

Low Coverage Whole Genome Sequencing (LcWGS)

Jordan Rodriguez
2022 Omics Workshop

Next Generation Sequencing Recap

- Fragmenting DNA/RNA into multiple pieces, adding adapters
- Sequencing the libraries
- Reassembling them to form a genomic sequence that we can analyze



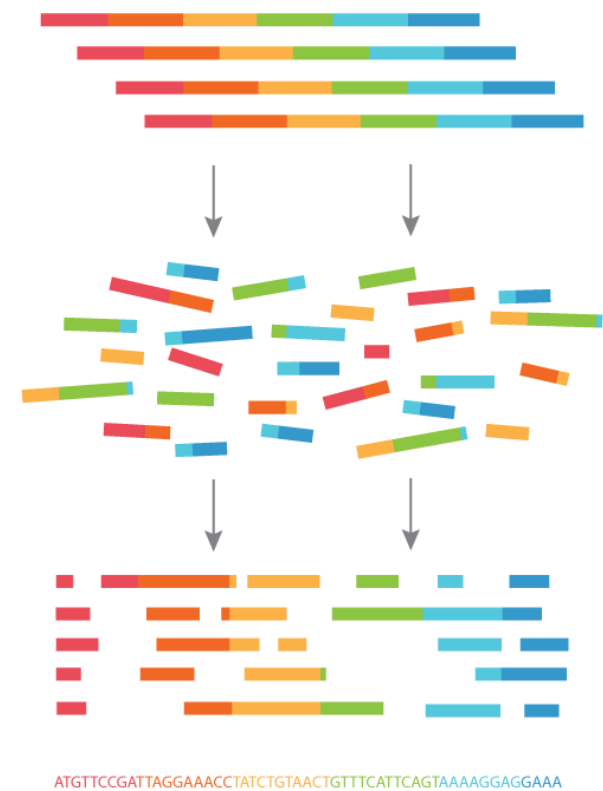
Prepare



Sequence



Analyze



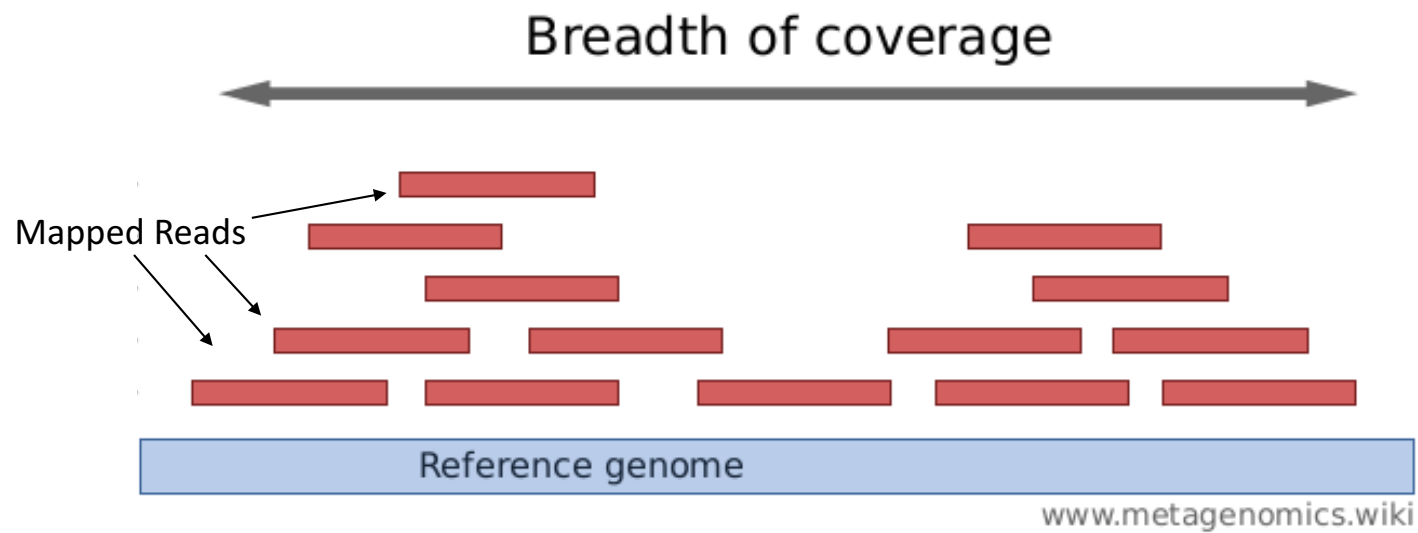


Sequence

IcWGS is a Specific Type of Next Generation Sequencing

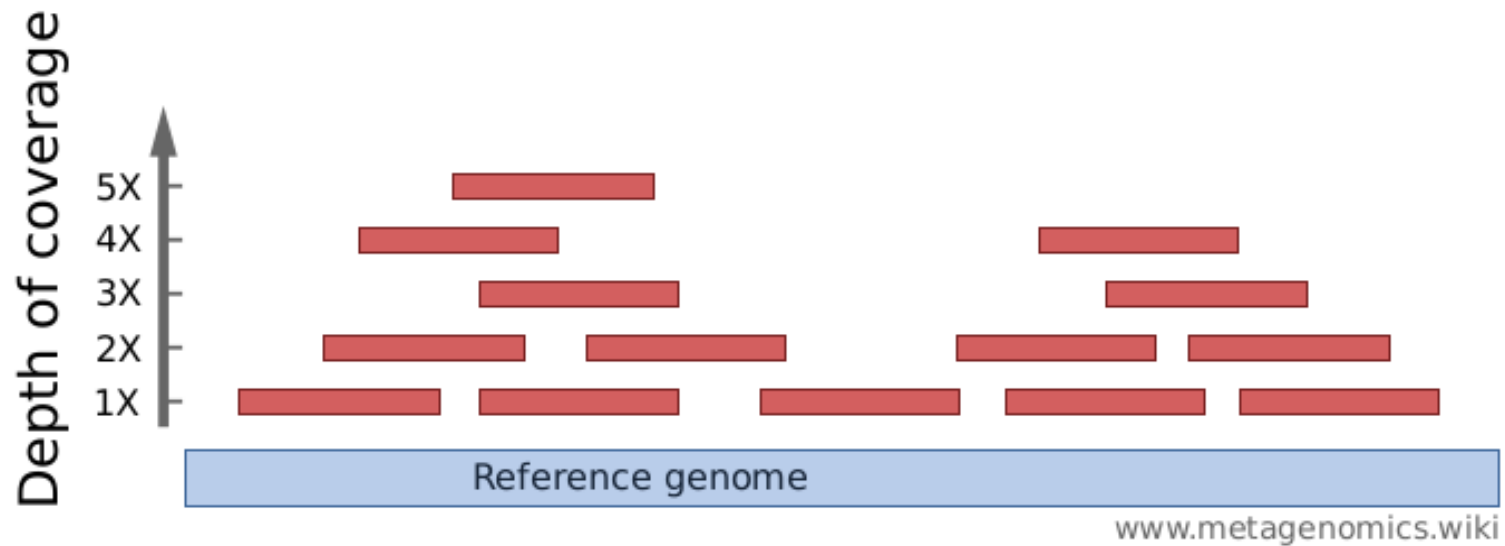
Coverage vs Depth

- What is the difference?

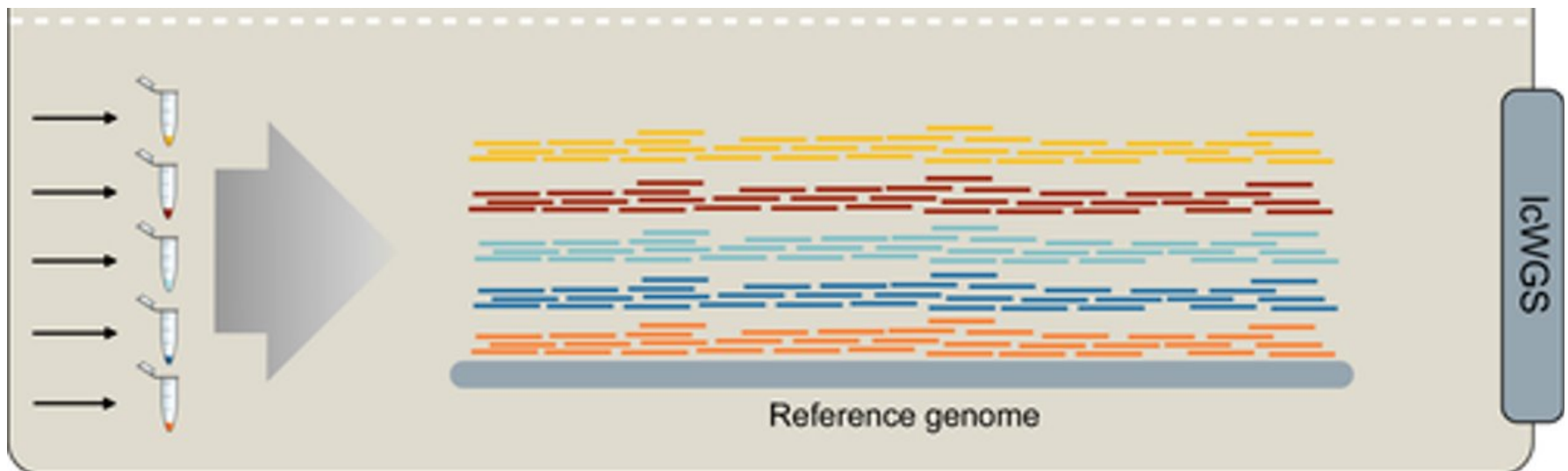


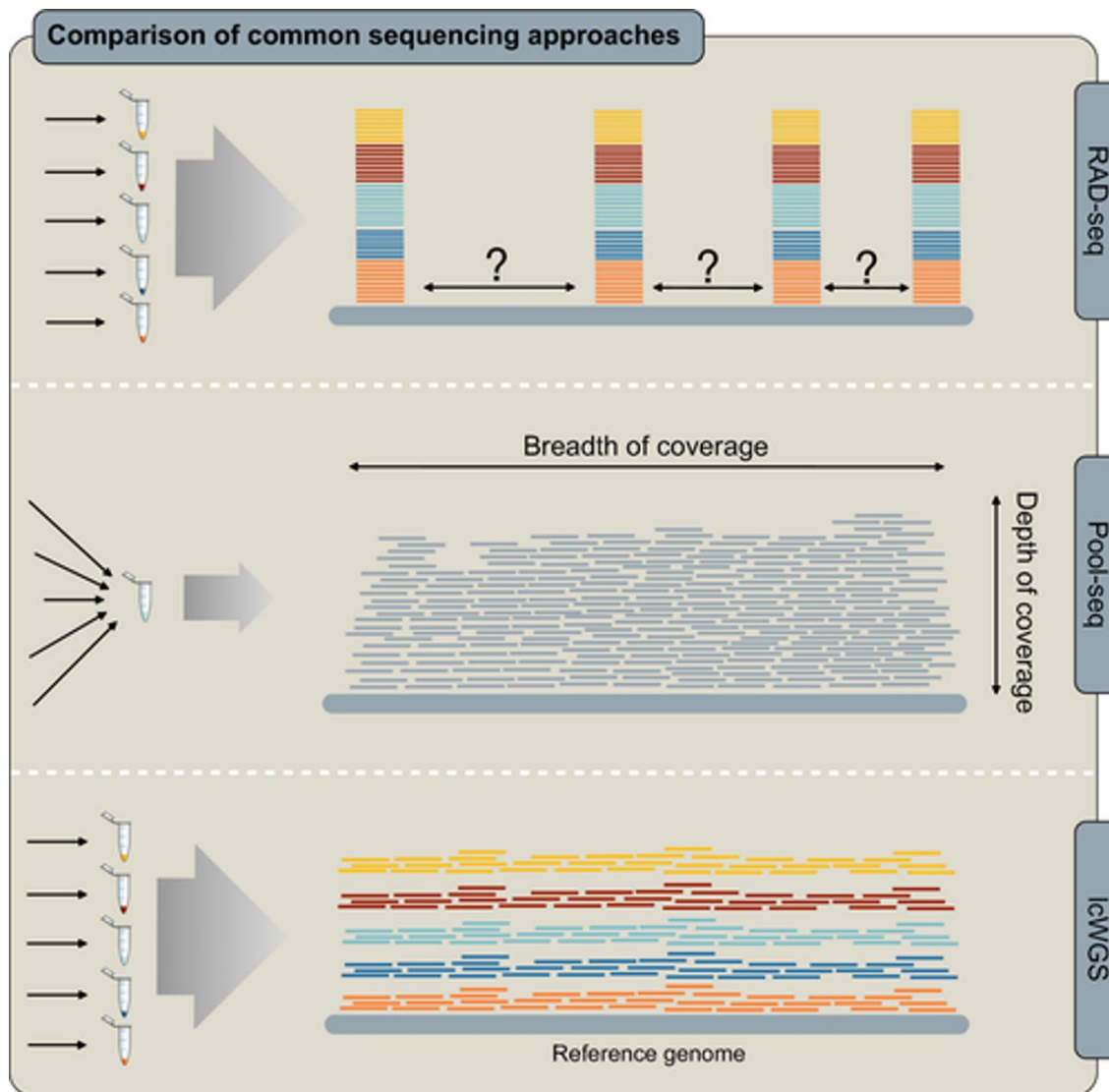
Coverage vs Depth

- What is the difference?



Low Mean Depth of Coverage





High depth from multiple individuals, but low breadth of coverage

High depth from one individual

Low depth from multiple individuals

How Low Can You Go?

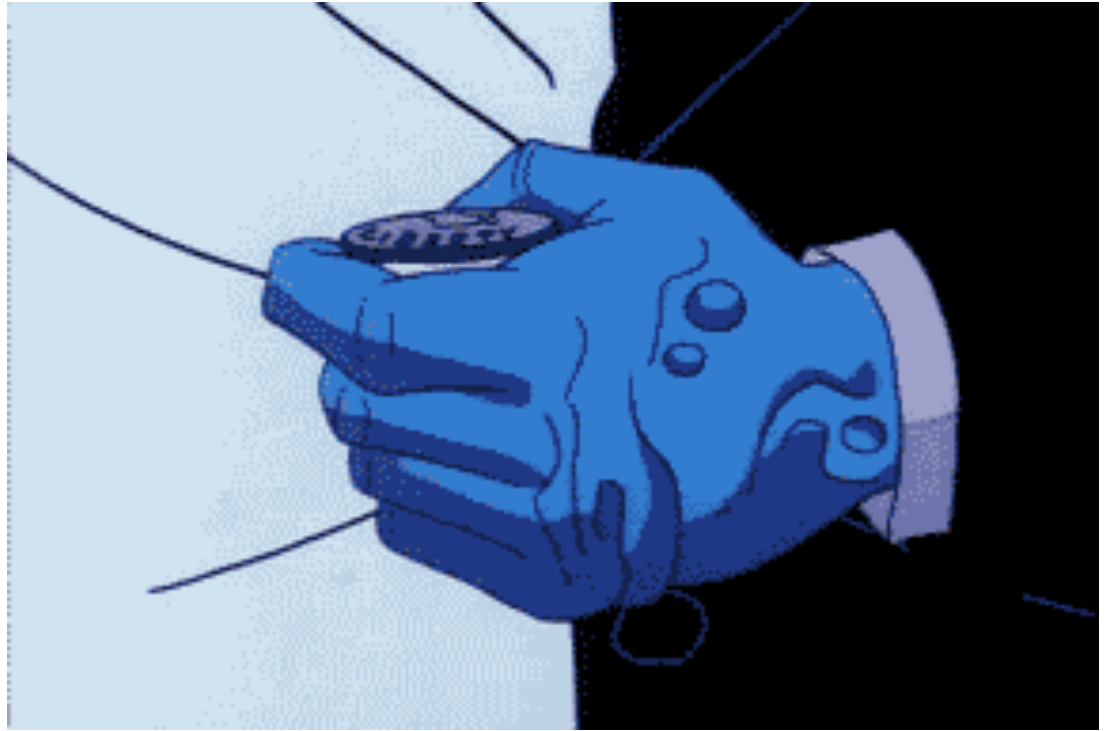
- What is the value of sequencing merely 0.1 – 3X?
- If one sample is **missing information** on a particular SNP, **we can recover it** through statistical imputation applied to the data set as a whole.



Credit: National Cancer Institute

Genotype Probabilities Over Called Genotypes

- With lcWGS, **read depth is too low** to confidently call genotypes
 - Think of a coin flip
 - 50 flips vs 3 flips
- Working under a **probabilistic statistical framework**



Statistical frameworks in place

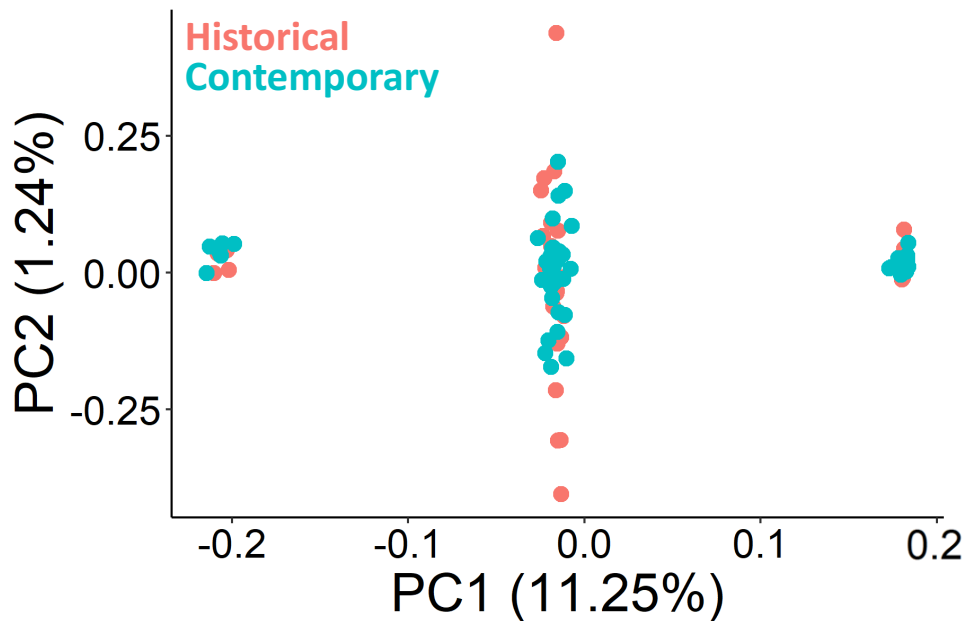
- 1cWGS requires **specialized analysis tools** that explicitly account for **genotype uncertainty**.
- **ANGSD**
- PCANGsd
- Fgvieria/NgsLD
- SNeP
- Can do things like:
 - Analyze next generation sequencing data !
 - Pairwise relatedness

Statistical frameworks in place

- lcWGS requires **specialized analysis tools** that explicitly account for **genotype uncertainty**.
- ANGSD
- **PCANGsd**
- Fgvieria/NGsLD
- SNeP
- perform multiple **population genetic analyses** using genotype likelihoods
 - Covariance matrix
 - Admixture estimations
 - Inbreeding coefficients (both per-individual and per-site)
 - HWE test

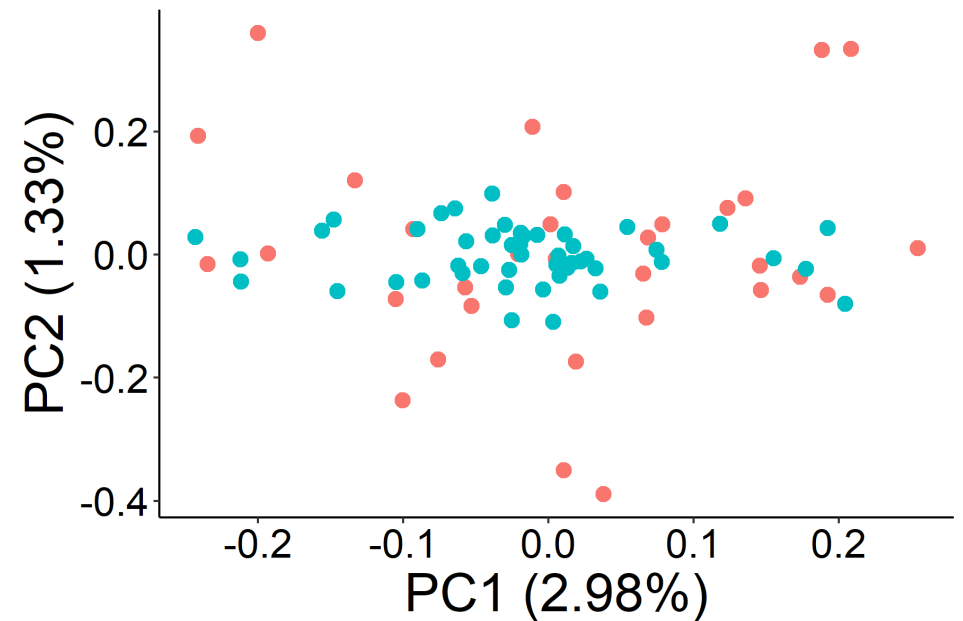
Results: Both Eras Similar – Same Population

CHR04 – Inversion?



CHR05, CHR06, CHR11, CHR14- follow this pattern

CHR01 – No Inversion



All other chromosomes follow this pattern, relatively (16 chromosomes)

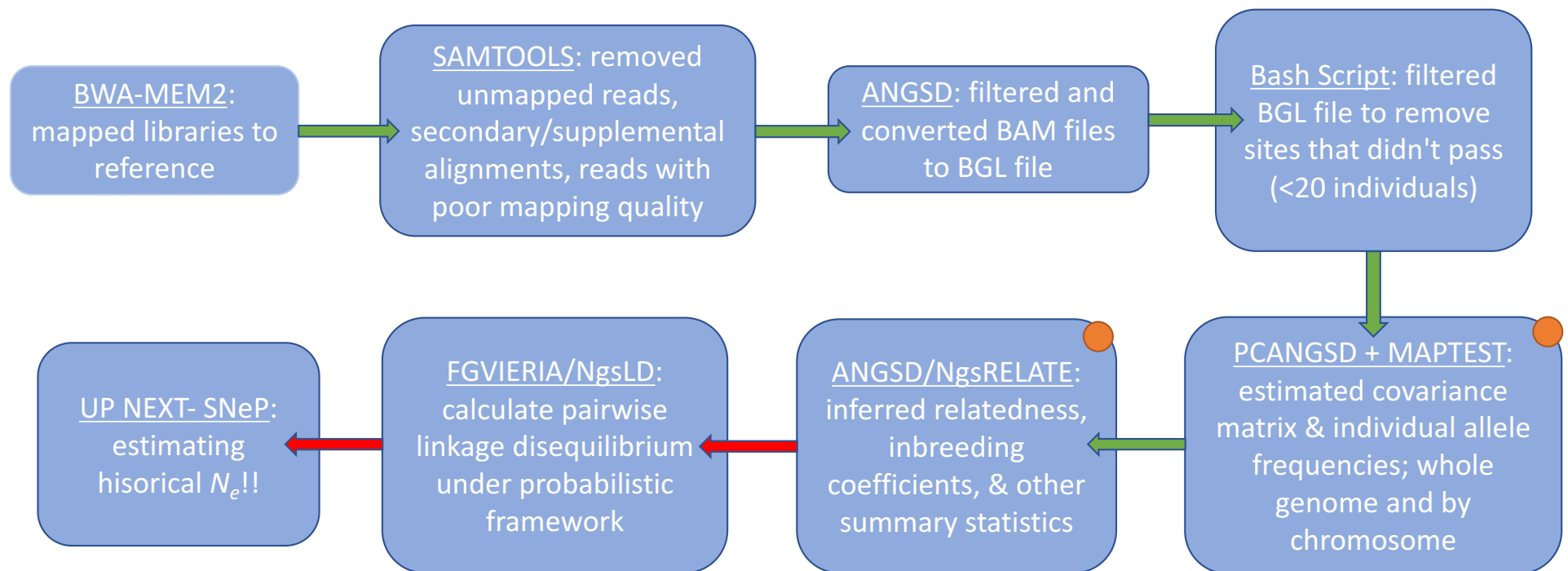
Statistical frameworks in place

- lcWGS requires **specialized analysis tools** that explicitly account for **genotype uncertainty**.
- ANGSD
- PCANGsd
- **Egvieria/NGsLD**
- SNeP
- program to estimate **pairwise linkage disequilibrium (LD)** taking the uncertainty of genotype's assignation into account. It does so by **avoiding genotype calling**

Statistical frameworks in place

- lcWGS requires **specialized analysis tools** that explicitly account for **genotype uncertainty**.
- ANGSD
- PCANGsd
- Fgvieria/NgsLD
- **SNeP**
- A tool that estimates changes in effective population size (N_e) from LD values calculated with other software
 - N_e – can describe the amount of genetic drift in a population

Applications of lcWGS



Benefits of lcWGS

- Powerful and Cost-efficient
 - Example: Library construction: 1/4x KAPA HyperPrep+ Kit
- Largely avoids the negative effects of DNA degradation and enzyme inhibitors in the preservative (French 2021)
- Possibility of population screening of the entire genome while retaining individual information.

IcWGS in action

- Watch how I've used IcWGS to sequence century old *Salarias fasciatus* specimens!



Let's play Kahoot!

- Join at **www.kahoot.it**
- or with the **Kahoot! App**