AlloSeq™ Assign®

Phasing of Data in AlloSeq Assign v1.0

Margot Duane
Stem Cell Transplant Recipient

CareDx Technical Support

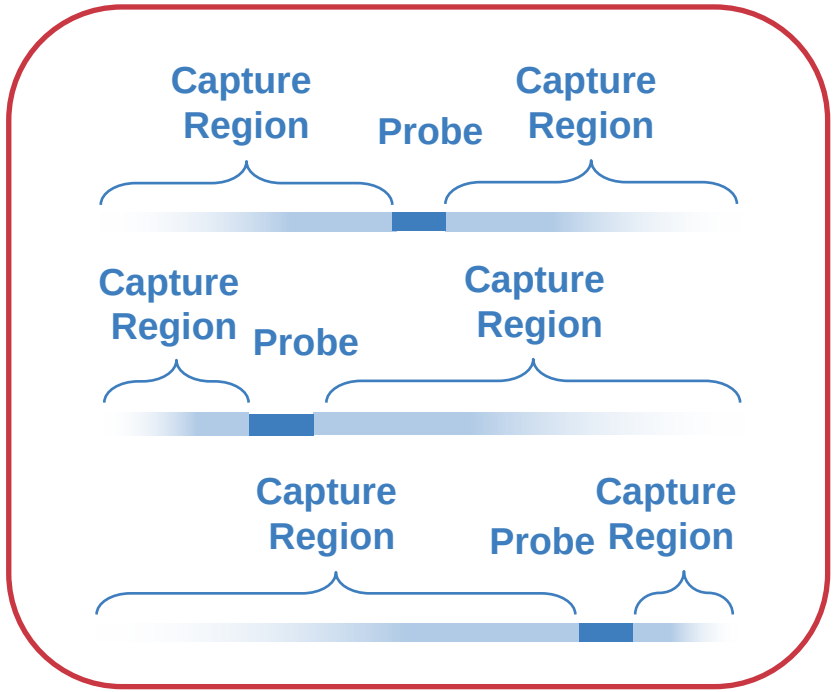
This material has been created for a global audience. Refer to all disclaimers for clinical claims and regulatory status in your region

AlloSeq Tx 17 and AlloSeq Assign are CE/IVD in the EU and the UK.
In the rest of the world, AlloSeq Tx 17 and AlloSeq Assign are for Research Use Only. Not to be used for diagnostic procedures.

TEC792-S_AlloSeq Assign 1.0_Phasing version 1.0 effective 14 Jun 22

CareDx®

- AlloSeq Tx has been optimised for Illumina Sequencers.
- Illumina sequencing is a “short read” sequencing platform that sequences 150 bp of both ends of a DNA fragment typically 500 bp in length*.
- AlloSeq Tx probes have a capture region which is 120 bp in length and will bind to any size DNA fragment, usually about 400-650 bp in length.



*DNA fragment length does not include the index adapters added during Index PCR reaction. The fragment size used to calculate final library concentration is generally around 800bp including the index adapters.

CareDx®

- Since most fragments average about 525 bp in length, and there are 2 reads (each of which extends 150 bp into either side of the fragment), there is usually about a 225 bp gap in coverage on each fragment.
- This gap in coverage can often be reconciled by linking polymorphisms

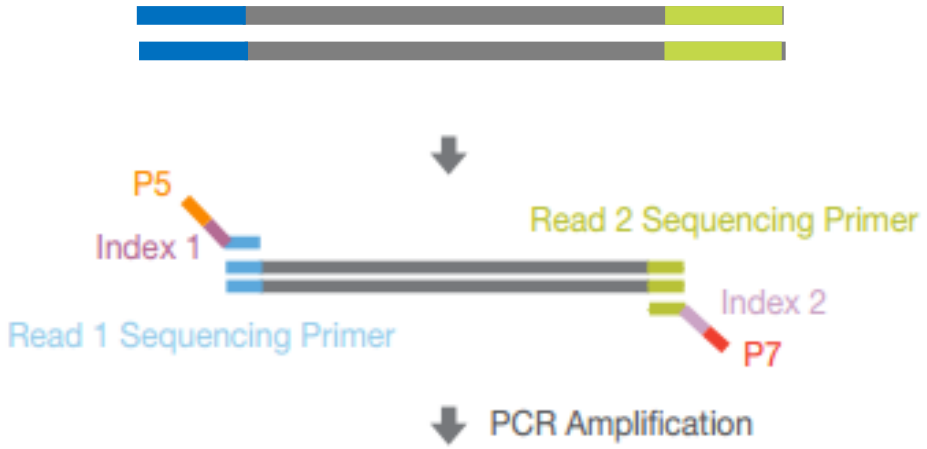
- In order to phase two polymorphisms, the distance between these polymorphisms must be within the length of the DNA fragments being sequenced.

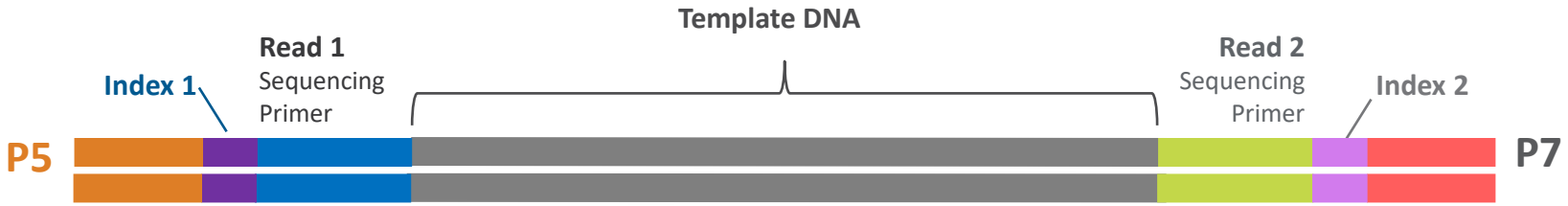
- Polymorphisms outside this region will not be phased.

- The inability to phase increases the risk of a report that includes a heterozygous ambiguity.

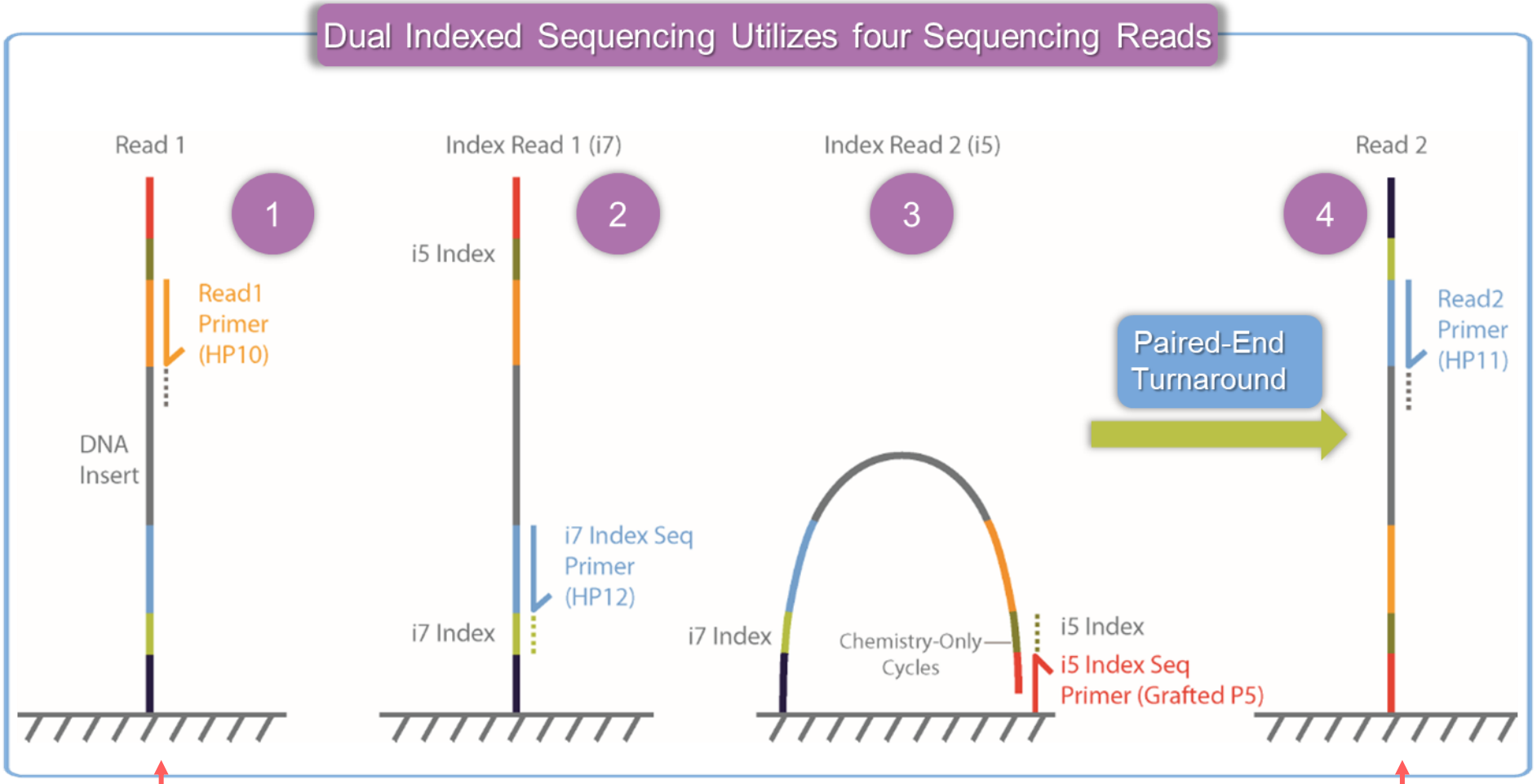
CareDx®

DNA is fragmented, and dual indexes are added via PCR





CareDx®

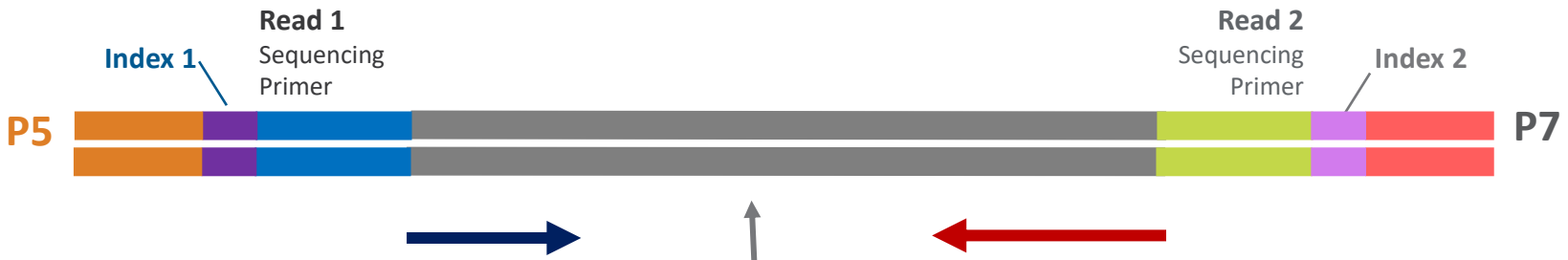


This read is the R1 fastq file

This read is the R2 fastq file

CareDx®

DNA is fragmented, and dual indexes are added via PCR



Read 1 sequencing primer binds, then sequences 150 bases towards the center of the fragment.

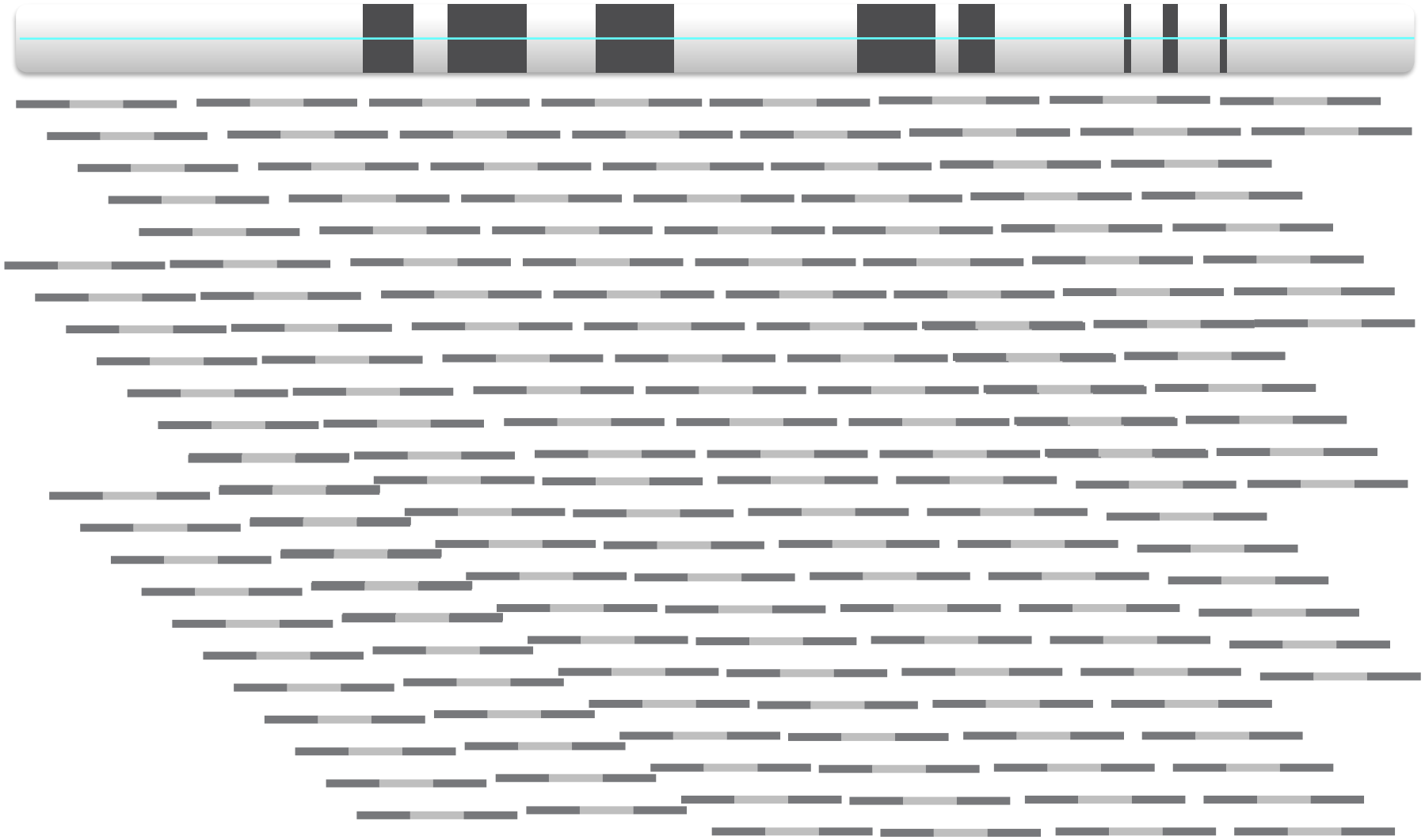
Read 2 sequencing primer binds, then sequences 150 bases towards the center of the fragment.

DNA at the center may not be sequenced. However, R1 and R2 can be attributed to the same fragment based on the X/Y coordinates of the cluster on the flowcell

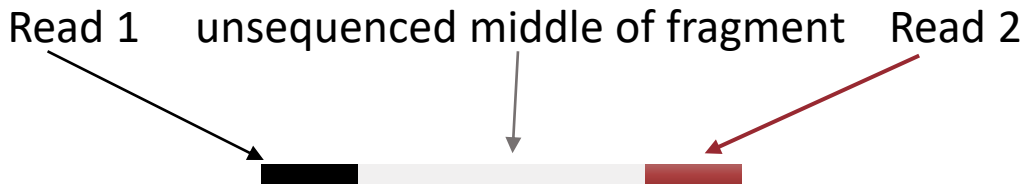
CareDx®

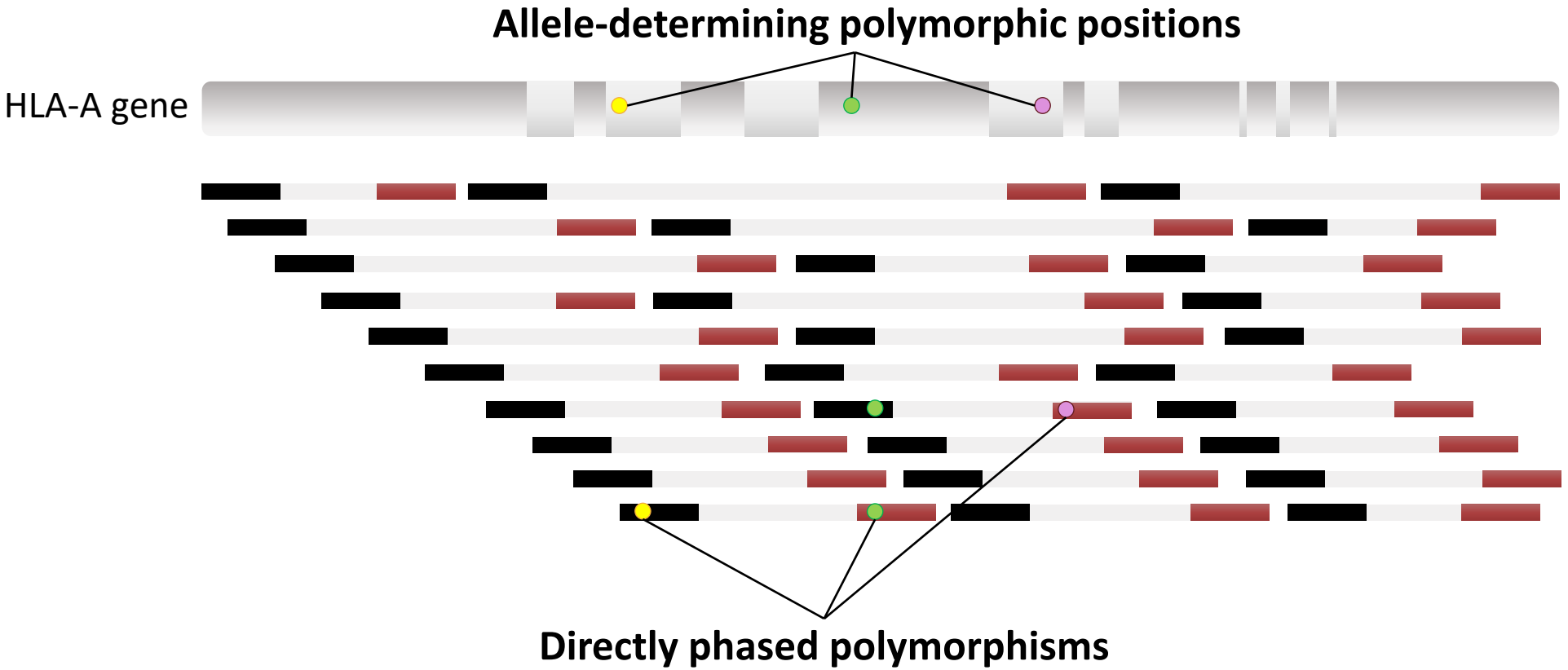
AlloSeqTx 17

Resulting fragments overlap across region captured by probes



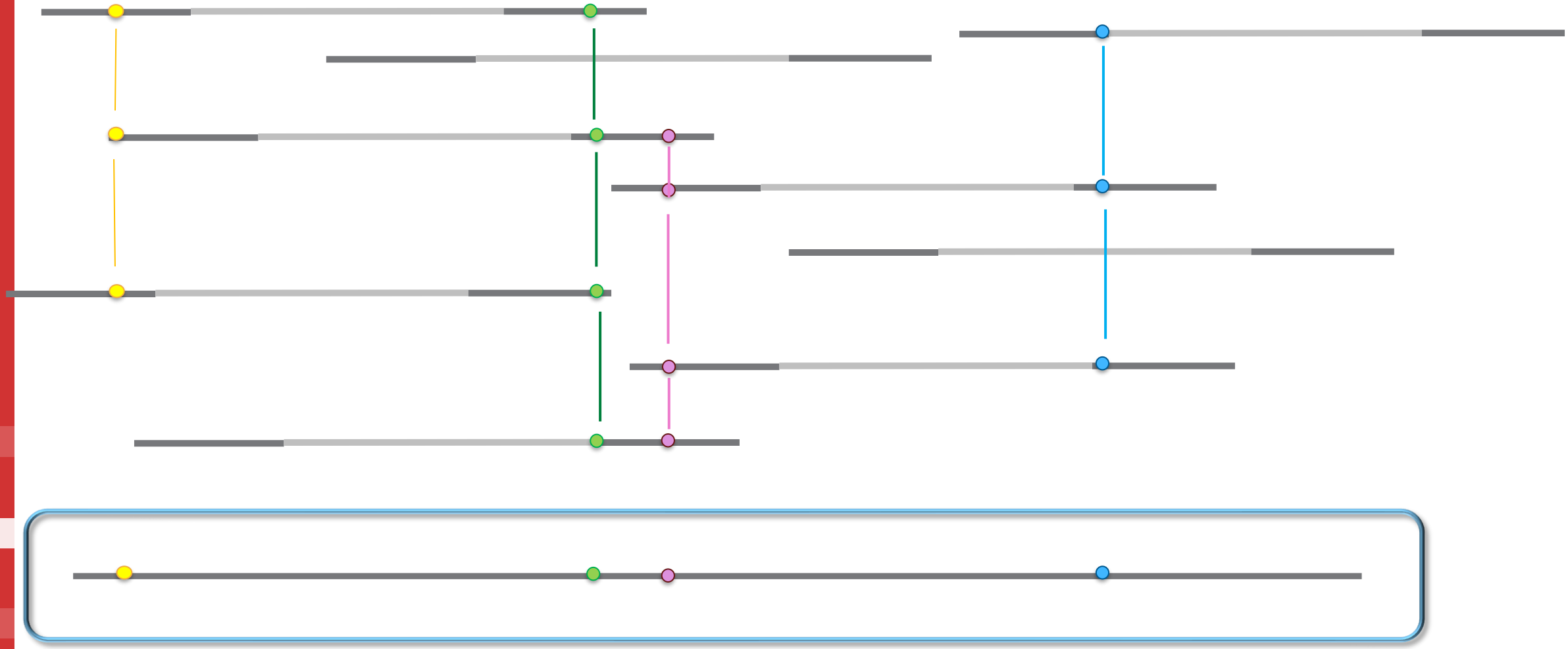
CareDx®


Read 1
unsequenced middle of fragment
Read 2


Allele-determining polymorphic positions
HLA-A gene
Directly phased polymorphisms

AlloSeqTx 17

Variants on either end of read are linked to create a single phased sequence



CareDx®

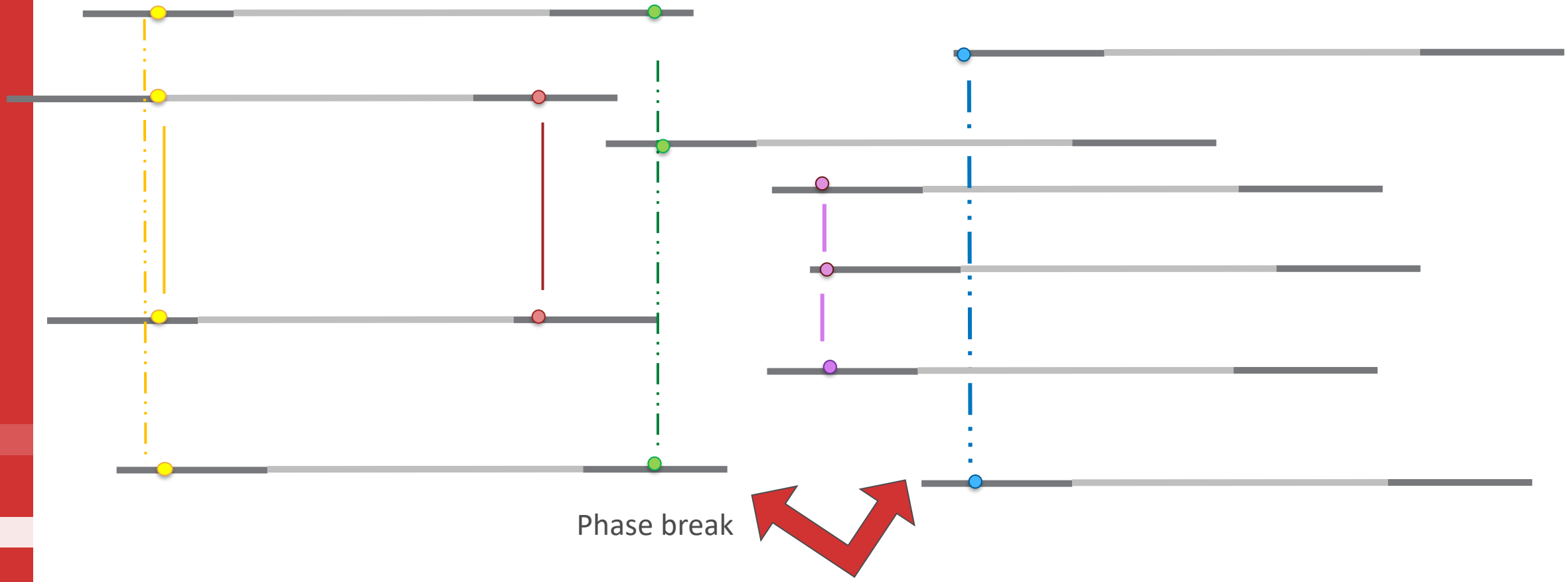
2 Causes of Phase Breaks

- 1. Polymorphic positions are too far apart for reads to span.
  - When the distance of polymorphic positions is greater than the fragment size, the data cannot be phased.
- 2. Gaps in probe coverage
  - In Class II, since the introns do not have probe coverage, the exons cannot be linked creating some ambiguous combinations.

- The following slides will illustrate these.

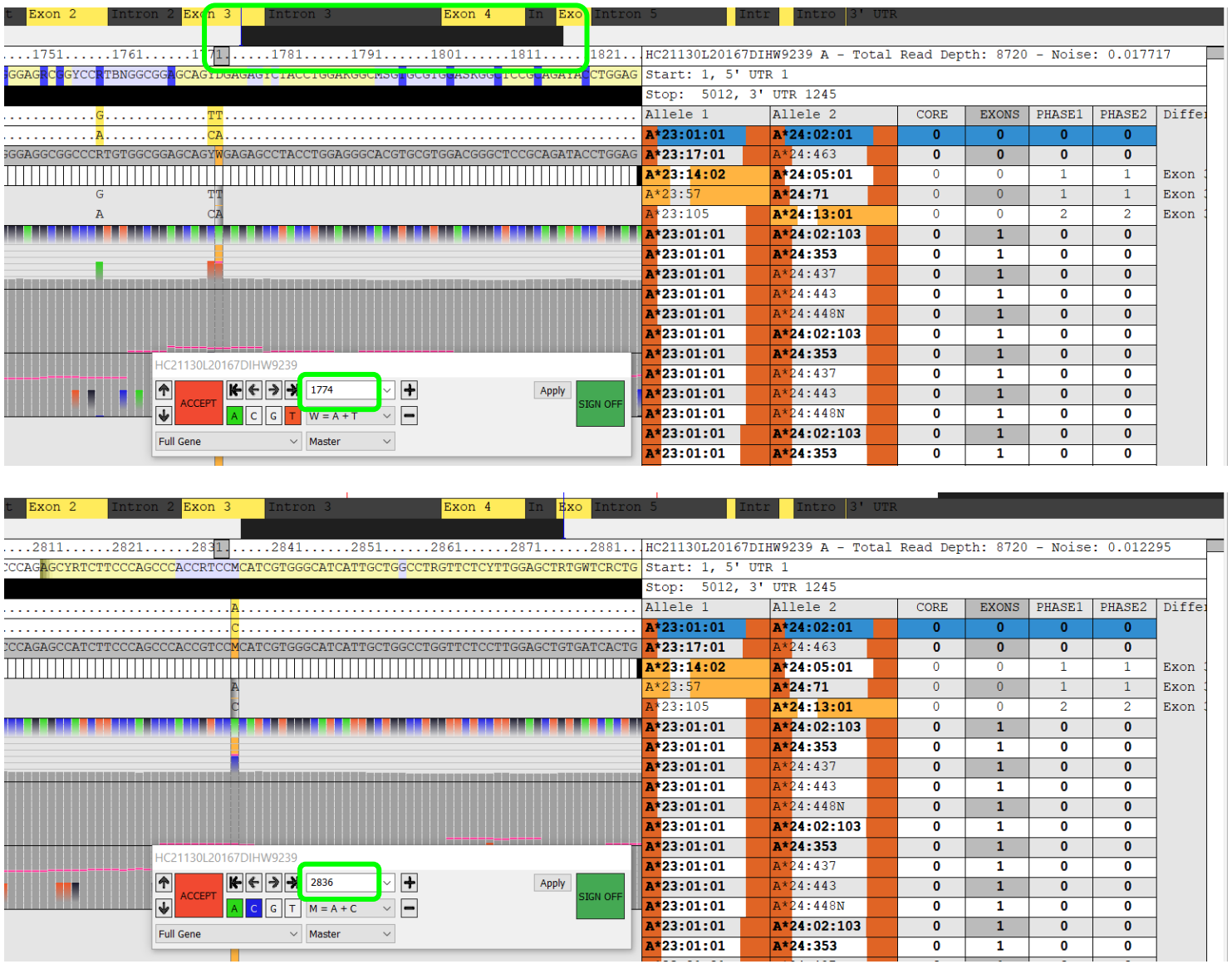
AlloSeqTx 17

When reads don't span variant positions, a phase break occurs.



Reads with purple and blue bases cannot be linked to the rest of the reads.

Phase Break due to lack of heterozygous positions

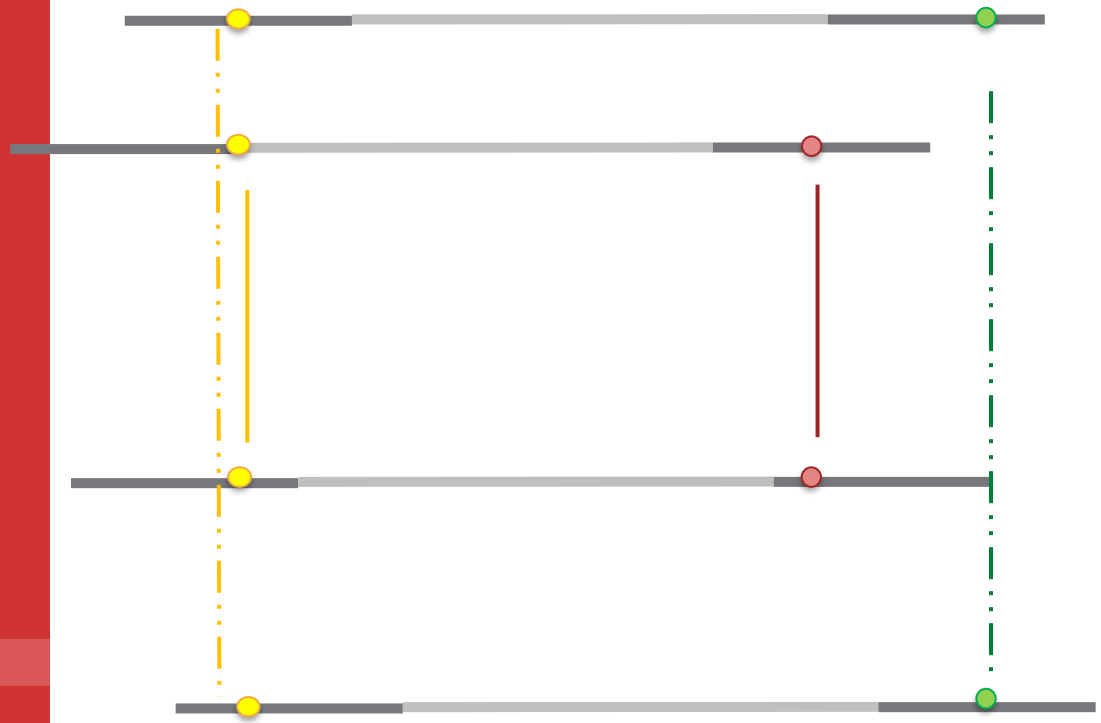


Heterozygous positions are ~1000 bp apart.
Typical fragment length is 450-650 bp long.
The distance between the heterozygous positions is longer than the fragment length.
This ambiguity would exist across all Illumina sequence-based assays.

CareDx®

AlloSeqTx 17

When there are breaks in coverage, a phase break may occur. The most common cause is in areas where there is no probe coverage (introns in Class II).

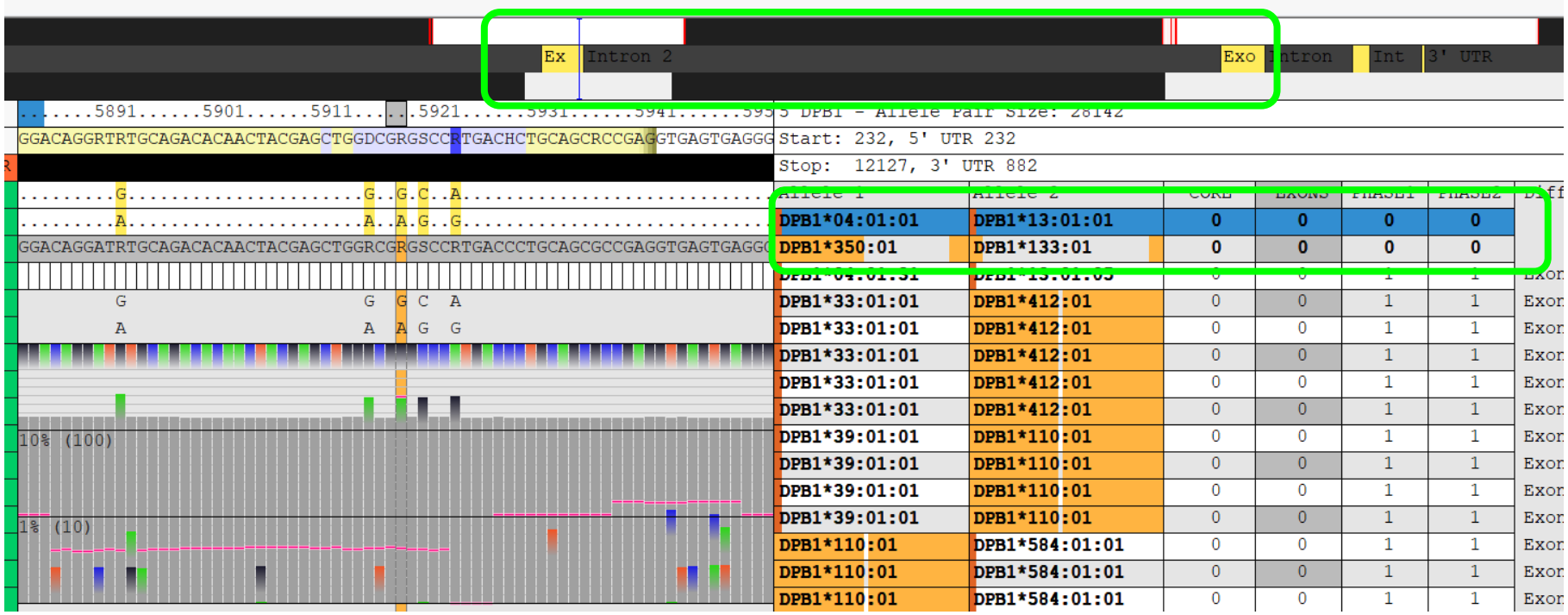


Phase break

No reads span this region.

CareDx®

Phase Break due to lack of probe coverage

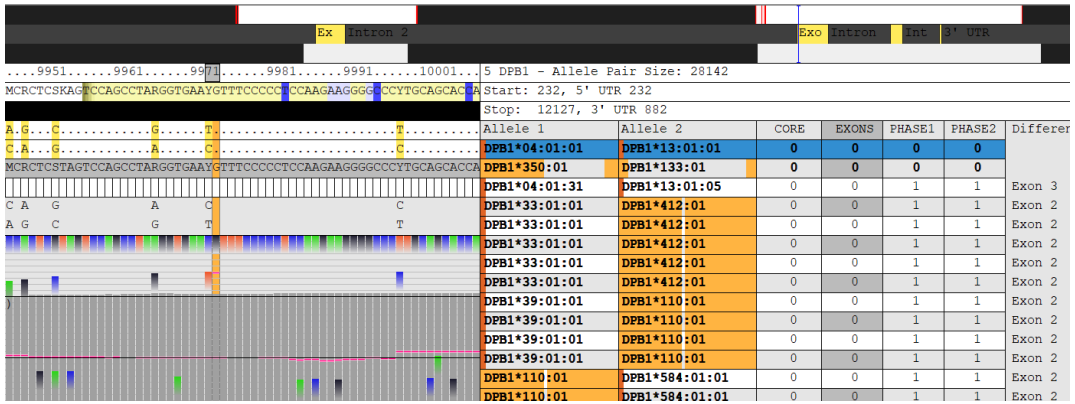


Since the probe coverage does not cover all of intron 2, some ambiguities will occur. These are generally same G or P group ambiguities.

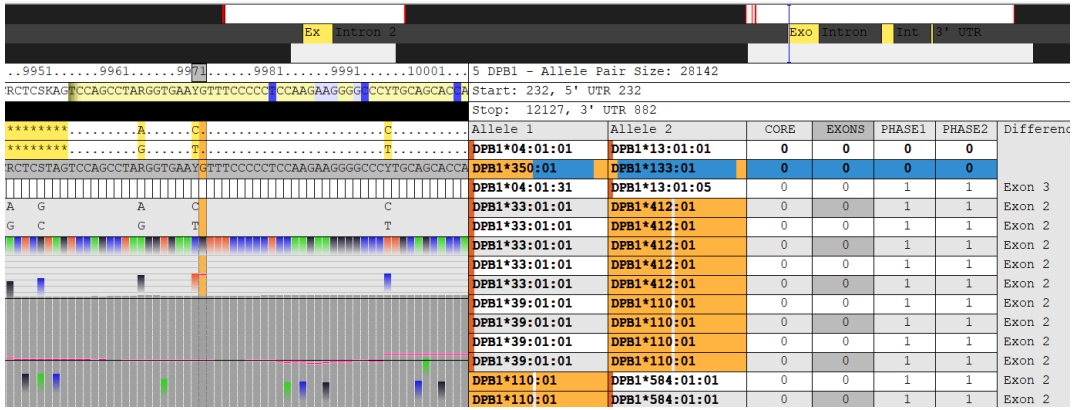
CareDx®

Phase Break due to lack of probe coverage

*04:04:01 and *350:01 are identical in exon 2, *13:01:01 and *133:01 are identical in exon 2.



In exon 3, *04:01:01 and *133:01 are identical, while *13:01:01 and *350:01 are identical.



Since the motifs are opposite between exon 2 & 3, and there is no coverage across the intron, the phase ambiguity is present.

CareDx®

Building of Phase blocks in AlloSeq Assign

CareDx®

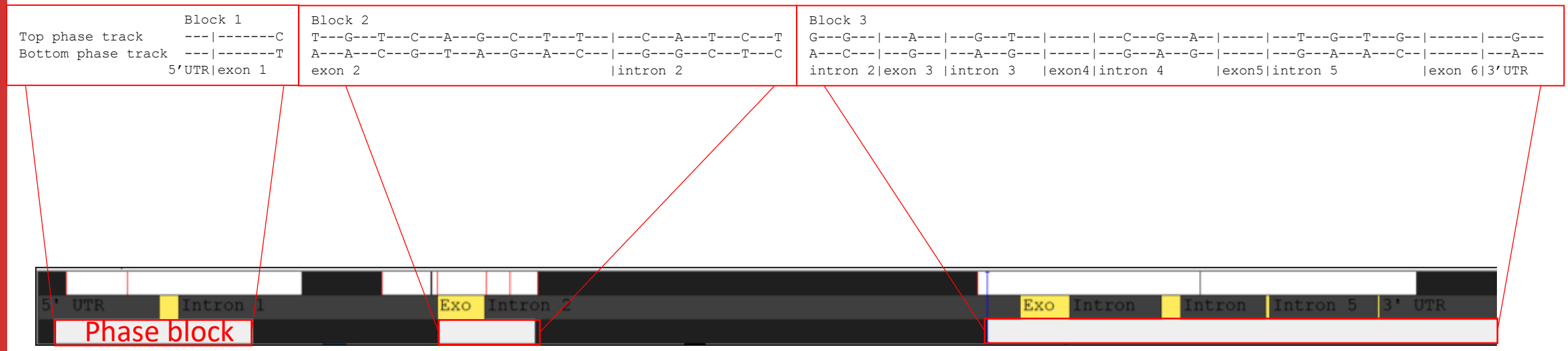
Building the phase in AlloSeq Assign v1.0.3

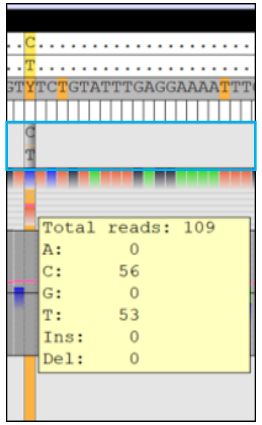
- The phasing is built in blocks from the 5’ end of the gene.
- Starting from the first heterozygous position, the base with the larger number of reads is assigned to the top phase track.
- Consecutive linked heterozygous bases are assigned to the same track as the first heterozygous position.
- Due to how bases are assigned to phase tracks, in some cases the phase tracks may not match the reference alleles 1&2.

CareDx®

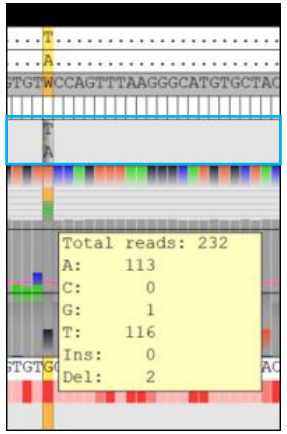
Building the phase in AlloSeq Assign v1.0.3

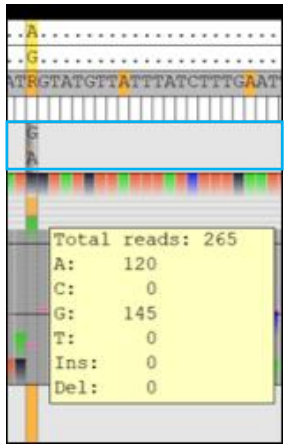
This sample has 3 phase blocks.





Phase tracks



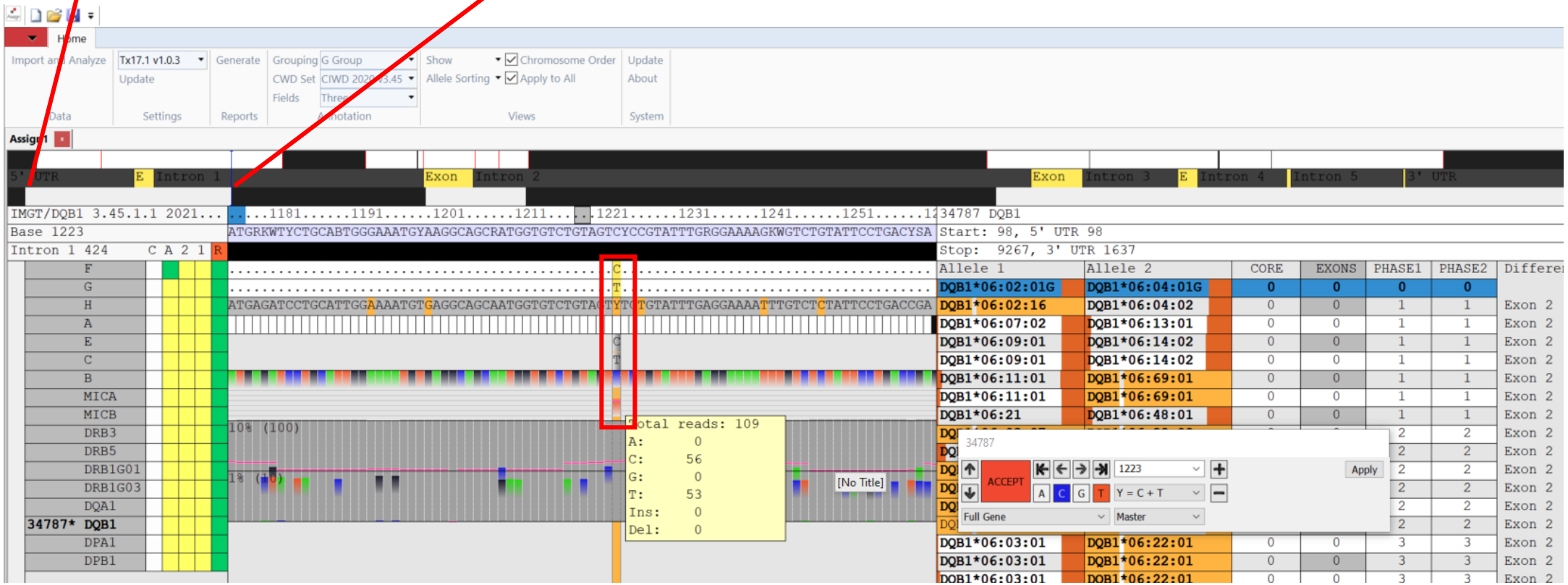


CareDx®

Phase block 1

```
Block 1
Top phase track              ---|-------C
Bottom phase track           ---|-------T
Allele 1(DQB1*06:02:01)      ---|-------C
Allele 2(DQB1*06:04:01)      ---|-------T
                                5'UTR|exon 1
```

- In this sample the first heterozygous position is at 1223.
- There are more reads present for the C, so it is assigned to the top phase track.
- Reference allele 1 matches the top phase track in this phase block.



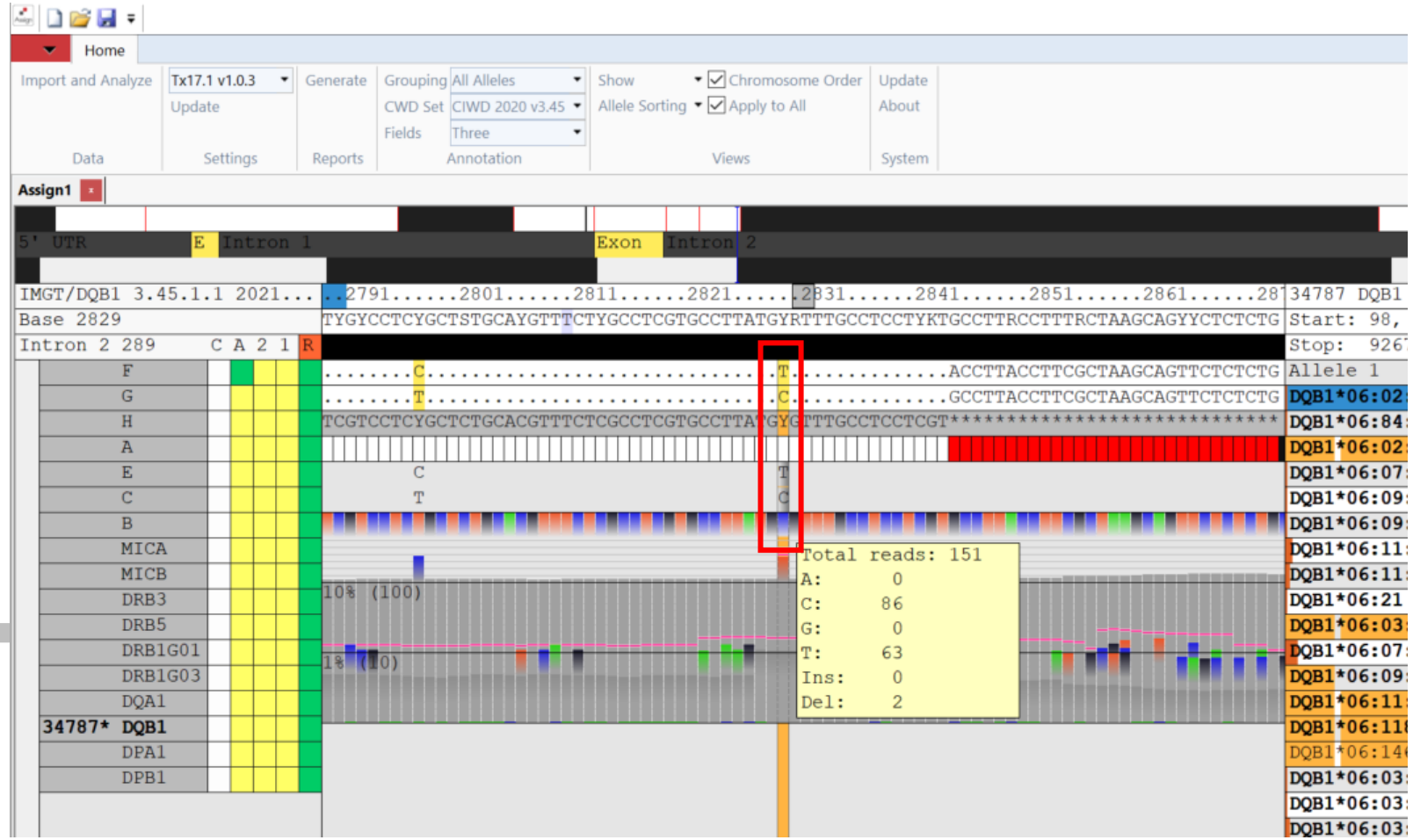
CareDx®

Phase block 2

- The next phase block for this sample starts at position 2283. T has the most reads, and is assigned to the top phase track.
- Reference allele 1 matches the top phase track in this phase block.

Phase block 2

- Consecutive linked heterozygous positions in the phase block are assigned to the phase tracks based on the first heterozygous position in the block.
- Position 2829 is the last heterozygous positioning the second phase block.

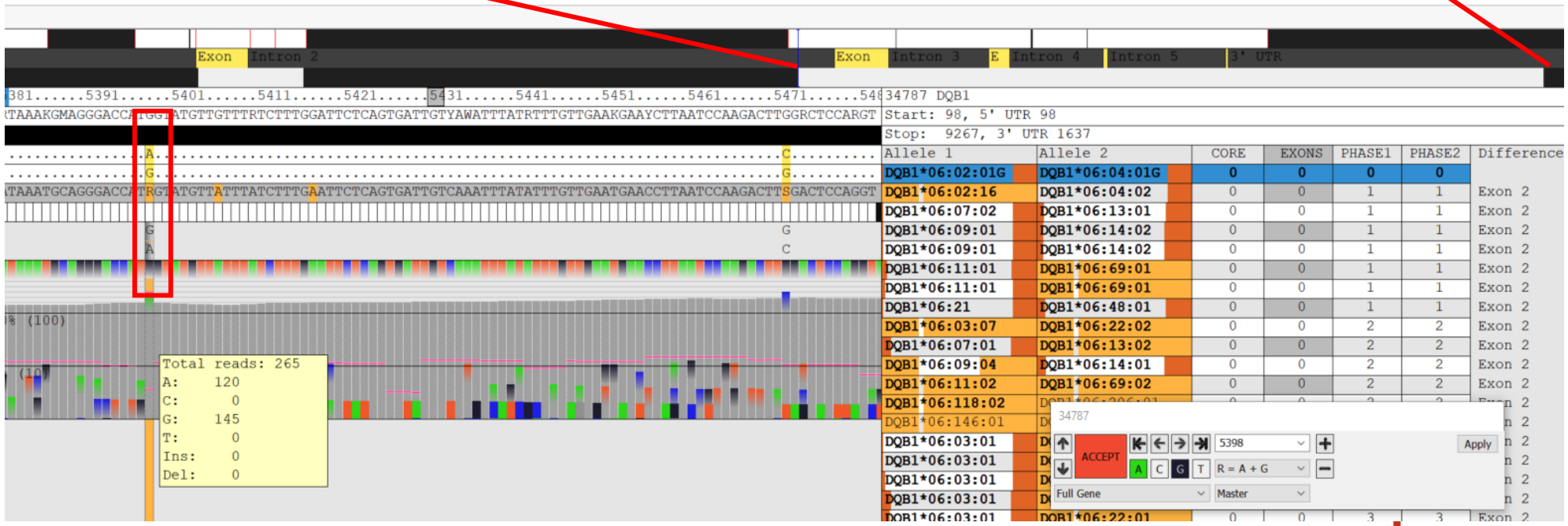


CareDx®

Phase block 3

- The next phase block for this sample starts at position 5398. G has the most reads, and is assigned to the top phase track.
- Reference allele 2 matches the top phase track in this phase block.

```
Block 3
Top phase track              G---G---|---A---|---G---T---|------|---C---G---A---|------|---T---G---T---G---|------|---G---
Bottom phase track           A---C---|---G---|---A---G---|------|---G---A---G---|------|---G---A---A---C---|------|---A---
Allele 1(DQB1*06:02:01)      A---C---|---G---|---A---G---|------|---G---A---G---|------|---G---A---A---C---|------|---A---
Allele 2(DQB1*06:04:01)      G---G---|---A---|---G---T---|------|---C---G---A---|------|---T---G---T---G---|------|---G---
                             intron 2|exon 3 |intron 3    |exon 4|intron 4       |exon 5|intron 5            |exon 6|3'UTR
```



CareDx®

Phase block 3

- As the G is assigned to the top phase track, all consecutive linked bases in phase block 3 are assigned to the top phase track.
- The phase has been correctly calculated, however due to how the phase block has been built the display appears inverted.