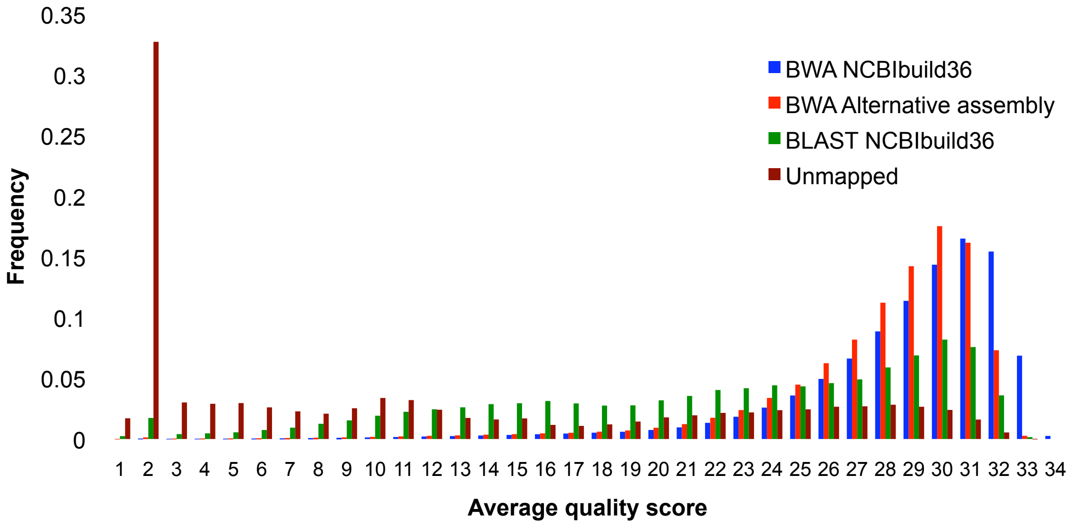
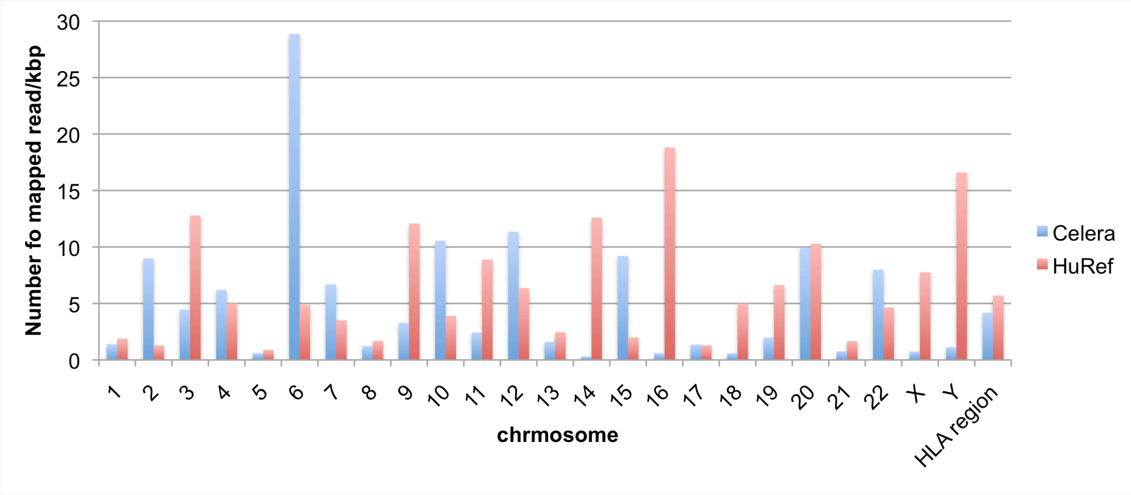
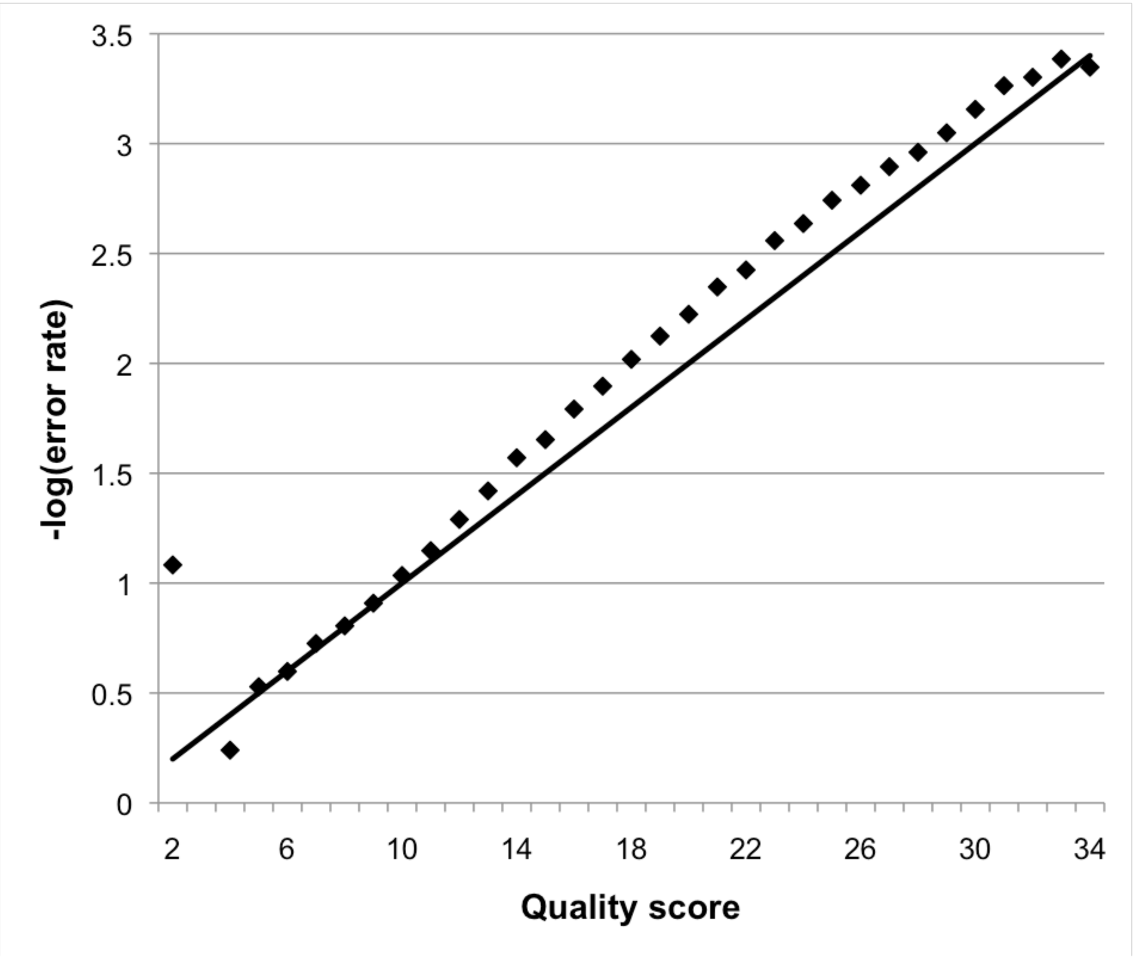


Supplementary Figure 1. **Distribution of quality score.** BWA NCBIbuild36; average quality score of the reads that were mapped to NCBI build 36. BWA Alternative assembly; average quality score of the reads that were mapped to alternative assembly. BLAST NCBIbuild36; average quality score of the reads that were mapped to NCBI build 36. Unmapped; average quality score of the unmapped reads.

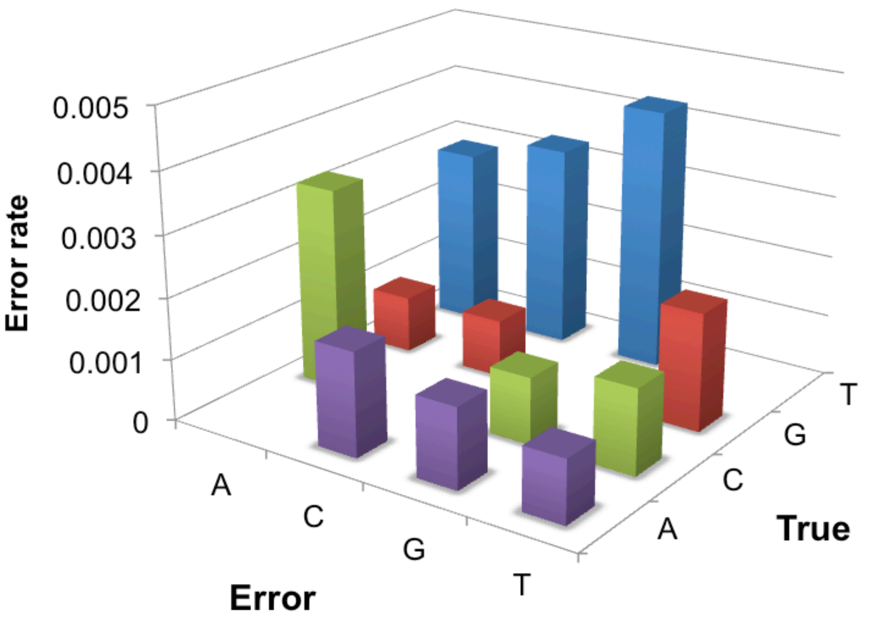


Supplementary Figure 2. **Number of mapped reads to alternative assemblies.** The numbers of mapped reads per kbp are shown.

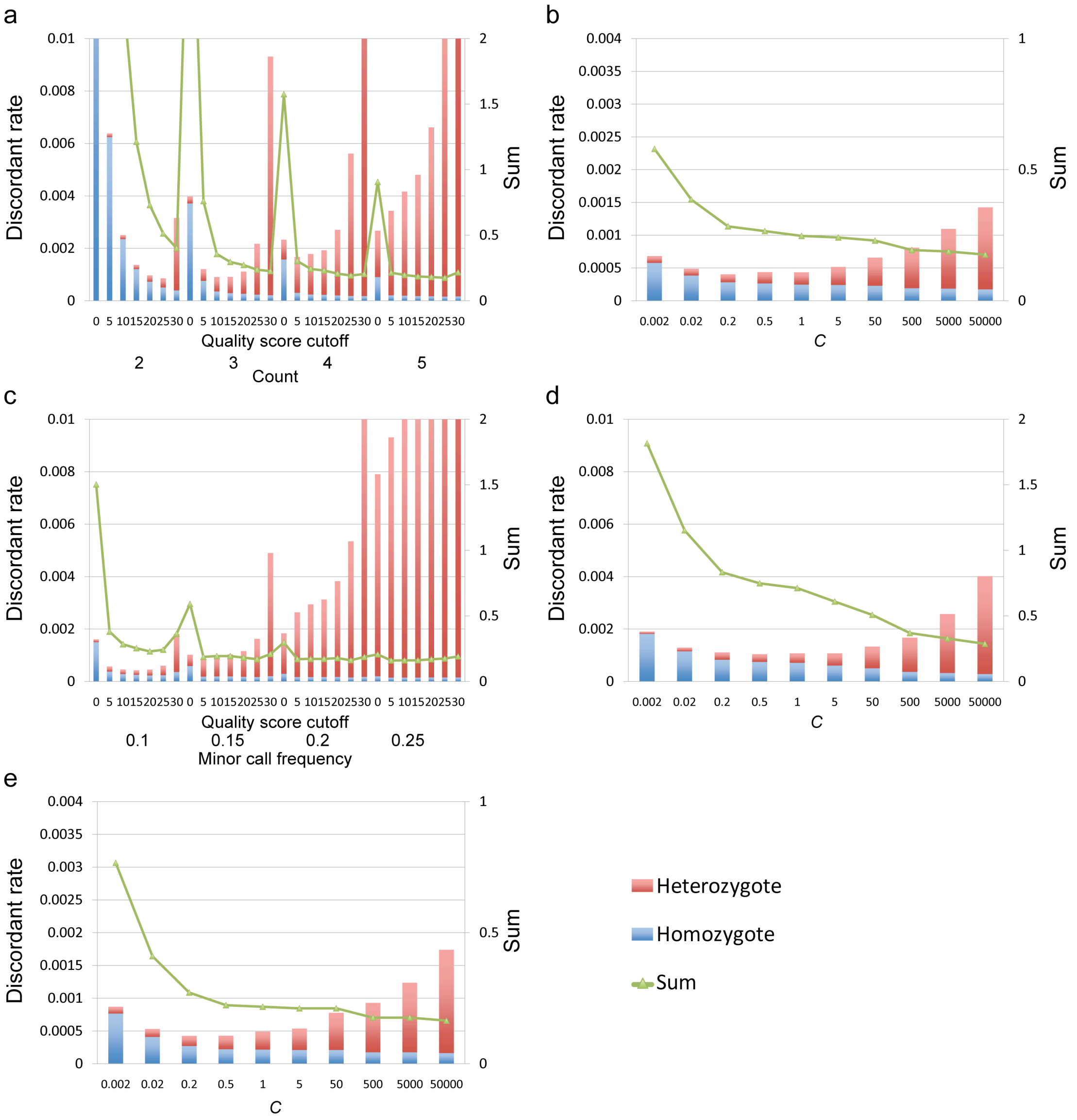


Supplementary Figure 3. **Correlation between observed and expected error rate.** Observed (squares) and expected (line) error rates based on chromosome X and Y are shown. Expected error rates were calculated from sequence quality score.

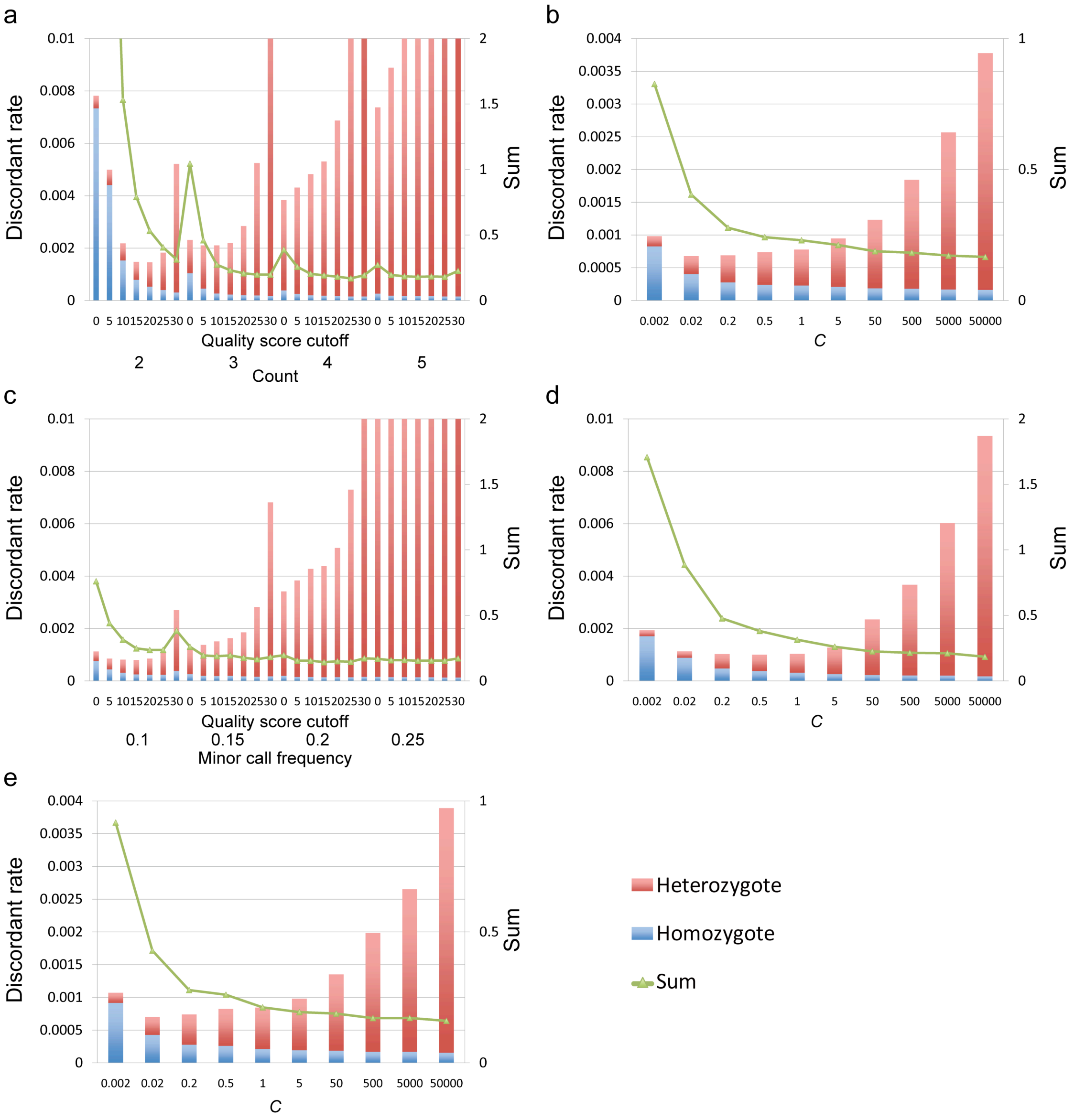
Nature Genetics: doi:10.1038/ng.691



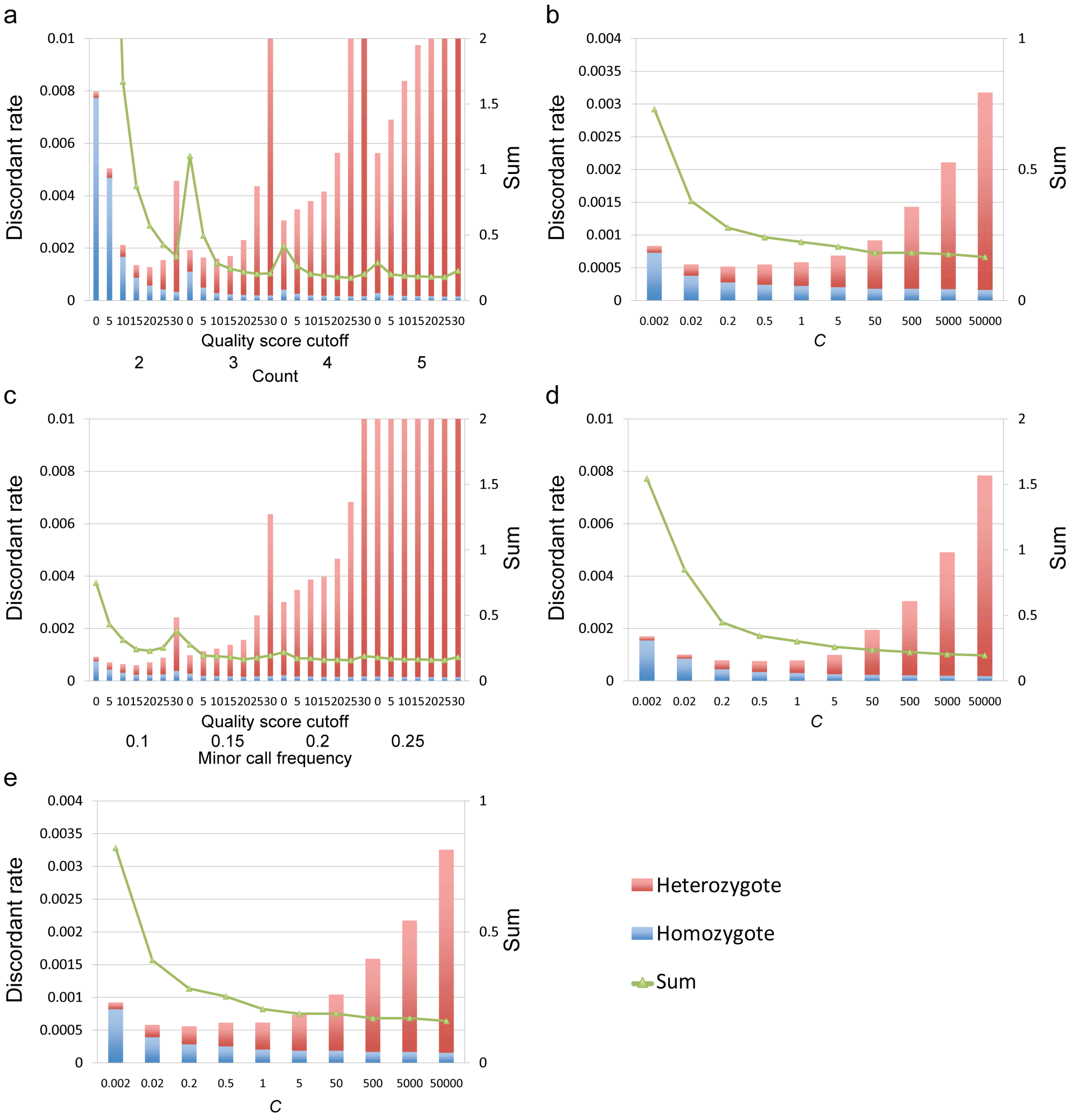
Supplementary Figure 4. **Result of sequencing error rate estimation.** Direction of sequencing errors were estimated from chromosome X and Y.



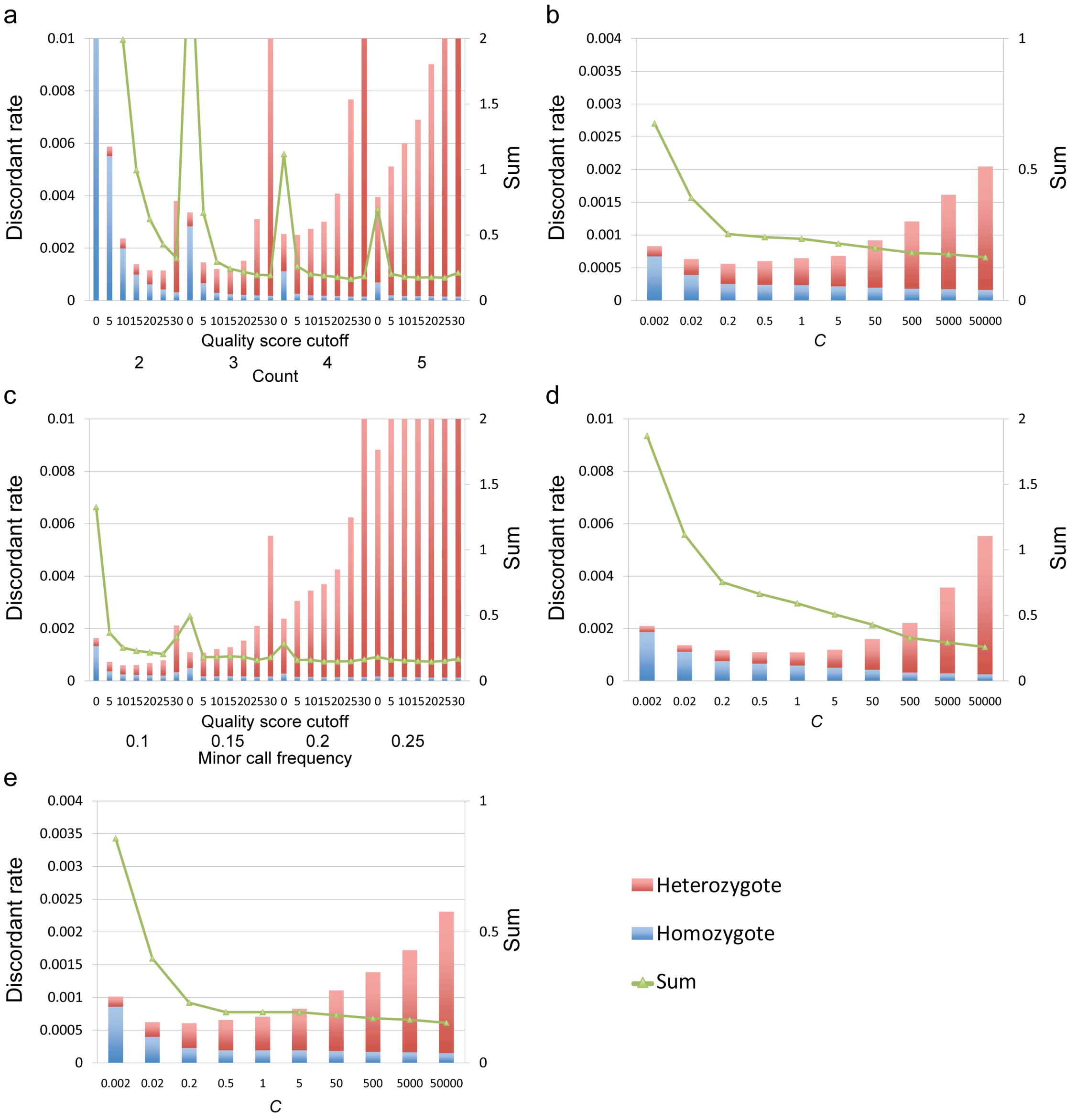
Supplementary Figure 5. **Comparison of SNP calling methods.** No read filtering was performed. (**a**) Counting method. (**b**) Bayesian decision method with quality score correction. *C* indicates the Bayesian-decision threshold. (**c**) Frequency method. (**d**) Bayesian decision method with a constant error rate. *C* indicates the Bayesian-decision threshold. (**e**) Bayesian decision method with no quality score correction. *C* indicates the Bayesian-decision threshold. Red bar; heterozygous discordant rate (D_{Ht}), blue bar; homozygous discordant rate (D_{Ho}) and green line; sum of both discordant rate (Sum = D_{Ht} + 1000*D_{Ho}). For counting and frequency method, we discarded calls whose quality score was smaller than a chosen cut-off value (Quality score cutoff).



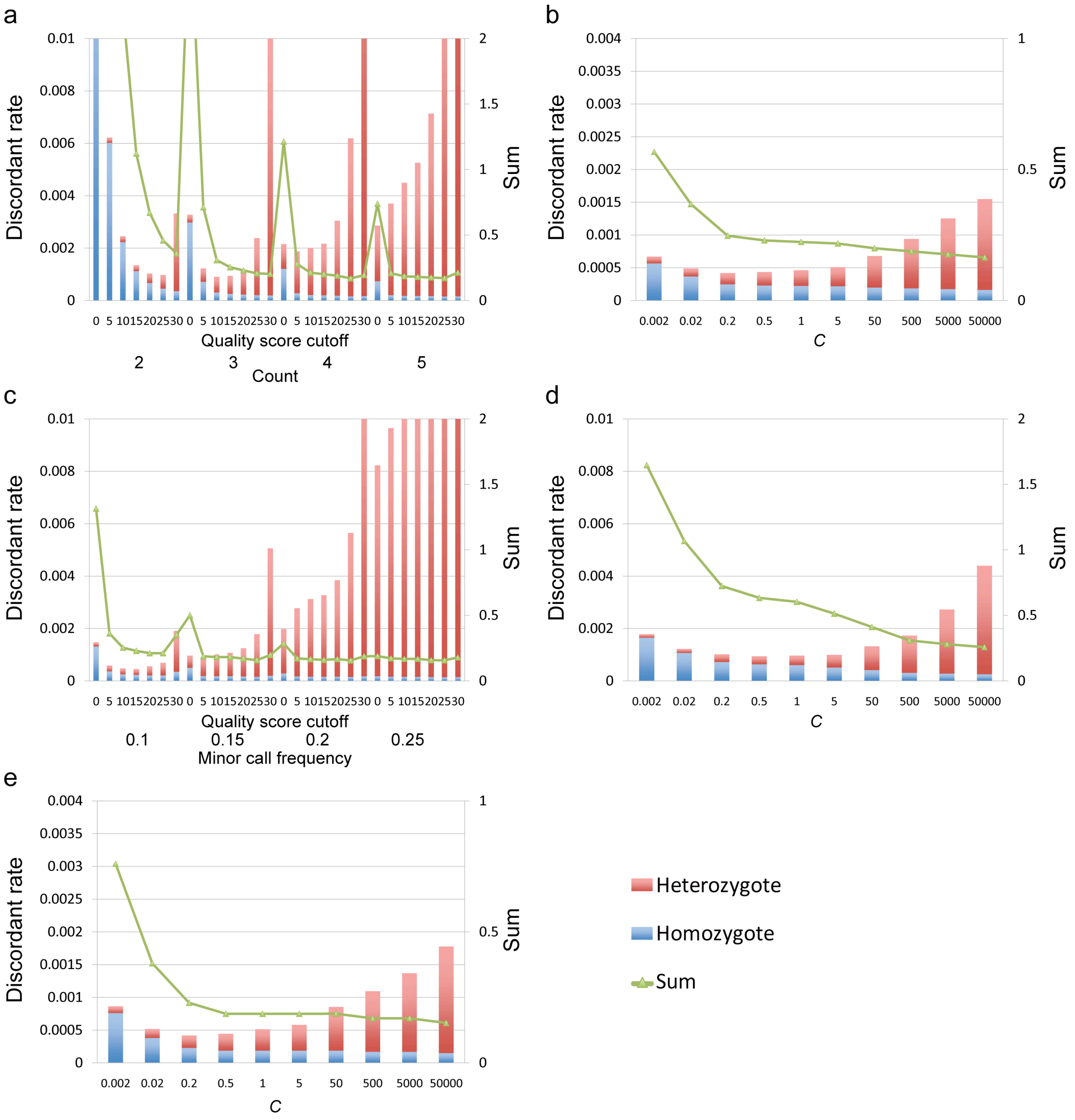
Supplementary Figure 6. **Comparison of SNP calling methods.** Read filtering; average base quality ≥ 20 and mapping quality ≥ 30. (**a**) Counting method. (**b**) Bayesian decision method with quality score correction. *C* indicates the Bayesian-decision threshold. (**c**) Frequency method. (**d**) Bayesian decision method with a constant error rate. *C* indicates the Bayesian-decision threshold. (**e**) Bayesian decision method with no quality score correction. *C* indicates the Bayesian-decision threshold. Red bar; heterozygous discordant rate (D_{Ht}), blue bar; homozygous discordant rate (D_{Ho}) and green line; sum of both discordant rates (Sum = D_{Ht} + 1000*D_{Ho}). For counting and frequency method, we discarded calls whose quality score was smaller than a chosen cut-off value (Quality score cutoff).



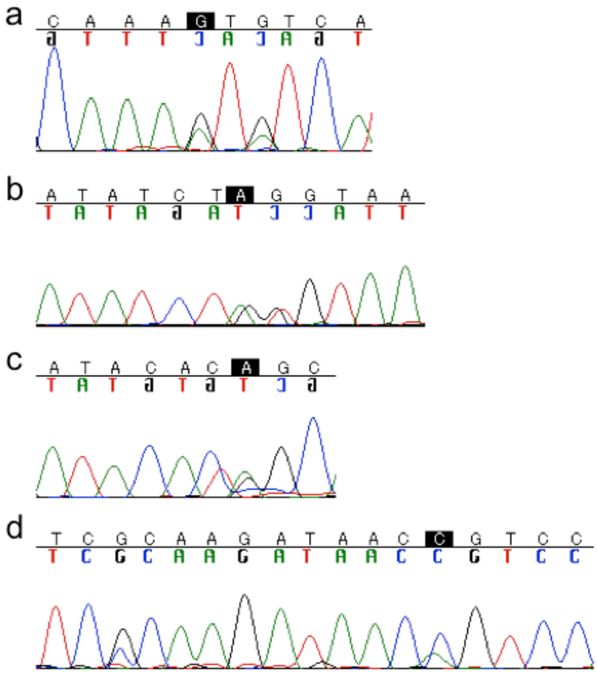
Supplementary Figure 7. **Comparison of SNP calling methods.** Read filtering; average base quality ≥ 20 and mapping quality ≥ 1. (**a**) Counting method. (**b)** Bayesian decision method with quality score correction. *C* indicates the Bayesian-decision threshold. **(c)** Frequency method. (**d**) Bayesian decision method with a constant error rate. *C* indicates the Bayesian-decision threshold. **(e)** Bayesian decision method with no quality score correction. *C* indicates the Bayesian-decision threshold. Red bar; heterozygous discordant rate (D_{Ht}), blue bar; homozygous discordant rate (D_{Ho}) and green line; sum of both discordant rates ($Sum = D_{Ht} + 1000*D_{Ho}$). For counting and frequency method, we discarded calls whose quality score was smaller than a chosen cut-off value (Quality score cutoff).



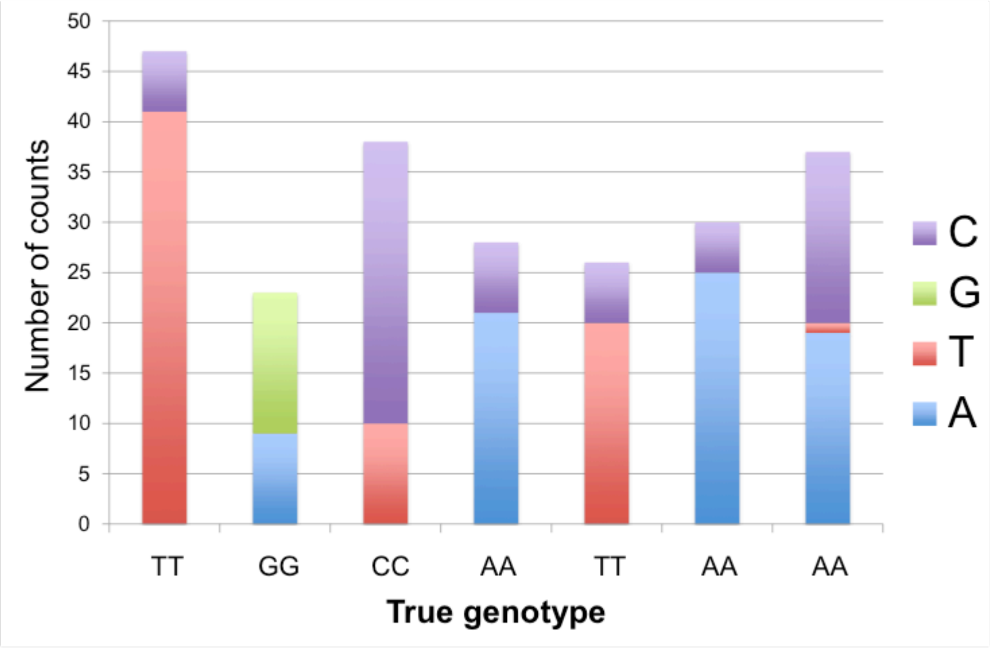
Supplementary Figure 8. **Comparison of SNP calling methods.** Read filtering; average base quality ≥ 10 and mapping quality ≥ 30. (**a**) Counting method. (**b**) Bayesian decision method with quality score correction. *C* indicates the Bayesian-decision threshold. (**c**) Frequency method. (**d**) Bayesian decision method with a constant error rate. *C* indicates the Bayesian-decision threshold. (**e**) Bayesian decision method with no quality score correction. *C* indicates the Bayesian-decision threshold. Red bar; heterozygous discordant rate (D_{Ht}), blue bar; homozygous discordant rate (D_{Ho}) and green line; sum of both discordant rates (Sum = D_{Ht} + 1000*D_{Ho}). For counting and frequency method, we discarded calls whose quality score was smaller than a chosen cut-off value (Quality score cutoff).



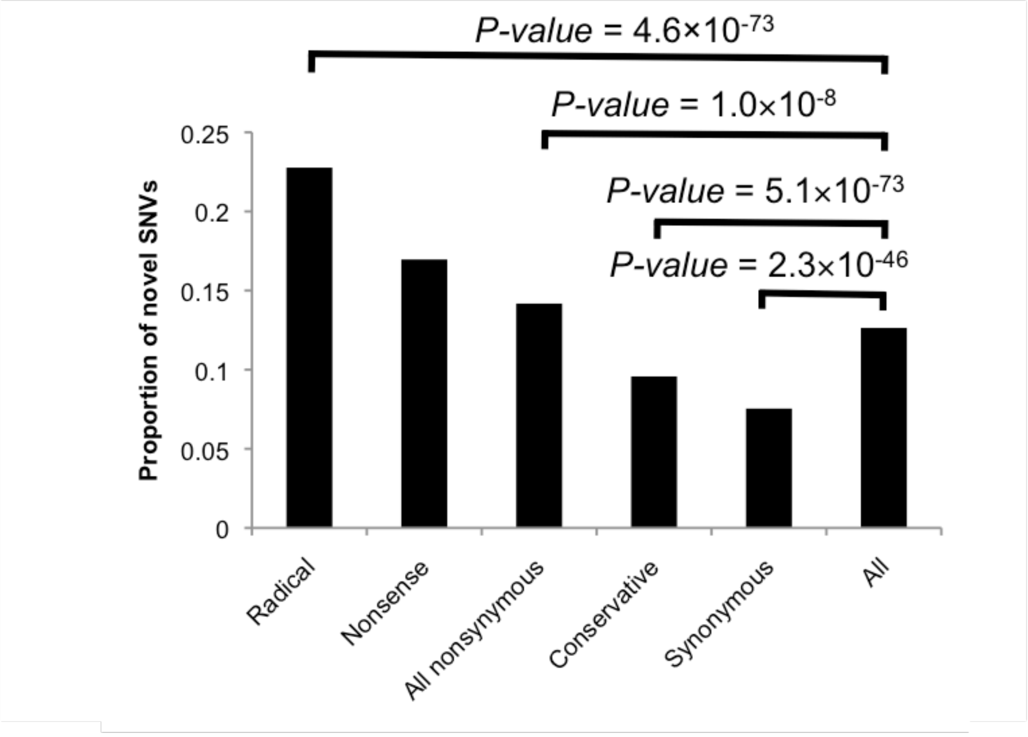
Supplementary Figure 9. **Comparisons of SNP calling methods.** Read filtering; average base quality ≥ 10 and mapping quality ≥ 1. (**a**) Counting method. (**b)** Bayesian decision method with quality score correction. *C* indicates the Bayesian-decision threshold. (**c)** Frequency method. (**d**) Bayesian decision method with a constant error rate. *C* indicates the Bayesian-decision threshold. (**e**) Bayesian decision method with no quality score correction. *C* indicates the Bayesian-decision threshold. Red bar; heterozygous discordant rate (D_{Ht}), blue bar; homozygous discordant rate (D_{Ho}) and green line; sum of both discordant rates (Sum = D_{Ht} + 1000*D_{Ho}). For counting and frequency method, we discarded calls whose quality score was smaller than a chosen cut-off value (Quality score cutoff).



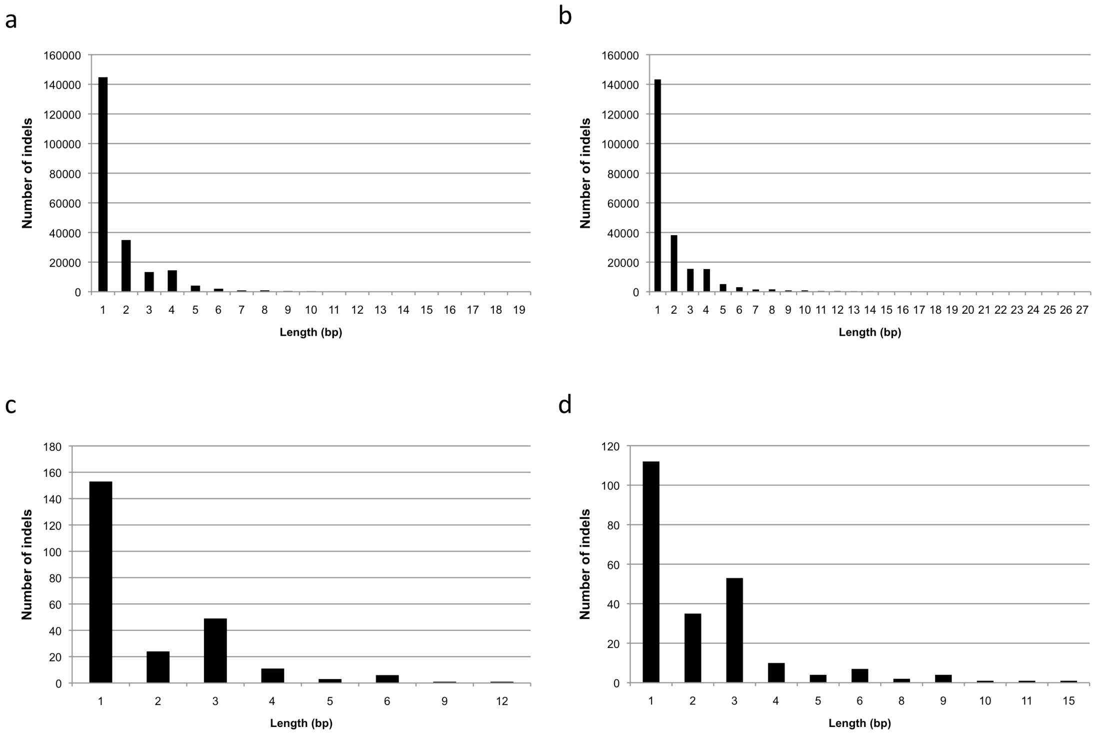
Supplementary Figure 10. **Result of Sanger sequencing near discordant SNPs.** (**a**) rs3828142. (**b**) rs2661665. (**c**) rs9914156. (**d**) rs4788728. Target SNPs were marked in black.



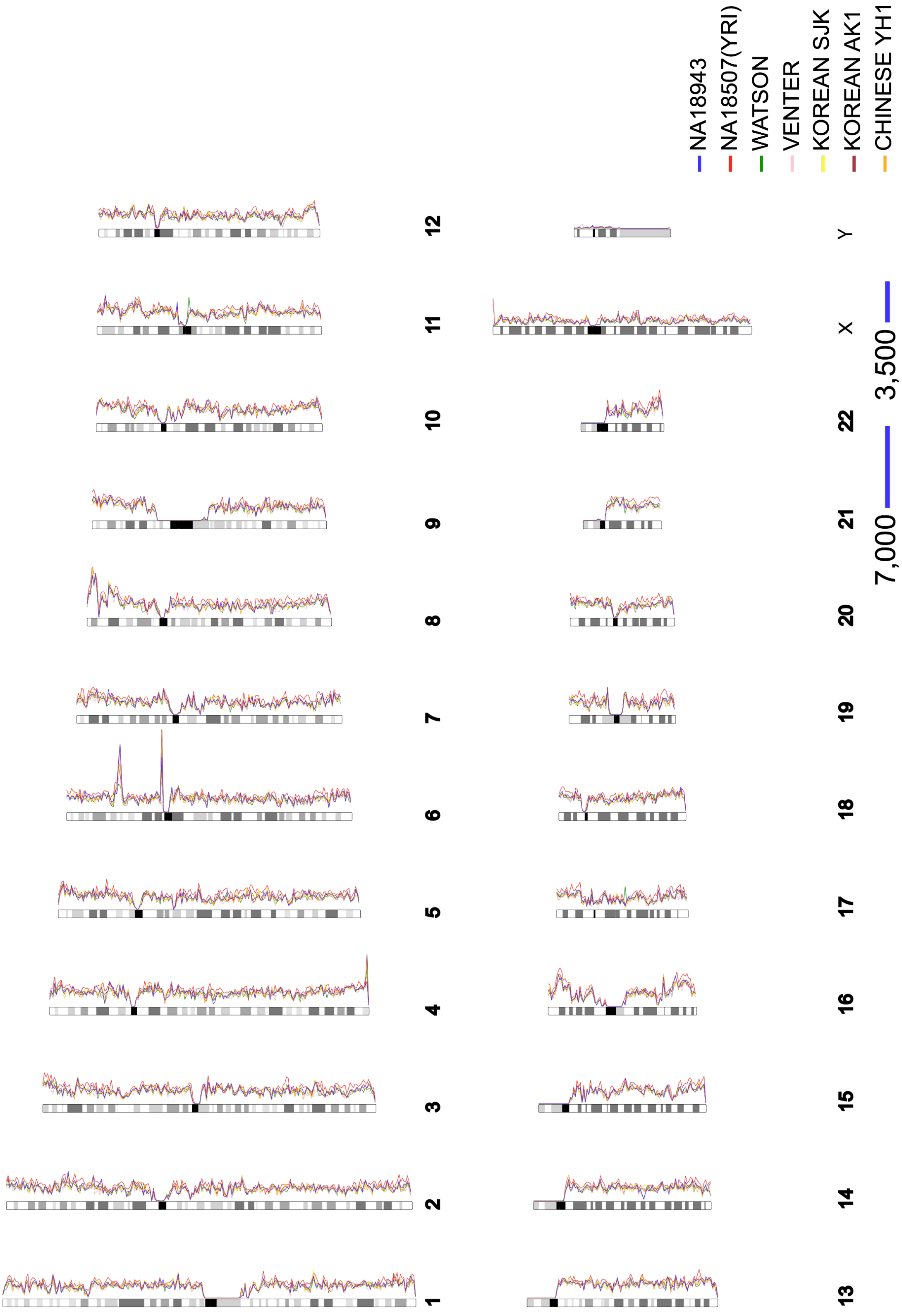
Supplementary Figure 11. **Observed number of calls on false positives of our SNV detection.** True genotype is based on Sanger sequencing method.



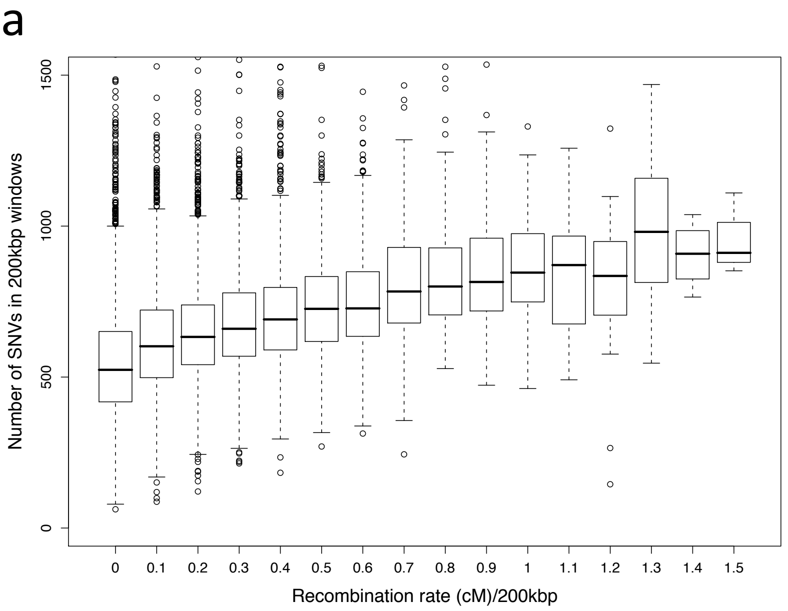
Supplementary Figure 12. **Fraction of novel SNVs in NA18943.** Number of novel and known SNVs (novel/all) were as follows; radical nonsynonymous 772/3,353, nonsense 19/112, all nonsynonymous 1,425/9,783, conservative nonsynonymous 634/6,318, synonymous 797/10,077, and all 395,940/3,132,608.

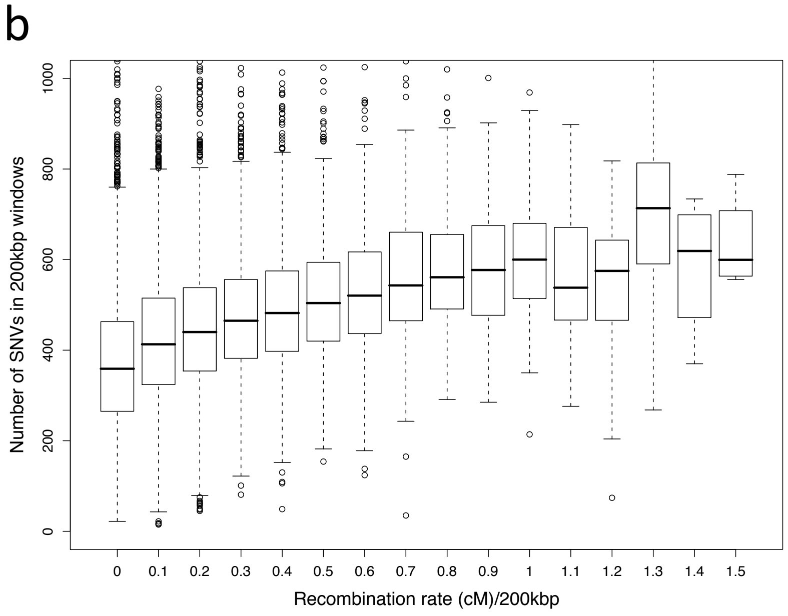


Supplementary Figure 13. **Length distribution of short indels.** (**a**) Insertions. (**b**) Deletions. (**c**) Insertions in coding region. (**d**) Deletions in coding region.

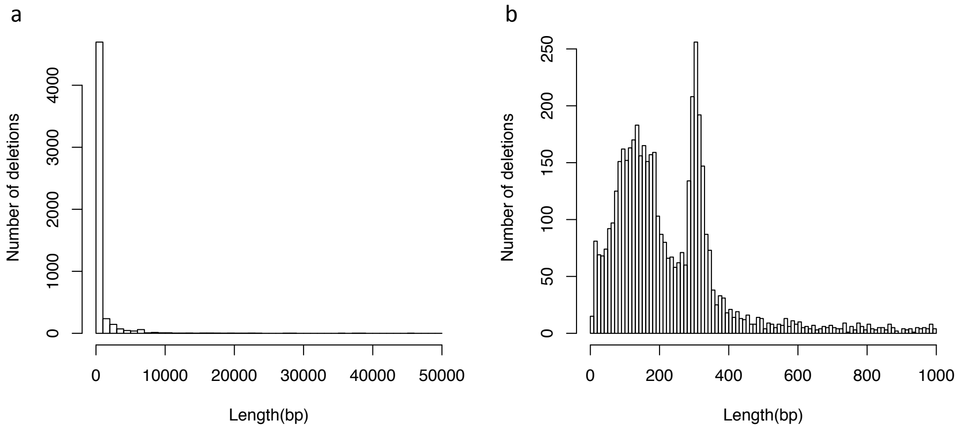


Supplementary Figure 14. **Distribution of number of SNVs within 1Mbp windows of seven individuals.** The vertical and horizontal axis represents position on each chromosome and the number of SNVs within 1Mbp windows, respectively. NA18943 (blue), NA18507 (pink), Watson (green), Venter (pink), SJK (yellow), AK1 (violet), and YH1 (orange). Scale bars at the bottom indicate the number of SNVs.

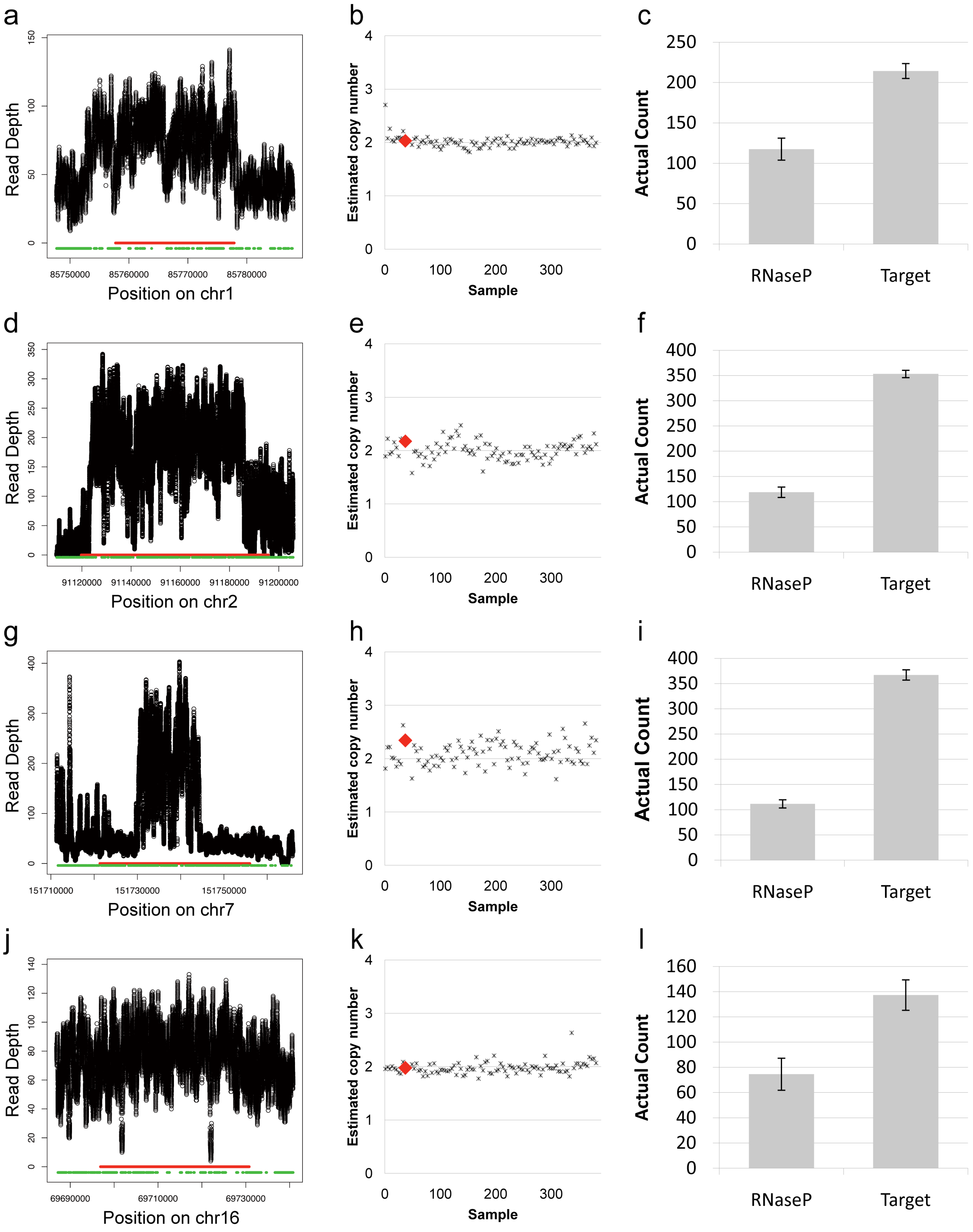




Supplementary Figure 15. **Relationship between recombination rate and SNV density within 200kbp.** (**a**) Seven individuals (NA18507, Watson, Venter, SJK, AK1, YH1 and NA18943). Significant correlation was observed (*P-value* $< 1.0\times10^{-16}$). (**b**) Four Asians (SJK, AK1, YH1 and NA18943). Significant correlation was observed (*P-value* $< 1.0\times10^{-16}$).

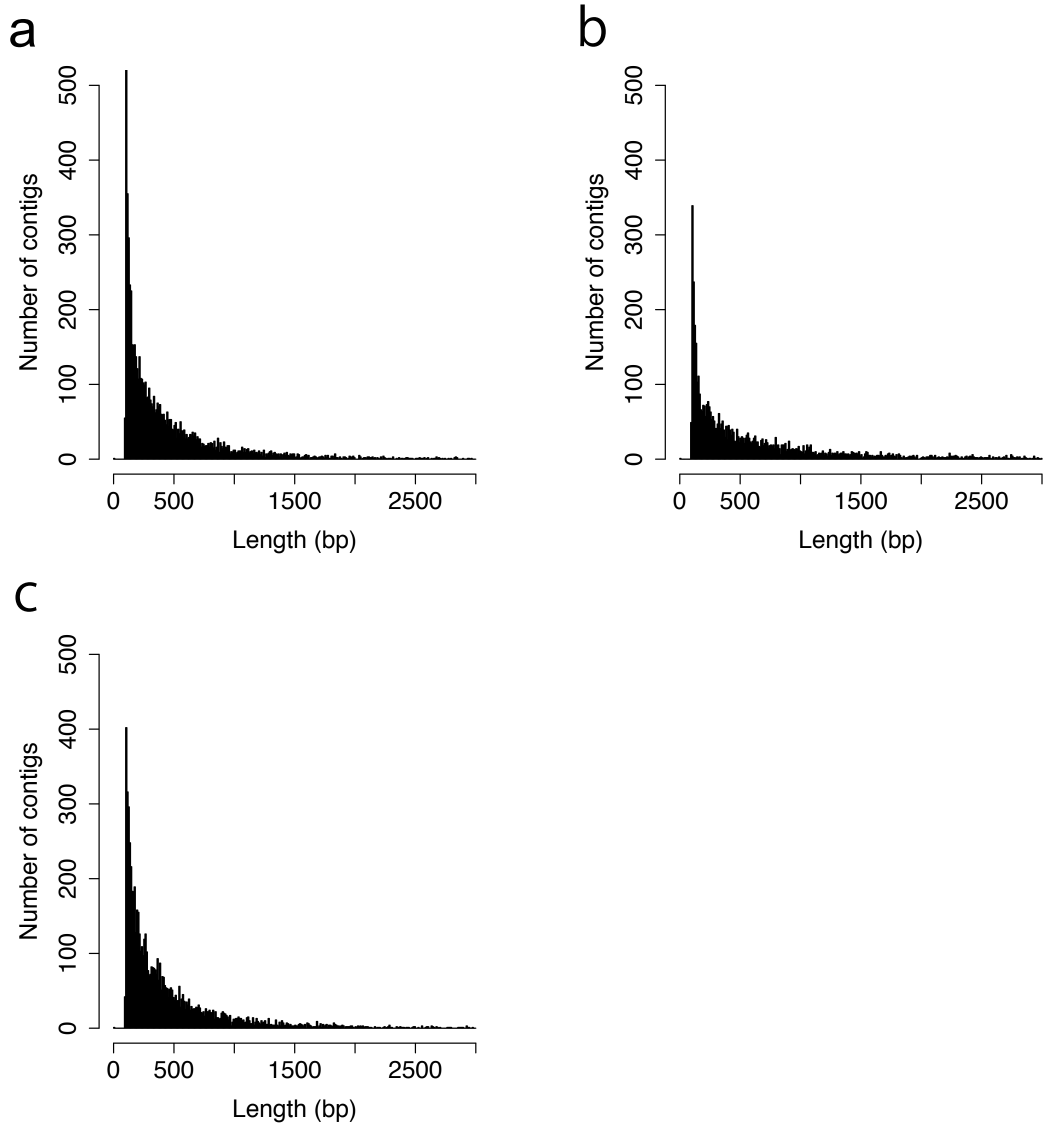


Supplementary Figure 16. **Length distribution of deletions.** (**a**) Length distribution of all deletions. (**b**) Length distribution of deletions with the length ≤ 1kbp.

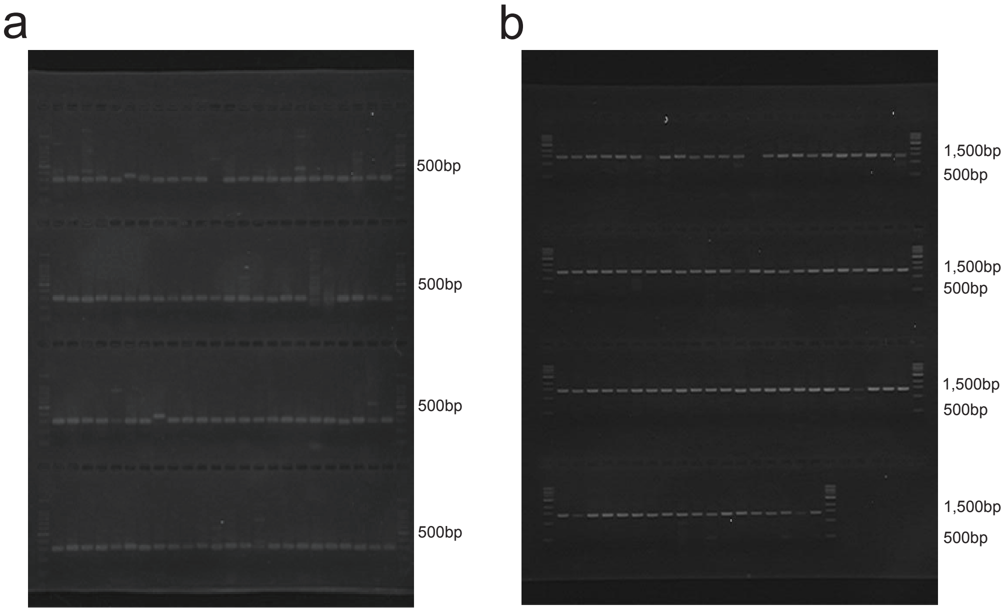


Supplementary Figure 17. **Digital PCR validation of the candidate copy number gain**. (a)-(c) Candidate region on chr1:85,757,716-85,777,862. Average read depth; 77.4. Estimated copy number by digital PCR; 4.3±0.4 (d)-(f) Candidate region on chr2:91,119,738-91,195,844. Average read depth; 161.6. Estimated copy number by digital PCR; 13.9±2.7. (g)-(i) Candidate region on chr7:151,721,324-151,755,980.

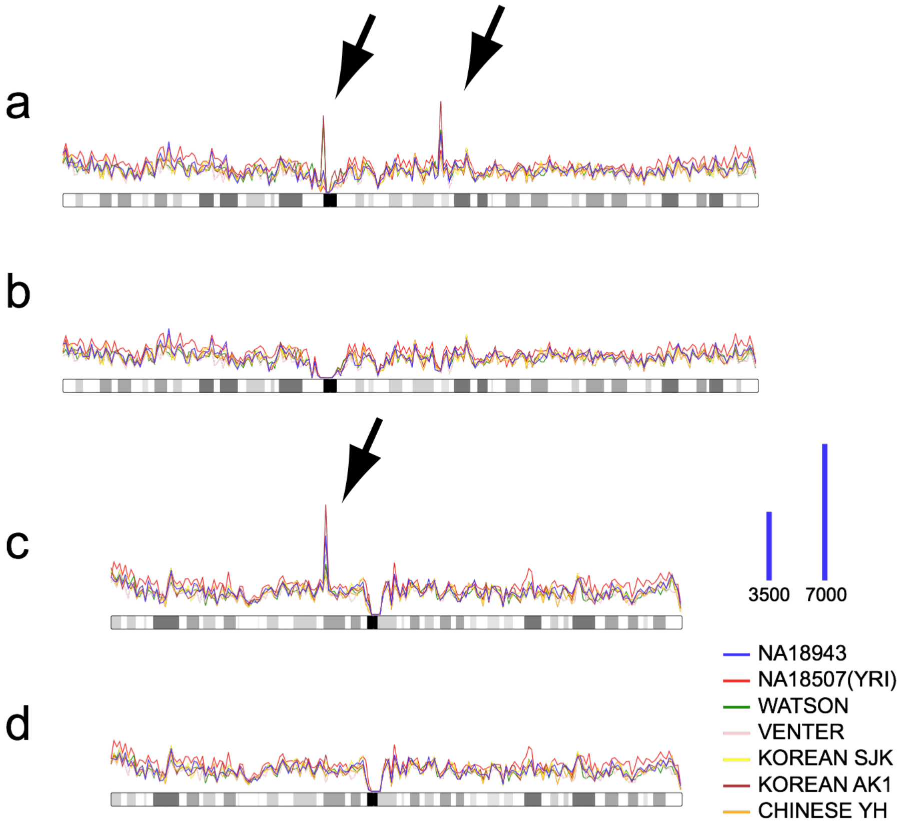
Average read depth; 146.7. Estimated copy number by digital PCR; 19.1±4.4. (j)-(l) Candidate region on chr16:69,711,614-69,730,788. Average read depth; 76.0. Estimated copy number by digital PCR; 4.2±0.5. (a), (d), (g), and (j) Read depth of the target region. Red bar; candidate region. Green bar; repeat masker suggested repeat region. (b), (e), (h), and (k) Comparison of copy number of NA18943 and HapMap samples by TaqMan. Red diamond represents NA18943. (c), (f), (i), and (l) Result of Digital PCR. In the all candidate regions, the TaqMan assay did not detect copy number differences among samples, however, digital PCR assay revealed that the candidates have increased copy number as the average read depth suggested.



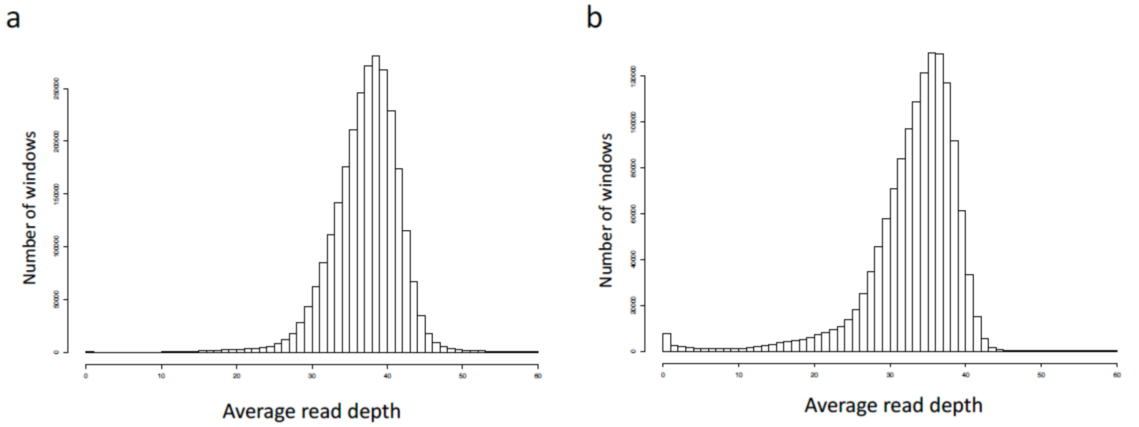
Supplementary Figure 18. **Distribution of the contig length generated by *de novo* assembly software.** (a)ABySS. (b)SOAPdenovo. (c)Velvet. The numbers of contigs with the length between 100bp and 3kbp are shown.



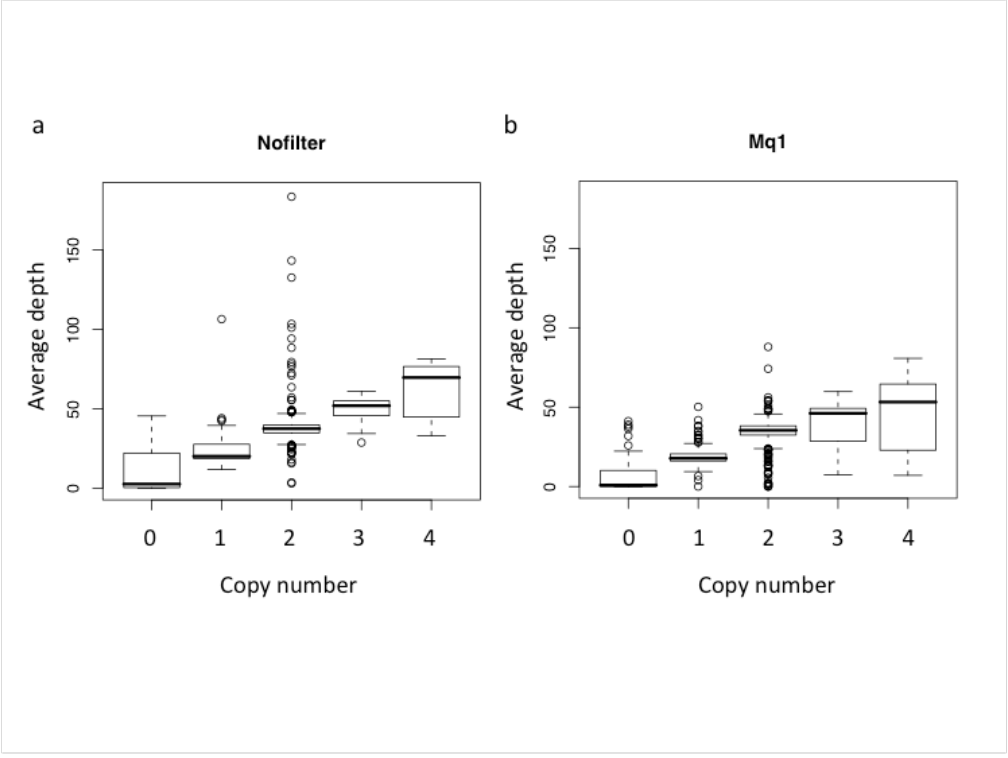
Supplementary Figure 19. **PCR validation of *de novo* contigs.** (a) PCR validations of the contigs with the length between 500bp and 1,000bp. PCR primers were designed to amplify 300-350bp fragments. Marker; 100bp ladder. (b) PCR validations of the contigs with the length more than 1,000bp. PCR primers were designed to amplify 1,300-1,450bp fragments. Marker length; 0.5, 1, 1.5, 2, 4, 5, 6, 8, and 10kbp.



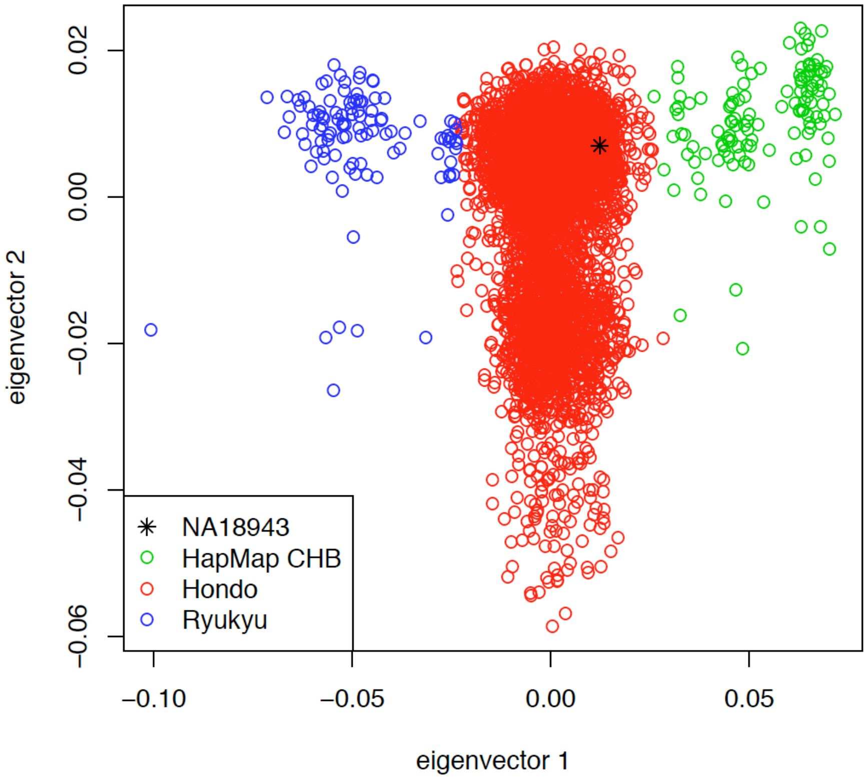
Supplementary Figure 20. **Distribution of the number of SNVs within 1Mbp windows of seven individuals.** (a) Distribution of number of SNVs on chr2. (b) Distribution of number of SNVs on chr2 after excluding SNVs in short repeat regions. Peaks that are indicated by the arrows are absent. (c) Distribution of number of SNVs on chr3. (d) Distribution of number of SNVs on chr3 after excluding SNVs in short repeat regions. Peak that is indicated by the arrow is absent. NA18507 (pink), Watson (green), Venter (pink), SJK (yellow), AK1 (violet), YH (orange) and NA18943 (blue) are shown. Scale bars indicate the number of SNVs.



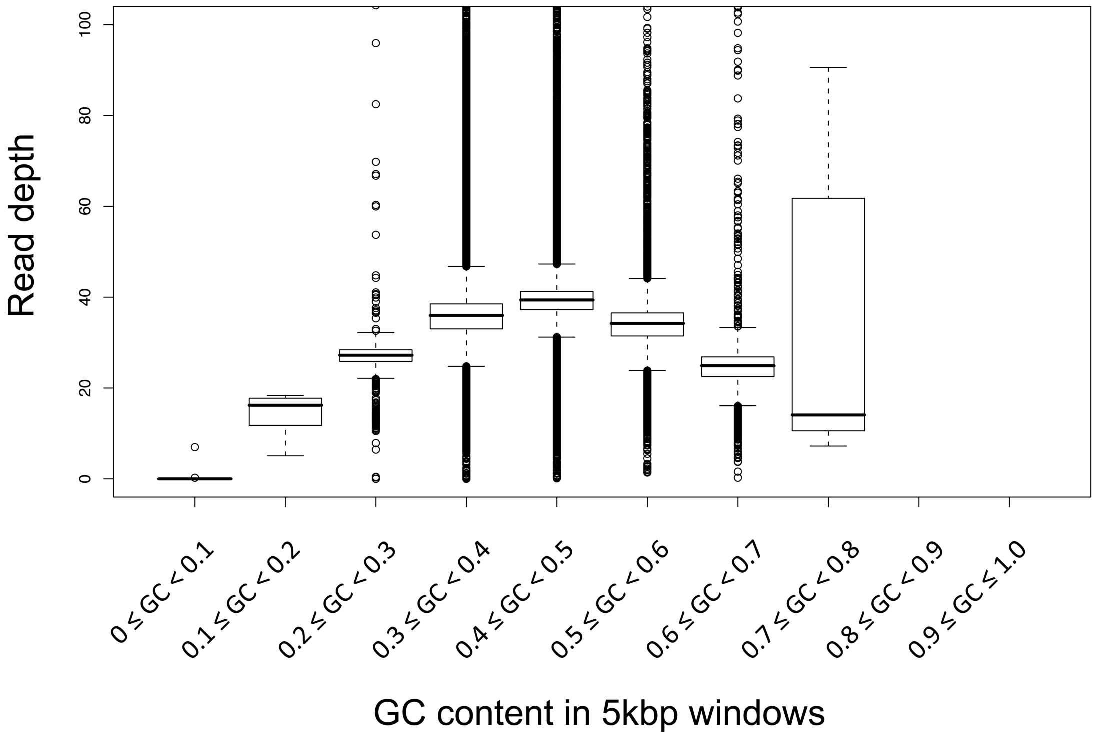
Supplementary Figure 21. **Distributions of average read depth in 5kbp windows.** (**a**) No read selection, (**b**) Read selection by mapping quality ≥ 1.



Supplementary Figure 22. **Copy number in McCarrol *et al* (2008) copy number polymorphism (CNP) region[1] vs. average read depth in each region.** (**a**) Average read depth was calculated using no read selection. Correlation coefficient = 0.54 (*P-value* $< 2.2\times10^{-16}$). (**b**) Average read depth was calculated from reads with a mapping quality ≥ 1. Correlation coefficient = 0.58 (*P-value* $< 2.2\times10^{-16}$).



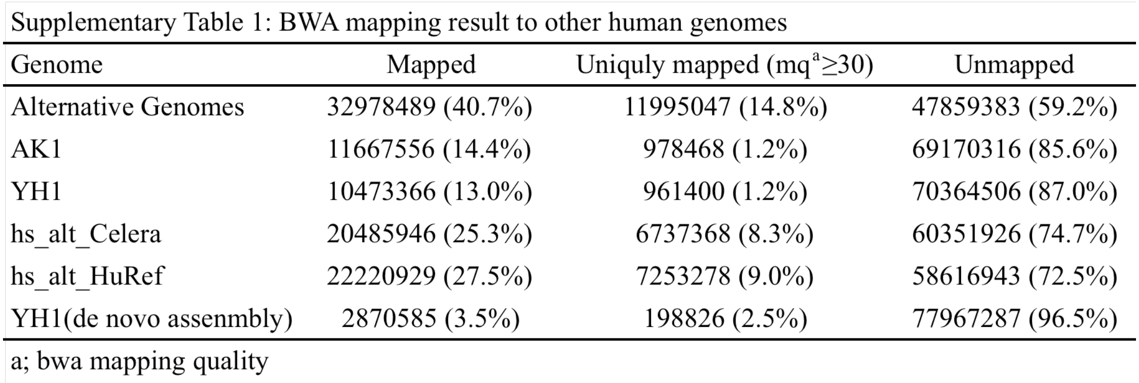
Supplementary Figure 23. **Analysis of the first two principal components.** Genetic relatedness of NA18943 to Japanese and Han Chinese populations were examined.



Supplementary Figure 24. **Relationship between GC content and average read depth.** GC content and read depth were calculated in 5kbp windows.

Reference

1. McCarroll, S.A. et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* **40**, 1166-1174 (2008).



Supplementary Table 1: BWA mapping result to other human genomes

Genome	Mapped	Uniquly mapped (mq[a]≥30)	Unmapped
Alternative Genomes	32978489 (40.7%)	11995047 (14.8%)	47859383 (59.2%)
AK1	11667556 (14.4%)	978468 (1.2%)	69170316 (85.6%)
YH1	10473366 (13.0%)	961400 (1.2%)	70364506 (87.0%)
hs_alt_Celera	20485946 (25.3%)	6737368 (8.3%)	60351926 (74.7%)
hs_alt_HuRef	22220929 (27.5%)	7253278 (9.0%)	58616943 (72.5%)
YH1(de novo assenmbly)	2870585 (3.5%)	198826 (2.5%)	77967287 (96.5%)

a; bwa mapping quality

Supplementary Table 6: Result of breakpoint detection

chr	position1[a]	position2[b]	type	mapping distance (kbp)	blast result
1	166779146	166814813	RR	35.74	Short repeat[c]
2	116694691	116698674	FF	4.06	Partial hit
3	80147837	80147167	RR	0.75	Breakopint
7	6204376	6204366	RF	0.06	Breakopint
7	149331264	153421572	RR	4090.38	Short repeat[c]
9	44333060	44194886	RR	138.22	Multiple hit
10	127187116	127180581	FF	6.61	Breakopint
10	127503337	127503640	RF	0.38	Breakopint
12	11450302	11411283	RF	39.07	Multiple hit
14	18082688	18077867	RF	4.89	No hit

a; Mapping position of a read of an anomalous pair

b; Mapping position of alternative read of an anomalous pair

c; Not sequencd due to short tandem repeat

Supplementary Table 7: Result of *de novo* assembly

Assembler	Kmer	Contigs	n50 (bp)	Maximum Length (bp)	Total Length (bp)	Scafolds	Reads Used
ABySS	25	17603	592	61196	3752402	821	2636407
SOAPdenovo	25	4826	1339	10374	3411209	1735	1276020
Velvet	25	8741	687	25072	3251121	10	2303705

Supplementary Table 8: Result of blast analysis of novel contigs

Category	ABySS		SOAPdenovo		Velvet	
	Number of Contigs	Length (bp)	Number of Contigs	Length (bp)	Number of Contigs	Length (bp)
Total	6535	3121855	4826	3411209	6617	3098957
Hs build36.3	0	0	6	4103	21	9728
Hs GRCh37	1294	579885	816	607696	1333	572149
Hs Alt	4613	2248301	3321	2573851	4515	2207731
Hs Other	353	192890	313	109129	461	195210
Chimpanzee	152	40727	147	46673	126	35622
Orangutan	25	6889	25	7548	27	7318
Rhesus Macaque	32	12190	18	9182	34	9857
Marmoset	5	3190	5	4108	6	2233
Herpesvirus 4	7	29528	86	34801	27	47727
Bos taurus	8	1282	16	3035	7	1107
Remaining	46	6973	73	11083	60	10275

Hs build36.3; Human genome build 36. Hs GRCh37; Human genome GRCh37. Hs Alt; Human genome alternative assemblies.
Hs Other; Other humna genome sequences in the NCBI database. Herpesvirus 4; Human herpesvirus 4.