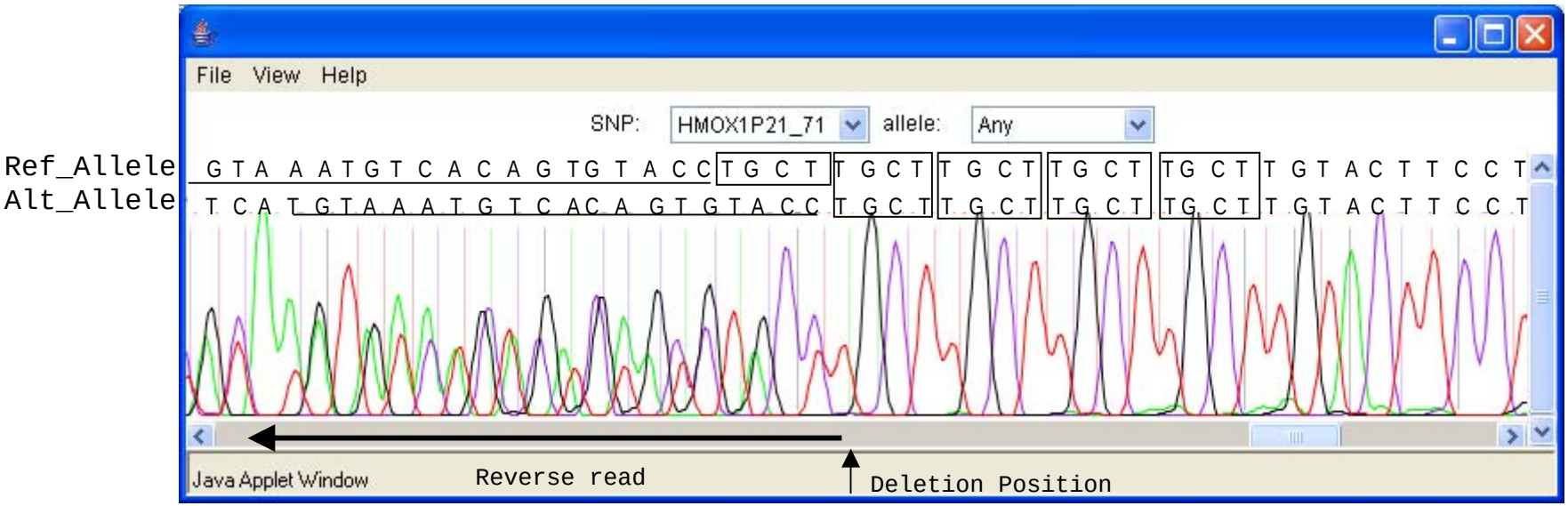
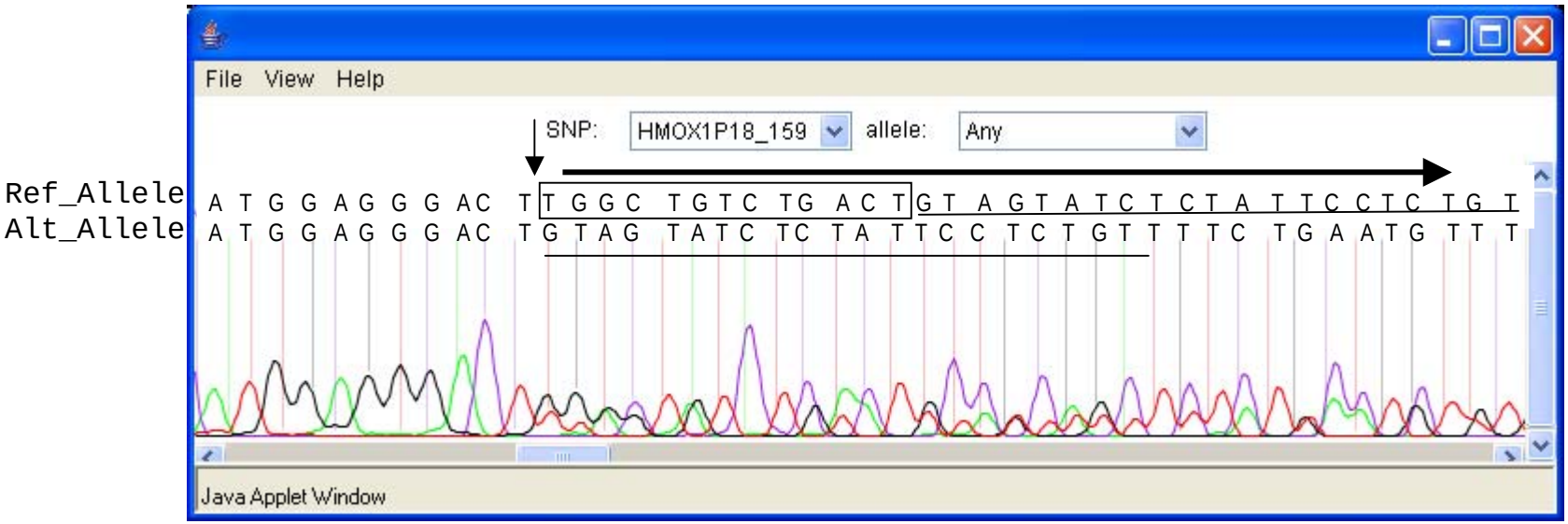
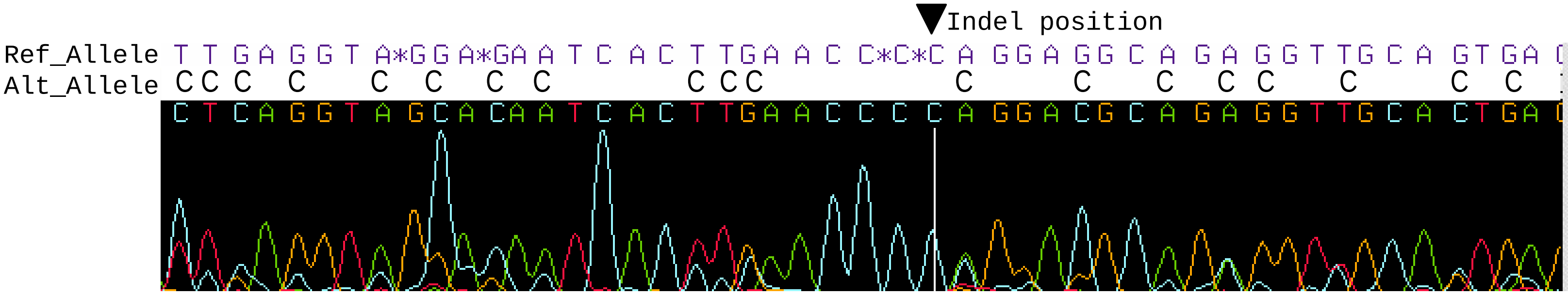


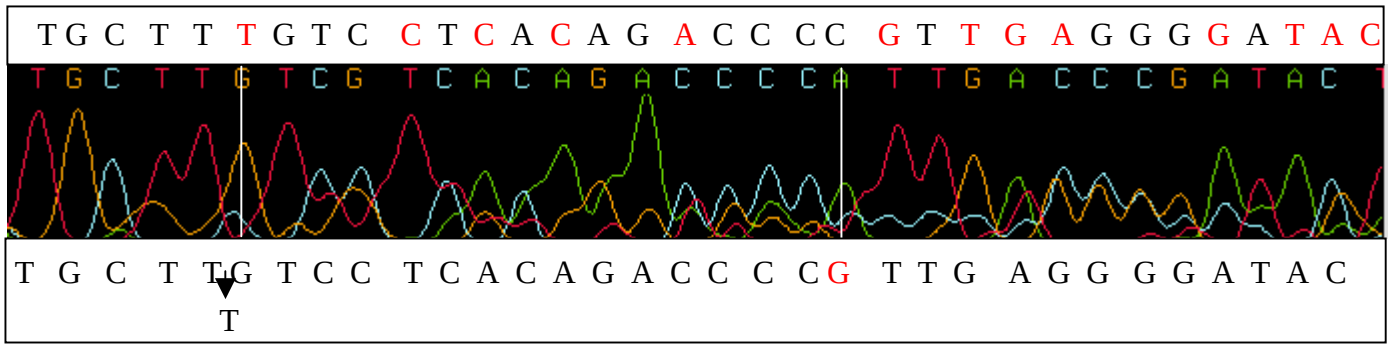
Supplementary Figure s1. The reverse read of sample NA17136 at the indel position of HMOX1P21_71/HMOX1_014568, a matching indel with different allele size by IndelDetector (HMOX1P21_71) and the SeattleSNPs (HMOX1_014568). HMOX1P21_71 was found to have a 4-bp deletion of TGCT by IndelDetector while HMOX1_014568 was reported to have a 5-bp deletion of CTGCT by the SeattleSNPs. The two alleles decoded by IndelDetector are labeled on top of the chromatogram. The matching bases after the deletion in the two alleles are underlined. The deleted bases that are present in the Ref_Allele but absent in the Alt_Allele is labeled by a box. There are 4 base pairs in the box, showing the deletion size is 4 and the deletion resides in a 4-bp repeat region. The display was generated by our trace viewer where the A, C, G, T residue is shown in green, purple, black and red color, respectively.



Supplementary Figure s2. The forward read of sample NA10831 at the indel position of HMOX1P18_159/HMOX1_012127, a matching indel with different allele size by IndelDetector and the SeattleSNPs. HMOX1P18_159 was found to have a 13-bp deletion by IndelDetector while HMOX1_01212 was reported to have a 1-bp deletion by the SeattleSNPs. The two alleles decoded by IndelDetector are labeled at the top of the chromatogram. The matching bases after the deletion in the two alleles are underlined. The deleted bases that are present in the Ref_Allele but absent in the Alt_Allele is labeled by a box and there are 13 bases in the box.



Supplementary Figure s3. Sequence chromatogram of sample NA10830 in indel CD26_021731. The indel position in this trace is highlighted. The indel was discovered only by SeattleSNPs group but not by IndelDetector. In SeattleSNPs' genotype report file, cd36a_prettybase.txt, this variation has heterogyzous +/- bi-allelic genotype in samples NA10830, NA17140 and NA17105. The NCBI trace archive only has trace for one sample, NA10830. In this sample, after subtracting the alleles that match the reference human genome (shown in the row Ref_Allele), the remaining signals (shown in the row Alt_allele) are all Cs, indicating that the alternative peaks are background noise rather than representing signals of an indel polymorphism.



Supplementary Figure s4. A read with no perfect match to the reference sequence in its peak-overlapping region. The two mismatches are highlighted with the white bars. The reference sequence is shown at the top and the bottom, a red base indicates a mismatch between the site and the reference base (e.g. neither the primary peak nor the secondary peak matches the reference base). The top shows the match status when considering the first mismatch is caused by a substitution change. The bottom shows the match status when the first mismatch is considered a deletion in the current read (displayed as an insertion in the reference).

A)

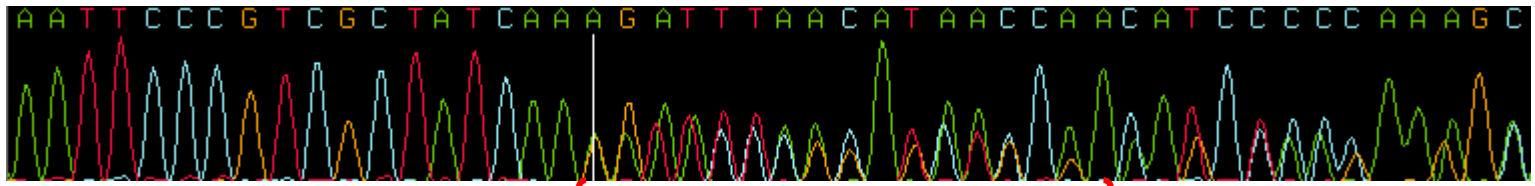
```
seq1    agaaagttaaaattcccgtcgctatcaaggaattaagagaagcaacatctccgaaagccaacaagga
seq2    ...........................------.c-------.cg.....................
```

B)

```
seq1    agaaagttaaaattcccgtcgctatcaaggaattaagagaagcaacatctccgaaagccaacaagga
seq2    ....................................--------aattcccg.....................
```

C)

```
hap1    agaaagttaaaattcccgtcgctatcaaggaattaagagaagcaacatctccgaaagccaacaagga
hap2    ...................................----------------.....................
                                                 aattcccg
```



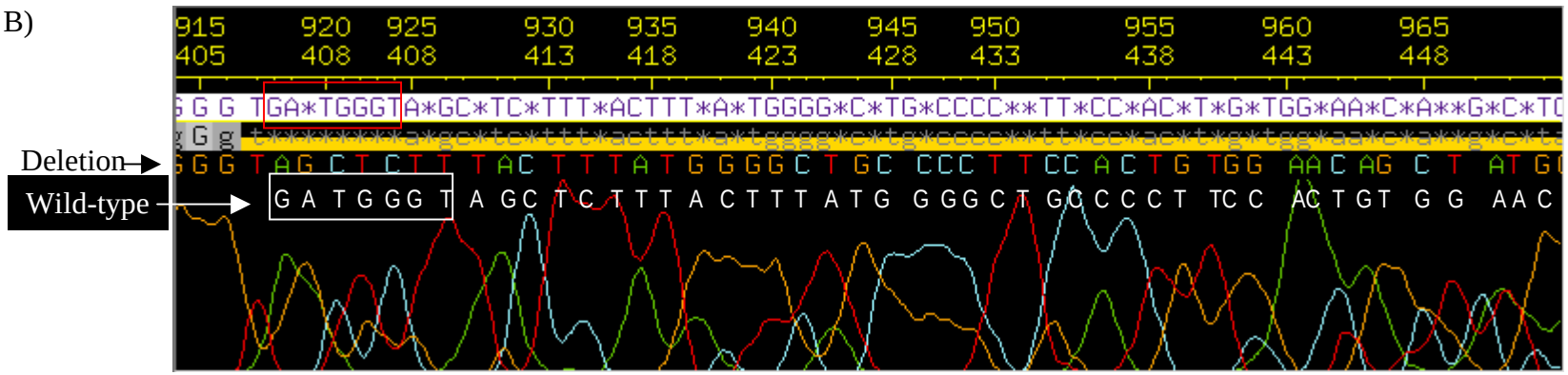
```
seq1 AATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGC
seq2 AATTCCCGTCGCTATCAAAATTCCGCATCTCCGAAAGCCAACAAGGC
```

Supplementary Figure s5. An unusual somatic deletion on EGFR exon 19. A) The optimum alignment between the two sequences decoded in the tumor sample H11318. seq1 matches the reference human genome while seq2 has the deletion. A matched base between the two sequences is shown as a dot ".”; a gap is shown as a dash "-“, and a mismatch is labeled with its basecall residue. The boxed region shows the area of mismatches and gaps. B) The non-optimum alignment created by merging the two gaps in A) into 9-bp gap. The 8-bp highly divergent region is shown in the rectangle. It matches perfectly to an 8-bp sequence 10bp upstream the gap which is shown in underline. C) The alignment that converts the 8-bp mismatch into an insertion and the 9-bp gap into a 17bp gap. The chromatogram of this region is shown below with two decoded sequences. The 17-bp deletion is marked in red parenthesis. The 8-bp replication in seq2 is underline with the insertion shown in a box.

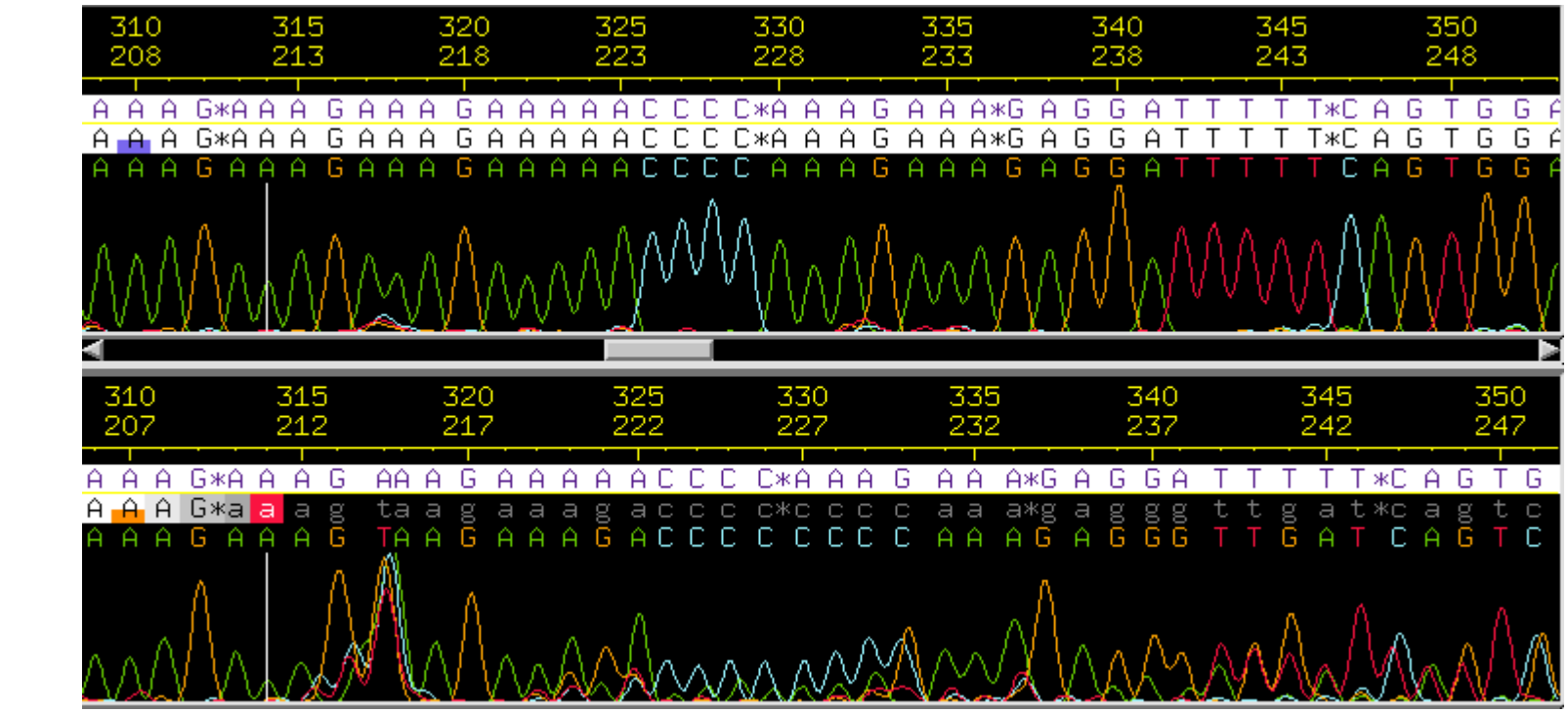
A)

```
                                    10        20        30        40        50        60        70        80        90        100
                                    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
template     > 383        ccttctgcaacagcctctgccacgacgtctcgggtgatgggtagctctttactttatggggctgccccttccactgtggaacagctctggccccggggaa 482
P072_P17_3   > 373        ..........................................-------.....................................................a..... 465
P072_P17_2   < 584        .a..ag.........g..g..t.....g.t...-------.....................c.............................. 492
```

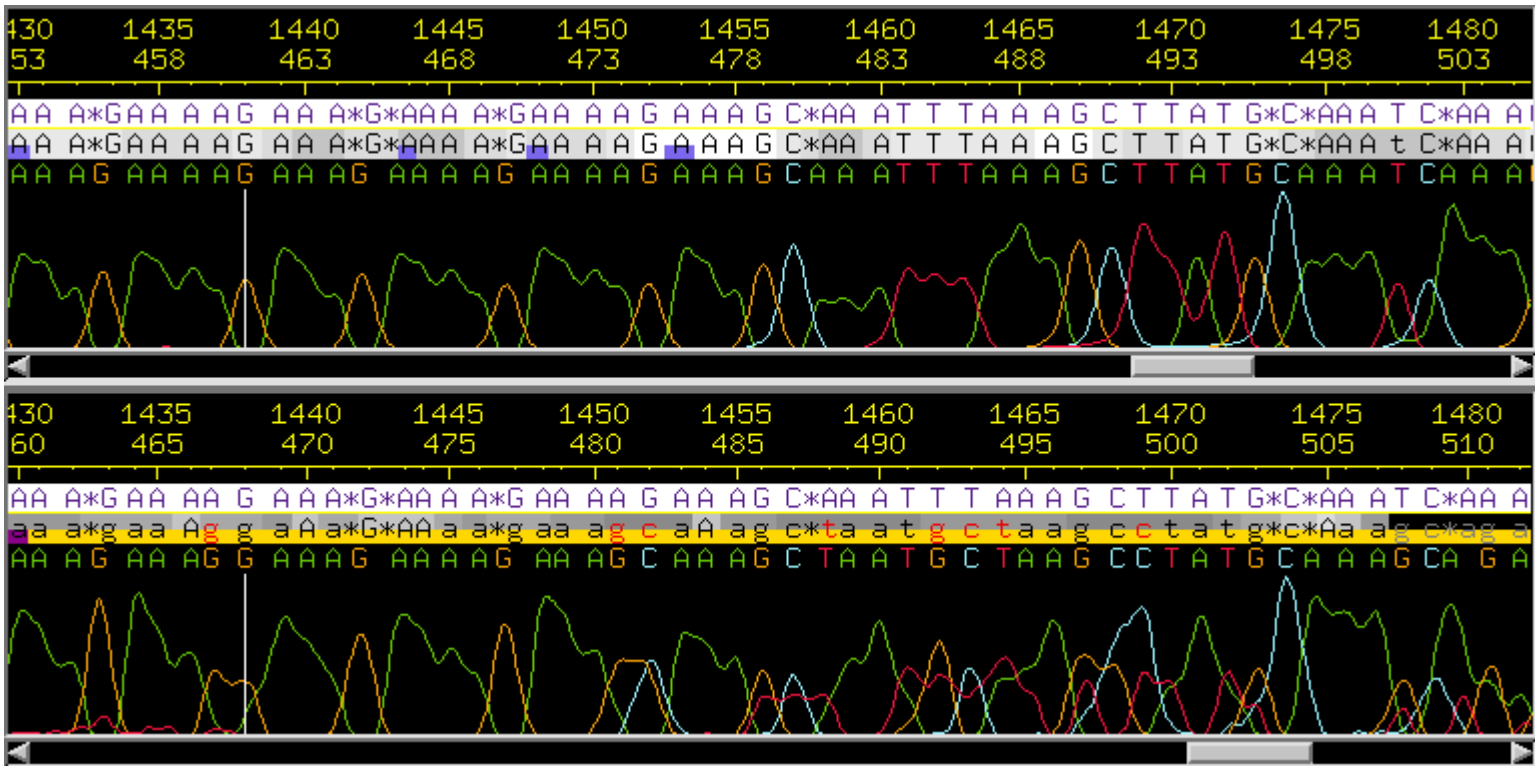
B)



Supplementary Figure s6. A 7-bp novel deletion (ALAD-003600) discovered by IndelDetector but not recorded in the genotype file of SeattleSNPs. The deletion was found in sample P072 and the NCBI trace uid (known as the ti numbers) for the forward and the reverse traces are 1002553127 (trace name ALADX0050P072.x.BT_079.1.scf
) and 1002443217 (trace name ALADX0051P072.y.BT_079.1.scf), respectively. A) The alignment of the forward and the reverse reads to the reference sequence. The alignment is displayed in the same format as that of Figure s5. The 7-bp deletion of gatgggt was found in both the forward and the reverse reads. B) The trace view of the forward read in the subregion highlighted in green in A). The primary peaks represent the deletion allele while the secondary peaks represent the wild-type allele (shown in white characters). The trace profile is consistent with the heterozygous genotype call made by IndelDetector



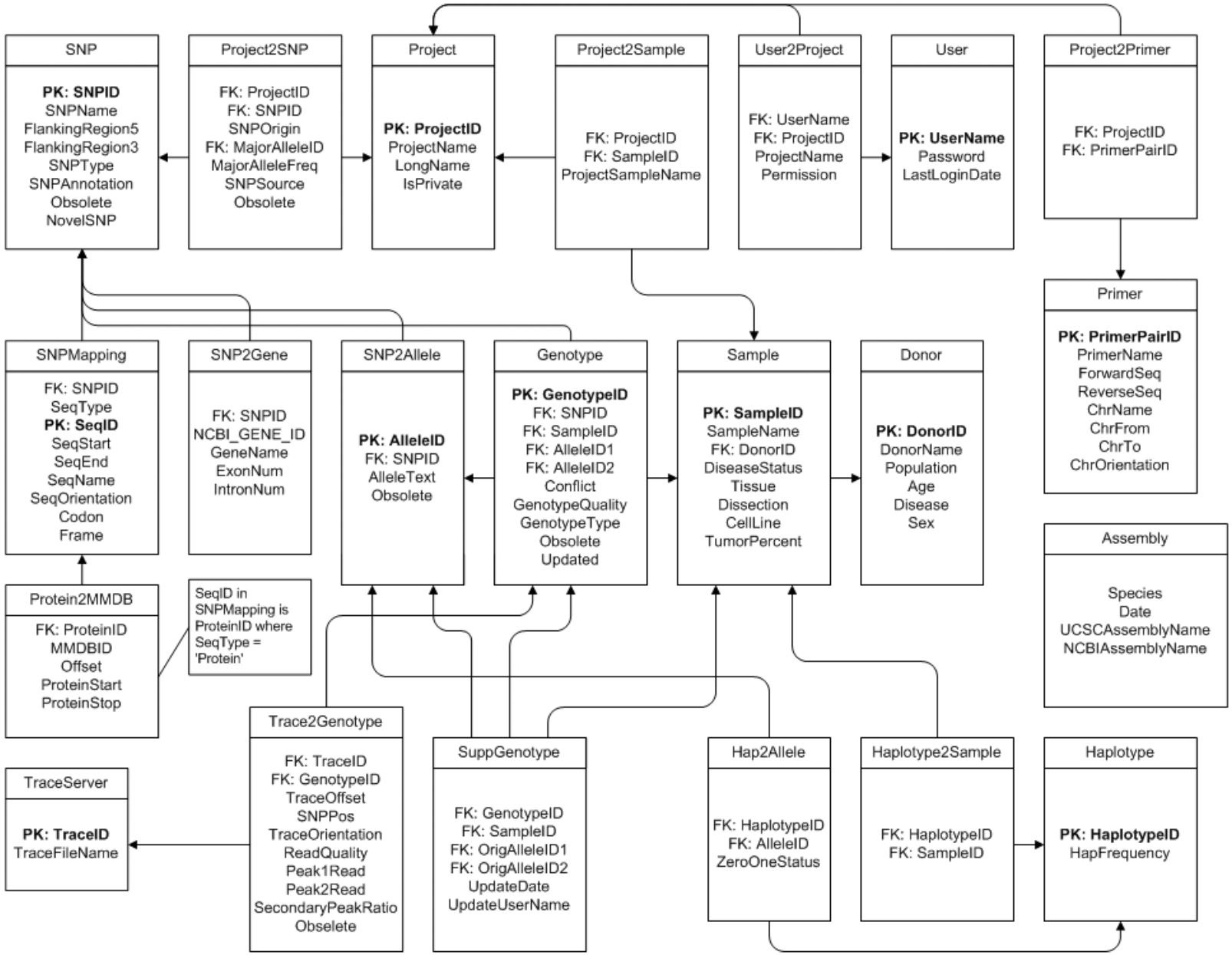
Supplementary Figure s7. A novel 4-bp insertion (SERPINE1-013511) discovered by IndelDetector in the reverse read of sample D037 (shown at the bottom panel) but recorded not in the genotype file of SeattleSNPs. The top panel displays the reverse read of a homozygous sample. The two alleles decoded from the heterozygous sample D037 are displayed at the bottom. seq1 represents the wild-type allele that matches the reference sequence while seq2 represents the indel allele that has the 4-bp insertion (labeled with a box). There is 1 base in seq1 (shown in red) that fails to match the reference sequence and 1 mismatch in seq2 (shown in red) to the reference sequence after the 4-bp insertion. In both cases the mismatches appear to be sequencing artifacts.



seq1 AA A G AA AG G AAA G AA A A G AA AA G AA AG C AA A T T T A A A G C T T A T G C A A A T C AA A

seq2 AA A G AA AG A A A A G AAA A G AA AG C AA AT T TA A A G C T T A T G C A A A T C AA A GA T G G

Supplementary Figure s8. A deletion (SERPINE1-002362) whose size was recorded as 1bp in the SeattleSNPs genotype file but found to be 5bp by IndelDetector. The top panel displays a homozygous sample while the bottom panel displays the forward read of the heterozygous indel sample E011. The two alleles decoded from the heterozygous sample D037 are displayed at the bottom. seq1 represents the wild-type allele that matches the reference sequence while seq2 represents the deletion allele. The 5-bp deletion is labeled in a box. seq2 aligns perfectly to seq1 after the 5-bp deletion.



Supplementary Figure s9. The CGWB variation database schema. Primary keys are indicated by the prefix "PK" and shown in bold, foreign keys by "FK". A foreign key - primary key relation between tables is shown by an arrow connecting the tables, pointing to the table that contains the primary key.