Wrap-up & Next Steps

1.) Explore data!

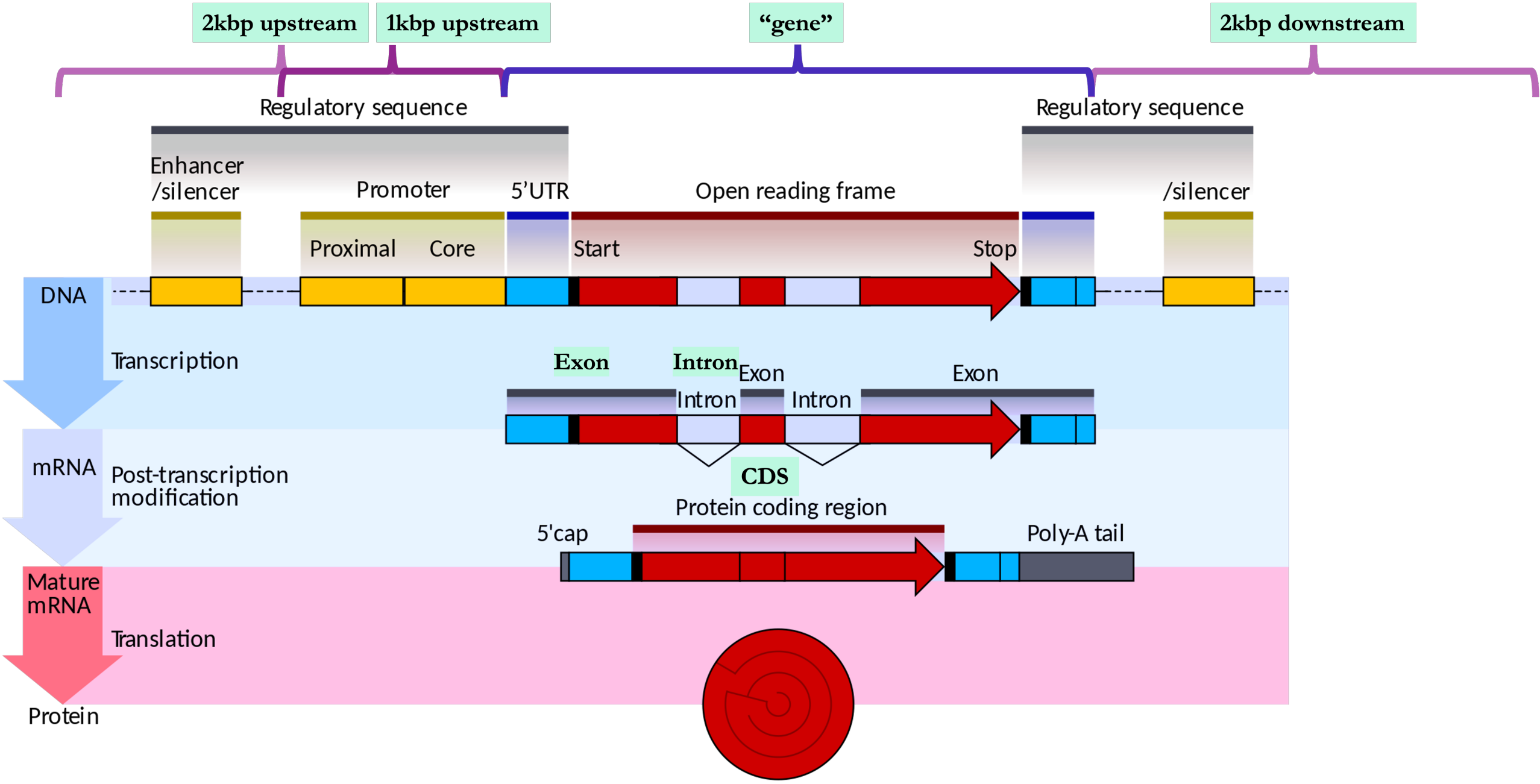
2.) Determine how to filter & prioritize SNPs for your research questions! (verbally)

- SNP filtering & prioritization questions
- Examples

GT-Seq Panel Prioritization

Q1: *Where* do you want your SNPs?

1A.) Candidate Gene Identification



GT-Seq Panel Prioritization

Q2: *How* do you want your SNPs?

- 1 SNP/amplicon, Bi-allelic only

- Multi-allelic where available

- Prioritize microhaplotypes

GT-Seq Panel Prioritization

Q3: *Which* loci?

- Any must-have candidate genes?
- **1 amplicon/gene** or more for certain genes?
- *Prioritization stats? (e.g., MAF, HWE, prioritize transitions vs transversions, etc.)
  or just try to balance/diversify among annotations?

Examples

Research Objectives:

Develop a gt-seq panel to aid in fisheries management

- stock identification by geographic basin
- migratory form (freshwater resident or anadromous)
- reproductive ecotype (stream-, shore-, deep-spawning)

Approach:

- SNPs ID'd from RADseq data (n= 7,347)
  - ➢ outlier loci to differentiate among migratory & reproductive ecotypes
  - ➢ neutral loci filtered to remove putative paralogs & loci that deviated from HWE

SNP Prioritization & Filtering:

- Weir and Cockerham's (1984) pairwise genetic differentiation (θ) (selected n= 550)
  - ➢ Sequentially ranked list by θ values (highest divergence) for geographic basin comparisons (n=200)
  - ➢ Sequentially ranked list by θ values (highest divergence) for migratory form comparisons (n=100)
  - ➢ Sequentially ranked list by θ values (highest divergence) for reproductive ecotype comparisons (n=250)
- Tested for pairwise linkage disequilibrium & removed less informative locus (n= 547)

Chang SL, Ward HGM, Russello MA (2021) Genotyping-in-Thousands by sequencing panel development and application to inform kokanee salmon (Oncorhynchus nerka) fisheries management at multiple scales. PLoS ONE 16(12): e0261966.
https://doi.org/10.1371/journal.pone.0261966

Examples

Research Objectives:

Develop a gt-seq panel to aid in genetic population monitoring

- non-invasive samples
- reliably differentiate individuals
- ascertain sex
- assess relatedness
- resolve population structure

Approach:

- SNPs ID'd from ddRADseq data & added 2 sex-determining loci (n = 411,094)

SNP Prioritization & Filtering:

- Min. depth = 10; min. genotype quality = 18 (n = 201,983)
- Max missing/locus = 50%; filtered for only biallelic SNPs (n = 79,906)
- MAF > 0.01 (n = 12,353)
- HWE (B-Y p value = 0.0059) (n = 11,144)
- Removed individuals with >70% missing data
- Successfully genotyped in ≥ 85% of individuals (n = 3,122)
- MAF > 0.25 (n = 689)
- SNPs thinned to 1 SNP/50,000 bp to reduce likelihood of linkage (n = 559)

Hayward et al. (2021) Genotyping-in-thousands by sequencing (GT-seq) of non-invasive fecal and degraded samples: a new panel to enable ongoing monitoring of Canadian polar bear populations. https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13583

GECO - Conservation Genomics

Research Objectives:

Develop a gt-seq panel of loci to aid in conservation management & long-term population monitoring

- Targeting genes important for reproductive fitness & survival:

Immune system function, reproductive timing, reproductive & predator avoidance behavior, reproduction, sperm development & morphology.

Example Utility:

- Quantify population genetic health (π; F_{IS}; N_e; HFC)
- Monitor trajectories of population genetic health metrics over time
- Monitor effects of marsh habitat management on breeding marsh population
- Test for inbreeding depression; population adaptive capacity
- Monitor strength of selection on adaptive loci over time
- Identify adaptive/maladaptive alleles/genotypes - important for translocations & captive breeding

Approach: - (Candidate Genes ID'd from literature & Gene Ontology terms; SNPs from WGS)

- SNPs must be variable at the population-level
- Prioritize SNPs in regulatory regions & non-synonymous AA changes in CDS
- Combination of loci with high & low heterozygosity (esp. latter if differ among species)?
- Weir and Cockerham's (1984) pairwise genetic differentiation (θ) among populations?
- Weighted distribution of SNPs among annotations (e.g., 60% immune-related, 20% reproduction…)

Wrap-up & Next Steps

1.) Explore data!

2.) Determine how to filter & prioritize SNPs for your research questions! (verbally)

- Goal:

  Have written summary/powerpoint slide with bullet points including:

  1.) Reiterate your research objectives for the panel loci. Include any planned analyses if known.

  2.) List of potential filtering & prioritization steps for your loci

  By next GECO meeting – February 14th, 2022 (~2 weeks)