Diagram Illustrating a Genetic Analysis Workflow that Uses ASAFE

Columns: Individuals

Rows: SNPs

Matrix of **Phased genotypes** for **Reference Panel 1** (e.g. Africans)

Columns: Individuals

Rows: SNPs

Matrix of **Phased genotypes** for **Reference Panel 2** (e.g. Europeans)

Columns: Individuals

Rows: SNPs

Matrix of **Phased genotypes** for **Reference Panel 3** (e.g. Native Americans)

Columns: Individuals

Rows: SNPs

Matrix of **Phased genotypes** for **Admixed Sample** (e.g. Hispanics)

Phasing program e.g. BEAGLE

Local ancestry calling program e.g. RFMIX

RFMIX only outputs results for SNPs with phased genotypes in ALL 3 reference panels AND the admixed sample. You can directly get ancestry specific allele frequencies for such SNPs by overlaying phased ancestries on re-phased genotypes.

Columns: Individuals

Rows: SNPs

Matrix of **Phased ancestries** for **Admixed Sample** (e.g. Hispanics)

Columns: Individuals

Rows: SNPs

Matrix of **Re-phased genotypes** for **Admixed Sample** (e.g. Hispanics)

Columns: Individuals

Rows: SNPs

Matrix of **Unphased genotypes** for **Admixed Sample** (e.g. Hispanics)

What about a SNP that is **NOT typed in a reference panel** (so phased reference genotype at the SNP is absent), or a SNP for which you have **dosage, a number in [0,2] equal to 2*p(11) + p(01) with posterior probabilities p(11) and p(01) from BEAGLE, calculated in the admixed sample** (so phased admixed genotype at the SNP is absent)?

Can use ASAFE.

What about a SNP that is **NOT typed in a reference panel** (so phased reference genotype at the SNP is absent),
or a SNP for which you have **dosage, a number in [0,2] equal to 2*p(11) + p(01) with posterior probabilities p(11) and p(01) from BEAGLE, calculated in the admixed sample** (so phased admixed genotype at the SNP is absent)?

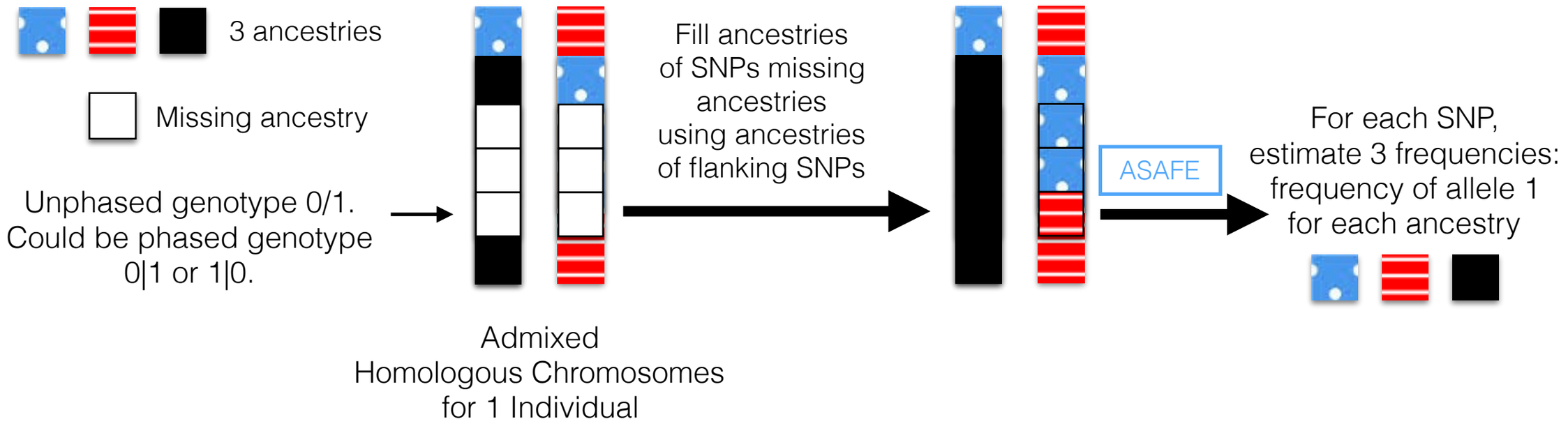

Figure 1. Data right before applying ASAFE: Phased ancestries and unphased genotypes for all bi-allelic SNPs with unphased genotype observed or imputed in all admixed individuals.
A genotype can be imputed like so: Round a dosage in [0,2] to 0, 1, or 2 corresponding to unphased genotypes 0/0, 0/1 i.e. 1/0, or 1/1 respectively, or take the highest probability genotype from BEAGLE.
Position along the chromosome is on the y-axis. Rows are SNPs.