AlloSeq Assign®

Phasing of Data in AlloSeq Assign v1.0

CareDx Technical Support

Margot Duane Stem Cell Transplant Recipient

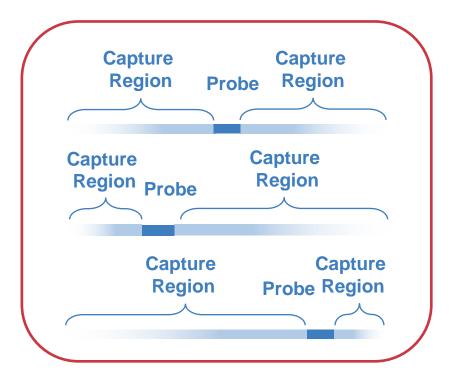
This material has been created for a global audience. Refer to all disclaimers for clinical claims and regulatory status in your region

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



- 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.



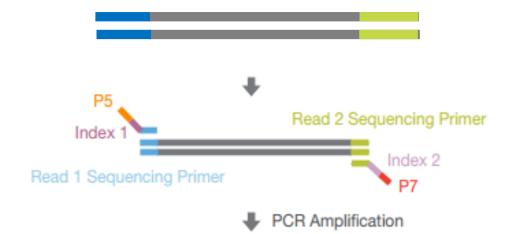
*DNA fragment length does not include the index adapters added during Index PCR reaction. The fragment size used to calculate final library concentration is generally around 800bp including the index adapters.

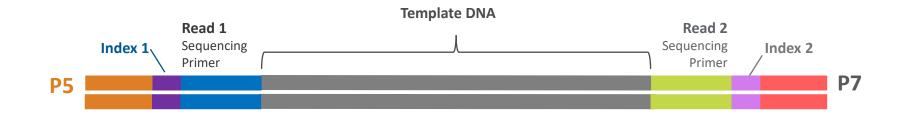


- Since most fragments average about 525 bp in length, and there are 2 reads (each of which extends 150 bp into either side of the fragment), there is usually about a 225 bp gap in coverage on each fragment.
- This gap in coverage can often be reconciled by linking polymorphisms
- In order to phase two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

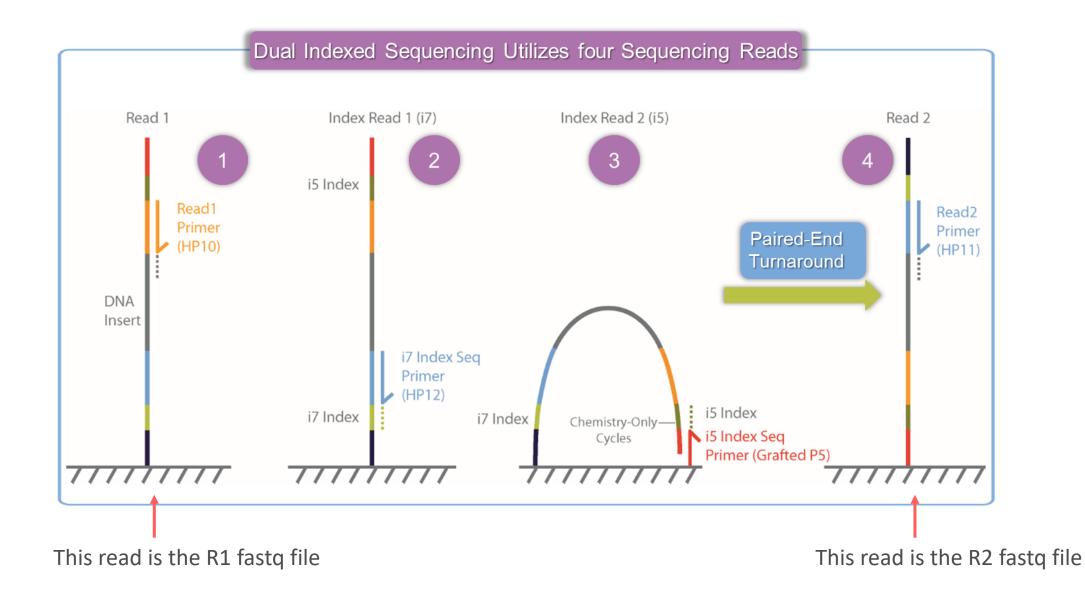


DNA is fragmented, and dual indexes are added via PCR



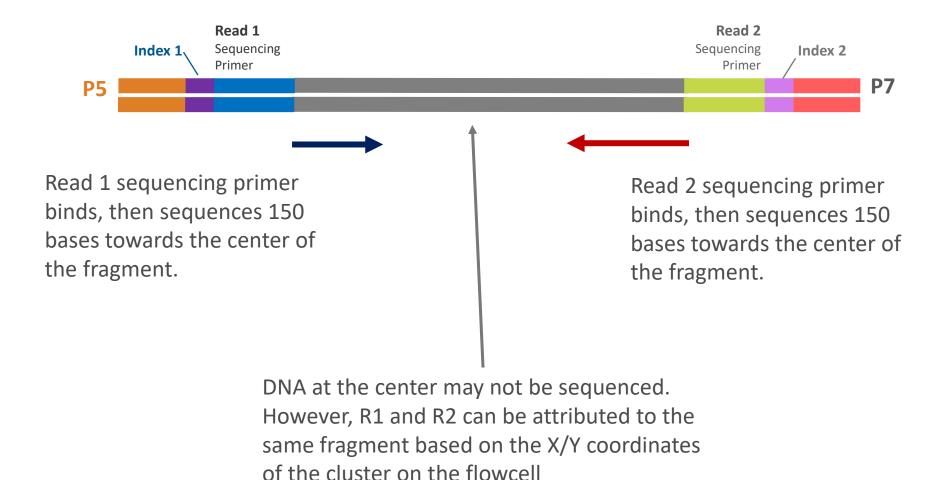








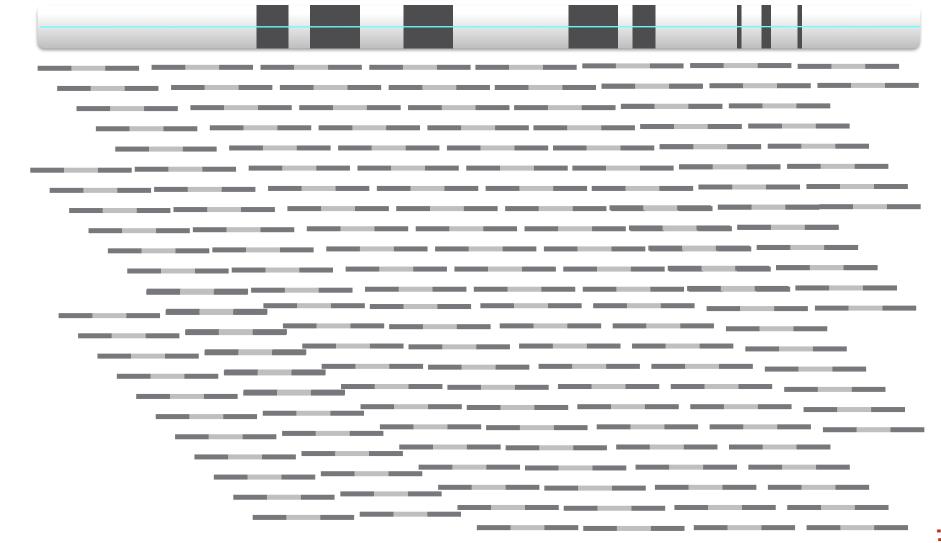
DNA is fragmented, and dual indexes are added via PCR



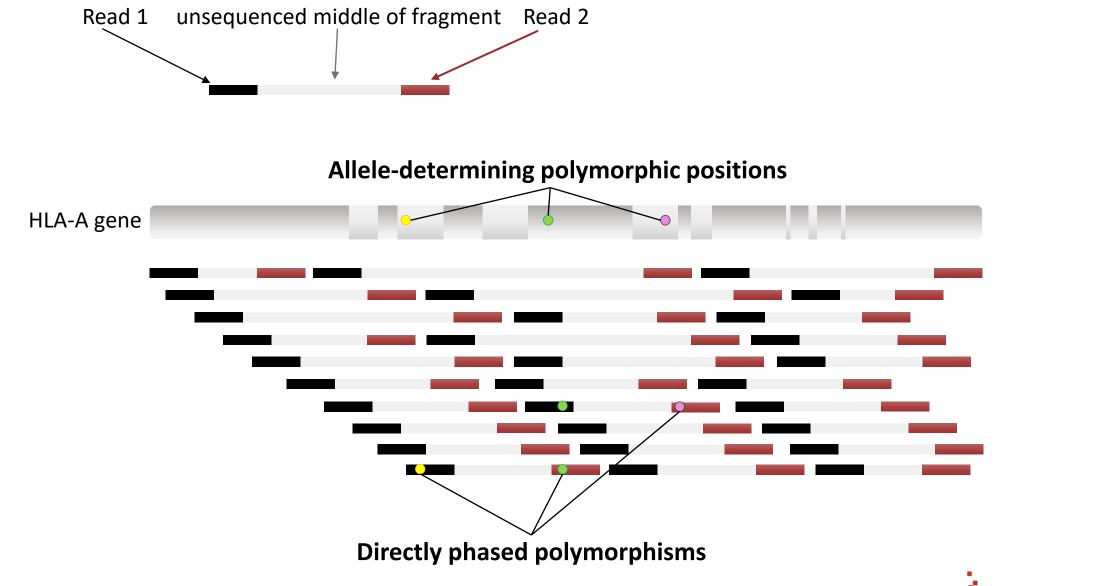




Resulting fragments overlap across region captured by probes



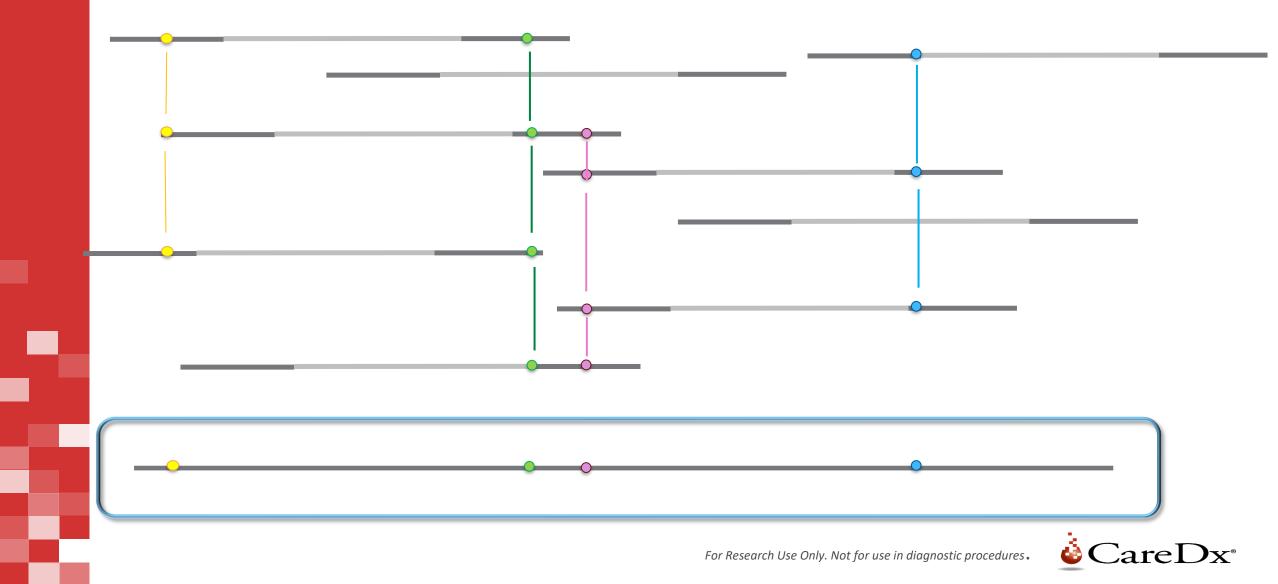








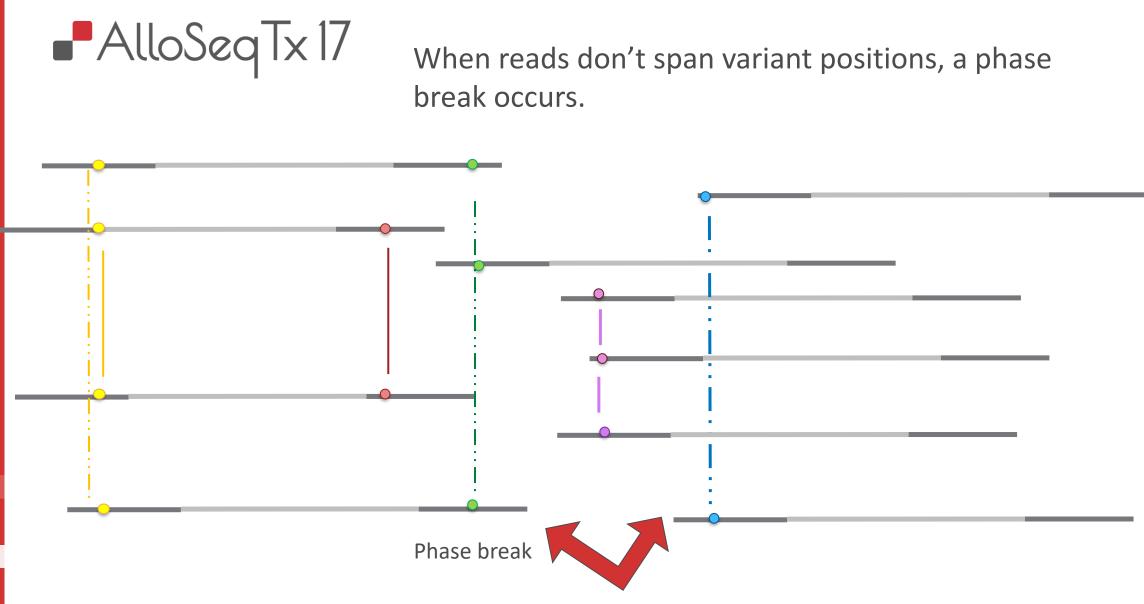
Variants on either end of read are linked to create a single phased sequence



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.





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



Phase Break due to lack of heterozygous positions

Exon 2 Intron 2 Exon 3 Intron 3 Exon 4 In Exo Intron 5 Intr In	tro 3' UTR				
175117611711781179118011811 1821HC21130L20167DIHW9239	A - Total Read Dep	th: 8720	- Noise	: 0.0177	17
GGAGRCCGYCCTBNGGCGGAGCAGTDGAGGGCICTACCTGGGAGGGCASCGCGCGGGASGGGCCGGGGAGGAGAGCGCGCGGGGGGGG					
Stop: 5012, 3' UTR 1.	.245				
G	.e 2 CORE	EXONS	PHASE1	PHASE2	Diffe:
A*23:01:01 A*24:		0	0	0	
GGAGGCGGCCCRTGTGGCGGAGCAGYWGAGAGCCTACCTGGAGGGCACGTGCGTGGACGGGCTCCGCAGATACCTGGAG A*23:17:01		0	0	0	
<u> </u>		0	1	1	Exon 3
G TT <u>A*23:57 <u>A*24:</u></u>		0	1	1	Exon (
A CA A*23:105 A*24:		0	2	2	Exon 3
	02:103 0	1	0	0	
A*23:01:01 A*24:		1	0	0	
A*23:01:01 A*24:		1	0	0	
A*23:01:01 A*24:		1	0	0	
A*23:01:01 A*24:		1	0	0	
	02:103 0	1	0	0	
HC21130L20167DIHW9239		1	0	0	
A*23:01:01 A*24:		1	0	0	
ACCEPT	•	1	0	0	
		1	0	0	
Full Gene V Master V	02:103 0	1	0	0	
A*23:01:01 A*24:	353 0	1	0	0	

HC21130L20167DIHW9239 A - Total Read Depth: 8720 - Noise: 0.012295 CCAGAGCYRTCTTCCCAGCCCACCRTCCMCATCGTGGGCATCATTGCTGGCCTRGTTCTCYTTGGAGCTRTGWTCRCT Start: 1, 5' UTR Stop: 5012, 3' UTR 1245 Allele 2 CORE PHASE2 Differ Allele 1 EXONS PHASE1 A*24:02:01 0 0 *23:01:01 0 0 *23:17:01 24:463 0 0 0 0 A*24:05:01 0 Exon A*23:14:02 0 1 1 A*24:71 0 0 Exon 2<mark>3:57</mark> 1 1 23:105 A*24:13:01 0 0 2 2 Exon A*23:01:01 A*24:02:103 0 1 0 0 *23:01:01 A*24:353 0 1 0 0 24:437 0 1 0 0 *23:01:01 24:443 0 1 0 0 *23:01:01 24:448N 0 1 0 0 *23:01:01 A*24:02:103 0 0 0 A*23:01:01 1 *23:01:01 A*24:353 0 1 0 0 24:437 0 1 0 0 A*23:01:01 A*23:01:01 24:443 0 1 0 0 *23:01:01 24:448N 0 1 0 0 0 *23:01:01 0 1 A*24:02:103 0 Full Gene *23:01:01 A*24:353 0 1 0 0

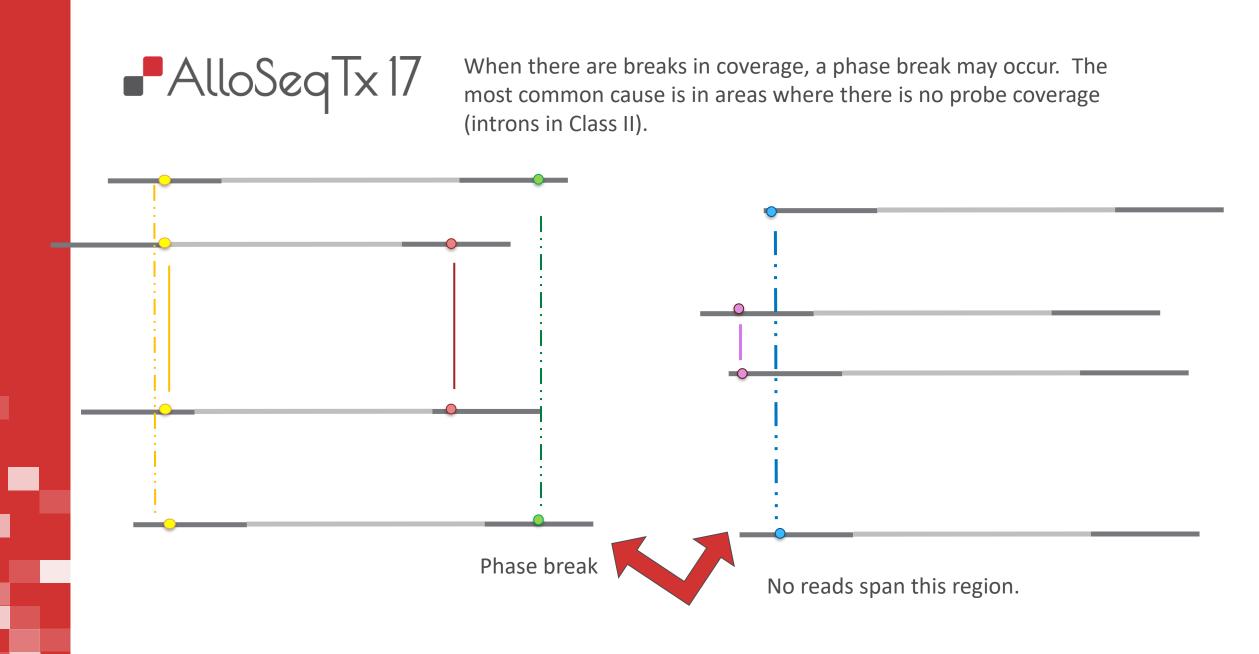
Heterozygous positions are ~1000 bp apart.

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.







Phase Break due to lack of probe coverage

	Ex Intro		35 DFB1 - Allele 1	2417 S128• /814/	Exo	Intron	Int	3' UTR	
GGACAGGRTRTGCAGACACAACTACGAGCTGGDC									
			Stop: 12127, 3'	UTR 882			-		
	. <mark>G.C</mark> <mark>A</mark>		Allele 1	Allele 2	CORE	GUOVE	THADLI	FIIADEZ	, Di€t
	. <mark>A.GG</mark>		DPB1*04:01:01	DPB1*13:01:01	0	0	0	0	
GGACAGGATRTGCAGACACAACTACGAGCTGGRC	G <mark>R</mark> GSCCRTGACCCTGCAGCGCCG	AGGTGAGTGAGG	DPB1*350:01	DPB1*133:01	0	0	0	0	
			DED1-04.01.31	DED1~13.01.03	0	0	1	1	ы к0:
G G	G C A		DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exo
A A	<mark>A</mark> G G		DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exo
			DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exo
			DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exo
			DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exo
10% (100)			DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exo
			DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exo
			DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exo
1% (10)			DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exo
			DPB1*110:01	DPB1*584:01:01	0	0	1	1	Exo
			DPB1*110:01	DPB1*584:01:01	0	0	1	1	Exo
			DPB1*110:01	DPB1*584:01:01	0	0	1	1	Exo

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



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.

					_		
Ex Intron 2			Exc	Intron	Int		
	015 DPB1 - Allele	Pair Size: 28142					
ICRCTCSKAG <mark>TCCAGCCTARGGTGAAYGTTTCCCCCC</mark> TCCAAGAAGGGG <mark>C</mark> CCYTGCAG	CACCA Start: 232, 5' U	TR 232					
	Stop: 12127, 3	UTR 882					
	Allele 1	Allele 2	CORE	EXONS	PHASE1	PHASE2	Differ
	DPB1*04:01:01	DPB1*13:01:01	0	0	0	0	
ICRCTCSTAGTCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGGGCCCYTGCAG	CACCA DPB1*350:01	DPB1*133:01	0	0	0	0	1
	DPB1*04:01:31	DPB1*13:01:05	0	0	1	1	Exon 3
C C C	DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exon 2
LGC GT T	DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exon 2
	DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exon 2
	DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exon 2
	DPB1*33:01:01	DPB1*412:01	0	0	1	1	Exon 2
	DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exon 2
	DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exon 2
	DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exon 2
	DPB1*39:01:01	DPB1*110:01	0	0	1	1	Exon 2
	DPB1*110:01	DPB1*584:01:01	0	0	1	1	Exon 2
	DPB1*110:01	DPB1*584:01:01	0	0	1	1	Exon 2

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

			_		_
Ex Intron 2	E	xo Intron	Int	3' UTR	
.99519961997199819991100015 DPB1 - Allele Pair Size:	28142				
RCTCSKAG <mark>TCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGGG</mark> CCYTGCAGCACCAStart: 232, 5' UTR 232					
Stop: 12127, 3' UTR 882					
*********	CORE	EXONS	PHASE1	PHASE2	Differe
**********	01:01 0	0	0	0	1
RCTCSTAGTCCAGCCTARGGTGAAYGTTTCCCCCTCCAAGAAGGGGCCCYTGCAGCACCA DPB1*350:01 DPB1*133:	01 0	0	0	0	1
DPB1*04:01:31 DPB1*13:0	1:05 0	0	1	1	Exon 3
A G A C C DPB1*33:01:01 DPB1*412:	01 0	0	1	1	Exon 2
G C G T T DPB1*33:01:01 DPB1*412:	01 0	0	1	1	Exon 2
DPB1*33:01:01 DPB1*412:	01 0	0	1	1	Exon 2
DPB1*33:01:01 DPB1*412:	01 0	0	1	1	Exon 2
DPB1*33:01:01 DPB1*412:	01 0	0	1	1	Exon 2
DPB1*39:01:01 DPB1*110:	01 0	0	1	1	Exon 2
DPB1*39:01:01 DPB1*110:	01 0	0	1	1	Exon 2
DPB1*39:01:01 DPB1*110:	01 0	0	1	1	Exon 2
DPB1*39:01:01 DPB1*110:	01 0	0	1	1	Exon 2
DPB1*110:01 DPB1*584:	01:01 0	0	1	1	Exon 2
DPB1*110:01 DPB1*584:	01:01 0	0	1	1	Exon 2

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



Building of Phase blocks in AlloSeq Assign



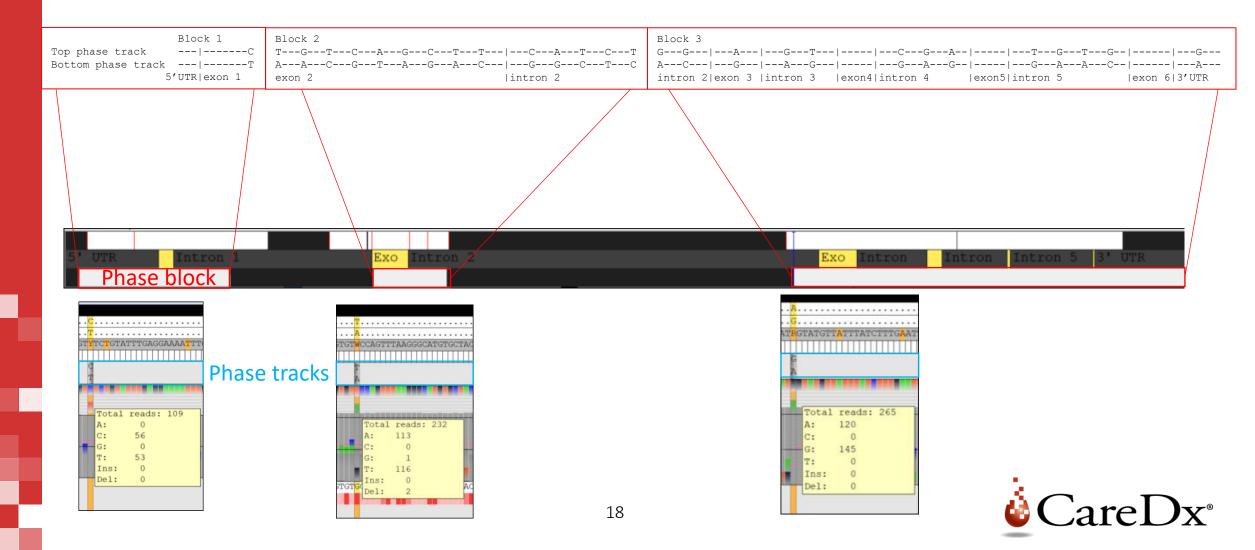
Building the phase in AlloSeq Assign v1.0.3

- 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.



Building the phase in AlloSeq Assign v1.0.3

This sample has 3 phase blocks.



Block 1 Top phase track	 In this sample the first h There are more reads prophase track. Reference allele 1 match 	resent for the C, so	it is assig	ned t		·
Home						
Import and Analyze Tx17.1 v1.0.3 • Generate Grouping G Group Show • Chron Update CWD Set CIWD 2022 + 5.45 • Fields Thres • Views	nosome Order Update to All About System					
Assign 1						
5 OTR E Intron 1 Exon Intron 2		Exon Intron 3 E Intr	on 4 Intron 5	3 '	UTR	
IMGT/DQB1 3.45.1.1 2021		787 DQB1				
	TCTGTAGTCYCCGTATTTGRGGAAAAGKWGTCTGTATTCCTGACYSA Sta					
Intron 1 424 C A 2 1 R		pp: 9267, 3' UTR 1637				
F		lele 1 Allele 2	CORE EXONS	-		Differe
		DQB1*06:02:01G DQB1*06:04:01G	0 0	0	0	Europ 0
H ATGAGATCCTGCATTGGAAAATGTGAGGCAGCAATGGTG		B1*06:02:16 DQB1*06:04:02 B1*06:07:02 DQB1*06:13:01	0 0	1		Exon 2
		B1*06:09:01 DQB1*06:14:02	0 0	1		Exon 2 Exon 2
		31*06:09:01 DQB1*06:14:02	0 0	1		Exon 2 Exon 2
B	~	B1*06:11:01 DQB1*06:69:01	0 0	1		Exon 2 Exon 2
MICA		31*06:11:01 DQB1*06:69:01	0 0	1		Exon 2 Exon 2
MICB		31*06:21 DQB1*06:48:01	0 0	1		Exon 2
DRB3 10% (100)	otal reads: 109			2		Exon 2
DRB5	A: U	34787		2		Exon 2
DRB1G01	C: 56	★ ★ ★ → → 1223 ~ +	4	apply 2		Exon 2
DRB1G03	G: [No Title]			2		Exon 2
DQA1	DOI			2		Exon 2
34787* DQB1	Ins: 0	Full Gene V Master V		2		Exon 2
DPA1		B1*06:03:01 DQB1*06:22:01	0 0	3	3 E	Exon 2
DPB1	DQI	B1*06:03:01 DQB1*06:22:01	0 0	3	3 E	Exon 2
	DO	31*06:03:01 DOB1*06:22:01	0 0	3	3 E	Exon 2

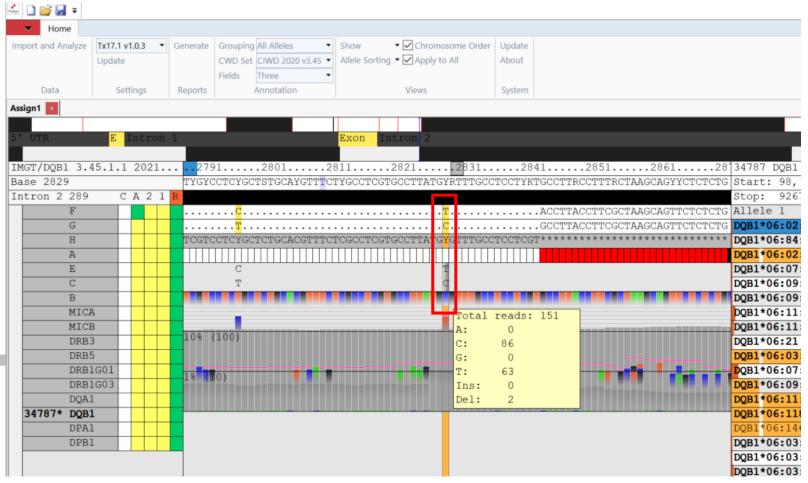


- The next phase block for this sample starts at position 2283. T has the most reads, and is assigned to the top phase track.
- Reference allele 1 matches the top phase track in this phase block.



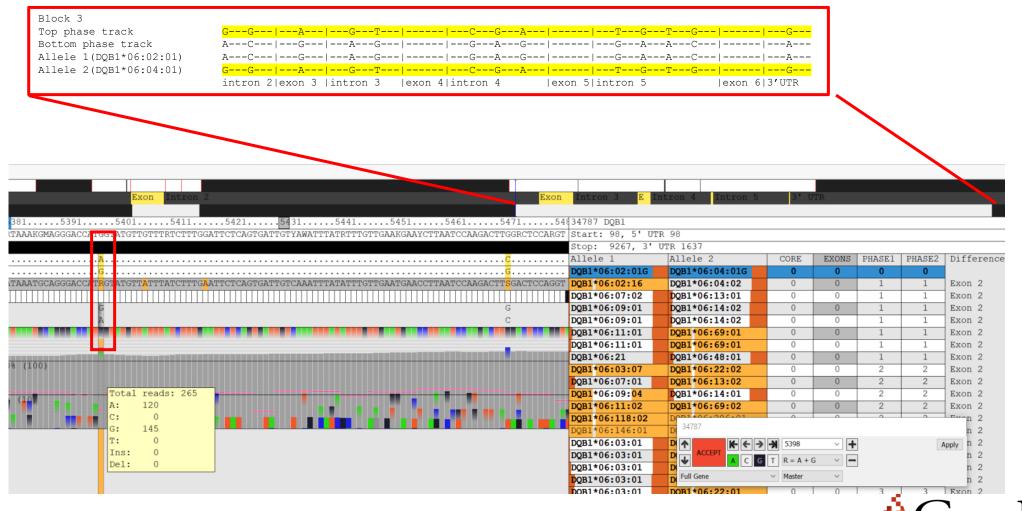


- Consecutive linked heterozygous positions in the phase block are assigned to the phase tracks based on the first heterozygous position in the block.
- Position 2829 is the last heterozygous positioning the second phase block.





- 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.



- 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.

Exon Intron 2		Exon	Intron 3 E In	tron 4 Intron 5	3' 0					
.5338153915401541154215431544154515461	5471	L548	DQB1							
${\tt TRTAAAKGMAGGGACCATGGTATGTTGTTTRTCTTTGGATTCTCAGTGATTGTYAWATTTATRTTTGTTGAAKGAAYCTTAATCCAAGACT$	TGGF	RCTCCARGT	Start: 98, 5' UTR	98						
Stop: 9267, 3' UTR 1637										
	. <mark>C</mark>		Allele 1	Allele 2	CORE	EXONS	PHASE1	PHASE2	Differen	
	. <mark>G</mark>		DQB1*06:02:01G	DQB1*06:04:01G	0	0	0	0		
TATAAATGCAGGGACC <mark>AT</mark> RG ^I ATGTT <mark>A</mark> TTTATCTTTG <mark>A</mark> ATTCTCAGTGATTGTCAAATTTATATTTGTTGAATGAACCTTAATCCAAGAC	T <mark>S</mark> GI	CTCCAGGT	DQB1*06:02:16	DQB1*06:04:02	0	0	1	1	Exon 2	
			DQB1*06:07:02	DQB1*06:13:01	0	0	1	1	Exon 2	
G	G		DQB1*06:09:01	DQB1*06:14:02	0	0	1	1	Exon 2	
A	С		DQB1*06:09:01	DQB1*06:14:02	0	0	1	1	Exon 2	
			DQB1*06:11:01	DQB1*06:69:01	0	0	1	1	Exon 2	
			DQB1*06:11:01	DQB1*06:69:01	0	0	1	1	Exon 2	
			DQB1*06:21	DQB1*06:48:01	0	0	1	1	Exon 2	
10% (100)			DOB1*06:03:07	DOB1*06:22:02	0	0	2	2	Exon 2	

	Stop: 9267, 3'	UTR 1637					
<mark>A</mark>	Allele 1	Allele 2	CORE	EXONS	PHASE1	PHASE2	Differend
	DQB1*06:02:01G	DQB1*06:04:01G	0	0	0	0	
TGCT TRTC CCTGCCCAGA	AT DQB1*06:02:16	DQB1*06:04:02	0	0	1	1	Exon 2
	DQB1*06:07:02	DQB1*06:13:01	0	0	1	1	Exon 2
G	DQB1*06:09:01	DQB1*06:14:02	0	0	1	1	Exon 2
A	DQB1*06:09:01	DQB1*06:14:02	0	0	1	1	Exon 2
	DQB1*06:11:01	DQB1*06:69:01	0	0	1	1	Exon 2
	DQB1*06:11:01	DQB1*06:69:01	0	0	1	1	Exon 2
	DQB1*06:21	DQB1*06:48:01	0	0	1	1	Exon 2
	DQB1*06:03:07	DQB1*06:22:02	3478	7	-	0	1
	DQB1*06:07:01	DQB1*06:13:02	54/6	(
	DQB1*06:09:04	DQB1*06:14:01	1	×	< → ≯	7685	~ +
	DQB1*06:11:02	DQB1*06:69:02	*	ACCEPT	CGT	R = A + G	× -
	DQB1*06:118:02	DQB1*06:206:01					
	DQB1*06:146:01	DQB1*06:186	Full G	ene	~ 1	Master	~

