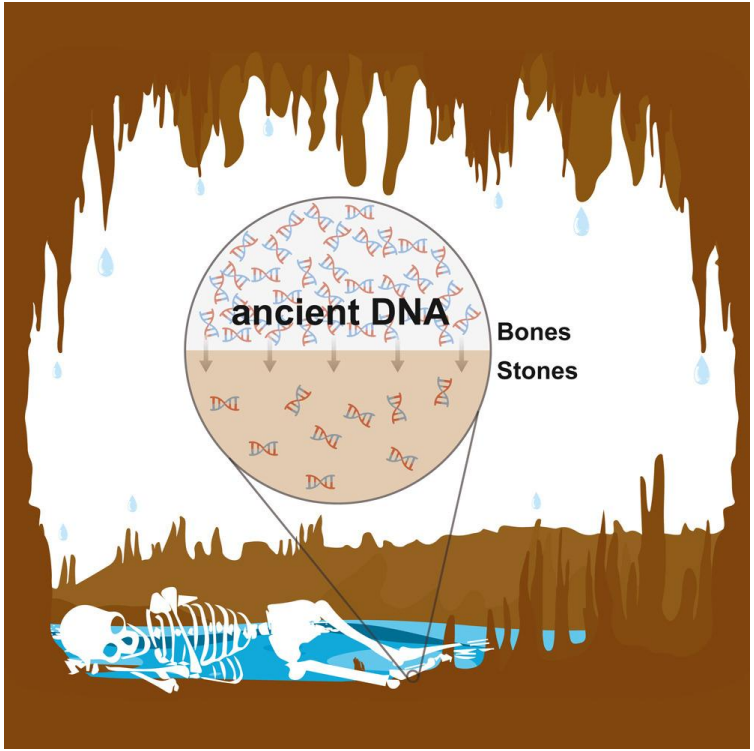
Building a Pipeline for Imputing Ancient Genomes

Kiran Kumar
Zoellner Lab

Ancient Genomes are Low Quality

- ➔ Precious samples with low-yield of aDNA
- ➔ Low coverage sequencing (< 1x) cost effective
- ➔ Deamination (C → T) chemical degradation leads to more genotyping errors

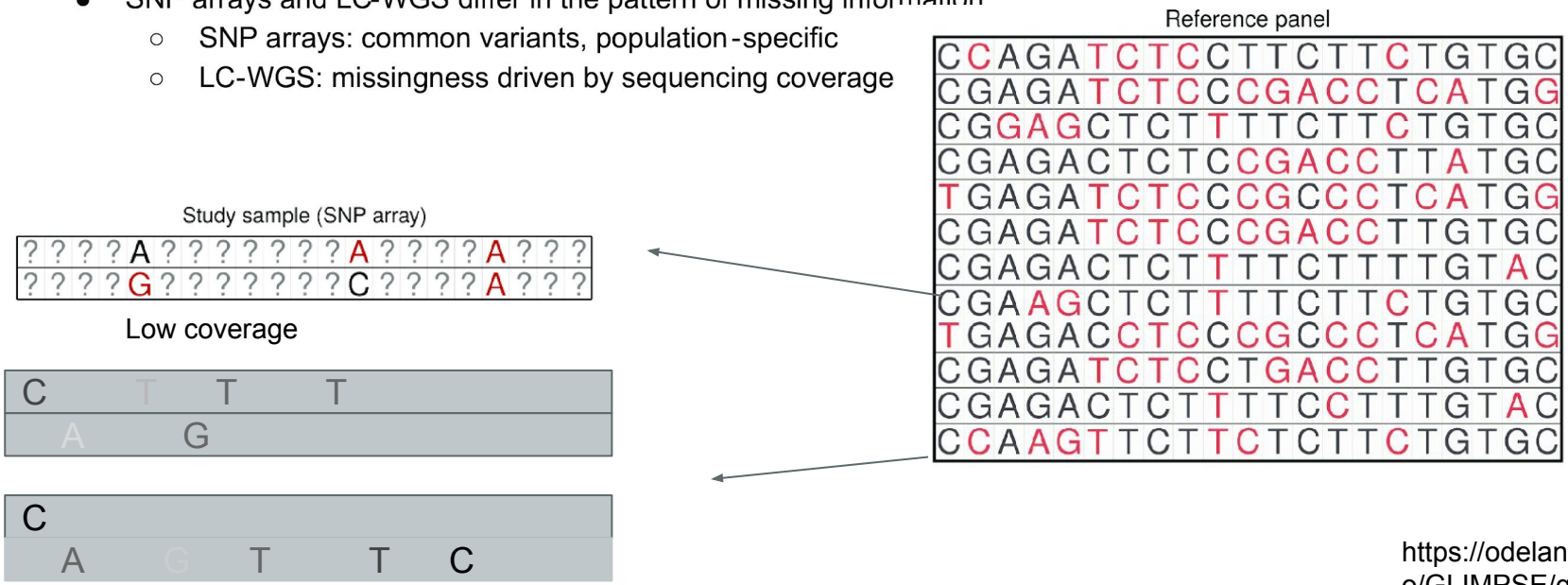
How do we improve this low quality data to answer questions about population history?



https://www.sciencedirect.com/science/article/pii/S2589004221013687#undfig1

Imputation for Low Coverage Genomes

- SNP arrays and LC-WGS differ in the pattern of missing information
  - SNP arrays: common variants, population-specific
  - LC-WGS: missingness driven by sequencing coverage



https://odelaneau.github.io/GLIMPSE/glimpse1/overview.html

Why Impute?

- ➔ Imputation improves the quality of genotypes for low coverage reads ("Refining") and fills in missing genotypes

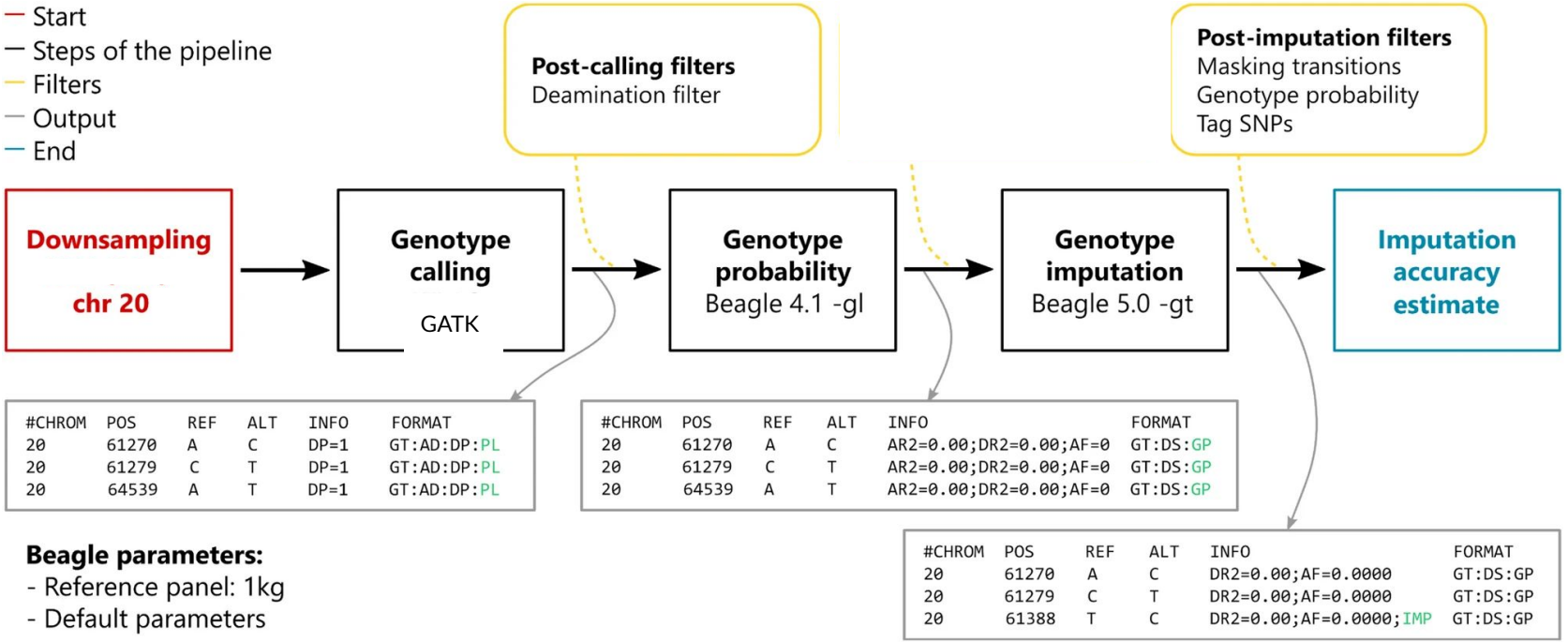
- ➔ Increases power/genotype density for almost any downstream analysis
  - ◆ GWAS,PCA, fine-mapping, population genetics analyses (runs of homozygosity, kinship, etc. …)

- ➔ Get diploid data from from < 1x coverage
  - ◆ On average 1x data only have either reads on maternal or paternal chromosomes

Ancient Samples Summary Statistics

Sample ID	Ancestry *ybp = years before present
I10873 “Shum Laka 2”	African ~8000 ybp
I10871 “Shum Laka 4”	African ~3000 ybp
I0103 “Early Neolithic”	European ~4500 ybp
I0054 “Late Neolithic”	European ~7000 ybp

Refinement Pipeline



Hui et al 2020 Nature Scientific Reports

Tools

Samtools (downsampling, check coverage)

Bcftools (data wrangling for vcf files)

GATK Haplotype Caller (calling genotypes from .bam files)

Beagle 4.1 and 5.0 (refinement and imputation)

Things I learned the way so you don't have to!

- ➔ Check the sequencing coverage of your data!
- ➔ Chop up files to use the cluster efficiently
- ➔ Divide scripts into smaller chunks for easier debugging/modification/documentation
- ➔ Make sure reference genome (hg19 vs hg37) alignment matches for target samples and reference panel

Thank you!

Zoellner Lab Members, Alumni, & Collaborators

Sebastian Zoellner

Jean Morrison

Yuhua Zhang

Kevin Liao

Noel McAllister

Andy Beck

Yichen Si

Jiongming Wang

Alicia Dominguez

Ruoyao Shi

Pedro Orozco Pino

Brian Browning

Keitan Yu

Simone Rubinacci