Supplementary Material

Diversity estimation

We determined diversity of the sequencing library by estimating the number of unique fragments that were sequenced. This was done by using a custom script kindly provided by Dr.R.K.Cheetam (Illumina, Inc.). Briefly, the numbers of unique read pairs (unique start and end positions) are calculated from different sample sizes ranging from 100,000 to 2.5 million read pairs. A regression analysis is then carried out based on the observed values to estimate (in billions) the number of unique DNA fragments that can be obtained by sequencing this library to a very high depth.

Comparison of data from $GAII_x$ and HiSeq 2000 sequencing

Uniformity of sequence data - We compared data quality and coverage uniformity from the two instruments based on several metrics. We also used a "normalized" HiSeq 2000 dataset for comparison, which was created by sampling reads at random from HiSeq 2000 flowcell B to make it equivalent to the $GAII_x$ data in terms of the average mapped depth.

First, we calculated the fraction of the whole genome that was covered by high quality bases. We limited our analysis to reads that remain after candidate molecular duplicates are removed, and only kept alignments with high mapping qualities (for reasons mentioned in the main manuscript). With roughly 118Gb of passing filter data, 97.13% and 97.35% of the whole genome was covered by at least one high-quality base (≥ Q20) in the $GAII_x$ and HiSeq 2000 datasets respectively (Supplementary Table 1 and Supplementary Fig. 2a). At ≥10×, the normalized HiSeq 2000 data covered 92.16% of the genome. In comparison the $GAII_x$ data covered 89.36% of the genome, suggesting slightly more uniform coverage with the HiSeq 2000 data.

To extend the analysis further, we looked at coverage of just the coding exons (defined as 34,068,542 non-redundant coding bases from the UCSC knownGenes

track excluding pseudo-autosomal regions) by high-quality bases. At ≥1×, sequencing using the HiSeq 2000 covered 95.64% of the CDS, 2.23% more of the coding bases than that by $GAII_x$ (Supplementary Table 1). At 10× or greater, the difference is more pronounced; less than 70% of the CDS bases were covered by data from $GAII_x$ while 76.23% was covered by HiSeq 2000-sequenced reads (Supplementary Table 1 and Supplementary Fig. 2b), indicating a better representation of CDS regions by data from the HiSeq 2000 run.

We suspected the reduced coverage in CDS regions was due to a G+C bias in the Illumina platform (as noted in previous reports), since a higher density of genes has been reported in GC-rich isochors (Mouchiroud et al. 1991; Vinogradov 2003; Zoubak et al. 1996). We investigated bias by looking for evidence of skew in representation of different G+C% regions of the genome in the normalized HiSeq 2000 data vs. $GAII_x$. We first calculated G+C% of non-overlapping 100bp/1kb windows from the reference genome. Windows were ordered on G+C% to then determine the expected fraction of the genome in each G+C% bin. Next, mapped reads from each dataset (normalized HiSeq 2000 and $GAII_x$) were categorized into one of the above G+C% bins to determine the observed fraction of reads in each one. Finally, we calculated the ratio of observed to expected fractions for the different G+C% regions of the genome. Indeed, we found that reads generated on the HiSeq 2000 were more uniformly distributed across various G+C% fractions and represented high G+C% regions of the genome better (Supplementary Fig. 3). While the differences were marginal for this library we have noted a more substantial reduction in G+C bias from more recent HiSeq 2000 runs (data not shown). The technical reasons for this difference may be attributed to a) improved fluidics of the cBot on which clusters for the HiSeq 2000 runs were generated compared to the cluster station used for the $GAII_x$ runs, or b) software/chemistry improvements, and not to a systemic bias in $GAII_x$ vs. HiSeq 2000 instruments.

Genotype calling and SNV detection – The slightly increased uniformity in coverage in the normalized HiSeq 2000 data is also evident in the fact that 91.74% of the

genome is callable compared to 88.82% in the $GAII_x$ dataset, based on the filters explained in the main text. Consequently, we detected 3,015,024 and 3,167,895 variants in the $GAII_x$ and normalized HiSeq 2000 datasets respectively (Supplementary Table 3). Concordance with the BeadChip data was equivalent for both datasets with the exception that we were able to call 3.15% more BeadChip positions with the HiSeq 2000 data (Supplementary Table 4).

Recalibration of genotype calls from samtools/bcftools using confidence measure

For a more equivalent evaluation (compared to using MPG) of the effect of score-to-depth confidence metric, we first generated genotype calls on the two 50x genomes without Base Alignment Quality (BAQ) computation for samtools mpileup. We only used bases with qualities Q20 or higher and reads with mapping qualities 30 or higher. We then determined the bcftools score equivalent to MPG=10 by generating a best-fit equation from a plot of bcftools vs MPG scores for ~1.5 million positions that are highly mappable (Supplementary Fig. 5a); this was calculated to be 63. Positions that met or exceeded this score in both genomes were then compared resulting in 17,068 discordant positions from 92.28% of callable sequence (Supplementary Table 2). A plot of coverage vs. quality scores for these discordant positions revealed a lower score-to-depth ratio compared to an equal number of random concordant ones (Supplementary Fig. 5b & 5c)

Since BAQ computation is recommended for samtools/bcftools in order to improve specificity, we repeated the above experiment and observed the number of discordant positions reduce to 3,925 (Supplementary Table 2). Further, recalibrating read mapping qualities for BWA alignments (using '-C 50') reduced this number to 895. The pattern of discordant and concordant positions' score-to-coverage ratios, however, did not change (Supplementary Fig 5d-i), illustrating that this score-to-coverage filter is not specific to MPG and can also be more broadly applied to other genotype callers such as samtools.

Simulation of MPG calls and binomial probabilities

We sought to theoretically determine the probability of correctly calling a heterozygous position with MPG at different depths of coverage. Let *N* be the number of reads that pileup at a known heterozygous position for a diploid genome. Given $X = \{A,C,G,T\}$, let $x,y \in X$ be the two alleles observed at this position. If *i* is the number of times allele *x* is observed and *p*=0.5 is the probability of observing *x* (for a normal genome with no mosaicism involved), then the probability of observing $x_i y_{N-i}$ is $\binom{N}{i} p^i (1-p)^{N-i}$. There are four possible calls that can arise for any combination of *N* and *i*: a) a confident homozygous call (MPG>=10 and confidence>=0.5), b) a weak homozygous call (MPG<10 or confidence<0.5), c) a confident heterozygous call, or d) a weak heterozygous call. Shown below are results from a simulation of genotype calls from MPG for possible values of *i* when *N*=10, along with their corresponding binomial probabilities:

i	MPG call	MPG score	Binom. prob.
0	*yy*	9	0.0009765625
1	*yy*	4	0.0097656250
2	*xy*	1	0.0439453125
3	*xy*	7	0.1171875000
4	*xy*	12	0.2050781250
5	*xy*	18	0.2460937500
6	*xy*	15	0.2050781250
7	*xy*	10	0.1171875000
8	*xy*	4	0.0439453125
9	*xx*	1	0.0097656250
10	*xx*	6	0.0009765625

A confident heterozygous call (c) is made when i=4,5,6 or 7, a weak homozygous call (b) arises when i=0,1,9 or 10 and a weak heterozygous call (d) arises when i=2,3, or 8. For N=10, a confident homozygous call is never made. Based on the outcomes shown above, the probability of not making a confident heterozygous call at 10× with a minimum MPG score of 10 and confidence measure of 0.5 is

$$\sum_{i=\{0,1,2,3,8,9,10\}} \binom{N}{i} p^i (1-p)^{N-i} = 0.226 \text{ or } 22.6\%$$

Supplementary Table 1. Comparison of Q20 coverage for $GAII_x$ vs HiSeq 2000 normalized

Dataset	Average mapped depth	Genome covered (%)		Coding exome covered (%)	
		1×	10×	1×	10×
$GAII_x$ (two flowcells)	34.2	97.13	89.36	93.41	67.42
HiSeq 2000 Normalized	34.2	97.35	92.16	95.64	76.23

Supplementary Table 2. Analysis of samtools/bcftools genotype calls to determine the proportion of genome callable and number of discordant positions in a comparison of identical genomes

samtools/bcftools parameters	bcftools score equivalent to MPG 10	hg18 callable in both genomes (%)	Number of discordant positions
MapQ 30	63	92.923	17,068
MapQ 30 + BAQ	63	92.829	3,925
MaqQ 30 + BAQ + C50	78	93.862	895

Supplementary Table 3. Variant calls from sequencing on $GAII_x$ and HiSeq 2000

Machine	Total variants detected	Autosomal variants detected	Autosomal Heterozygotes	Autosomal Homozygotes	Autosomal Het/Hom Ratio
$GAII_x$	3,015,024	2,942,634	1,855,442	1,087,192	1.71
HiSeq 2000 normalized	3,167,895	3,094,243	1,948,143	1,146,100	1.70
HiSeq 2000 FC-A	3,055,069	2,982,964	1,879,511	1,103,453	1.70
HiSeq 2000 FC-B	3,244,003	3,170,264	1,975,444	1,194,820	1.65
$GAII_x$ + HiSeq 2000 FC-AB	3,502,076	3,425,676	2,110,162	1,315,514	1.60

Supplementary Table 4. Comparison of genotype calls from sequencing and Illumina 1M Duo BeadChip. Coverage and overall concordance are shown for each dataset based on 1,096,530 filtered positions. Discordance is categorized as heterozygote calls by sequencing but not by BeadChip, heterozygote calls by BeadChip but not by sequencing and other (e.g. two different heterozygote calls, or reference and homozygous non-reference, etc.). Also shown is the concordance of positions that were called heterozygous by sequencing. BeadChip positions were filtered based on "hidden SNPs" (see Methods) identified on the largest dataset ($GAII_x$ + HiSeq 2000 A-B).

Machine	BeadChip positions callable (%)	Overall concordance (%)	Overall discordance			Concordance of het. positions from seq. (%)
			Het. call by seq. only (%)	Het. call by array only. (%)	Other (%)	
$GAII_x$	91.72	99.961	0.011	0.013	0.015	99.938
HiSeq 2000 normalized	94.87	99.961	0.011	0.013	0.015	99.937
HiSeq 2000 FC-A	94.10	99.962	0.01	0.014	0.014	99.934
HiSeq 2000 FC-B	96.45	99.960	0.01	0.014	0.016	99.932
$GAII_x$ + HiSeq 2000 FC-AB	99.24	99.956	0.012	0.016	0.016	99.926

Supplementary Table 5. Percentage of genome and CDS callable in datasets generated by sampling read-pairs from the 102× genome (see Table 1, "HiSeq 2000+GAII$_x$"). Each successive dataset has an additional 5× average mapped depth. Column 2 shows reduction in the average mapped depth when MapQ filter (≥30) is applied.

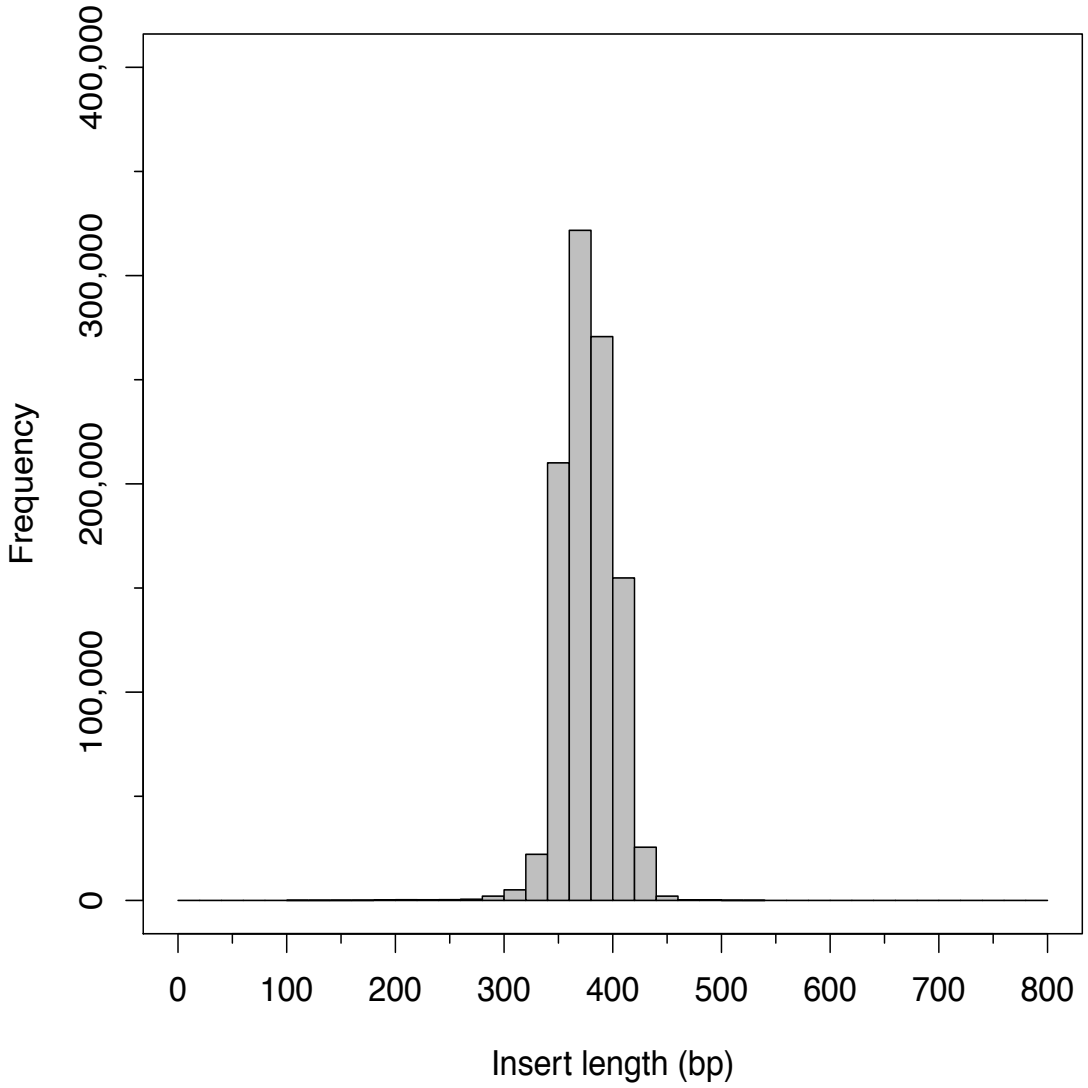
Average mapped depth	Average filtered mapped depth (MapQ ≥ 30)	Genome callable (%)	CDS callable (%)
5×	4.4×	5.04	2.53
10×	8.8×	34.56	20.48
15×	13.3×	66.07	43.50
20×	17.7×	80.23	56.77
25×	22.2×	86.49	64.66
30×	26.6×	89.7	70.00
35×	31.1×	91.54	73.96
40×	35.5×	92.7	77.03
45×	39.9×	93.48	79.43
50×	44.4×	94.03	81.36
55×	48.8×	94.44	82.93
60×	53.3×	94.75	84.23
65×	57.7×	95.0	85.33
70×	62.2×	95.2	86.26
75×	66.6×	95.36	87.06
80×	71×	95.5	87.76
85×	75.4×	95.61	88.37
90×	79.8×	95.71	88.91
95×	84.2×	95.8	89.39
100×	88.6×	95.88	89.81

Supplementary Table 6. Summary of variants detected. Column 3 shows the number of common variant positions between dataset (N-5)× and N×, and columns 4-5 show changes observed in variant calls

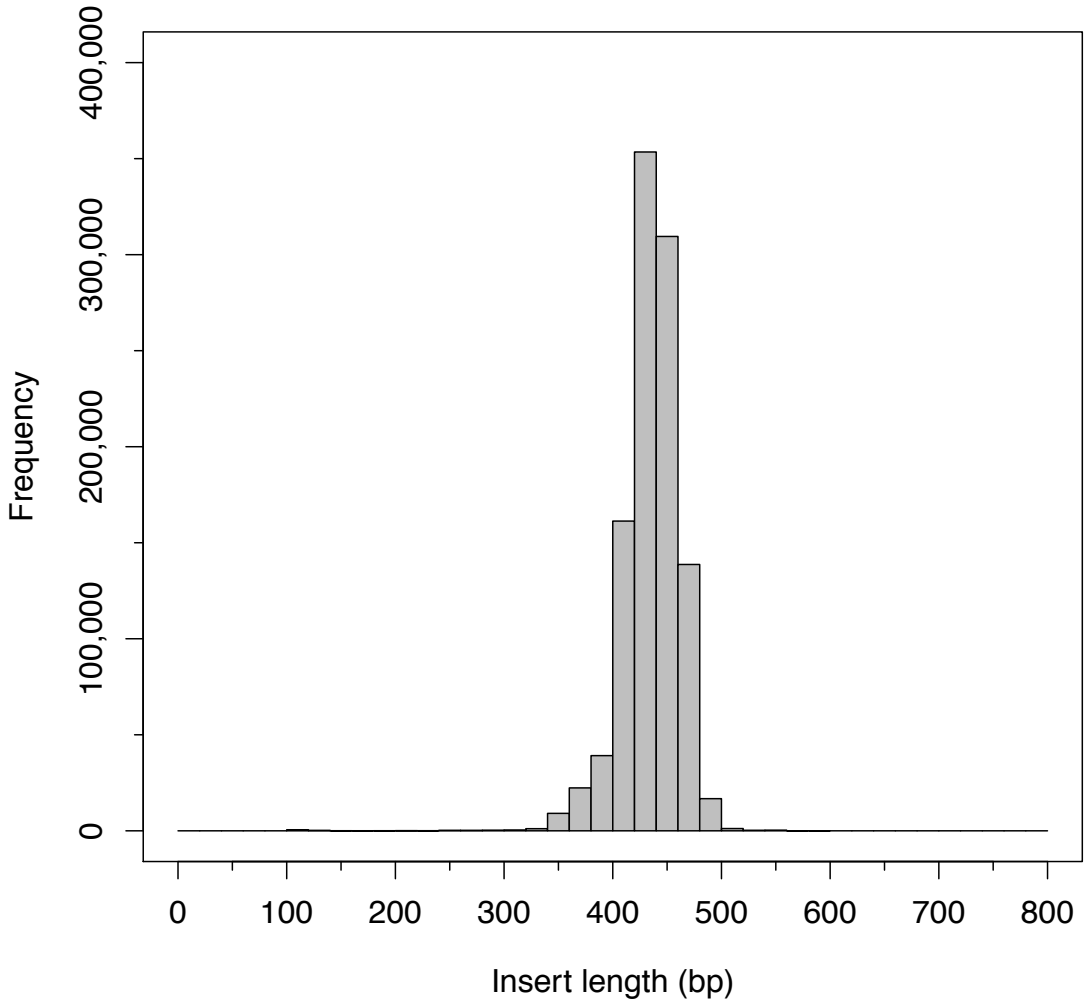
Dataset	Total variant positions	Common variant positions	Non-variant to variant changes	Variant to non-variant changes	Percentage of total variants seen at 100x
5×	177,757	--	--	--	5.08
10×	942,695	167,393	775,302	7,943	26.95
15×	1,847,532	919,832	927,700	22,863	52.81
20×	2,464,862	1,822,382	642,480	25,150	70.45
25×	2,805,840	2,442,105	363,735	22,757	80.2
30×	3,004,520	2,786,282	218,238	19,558	85.88
35×	3,129,807	2,987,541	142,266	16,979	89.46
40×	3,214,476	3,114,713	99,763	15,094	91.88
45×	3,274,782	3,200,782	74,000	13,694	93.61
50×	3,319,872	3,262,322	57,550	12,460	94.89
55×	3,354,921	3,308,392	46,529	11,480	95.9
60×	3,382,920	3,344,144	38,776	10,777	96.7
65×	3,406,410	3,372,680	33,730	10,240	97.37
70×	3,425,862	3,396,634	29,228	9,776	97.92
75×	3,441,865	3,416,482	25,383	9,380	98.38
80×	3,456,415	3,433,023	23,392	8,842	98.8
85×	3,468,665	3,447,495	21,170	8,920	99.15
90×	3,479,868	3,460,206	19,662	8,459	99.47
95×	3,489,687	3,471,564	18,123	8,304	99.75
100×	3,498,512	3,481,558	16,954	8,129	100

Supplementary Figure 1. Distribution of insert lengths for libraries sequenced in this study. Approximately one million mapped read pairs were sampled from one lane each from the individual $GAII_x$ runs. The estimated insert lengths were determined to be 378bp ± 24 s.d. (a) and 436bp ± 26 s.d. (b).

1a.

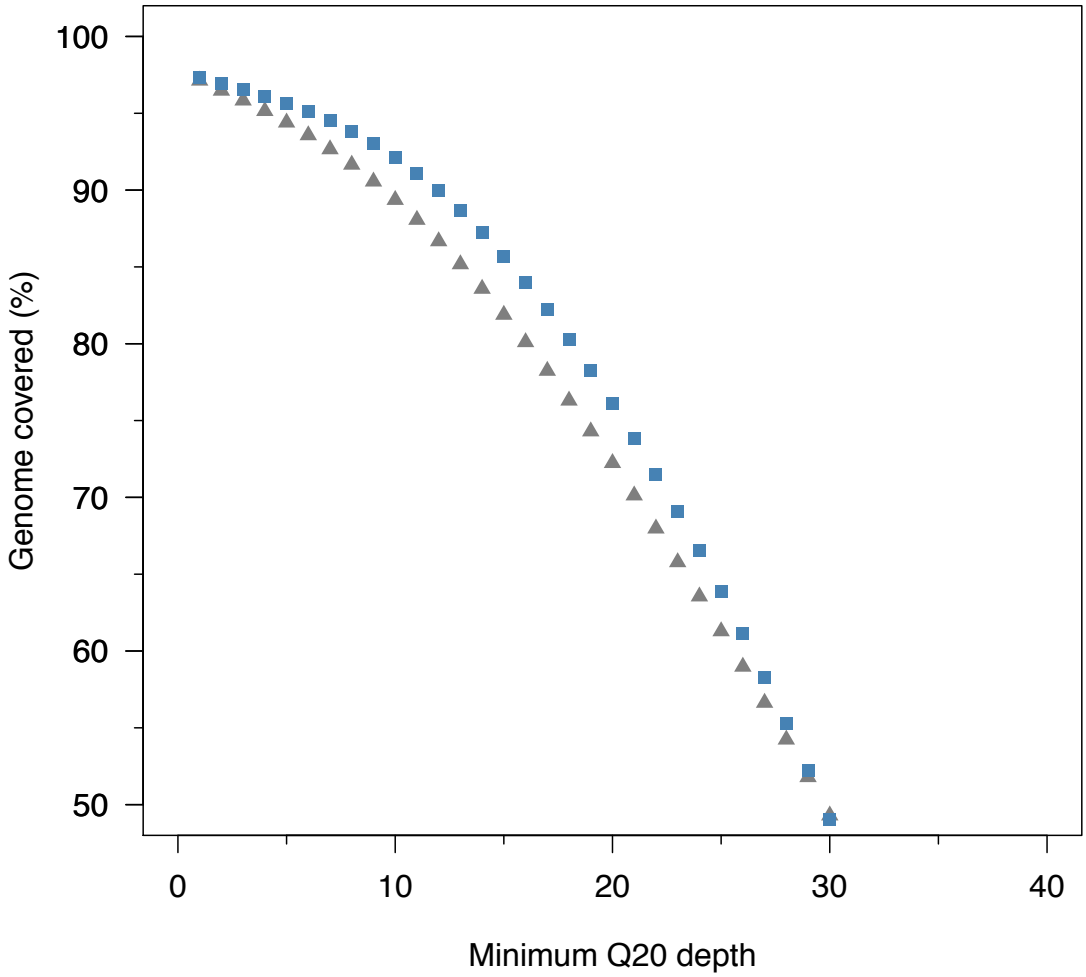


1b.

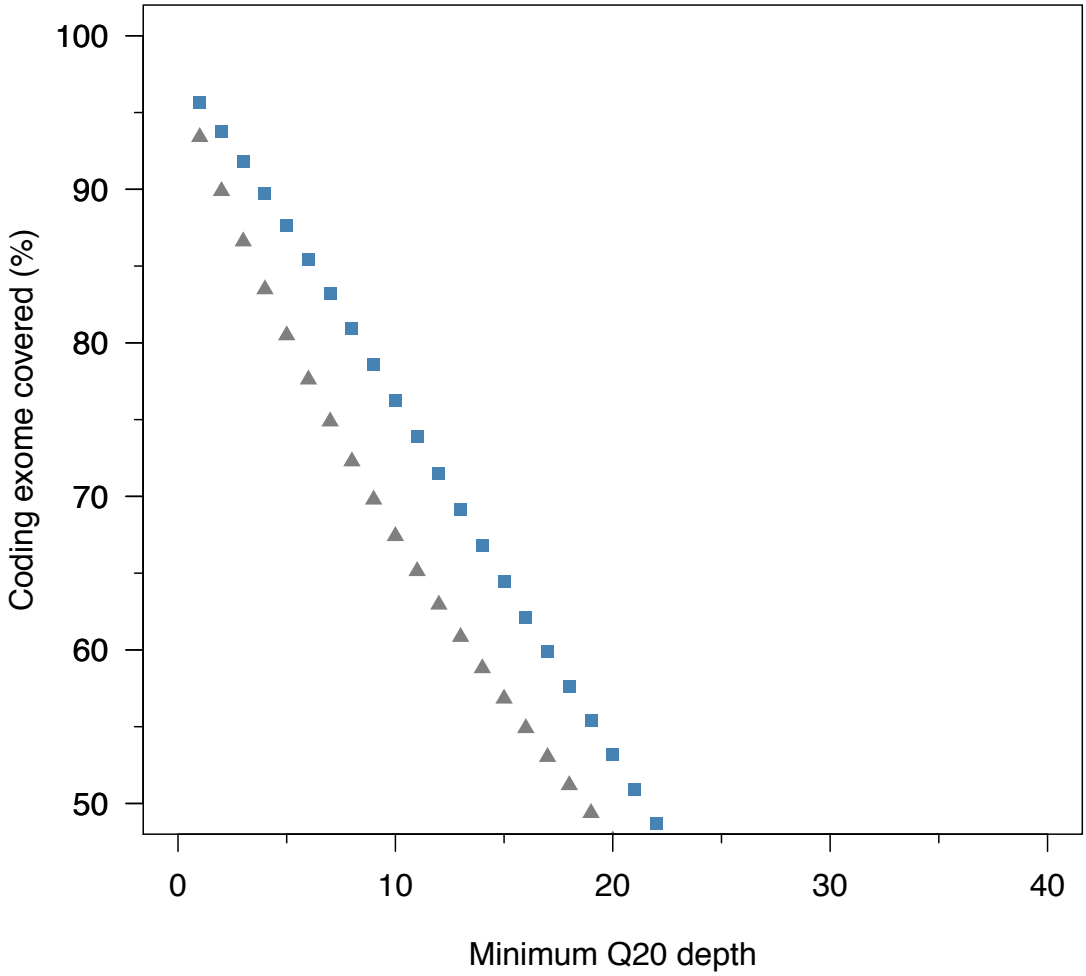

Frequency
0
100,000
200,000
300,000
400,000
0
100
200
300
400
500
600
700
800
Insert length (bp)

Supplementary Figure 2. Comparison of breadth vs. depth of whole genome coverage for GAIIx and HiSeq 2000. The x-axis represents the minimum Q20 depth and the y-axis represents the fraction of (a) genome and (b) coding exome covered at that depth. To calculate percentages, the total size of hg18 build and the total number of non-redundant coding bases from UCSC knownGenes table (2,852,680,119 bp and 34,068,542 bp respectively) were used. Gaps and pseudo-autosomal regions (PAR) were excluded. Values were plotted for a normalized HiSeq flowcell (blue square) whose average mapped depth is equivalent to that of the $GAII_x$ dataset (grey triangle).

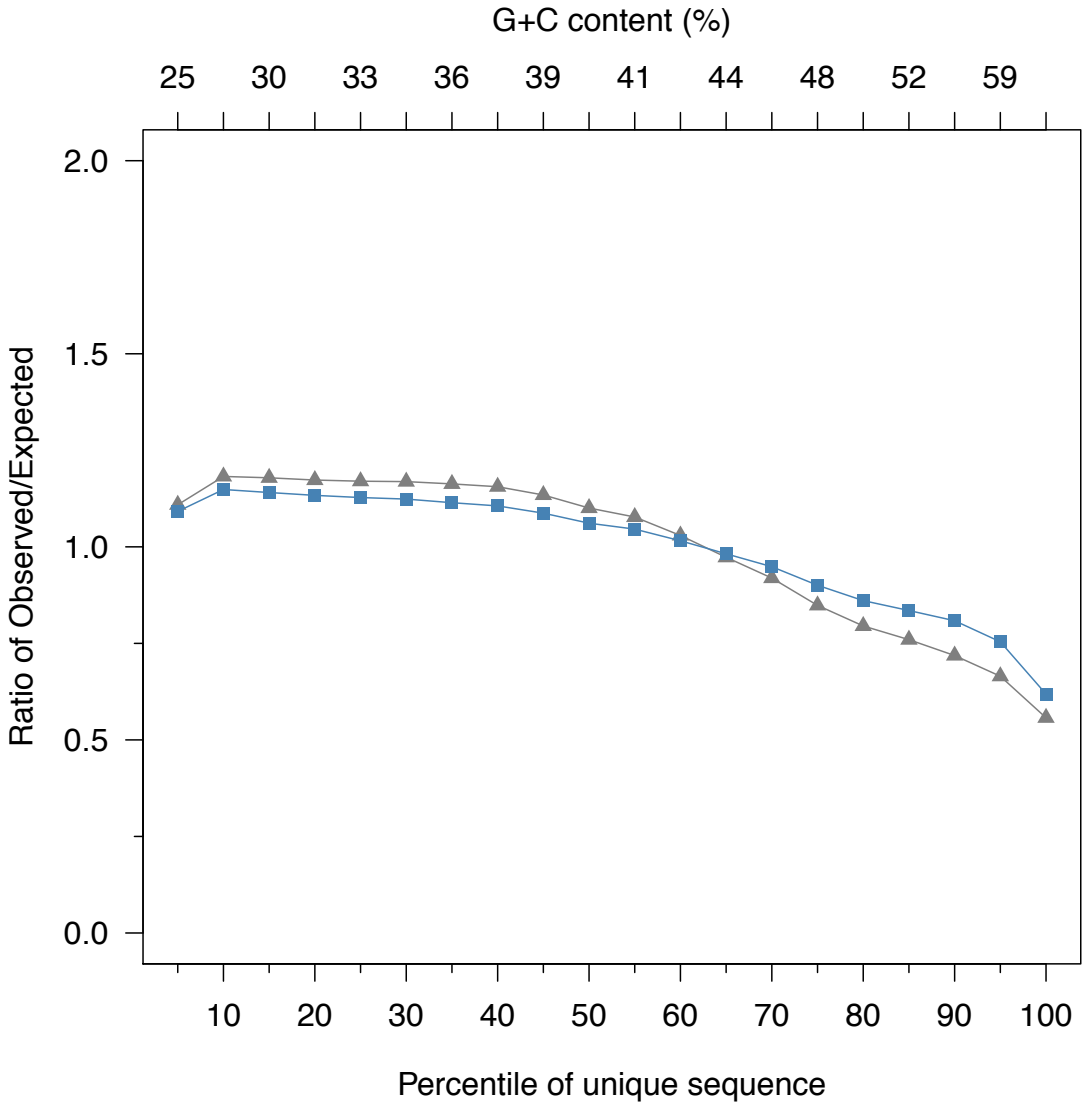
2a.



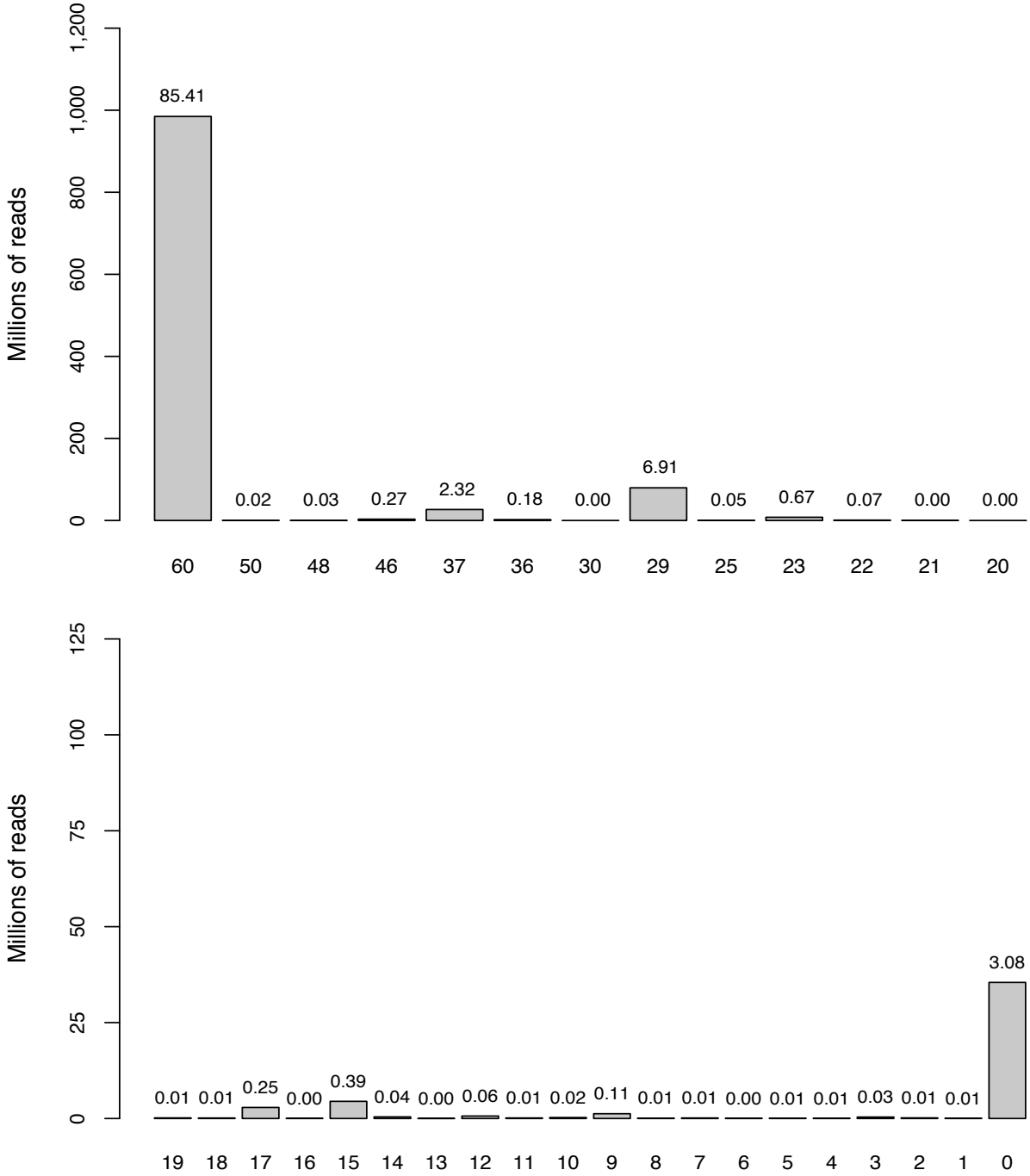
2b.


100
90
80
70
60
50
Coding exome covered (%)
0
10
20
30
40
Minimum Q20 depth

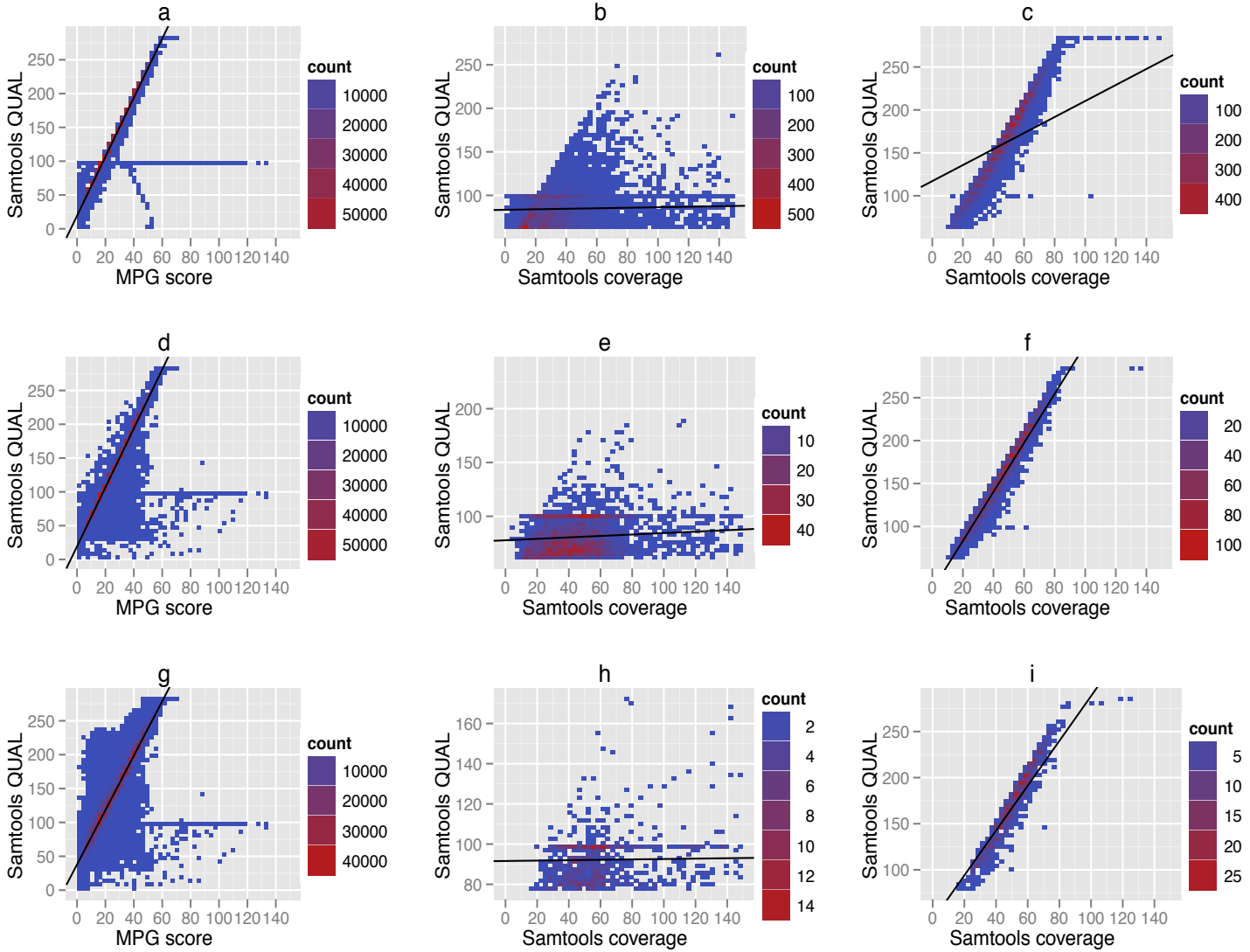
Supplementary Figure 3. Evaluation of data from HiSeq 2000 and $GAII_x$ for GC-bias. X-axis represents twenty different G+C% bins each with an equivalent number of 100bp genomic windows with corresponding G+C% content, and y-axis represents the ratio of the observed to expected fraction of 100bp reads that fall in the different G+C% bins. Both datasets show under-representation of higher G+C% regions of the genome. However, the normalized HiSeq 2000 dataset (blue square) exhibits a more uniform coverage compared the to $GAII_x$ data (grey triangle).



Supplementary Figure 4. Distribution of mapping qualities for reads from one HiSeq 2000 flowcell aligned to the reference genome. X-axis shows mapping qualities and y-axis shows the number of reads in millions. Numbers on each bar represent the percentage of reads aligned with the indicated mapping quality.

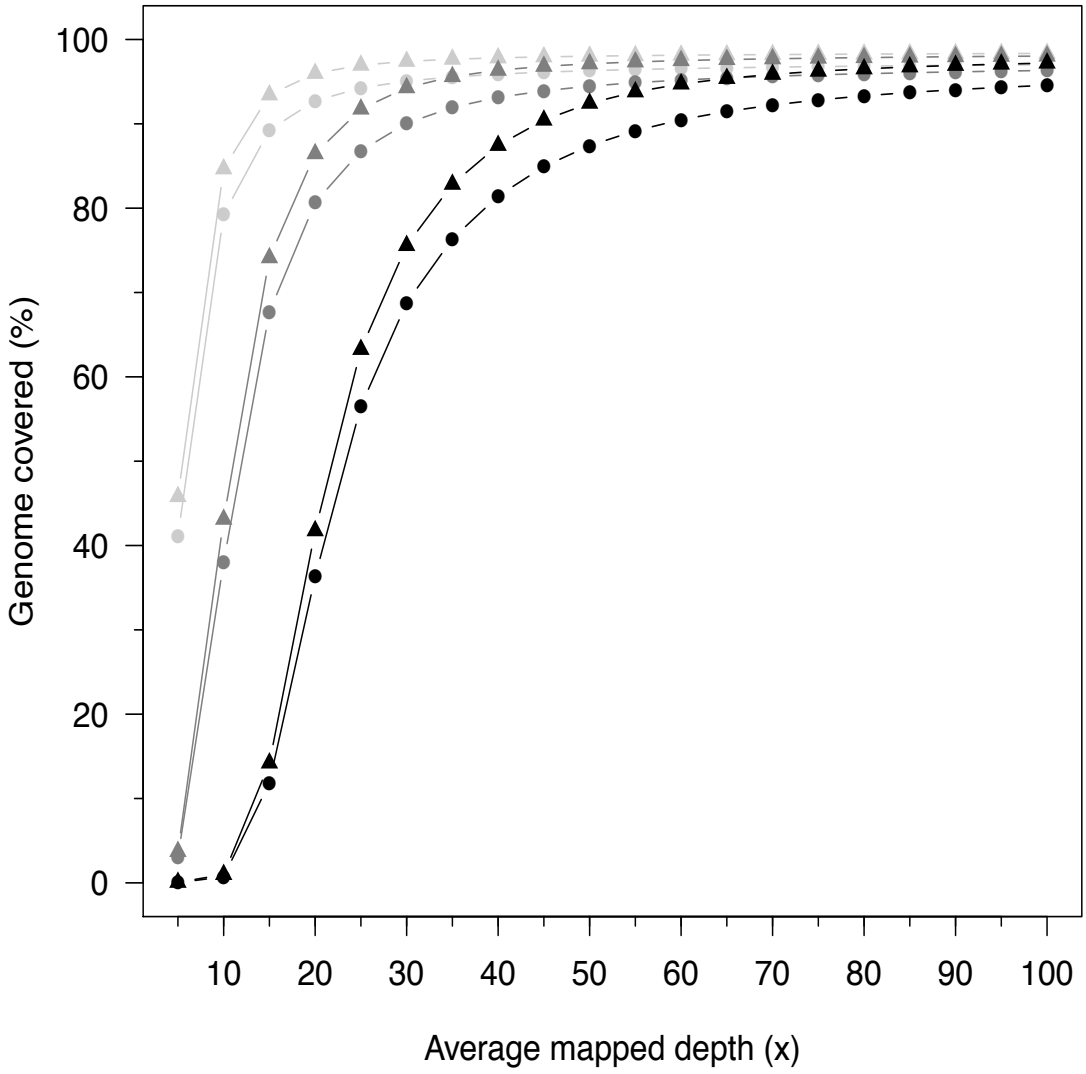


Supplementary Figure 5. Analysis of identical genomes using `samtools/bcftools`. Genotype calls were first made with MapQ>=30 (a-c) followed by the successive incorporation of BAQ (d-f) and MapQ recalibration for `mpileup` using '–C50' (g-i). To determine the score threshold for comparing two 50x genomes, the bcftools score equivalent for MPG=10 was determined by a best-fit procedure (a,d,g); values are shown in Supplementary Table 2. Genotype qualities vs coverage were plotted for discordant positions (b, e, h) and an equal number of concordant positions (c, f, i). Black solid lines indicate best-fit to the data.

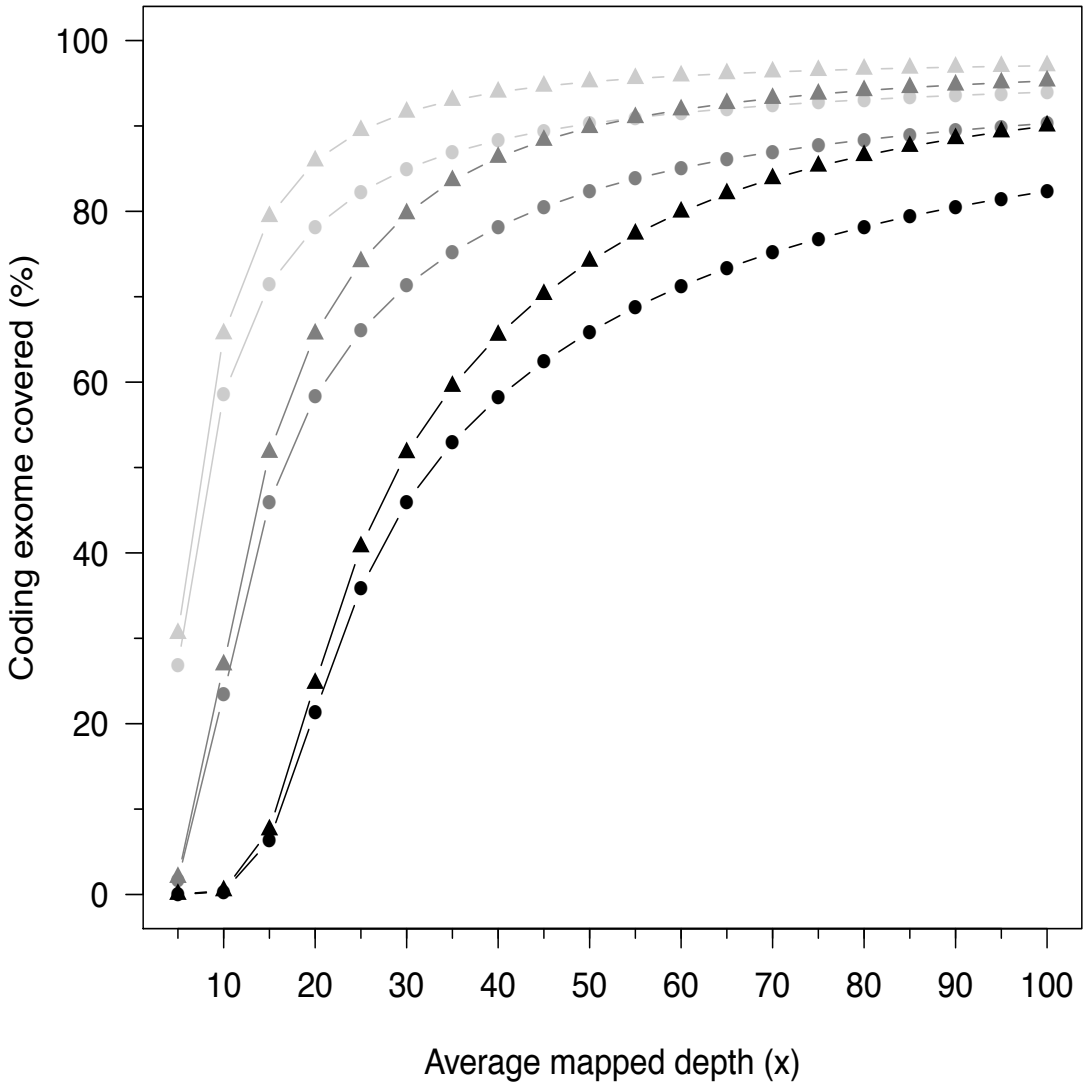


Supplementary Figure 6. Genome and coding exome coverage as a function of average mapped depth. Proportion of genome (a) and coding exome (b) covered by ≥5× (light grey), ≥10× (dark grey) and ≥20× (black) using all uniquely aligned reads (MapQ > 0, filled triangles) or confidently aligned reads (MapQ ≥ 30, filled circles)

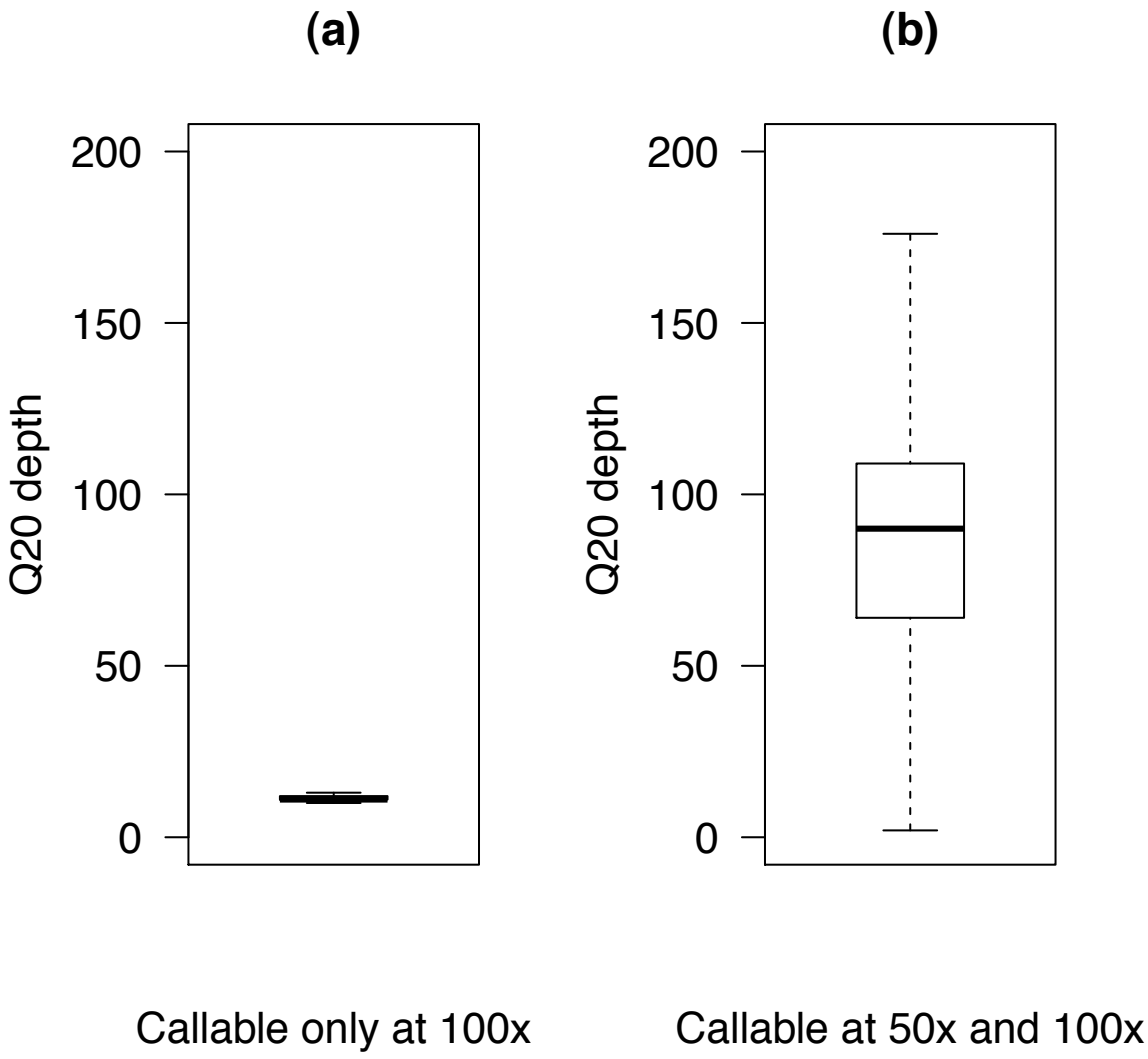
6a.



6b.


Coding exome covered (%)
100
80
60
40
20
0
10
20
30
40
50
60
70
80
90
100
Average mapped depth (x)

Supplementary Figure 7. Distribution of the number of bases that cover positions that are callable only at 100× (a) and positions that callable at 50× and 100× (b). Boxplots reflects high-quality bases only (≥Q20).



GERALD SUMMARIES FOR GAII$_x$ RUNS

Run 1, Read 1

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	4161323	405724 +/- 35655	343343 +/- 28917	465 +/- 28	82.27 +/- 5.46	84.69 +/- 1.84	85.88 +/- 0.18	339.68 +/- 4.88	0.85 +/- 0.22
2	4280459	414463 +/- 22971	353173 +/- 17646	471 +/- 12	80.94 +/- 0.93	85.25 +/- 1.05	85.94 +/- 0.00	335.84 +/- 4.04	0.88 +/- 0.14
3	4272501	415100 +/- 23923	352516 +/- 17926	450 +/- 13	81.02 +/- 0.89	84.97 +/- 1.06	85.94 +/- 0.07	336.03 +/- 2.75	0.88 +/- 0.11
4	4385049	437796 +/- 7133	361802 +/- 4410	426 +/- 20	80.96 +/- 1.20	82.65 +/- 0.78	85.93 +/- 0.02	333.70 +/- 3.60	0.92 +/- 0.13
5	4277677	432384 +/- 21500	352943 +/- 14858	427 +/- 11	81.73 +/- 0.78	81.68 +/- 1.65	85.92 +/- 0.08	331.22 +/- 2.79	0.98 +/- 0.11
6	4267528	433774 +/- 6880	355065 +/- 3934	435 +/- 9	80.77 +/- 1.02	81.87 +/- 1.08	85.96 +/- 0.06	335.05 +/- 3.01	0.88 +/- 0.12
7	4300586	454727 +/- 7793	354833 +/- 4071	450 +/- 7	81.03 +/- 1.28	78.06 +/- 1.66	85.89 +/- 0.19	332.11 +/- 2.53	0.92 +/- 0.11
8	4270283	455052 +/- 8413	352333 +/- 3195	493 +/- 68	77.21 +/- 4.45	77.45 +/- 1.56	85.84 +/- 0.28	326.77 +/- 4.50	0.95 +/- 0.09

Run 1, Read 2

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	4161323	405724 +/- 35655	343343 +/- 28917	330 +/- 18	82.00 +/- 3.55	84.69 +/- 1.84	84.92 +/- 0.27	330.40 +/- 4.91	0.92 +/- 0.18
2	4280459	414463 +/- 22971	353173 +/- 17646	322 +/- 7	80.34 +/- 1.16	85.25 +/- 1.05	84.92 +/- 0.11	325.32 +/- 3.71	1.07 +/- 0.13
3	4272501	415100 +/- 23923	352516 +/- 17926	312 +/- 4	81.15 +/- 1.04	84.97 +/- 1.06	84.97 +/- 0.11	326.19 +/- 2.01	1.03 +/- 0.06
4	4385049	437796 +/- 7133	361802 +/- 4410	287 +/- 16	80.19 +/- 1.13	82.65 +/- 0.78	84.88 +/- 0.14	321.61 +/- 3.12	1.14 +/- 0.08
5	4277677	432384 +/- 21500	352943 +/- 14858	294 +/- 10	80.58 +/- 1.91	81.68 +/- 1.65	84.83 +/- 0.09	319.54 +/- 2.60	1.19 +/- 0.08
6	4267528	433774 +/- 6880	355065 +/- 3934	318 +/- 7	80.39 +/- 1.40	81.87 +/- 1.08	84.86 +/- 0.13	321.36 +/- 4.22	1.13 +/- 0.13
7	4300586	454727 +/- 7793	354833 +/- 4071	318 +/- 9	81.62 +/- 1.21	78.06 +/- 1.66	84.02 +/- 7.74	315.99 +/- 29.17	1.14 +/- 0.12
8	4270283	455052 +/- 8413	352333 +/- 3195	253 +/- 46	79.31 +/- 10.47	77.45 +/- 1.56	83.07 +/- 10.86	303.14 +/- 41.03	1.39 +/- 0.22

Run 2, Read 1

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	6060605	387468 +/- 7189	334470 +/- 12906	449 +/- 38	82.60 +/- 18.50	86.32 +/- 3.02	85.64 +/- 0.32	464.16 +/- 18.46	2.71 +/- 0.62
2	6239769	397213 +/- 21917	344358 +/- 17851	458 +/- 11	80.59 +/- 1.92	86.72 +/- 0.52	85.76 +/- 0.00	431.98 +/- 15.39	3.96 +/- 0.67
3	6212200	395257 +/- 5494	342836 +/- 5754	451 +/- 11	80.65 +/- 1.81	86.74 +/- 0.48	85.54 +/- 0.28	416.94 +/- 18.27	4.70 +/- 0.90
4	6419286	410974 +/- 19965	354265 +/- 15855	436 +/- 36	81.46 +/- 8.07	86.22 +/- 0.52	85.72 +/- 0.13	415.78 +/- 27.58	4.79 +/- 1.36
5	6386435	411906 +/- 20267	352452 +/- 15233	447 +/- 12	80.42 +/- 2.72	85.60 +/- 1.02	85.66 +/- 0.38	405.42 +/- 34.28	5.15 +/- 1.62
6	6512883	423487 +/- 5226	359430 +/- 4260	453 +/- 10	80.44 +/- 3.07	84.88 +/- 0.51	85.72 +/- 0.13	414.25 +/- 20.02	4.62 +/- 0.93
7	6905867	440011 +/- 11663	381118 +/- 7117	487 +/- 12	82.39 +/- 2.07	86.63 +/- 0.80	85.56 +/- 0.10	445.41 +/- 14.12	3.31 +/- 0.59
8	6614173	430158 +/- 11997	365020 +/- 38031	546 +/- 72	78.12 +/- 6.28	84.83 +/- 8.54	85.78 +/- 1.41	444.19 +/- 43.21	3.23 +/- 2.41

Run 2, Read 2

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	6060605	387468 +/- 7189	334470 +/- 12906	257 +/- 12	81.21 +/- 3.02	86.32 +/- 3.02	84.43 +/- 0.33	425.79 +/- 19.90	4.23 +/- 0.67
2	6239769	397213 +/- 21917	344358 +/- 17851	252 +/- 10	79.67 +/- 2.48	86.72 +/- 0.52	84.46 +/- 0.17	397.18 +/- 14.60	5.87 +/- 0.78
3	6212200	395257 +/- 5494	342836 +/- 5754	253 +/- 21	78.84 +/- 2.78	86.74 +/- 0.48	83.77 +/- 1.90	377.27 +/- 18.73	7.00 +/- 1.08
4	6419286	410974 +/- 19965	354265 +/- 15855	236 +/- 22	79.13 +/- 5.94	86.22 +/- 0.52	84.36 +/- 0.22	376.78 +/- 21.39	7.40 +/- 1.49
5	6386435	411906 +/- 20267	352452 +/- 15233	231 +/- 18	78.65 +/- 3.14	85.60 +/- 1.02	83.45 +/- 7.71	363.30 +/- 54.27	7.78 +/- 2.47
6	6512883	423487 +/- 5226	359430 +/- 4260	252 +/- 14	78.91 +/- 2.55	84.88 +/- 0.51	84.31 +/- 0.56	382.33 +/- 15.47	6.64 +/- 1.01
7	6905867	440011 +/- 11663	381118 +/- 7117	302 +/- 12	80.15 +/- 2.50	86.63 +/- 0.80	84.86 +/- 0.11	424.44 +/- 11.10	4.47 +/- 0.58
8	6614173	430158 +/- 11997	365020 +/- 38031	326 +/- 37	79.51 +/- 2.71	84.83 +/- 8.54	85.25 +/- 0.21	442.22 +/- 13.41	3.59 +/- 0.58

GERALD SUMMARIES FOR HiSeq 2000 Run

Flowcell 1, Read 1

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	7465352	2722965 +/- 334884	2309824 +/- 242000	789 +/- 41	84.67 +/- 4.37	85.09 +/- 3.59	83.75 +/- 1.02	318.05 +/- 13.79	1.23 +/- 0.56
2	7880521	2829797 +/- 260425	2438280 +/- 203581	789 +/- 29	85.90 +/- 1.97	86.23 +/- 1.02	84.51 +/- 0.00	333.46 +/- 3.56	0.73 +/- 0.08
3	7873545	2825209 +/- 252960	2436121 +/- 196000	785 +/- 25	85.49 +/- 1.56	86.30 +/- 1.01	84.21 +/- 0.00	330.50 +/- 2.25	0.73 +/- 0.09
4	7932502	2847318 +/- 253022	2454363 +/- 194962	768 +/- 26	85.93 +/- 1.05	86.27 +/- 1.04	84.22 +/- 0.00	330.61 +/- 1.05	0.73 +/- 0.05
5	7955647	2859571 +/- 256167	2461524 +/- 199968	803 +/- 31	85.14 +/- 0.67	86.15 +/- 0.97	84.12 +/- 0.00	330.34 +/- 0.98	0.71 +/- 0.05
6	7969991	2859311 +/- 285743	2465962 +/- 221950	794 +/- 33	85.22 +/- 0.62	86.33 +/- 1.05	84.03 +/- 0.00	329.86 +/- 1.08	0.72 +/- 0.05
7	8051162	2905596 +/- 291568	2491077 +/- 221965	806 +/- 35	85.61 +/- 0.55	85.83 +/- 1.18	84.21 +/- 0.00	330.50 +/- 0.92	0.72 +/- 0.05
8	6669846	2340970 +/- 246412	2063690 +/- 199831	774 +/- 36	86.16 +/- 0.68	88.24 +/- 0.95	84.29 +/- 0.09	332.21 +/- 1.00	0.71 +/- 0.05

Flowcell 1, Read 2

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	7465352	2722965 +/- 334884	2309824 +/- 242000	637 +/- 28	85.60 +/- 1.74	85.09 +/- 3.59	81.89 +/- 4.15	288.39 +/- 28.56	2.68 +/- 1.14
2	7880521	2829797 +/- 260425	2438280 +/- 203581	639 +/- 24	87.25 +/- 1.19	86.23 +/- 1.02	83.59 +/- 0.30	311.79 +/- 5.41	1.74 +/- 0.18
3	7873545	2825209 +/- 252960	2436121 +/- 196000	631 +/- 19	88.00 +/- 0.98	86.30 +/- 1.01	83.08 +/- 0.24	309.02 +/- 5.74	1.77 +/- 0.18
4	7932502	2847318 +/- 253022	2454363 +/- 194962	618 +/- 22	87.63 +/- 0.80	86.27 +/- 1.04	83.08 +/- 0.08	309.26 +/- 5.16	1.79 +/- 0.17
5	7955647	2859571 +/- 256167	2461524 +/- 199968	630 +/- 19	88.24 +/- 0.98	86.15 +/- 0.97	83.10 +/- 0.15	309.58 +/- 4.93	1.78 +/- 0.14
6	7969991	2859311 +/- 285743	2465962 +/- 221950	633 +/- 24	88.10 +/- 0.77	86.33 +/- 1.05	82.97 +/- 0.00	308.76 +/- 5.36	1.78 +/- 0.17
7	8051162	2905596 +/ 291568	2491077 +/- 221965	637 +/- 27	88.28 +/- 0.91	85.83 +/- 1.18	83.21 +/- 0.17	306.63 +/- 5.57	1.84 +/- 0.20
8	6669846	2340970 +/- 246412	2063690 +/- 199831	619 +/- 29	88.56 +/- 0.72	88.24 +/- 0.95	83.28 +/- 0.34	307.80 +/- 3.57	1.78 +/- 0.17

Flowcell 2, Read 1

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	8503919	3156095 +/- 309726	2631163 +/- 211107	1046 +/- 32	85.39 +/- 1.24	83.52 +/- _1.71	84.49 +/- 0.19	338.43 +/- 0.53	0.63 +/- 0.02
2	9162450	3492224 +/- 343387	2834916 +/- 213132	989 +/- 33	84.49 +/- 1.27	81.36 +/- 2.01	84.52 +/- 0.16	336.30 +/- 1.07	0.69 +/- 0.06
3	9199376	3500787 +/- 353401	2846341 +/- 219297	995 +/- 28	85.90 +/- 2.75	81.50 +/- 2.05	84.59 +/- 0.07	338.19 +/- 0.83	0.64 +/- 0.01
4	9211378	3496419 +/- 351885	2850055 +/- 223516	1002 +/- 30	85.37 +/- 1.90	81.70 +/- 2.03	84.42 +/- 0.17	330.85 +/- 5.70	0.68 +/- 0.05
5	9226257	3506044 +/- 387860	2854659 +/ 247092	1087 +/- 44	83.68 +/- 1.30	81.64 +/- 2.15	84.65 +/- 0.10	337.83 +/- 0.65	0.64 +/- 0.03
6	9208235	3498983 +/- 338971	2849082 +/- 211385	1039 +/- 36	83.84 +/- 1.21	81.61 +/- 2.08	84.95 +/- 0.09	338.88 +/- 1.41	0.66 +/- 0.07
7	9100304	3451749 +/- 373225	2815688 +/- 235775	1041 +/- 31	84.20 +/- 1.10	81.79 +/- 2.29	84.44 +/- 0.00	337.11 +/- 0.65	0.63 +/- 0.00
8	9075332	3449501 +/- 340891	2807961 +/- 211607	1025 +/- 35	84.28 +/- 0.99	81.59 +/- 2.17	84.47 +/- 0.12	336.96 +/- 0.61	0.65 +/- 0.02

Flowcell 2, Read 2

Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	8503919	3156095 +/- 309726	2631163 +/- 211107	756 +/- 34	87.00 +/- 0.92	83.52 +/- 1.71	83.06 +/- 0.69	298.60 +/- 24.96	1.53 +/- 0.46
2	9162450	3492224 +/- 343387	2834916 +/- 213132	723 +/- 30	86.65 +/- 1.04	81.36 +/- 2.01	83.31 +/- 0.11	312.61 +/- 1.91	1.66 +/- 0.14
3	9199376	3500787 +/- 353401	2846341 +/- 219297	743 +/- 29	86.27 +/- 0.74	81.50 +/- 2.05	83.32 +/- 0.39	308.32 +/- 8.41	1.69 +/- 0.35
4	9211378	3496419 +/- 351885	2850055 +/- 223516	750 +/- 32	85.94 +/- 0.75	81.70 +/- 2.03	83.17 +/- 0.20	313.37 +/- 1.06	1.56 +/- 0.06
5	9226257	3506044 +/- 387860	2854659 +/- 247092	798 +/- 42	85.97 +/- 0.47	81.64 +/- 2.15	83.58 +/- 0.13	316.36 +/- 2.00	1.46 +/- 0.08
6	9208235	3498983 +/- 338971	2849082 +/- 211385	760 +/- 31	85.85 +/- 0.61	81.61 +/- 2.08	83.73 +/- 0.21	310.22 +/- 7.02	1.55 +/- 0.07
7	9100304	3451749 +/- 373225	2815688 +/- 235775	759 +/- 28	86.05 +/- 0.47	81.79 +/- 2.29	83.22 +/- 0.11	313.84 +/- 2.46	1.55 +/- 0.09
8	9075332	3449501 +/- 340891	2807961 +/- 211607	740 +/- 33	86.28 +/- 0.71	81.59 +/- 2.17	83.20 +/- 0.17	312.92 +/- 2.19	1.58 +/- 0.10