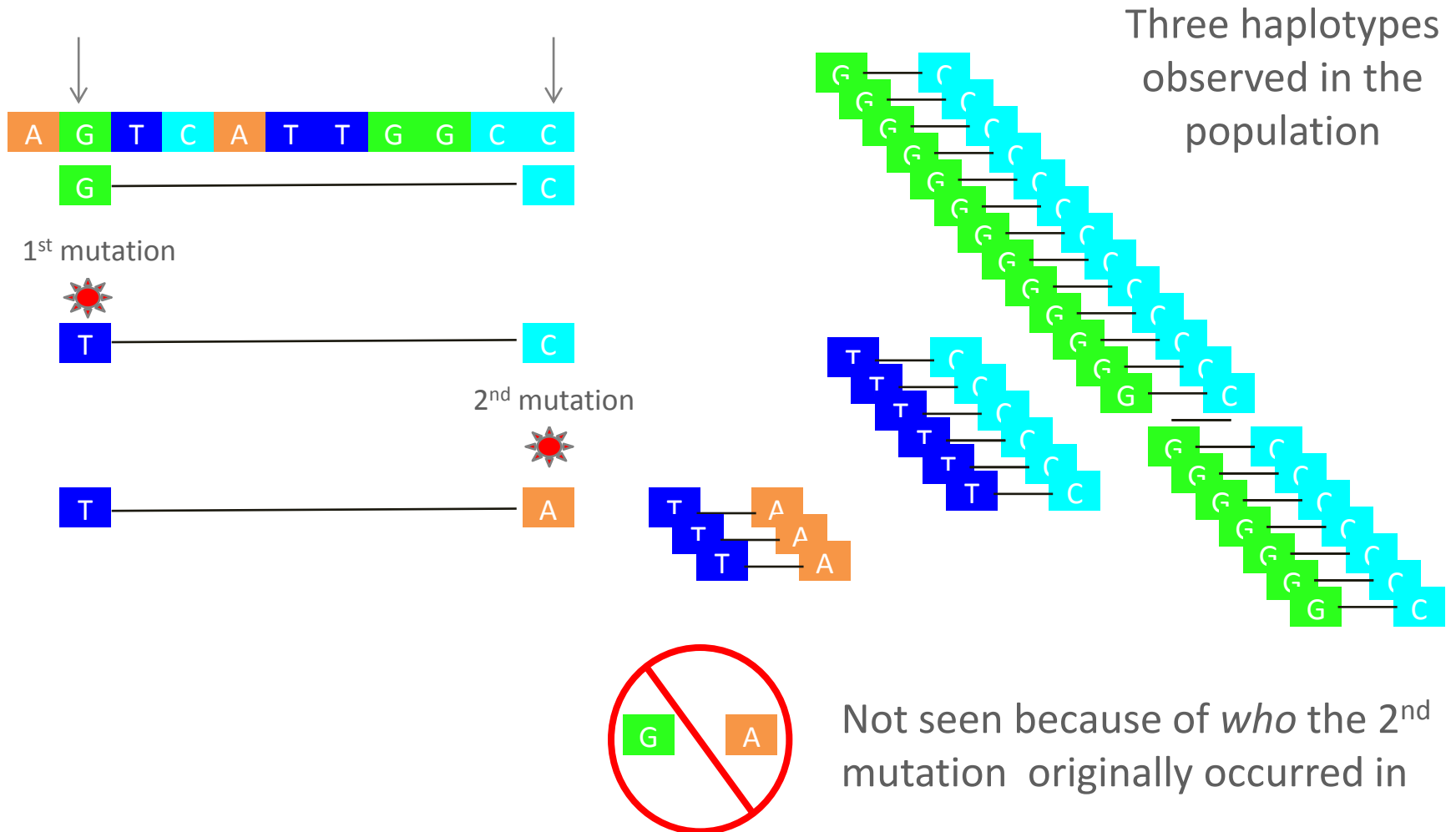


MODULE 5: Architecture of human genetic variation

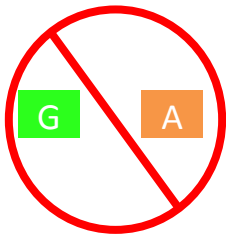
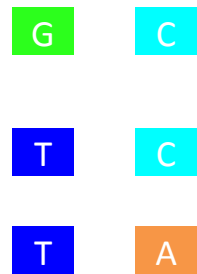
How haplotypes (co-inherited alleles) arise



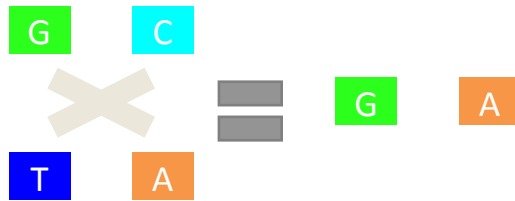
Chromosomal recombination breaks up haplotypes

Before
recombination

3 haplotypes



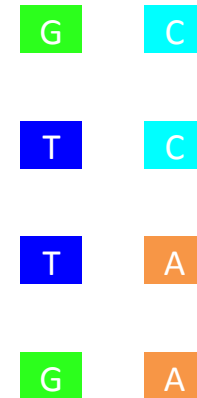
*Chromosomal
recombination*



After

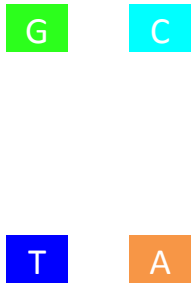
recombination

4 haplotypes

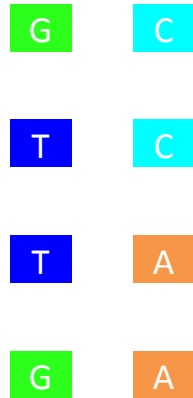


Linkage disequilibrium

Two SNPs in strong LD



Two SNPs with no LD



- Alleles inherited in a haplotype are said to be in 'linkage disequilibrium' (LD)
- LD is stronger when distance between variants is short
- LD is shaped by recombination

LD allows prediction of alleles

- For two variants in strong LD, alleles at one location provide information about alleles at another location

Variants in strong LD

G C

T A

Variants with no LD

G C

T C

T A

G A

Predict the missing allele using information from the variants in strong LD

T ?

G ?

Now try again for the variants with no LD

T ?

G ?

Macro-level haplotype structure in the human genome

- Haplotype blocks punctuated by regions of recombination

ATTGCCGATACGGGACTTAACGACTAACCAACACTAGGCAGATCGACCAGATCGACGTAGCCAGCTTA



block1
84kb

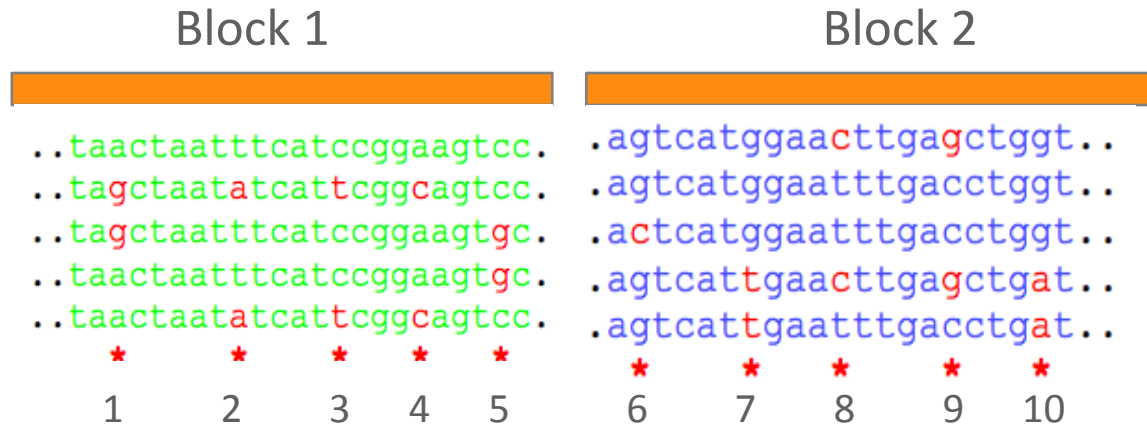
block2
3kb

block3
14kb

block4
30kb

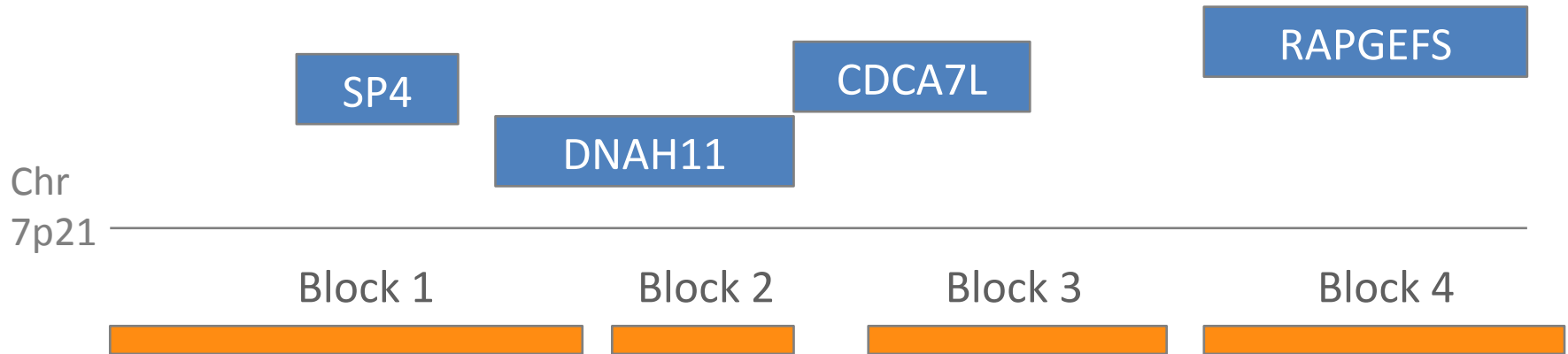
block5
25kb

Strong correlation within a haplotype block, but not between blocks

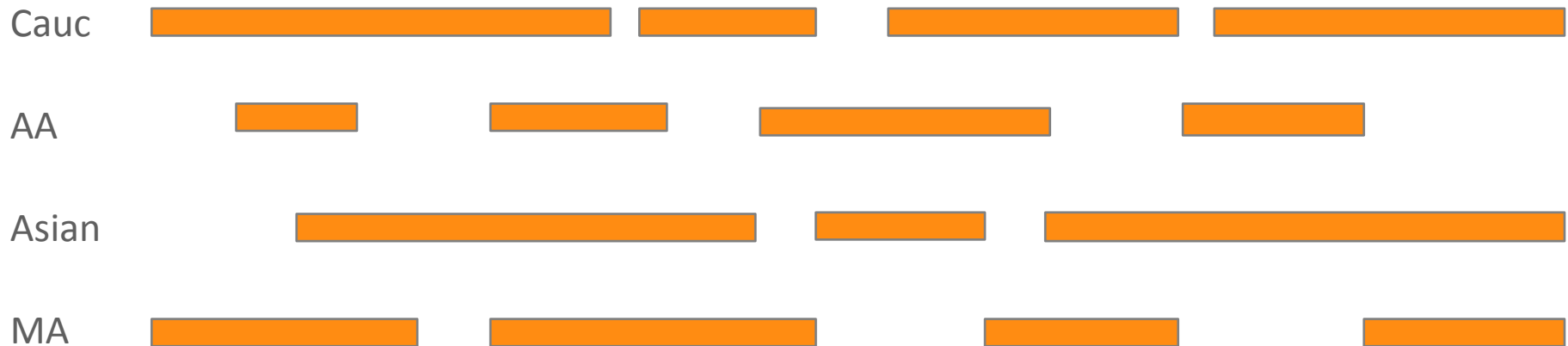


- Do you see any patterns in Block 1? Block 2?
- Which SNPs are in LD?
- Can any SNPs on Block 1 predict SNPs on Block 2?

Haplotype blocks are independent of genes



Haplotype blocks vary by race



Key points

- Haplotype blocks are regions within which there is strong LD, or correlation between variants
- One variant can capture information about another variant, either known or unknown
- Implications for GWAS (Lecture 4)

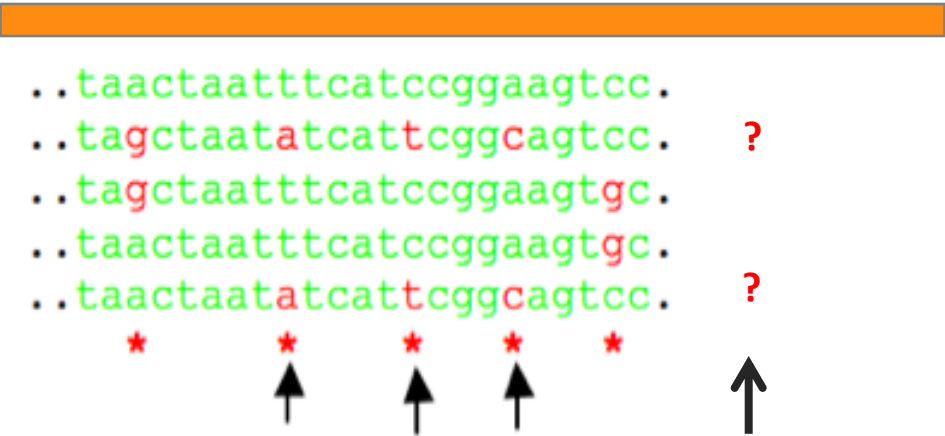


Diagram illustrating a haplotype block with five variants marked by red stars and arrows. The variants are located at positions 1, 2, 3, 4, and 5 of the haplotype sequences. The sequences are:

```
..taactaatttcacccggaagtcc.  
..tagctaataatcattcggcagtcc. ?  
..tagctaatttcacccggaagtgc.  
..taactaatttcacccggaagtgc.  
..taactaataatcattcggcagtcc. ?
```

The variants are marked by red stars at positions 1, 2, 3, 4, and 5. Arrows point to the positions of the variants. A question mark is present at the end of the second and fifth sequences.