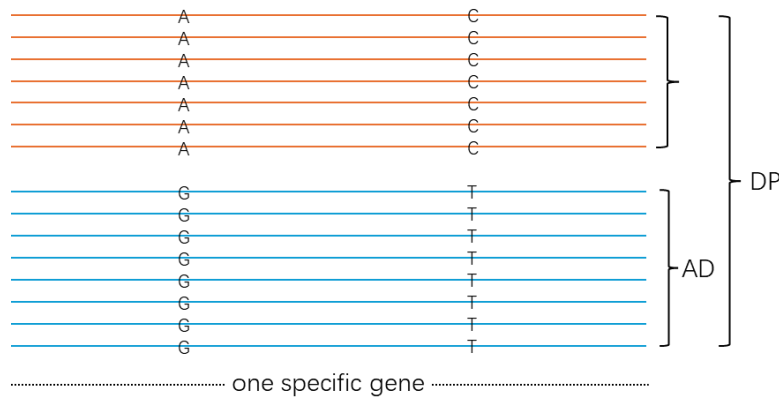
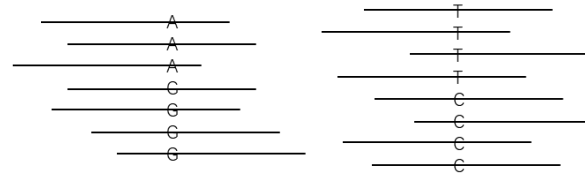
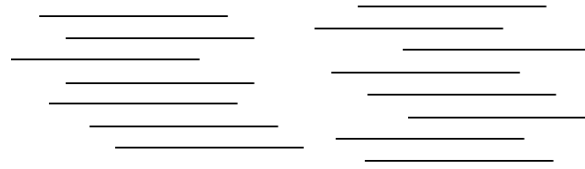


BAM file(s)



VCF file

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
.....									
1	20230403	.	A	G	.	.			
1	20240719	.	T	C	.	.			
.....									

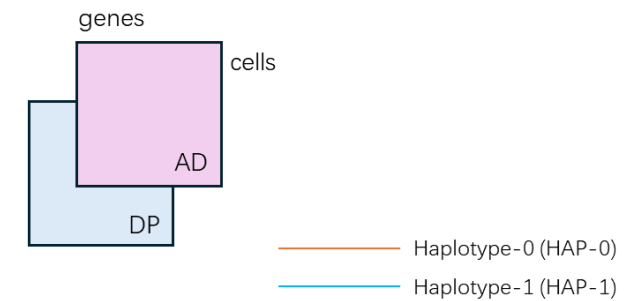
**pileup:** (1) count the REF & ALT alleles;  
(2) do genotyping and keep heterozygous SNPs only.

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
.....									
1	20230403	.	A	G	.	.	AD=40;DP=70	GT	0/1
1	20240719	.	T	C	.	.	AD=40;DP=80	GT	0/1
.....									

**phasing:** (1) do reference phasing with Eagle2;  
(2) to obtain the haplotype information of REF & ALT alleles.

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
.....									
1	20230403	.	A	G	.	.	AD=40;DP=70	GT	0 1
1	20240719	.	T	C	.	.	AD=40;DP=80	GT	1 0
.....									

**snp2gene** (1) aggregate haplotype-specific counts of all SNPs in a gene;  
(2) output cell x gene AD (HAP-1) and DP (both HAPs) count matrices.



xcltk – preprocessing BAF signals

Version: 0.3.1

Usage: xcltk baf [options]

#### Options:

--label STR	Task label.
--sam FILE	Comma separated indexed BAM/CRAM file(s).
--samList FILE	A file listing BAM/CRAM files, each per line.
--barcode FILE	A plain file listing all effective cell barcodes, for droplet-based data, e.g., 10x Genomics.
--sampleList FILE	A plain file listing sample IDs, one ID per BAM, for well-based data, e.g., SMART-seq.
--snpvcf FILE	A VCF file listing all candidate SNPs.
--region FILE	A TSV file listing target features. The first 4 columns are: chrom, start, end (both 1-based and inclusive), name.
--outdir DIR	Output dir.
--gmap FILE	Path to genetic map provided by Eagle2 (e.g. Eagle_v2.4.1/tables/genetic_map_hg38_withX.txt.gz).
--eagle FILE	Path to Eagle2 binary file.
--panelDir DIR	Directory to phasing reference panel (BCF files).
--version	Print version and exit.
--help	Print this message and exit.

#### Optional arguments:

--cellTAG STR	Cell barcode tag; Set to None if not available [CB]
--UMITag STR	UMI tag; Set to None if not available [UB]
--ncores INT	Number of threads [1]

#### Notes:

1. One and only one of `--sam` and `--samList` should be specified.
2. For well-based data, the order of the BAM files (in `--sam` or `--samList`) and the sample IDs (in `--sampleList`) should match each other.
3. For bulk data, the label (`--label`) will be used as the sample ID.