Supplemental Table S1. First column: Counts of SNPs from trio analysis. Second and Third column: Counts of trio phased SNPs that match/do-not-match phasing assigned by Illumina Platinum genomes. Fourth column: Counts of SNPs not found in Illumina Platinum Genomes.

Trio phased SNP counts		Trio phased SNP counts (phasing compared with ILLUMINA Platinum Genomes)		
		Matched	Not Matched	Not Found
1868 in-phase	1762 match TruSeq in-phase	1,712	0	50
	106 do not match TruSeq in-phase	67	0	39
664 out-of-phase	662 match TruSeq out-of-phase	642	3	17
	2 do not match TruSeq out-of-phase	1	1	0

Supplemental Table S2. Counts of TruSeq phased SNPs (for the entire set of 262 heterozygous deletions) that could and could not be phased by trio analysis.

	TruSeq SNPs	Phased by trio	Not phased by trio	Percentage not phased
In-phase	2,185	1,852	333	15.2±0.77%
Out-of-phase	1,065	897	168	15.8±1.12%

Supplemental Table S3. Counts of TruSeq phased SNPs (for 220 heterozygous deletions, phased using trio analysis) that could and could not be phased by trio analysis.

	TruSeq SNPs	Phased by trio	Not phased by trio	Percentage not phased
In-phase	1,889	1,765	124	6.6±0.57%
Out-of-phase	867	769	98	11.3±1.08%

Supplemental Table S4. Counts of SNPs/indels flanking NH deletions (likely generated by replication based mechanisms) and other non-replication based deletion events. The counts are for variants phased by trio analysis. 10 kbp flanking windows were considered. Deletions in HLA loci were removed from counting. The difference in ratios for SNPs/indels for NA12878 is highly statistically significant (p-value < 10^{-3}, two sample proportion test). For NA19240, the difference in ratios for SNPs is marginally significant (p-value = 0.02, two sample proportion test); and not statistically significant for indels (p-value = 0.90, two sample proportion test).

		Number of Deletions	SNPs			Indels		
			In-phase	Out-of-phase	Ratio	In-phase	Out-of-phase	Ratio
NH (replication based)	NA12878	91	1,969	638	3.09	380	120	3.17
	NA19240	141	2,617	1,218	2.15	515	283	1.82
Not replication based	NA12878	207	3,098	1,663	1.86	711	359	1.98
	NA19240	315	5,839	2,987	1.96	1,191	662	1.80

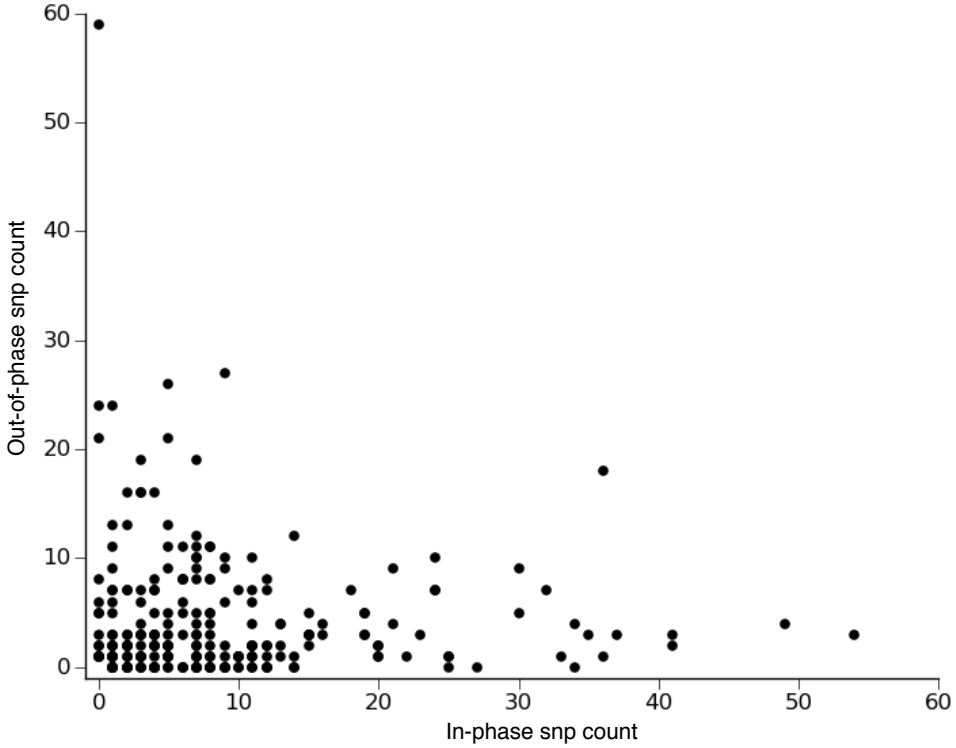
Stratification of SNPs and Indels between NH (replication based) and other deletion events:

We classified the deletion events by their mechanisms of origin. For NA19240, we have 457 het, phased deletion events: 349 NH (76.4%), 75 are NAHR (16.4%), 14 are TEI (3.1%), 14 are VNTR (3.1%), and 5 (1%) are classified as unsure.

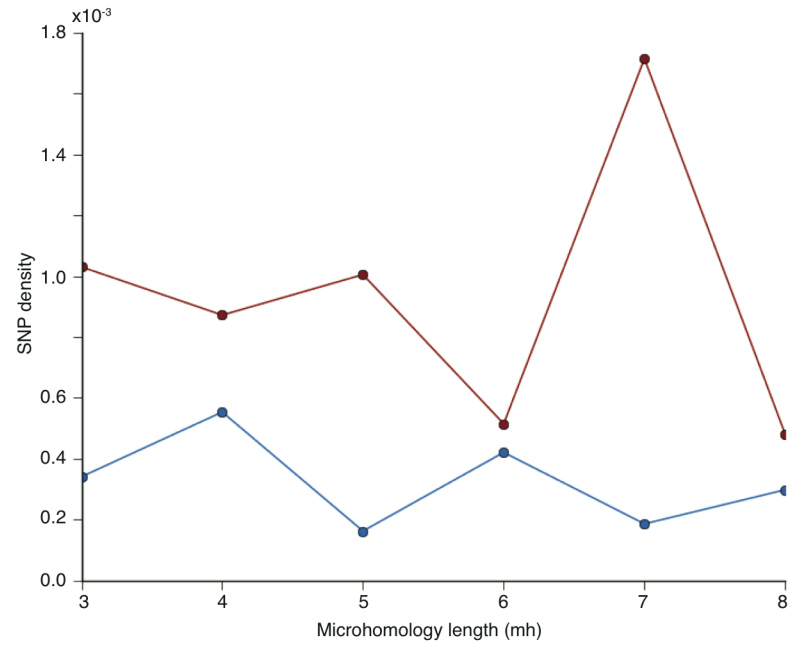
For NA12878, we have 300 het, phased deletion events: 221 NH (73.67%), 43 are NAHR (14.3%), 20 are TEI (6.7%), 10 are VNTR (3.3%), and 6 (2%) are classified as unsure.

Supplemental Table S5. T_i/T_v ratios for in-phase and out-of-phase SNPs around deletion breakoints, for the two individuals. SNPs were phased by trio analysis.

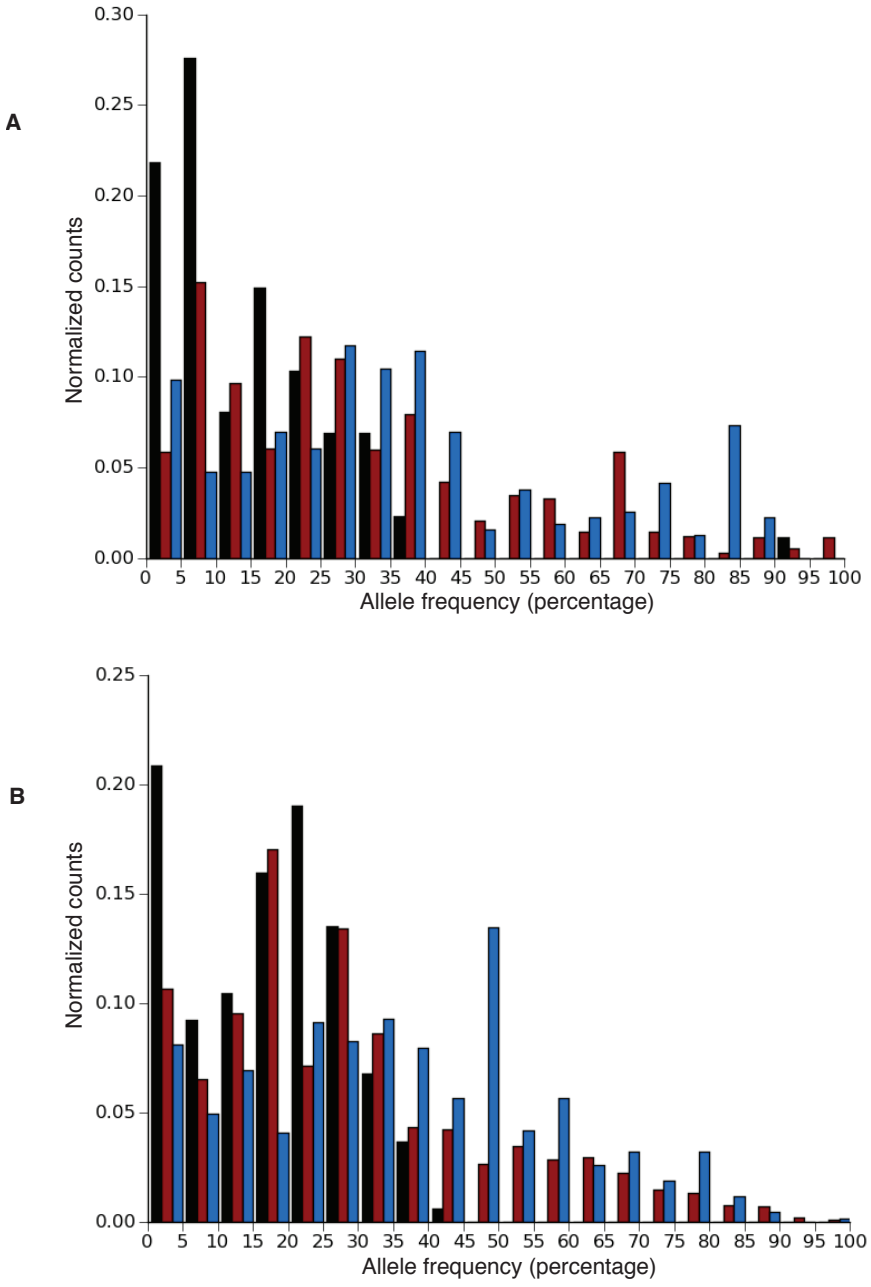
Individual	T_i/T_v ratio				
	10 kbp		50 kbp		Genome wide
	In-phase	Out-of-phase	In-phase	Out-of-phase	
NA12878	2.07	2.00	2.01	1.97	2.08
NA19240	2.06	2.08	2.04	2.06	2.08



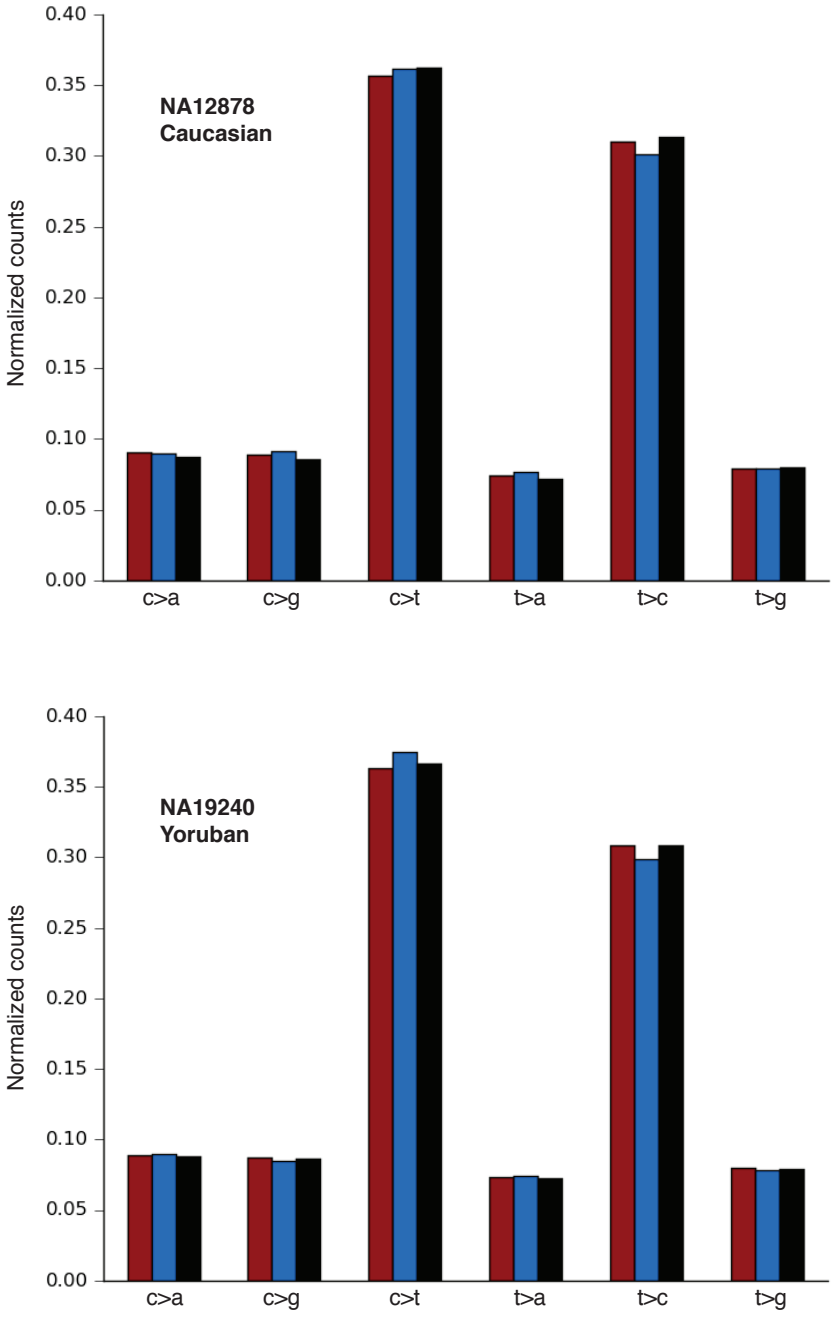
Supplemental Figure S1. In-phase and out-of-phase SNPs counts for deletions. Each dot represents a deletion event considered in the TruSeq analysis. 250 of the 262 heterozygous deletions have SNP counts. Total number of in-phase SNPs is 2,185; and out-of-phase SNPs is 1,065.



Supplemental Figure S2. In-phase (dark red) and out-of-phase (pure blue) SNP densities for NH deletions with different microhomology lengths (TruSeq analysis). The plot is for deletions classified as likely generated by replication-based mechanisms. For mh ≥ 8, we lumped all the deletions and the associated SNPs together. For the density, we took flanking windows of size 10 kbp.



Supplemental Figure S3. AF distribution for in-phase TruSeq SNPs (dark red), out-of-phase TruSeq SNPs (pure blue), and for deletions (black). (A) The distributions are for 94 deletions (those likely generated by replication-based mechanisms) and associated SNPs **(B)** The distributions are for the remaining 168 deletion and associated SNPs.



Supplemental Figure S4. Normalized mutational profile for SNPs obtained from trio phasing. The profile is obtained for in-phase (dark red) and out-of-phase (pure blue) SNPs (within 50 kbp flanking windows), for the two individuals. The six possible transversions and transitions are shown, with the normalized count on the vertical axis. The third bar (black) represents personal (for each subject) genome wide profile based on heterozygous SNPs.