabeled sulfur or radiolabeled phosphorus then specifically are marked for either protein or DNA, respectively.
- Radioactivity stays with the protein, entering the infected bacteria only when it's phosphorus that's labeled and staying with the phage otherwise.
- Progeny phages are labeled if and only if the parent phages were made with radiolabeled phosphorus, not DNA, showing that ***DNA****, but not protein, is* ***heritable***

Q11. What's the nickname for the Hershey-Chase experiment? Why?

Answer: The Hershey-Chase experiment is nicknamed the "blender experiment" b/c of use of the blender to remove the phages from the bacteria.

Q12. Which experiment (and when) showed that ***RNA*** *can be the material of* ***heredity***?

Answer: 1955's **Fraenkel-Conrat** experiment showed that RNA can encode heredity.

Q13. Which virus was used to show that ***RNA*** *can be the material of* ***heredity***?

Answer: Tobacco mosaic virus (TMV)

Q14. How was RNA shown to be capable of being heredity's material?

Answer:

- **TMV** in the 1955 **Fraenkel-Conrat experiment** was used in two strains, and DNA from one was mixed with protein of the other to make 2 kinds of chimeric virus particles.
- Progeny viruses had the protein coats matching that of the strain from which the parent's RNA was from (i.e., the chimeric state lasted only one generation and was not heritable.)

11 General Membrane Biology

Q1. In what important electrostatic way do animal/eukaryotic cell membranes differ from microbe membranes?

Answer: ***Charge*** **distribution**: microbes have fairly randomly distributed charge w.r.t. cytoplasmic vs. exoplasmic side. In contrast, *animal cells tend to* ***organize*** the charge distribution.

Q2. How does the main electrostatic difference between eukaryotic cell membrane and microbe cell membrane relate to **defensins** of *innate immunity*?

Answer:

- Defensins work basically by boring a hole in the plasma membrane, using positive charges.

- Microbes are vulnerable to defensins since they have some negative charge on the exoplasmic side, since their charge distribution is largely random.
- Animal/eukaryotic/host cells are protected from their own defensins because they tend to organize the negative charges of the lipid bilayer on the cytoplasmic membrane side, such that few, if any, are exposed exoplasmically to the defensins.

Q3. *How* do animal/eukaryotic cells regulate the electrostatic properties of their membranes?

Answer: **Flippases,** or *phospholipid translocases*: these enzymes facilitate the "flipping" of a phospholipid from one face of the bilayer to the other, which would otherwise be energetically disfavored.

Q4. *What molecules* (**phosphilipids,** specifically) are heavily involved in membrane biology? Name 3. Which in particular is involved in charge distribution?

Answer:

- Phosphatidyl*serine*
- Phosphatidyl*choline*
- Phosphatidyl*ethanolamine*

Q5. Describe the basic biochemical structure of the 3 main phosphilipids mentioned in *Molecular Biology of the Cell* $(566 - 567)$?

Answer:

- Distinguishing group (e.g., choline): this, the phosphate group, and glycerol jointly comprise the ***hydrophilic*** **"head" group**
- **Phosphate** group: a sort of linker
- **Glycerol**: all three hydroxyl groups participate in dehydration synthesis / condensation reactions, with one each going to fatty acids and the third to the phosphate group.
- **Fatty acids**: a pair of fatty acid chains, *one unsaturated*, comprise the ***hydrophobic*** **"tail"**

12 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.

13 HIV

Q1. Why is HIV so interesting to study for alternative splicing?

Answer:

- The entire HIV proteome comes from a **single *primary transcript***, so alternative splicing is an essential component of HIV gene expression.
- The structural conformation of HIV RNA plays a significant role in how it's spliced, so the virus provides an interesting perspective on the link between structure and splicing.

Q2. Which protein, made in the early phase, plays a major role in late phase and binds to a particular subsequence?

Answer: The **Rev** protein is made early and binds Rev late, helping get gRNA out of the nucleus.

Q3. Roughly where in the HIV genome is the RRE?

Answer: The RRE is near the 3' end of the genome.

Q4. What's the **RRE**?

Answer: **Rev response element**: this is a regulatory sequence that the Rev protein binds to.

Q5. What process is the sole determinant of whether an mRNA will be able to code for the protein *Tat*? What region of the HIV genomes is most relevant for this for Tat in partciular?

Answer:

- Alternative splicing entirely determines whether *Tat* will be made; without the use of a particular splice site, Tat is not made.
- It's the **A3 splice site** of HIV that determines whether Tat will be made, and that (A3 SS) region of the RNA molecule exists in a couple different major structural conformations with differential accessibility of the site to the splicing machinery.

Q6. Which regions of the HIV genome are both especially divergent in structural conformation and have especially stable alternative conformations?

Answer: Both the 5' UTR and the **Rev response element (RRE)** are have stable, divergent structures, as revealed by DMS-MaPseq and application of the DREEM algorithm.

Q7. What clever trick must HIV pull toward the end of its lifecycle?

Answer:

- The *integrated* ***provirus*** (DNA) form of the genome is the template for RNA synthesis (i.e., HIV uses the *reverse transcriptase* strategy rather than the RNA-dependent RNA polymerase (RdRP) as the biosynthetic machinery to account for host cells' lack of it.
- Since the **full-length *primary* transcript** is not only mRNA for alternative splicing and protein synthesis, but also the viral genome, the virus *must export* ***unspliced*** *RNA* from the nucleus.
- Getting unspliced RNA out of the nucleus is tricky becuase ordinarily the exportins that interact with nuclear pore complex proteins to license export are applied in the course of splicing; HIV must find an alternative.

Q8. What are "elite controllers", and why are they interesting to study?

Answer:

- "Elite controllers" are HIV patients that *keep virus down* at very low levels without the assistance of antiretroviral therapy (ART).
- "Elite controllers" are interesting to study for the insight they may bring to the disease biology and pathogenesis, as well in a broader sense for the connection between robustness/flexibility of the adaptive immune response to rapidly mutating pathogens.
- The known *protective HLA allele* (HLA-B57) greatly enriched in these individuals is known to present a narrower range of host peptides to T cells during their development in the thymus, leading to a cognate TCR pool restricted to this allele's MHC molecule that tends to depend on fewer individual, but strong peptide contacts.

14 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.

15 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.

16 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**

17 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.

18 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.**

19 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.*

20 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.

21 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.

22 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 template.
- For **meiotic** ***recombination,*** a primary concern is *generation of genetic* ***diversity,*** so the homologous chromosome is favored as recombination partner.

23 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.

24 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.

25 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.

26 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.

27 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 $gRNA = mRNA$.

Q5. What are the 2 main types of viral **transcription regulatory frameworks**? Describe each.

Answer:

- ***Cascade* regulation**: Genes may have a *serial* expression structure, in space and/or in time. One or more "upstream" gene products activates/licenses a "downstream" gene.
- ***Feedback* regulation**: One gene may ***auto****regulate*; this could be negative or positive; more about *how much* than abotu ON/OFF.

Q6. What are some examples of **modular organization** (in *function* and in *sequence* of sequence-specific transcriptional activators?

Answer:

- **DNA binding**: most N-terminal region of the protein; often a *zinc finger*, a *helix-turn-helix* motif, or an especially *basic (ph)* region
- **Dimer formation**: often implemented via *leucine zipper*
- **NLS**: nuclear localization sequence, used to direct a transcription factor back to the nucleus to do its job
- **Activation**: most C-terminal region; often a *acidic* or rich in glutamine, proline, or isoleucine

Q7. What are 3 common ***molecular* implementations** of the **DNA binding** portion of a transcriptional activator?

Answer:

- *Zinc finger*:
- *Helix-turn-helix*: a particular structural motif
- *Basic*: a region enriched for basic amino acids

Q8. What's the primary purpose of early vs. late phases? That is, why have phases?

Answer: Partitioning the life cycle compartmentalizes gene expression according to life cycle timing, avoids the otherwise wasteful output, and economizing the life cycle.

Q9. What kind/class of protein is particularly common in the early phase?

Answer: Transcriptional activators and regulatory proteins are especially likely to be in the early phase.

Q10. What kind of proteins are especially associated with the late phase? Why?

Answer: Structural proteins are typically late-phase, since they're only needed for particle assembly when it's time to infect other cells.

Q11. What viral **life cycle *molecular process*** is especially associated with coupling to the activation of *extbflate-phase transcription*?

Answer: A common viral strategy is to *couple* ***late****-phase transcription to* ***genome replication***, as replication signals that it's about time to *form new particles* and infect other cells.

Q12. **SV40**: Descibe the general transcriptional regulatory strategy of SV40.

Answer:

- **SV40** has a *circular genome* with just a ***single, bidirectional*** *promoter* that drives transcription.
- **SV40** *physically bifurcates transcriptional phases*; that is, the early- and late-phase transcriptionl programs are implemented by *procession in* ***opposite directions*** from the single promoter.
- **SV40** may use a *micro-RNA*-based strategy to suppress expression of early-phase transcripts during the late phase, with the late-phase transcriptional direction encoding a miRNA that seems to antagonize early-phase gene translation.

Q13. Which host transcriptional activator does adenovirus hijack to drive transcription? How does it do it? What can that cause for the host?

Answer:

- Adenovirus **E1A** protein *boots* ***retinoblastoma (Rb)*** protein from **E2f**, which adenovirus uses to drive transcription.
- ***Rb*** *recruits HDACs*, so the eviction of Rb from a locus an allow histone acetylation and gene activation, which can promote *cell division and* ***cancer***.

Q14. SV40 and adenovirus share in common the regulatory trait that ***DNA synthesis*** *helps stimulate transition to* ***late-phase*** *transcription.* Desribe the main difference between SV40 and adenovirus in late-phase transition.

Answer: Adenovirus uses two proteins (**IVa2** and **L4**) as positive regulators of late-phase transcription, rather than just the DNA synthesis based stimulation of transition to late-phase transcription.

Q15. What's a distinguishing feature of herpesvirus infection, relative to SV40 and to adenovirus, and with respect to "getting started" by initiating the life cycle and transciptional program?

Answer: **Pre-prepared protein**: Unlike SV40 and adenovirus, which must make *make* a protein to get started, herpesvirus comes packaged with the **VP16** protein pre-made, between the capsid and the envelope.

Q16. Why is *escape from the nucleus* so vitally important in the life of many mRNA molecules?

Answer: For much of the viral life cycle, mRNA needs to be *translated*, and the *ribosomes are in the cytosol*. This means that the mRNA molecules must be exported from the nucleus for translation.

Q17. How do RNA molecules normally make it out of the nucleus and into the cytosol?

Answer: RNA, like proteins and other macromolecules, is *too large to passively translocate* from nucleus to cytosol; instead, RNA molecules often have *nuclear export signals* that are then *bound by* ***exportin*** proteins medaite passage by *interactive with the* ***nuclear pore complex (NPC)*** proteins.

Q18. Why must some viral RNA make it out of the nucleus unspliced?

Answer: If a virus has a RNA genome and assembles into particles in the cytosol, then the full-length RNA genome must get out into the cytosol for packaging into new virus particles.

Q19. How does the discovery of **alternative splicing** relate to virology?

Answer:

- Alternative splicing was discovered in **adenovirus**, as researchers realized that mature (spliced), cytosolic RNA tended to be smaller than many of the younger RNA molecules in the nucleus.
- Upon closer examination with the electron microscope, it was discovered that adenovirus mRNA-DNA duplexes had large DNA loops not hybridized to the RNA; these were introns that had been spliced out.

Q20. How does SV40 leverage a host cell protein to regulate part of the timing of its lifecycle / transcriptional program?

Answer: SV40 uses a host cell protein as a repressor that binds late-phase genes.

Q21. How does SV40 *couple transcription regulation to genome replication*?

Answer:

- SV40 uses a **dilution**-like mechanism to couple DNA synthesis (genome replication) to transcription program timing.
- Specifically, the *level of the host cell protein remains constant.* Once SV40 genome replication begins, though, the ration of repressor to genome copies declines, and there's an increasing availability of DNA molecules that lack repressive binding of the host cell protein to late-phase genes.

Q22. For viruses, what's a major advantage of splicing?

Answer: As with host cells, *splicing generates* ***transcriptome diversity***. The **genome economy** this facilitates, allowing the virus a compact but information-rich genome, is valuable.

Q23. What's the general order of operation of the main proteins involved in miRNA processing?

Answer:

- 1. **DROSHA** forms the initial, *primary micro-RNA (pri-miRNA).*
- 2. **DICER** processes pri-miRNA to produce mature miRNA.
- 3. **Argonaute** binds miRNA and is related to effector function.

Q24. How are ***circular* RNA** molecules generated (by what molecular process)?

Answer: *Backsplicing* generates circular RNA molecules, by joining the 3' end of a downstream exon to the 5' end of an upstream exon.

28 Virology2020 L09 Reverse transcription and integration

Q1. Affirm or refute: any virus that contains reverse transcriptase (RT) is a retrovirus.

Answer: *False*: *Retroviridae* is Baltimore Class VI, but Baltimore Class VII also has RT.

Q2. Affirm or refule: any virus that contains reverse transcriptase (RT) has a *RNA* genome.

Answer:

- *False*: Baltimore Class VII viruses have RT but use a dsDNA genome (dsDNA-RT viruses)
- The dsDNA-RT (Class VII) families are *Hepadnaviridae* and *Caulimoviridae.*

29 random virology

Q1. Give at leas two different solutions to the problem of *targeting a* ***protein*** *to a* ***membrane.***

Answer:

- Retroviruses use **myristate,** a lipid / fatty acid compound, conjugated onto the 5' end of a protein after translation
- Some viruses (and cell proteins) us a **signal sequence** of amino acids within the protein itself.

30 TWiV96 Making Viral DNA

Q1. Main steps: What are the 4 main steps of viral DNA replication?

Answer:

- 1. *Recognition*: DNA replication is *spatially nonrandom*; that is, it begins at one or more defined places within the virus genome.
- 2. *Initiation*: (almost) all DNA synthesis requires a *primer*, and the mechanics/logistics of initiation vary largely w.r.t. what priming strategy a virus uses.
- 3. *Elongation*: Once replication has been initiated, polymerase(s) elongate(s) DNA; in the case of a replication fork, there's typically quite different enzymatic machinery for each of the strands (leading and lagging).,
- 4. *Termination*: Typically (?) by simply running out of template, replication eventually ceases.

Q2. Priming: What are the 3 main strategies for *priming* DNA synthesis? How does this fundamentally differ from RNA viruses?

Answer:

- ***RNA*** *priming*: Some(Specialized *DNA-dependent RNA polymerases* called **primases** can synthesize a short RNA oligomer complementary to a DNA template, providing the essential free 3'-OH for DNA polymerase to prime off of.)
- ***DNA*** *priming*: Some(Some viruses build the free 3'-OH for DNA polymerase right into their genome.)
- ***Protein*** *priming*: Some(Some viruses have a DNA-binding protein that coordinates with DNA polymerase and possible other proteins to ignite DNA synthesis.)
- RNA synthesis often doesn't required a primer, so this is a less essential aspect of consideration for RNA viral study.

Q3. Provide an example of each of the three main priming strategies.

Answer:

- ***RNA*** *priming*: Some(Eukaryotic, linear chromosomes use RNA priming (**primase**, a specialize type of *DNA-dependent RNA polymerase*, or DdRp))
- ***DNA*** *priming*: Some(*Parvoviridae* members like AAV have an exposed 3'-OH at the tail of each inverted terminal repeat (ITR) hairpin end, so the primer's built right into the genome.)
- ***Protein*** *priming*: Some(*Adenovirus* uses protein priming, with **preterminal protein (pTP)** working in conjunction with polymerase.)

Q4. *Elongation* examples: Provide example viruses for each of the two main elongation stratgies.

Answer:

- **Replication fork**: Some(SV40)
- **Strand displacement**: Some(*Parvoviridae* and *Adenoviridae*)

Q5. *Alternative elongation strategy*: What's a major alternative strategy used for elongation during RNA replication, in a way *combining the two major strategies*?

Answer: "Rolling circle" replication is a strategy used by many viruses for DNA genome replication.

Q6. *Rolling circle* replication: provide an example of a virus/es that use *rolling circle* replication.

Answer:

- Human herpesvirus 6 (HHV-6)
- Human papillomavirus-16 (HPV-16)

Q7. Universal truths: What are several things about viral DNA replication that always (or nearly always) hold true?

Answer:

- *Priming*: DNA synthesis (nearly) always requires a primer.
- *Directionality*: DNA synthesis always occurs in the 5'-to-3' direction, reading the template 3'-to-5'.

- *At least one protein*: All viruses need the host to make at least one protein for them. The virus can encode a DNA polymerase, and if it doesn't it needs a protein to "orchestrate the host" to get viral DNA made rather than the excess of host DNA.

Q8. Fundamental problem: What fundamental problem arises for *linear DNA* genome replication? Why for *DNA*?

Answer:

- *The 5' end problem* refers to the difficulty of copying the DNA at the 5'-most end of the template.
- The 5' end problem arises specifically in *DNA* viruses because of the *priming requirement*. RNA viruses don't in general face the same challenge.

Q9. For humans (and mostly all other linear genome eukaryotes?) what's the solution to the 5' end problem?

Answer: Telomeres solve the 5' end problem by signaling the end of chromosomes, with the support of specific sequence maintenance and DNA-binding enzymes and other proteins.

Q10. Bare (protein) necessities: what kinds of proteins provide the minimal activities needed for viral DNA genome replication?

Answer:

- *Recognition*: the virus needs at least one protein to recognize its ori(s), so **origin-binding protein(s)**.
- *Unwinding*: the virus needs **helicase(s)** to unwind the DNA.
- *Elongation*: the virus needs **DNA polymerase(s)** for the main work of DNA synthesis.
- *Chew back*: The virus also needs **exonucleases** to chew back DNA.
- *Metabolism*: The virus depends on various **nucleic acid metabolic enzymes**.

Q11. What's the general relationship between genome size and independence from, vs. reliance on, the host cell?

Answer: In general, *smaller viruses "orchestrate"* while *larger viruses "encode"*; that is, a smaller genome encodes protein(s) that manipulate host machinery while a larger genome encodes machinery specifically for the virus.

31 Molecular Popgen

Q1. What are the two main **axes/forms of violation of the Hardy-Weinberg law's conditions?** Equivalently, what are the two main kinds of conditions?

Answer:

- **Internal** to the population: endogenous forces like *nonrandom mating* and various molecular mechanistic biases.
- **External** to the population: exogenous forces include *gene flow*, e.g. from *migration* or *horizontal gene transfer*, or *natural selection.*

Q2. What are some broad kinds of **molecular mechanistic biases** that cause violation of Hardy-Weinberg conditions?

Answer:

- *Gametogenesis*: certain alleles may skew the distribution of alleles by meiotic drive, whether by *covariance* or by *disproportionality.*
- *Fertilization*: certain alleles may give a particular gamete an unusually high or low probability of winning the Fertilization race.

32 Hardy Weinberg

Q1. What are the **Hardy-Weinberg** law's three *fundamental consequences/conclusions,* that hold when conditions are met?

Answer:

- 1. Allele frequencies never change
- 2. Genotype frequencies change, at most, only in the first mating, remaining constant thereafter.
- 3. Allele frequencies and genotype frequencies are mutually/bidirectinoally determined, so that knowledge of value(s) in one domain directly maps to knowledge of certain value(s) in the other domain.

Q2. What's the general mathy strategy for proving the Hardy-Weinberg law?

Answer: The Hardy-Weinberg is logically proved by, for each progeny/offspring case, taking a *weighted average of conditional probabilities*, with offspring genotype conditions on parental mating genotypes, weighted by the mating's probability.

Q3. What's a useful frame in which to prove the Hardy-Weinberg law?

Answer: It's useful to frame the proof by considering and classifying cases by ***number*** *and* ***kind*** *of heterozygotes* in each mating.

Q4. **Assumptions and justification**: what assumption(s) justify the computation of *genotype frequency* from allele frequency (i.e., a simple product)?

Answer: **Conjectured *independence***: it's the assumed independence, both during *gametogenesis* and during *fertilization*, that justifies this. Specifically, alleles are equally likely to make it into gametes, and gametes of different haplotypes are equally likely to win the "ferilization race."

Q5. **Assumptions and justification**: what assumption(s) about the population dynamics, structure, etc. justify the doubling the *genotype probability* to map from ordered to unordered (e.g., $f(A_iA_j) = 2P_{ij}$)?

Answer: ***Monoecious*, or *balanced allele source frequencies***: If we regard the two "slots" of a diploid genotype as maternal and paternal alleles, then doubling the ordered probability is justified by the assumed equiprobability of a particular allele coming from either parent.

Q6. **Assumptions and justification**: what assumption(s) justify the use of symmetry in computation of *mating probavility* (multiplying by 2))?

Answer: ***Monoecious*, or *balanced genotype frequencies***: "reflecting" the genotypes in the mating and as such doubling the mating probability to go from orered to unordered reflects the assumption of either a monoecious population (no distinction between parents), or *equal genotype distributions* between the population's sexes.

33 HIV Nature 2020 RNA structure DREEM

Q1. Which important part of the HIV genome seems to exist in a roughly 2:1 or 3:1 ratio of 4- and 5-stem structures?

Answer: The **Rev response element (RRE)** exists mainly in this ratio of stem counts.

Q2. What two tools were combined to conduct this study?

Answer:

- **DMS-MaPseq** for the biological, experimental work (methyl adducts to the exposed A and C bases, then random mutations from the RT)
- **DREEM** algorithm for inferring structure from the random mutation data coming out of DMS-MaPseq

Q3. What are the two main challeges with the use of DMS-MaPseq, related to the biochemistry of the experiment?

Answer:

- *Random, low-frequency mutation*: the mutations from the TGIRT-III reverse transcriptase are random and at only $2-10\%$ of methylated sites.

- *Sequence context dependence*: the frequency of the RT's mutations that provide the accessibility signal depends on the sequence context.

Q4. What does this study show about the relationship between biological context (in vitro, in virion, and in infected cells) about RRE structure?

Answer: Proportions of the two main RRE structural variants were consisitent across biological contexts, suggesting that the structure is relatively invariant w.r.t. context.

Q5. Which site in the HIV genome was used to study alternative splicing?

Answer: The **A3 splice site**'s structure was studied to assess the structural impact on alternative splicing?

Q6. What are the main findings regarding structure's effect on alternative splicing?

Answer:

- A3 splice site exists in two main conformations, one with the splice site exposed and one with it hidden in a base-paired structure.
- Other sequences near the A3 splice site (e.g., splice enhancers) also strongly influence splice site use by their structural conformation.
- A3 splice site usage is heavily influenced by structure, with the self-paired structure much less amenable to splicing.

Q7. How was structure manipulated to investigate alternative splicing?

Answer: Mutations were introduced that stabilized the self-base-paired loop structure that hides the A3 splice site.

Q8. Which HIV protein's mRNA level is entirely regulated by the A3 splice site and therefore very affected by structure there?

Answer: The *Tat* protein's mRNA only arises when the A3 splice site is used.

34 HIV Nature 2010 thymic selection and protective MHC allele

Q1. What three immunodevelopmental processes/phenomena are crucial for understanding this paper?

Answer:

- MHC restriction
- *Positive* selection of T cells during development in thymus
- *Negative* selection of T cells during development in thymus

Q2. What's the name/phrase for the HIV patients who keep virus low even without anti-retroviral therapy (i.e., the human subjects of this study)?

Answer: "Elite controllers" are HIV patients capable of suppressing viral load even without ART assistance.

Q3. What's the name of the protective MHC allele?

Answer: HLA-B57

Q4. Why is the protective MHC allele protective?

Answer:

- This allele binds fewer self-peptides than most, implying that it presents a less diverse
- Presentation of fewer self-peptides during thymic selection by this allele leads to a batch of cognate TCRs enriched for few, but strong, peptide-contacting residues
- A TCR pool reliant on relatively few (1-3) individual contacts with a peptide is inherently more *robust to mutations,* as a given amino acid substitution is less likely to lower the interaction between the TCR and the peptide:MHC complex enough to ablate recognition by the T cell.

Q5. What important facts link the scarcer self-peptide presentation by the protective MHC allele to presentation of a less diverse pool of self-peptide?

Answer:

- The subset of the human proteome presented as self-peptide during thymic selection is small relative to the total proteom
- For MHC to bind less self-peptide yet present just as diverse a subset to T cells during thymic selection, we'd need unusually many "collisions" for the protective allele (greater frequency of *jointly* TCR- and MHC-compatible residues).
- Assuming independence between MHC allele and TCR compatibility, then, there's a vanishingly small probability of the protective allele "making up for" its poor self-peptide binding by binding equally many TCR-compatible peptides as MHC proteins from more broadly self-binding alleles.

Q6. Does this "protective allele" phenomenon appear to **generalize**? If so, how/to what?

Answer: Yes, the HLA-B57 allele also appears to confer some protection agains **hepatitis C virus (HCV)**.

Q7. How/why does less broad self-peptide presentation by a MHC molecule yield a cognate TCR pool for that MHC that's more "tolerant of change," or "robust to mutations" (i.e., better for HIV control)?

Answer:

- ***Negative* selection** during thymocyte development eliminates overly self-reactive T cells (those binding to peptide:MHC complex too strongly, above some threshold).
- As more self-peptide is presented by a MHC, the probability rises for each compatible T cell that it will react too strongly with a complex and be eliminated.
- More self-peptide presentation by a MHC, then, tends to yield a cognate TCR pool enriched for weaker interactions.
- To still be positively selected despite individual residue's relative weak interaction with peptide, the TCRs for a broadly-self-presenting MHC tend to rely on more individual contacts with the peptide.
- Reliance by the TCR on more individual contacts with peptide means a given AA substitution in the peptide is more likely to ruin the recognition by the T cell.

35 Influenza Nature 2019 MHC classII receptor for novel bat IAV

Q1. What's flu's surface protein that it uses for entry? What's the usual receptor? What's the receptor for these two new bat IAVs (H17N10 and H18N11)?

Answer:

- Flu uses the *hemagluttinin (HA)* protein to bind to a cell surface molecule.
- Normally, flu uses *sialic acid,* an acidic sugar on cell surfaces, as its receptor; there's a connection between flu strain, host species, and sialic acid linkage preference (i.e., which bonds in the sialic acid the hemagglutinin prefers to bind).
- For these bat viruses, the receptor is not a sugar (like sialic acid), but rather a protein (MHC class II).

Q2. What's a a particularly striking difference between flu's typical receptor and the receptor for these bat viruses?

Answer: Flu's normal receptor is an acidic sugar, but in these bat viruses it was thought previously, and shown here, to be a protein.

Q3. Which two methods were used to screen *broadly* for candidate genes

Answer:

- *Differential expression* contrasted gene expression from three cell cell lines susceptible to bat IAV with gene expression from three cell lines not susceptible to the virus, looking for genes up in all three susceptible cell lines relative to the non-susceptible lines. Candidates were filtered based on Gene Ontology search for plasma membrane association/function.
- *Genome-wide CRISPR screening* was used transfecting *Cas9*-expressing cells with lentiviral vectors carrying a bunch of guide RNAs, and then cells were selected for resistance to infection by VSV-H18 (VSV expressing one of the bat flu hemagluttinins). Then analysis showed which guide RNAs were in those cells (so which genes were knocked down).

Q4. What method was used to *hone in on the target* gene(s)?

Answer: With the genes up in all three susceptible cell lines (relative to non-susceptible) in hand, the investigators used **RNAi** (specifically, with **siRNA**) directed against the target genes.

Q5. In what main ways was it demonstrated definitively that MHC class II is the receptor for these new bat flu viruses?

Answer:

- Genetic knockout of MHC class II is protective against entry of H18 VLPs (controlled by H1N1 VLPs still showing ability to enter).
- siRNA knockdown (maybe) of MHC class II impedes H18 VLP entry.
- MHC class II expression from lentiviral vectors restores susceptibility to entry by H18 VLPs.

Q6. In what main ways was cell the link between the receptor both strengthened and generalized?

Answer:

- Many different species HLA-DR molecules confer susceptibility on HEK293 cells (express both chains for that MHC molecule) in the HEK293 cells and infect with VSV or VLPs with the bat flu hemagluttinins H18.
- Different MHC alleles also confer susceptibility (e.g., HLA-DRQ and HLA-DRP).

Q7. Cross-species transmissibility: What were the main findings about susceptibility in *mice* to the actual bat viruses?

Answer:

- It appears that *mice* are indeed susceptibile, but virus replication is *confined to the* ***upper respiratory tract***, and really only to a region of the **epithelium**.
- MHC class II deficiency is protective in mice.
- The link between MHC class II and susceptibility is strengthened by ablating MHC class II specifically on endothelial cells in the upper respiratory tract (basically the only replicatibe location) and demonstrating lack of plaque formation (i.e., no antigen-presenting cells still have MHC class II, but virus doesn't replicate).