Cells (pg DNA)	Effective coverage per allele	Usable wells	% bases read	Shared wells
10 (66)	3x	9.5	95	7.7
15 (99)	2x	14	86	9.3
20 (132)*	1.5x	18	78	9.9
30 (198)**	1x	26	63	9.3
40 (264)	0.75x	32	53	8.1
60 (396)	0.5x	44	39	6.0

Supplementary Table 1. Optimal number of cells for a predefined number of reads. Usable wells are calcualted by taking into account wells having more than one fragment with the same haplotype or mix of both haplotypes. % bases read in each molecule is calcualted using binomial distribution on effective coverage; Shared wells = (usable wells) x (%bases read)2 X 0.9. 0.9 is the fraction of two loci separated by 10kb that are witihin the same 100kb fragment. *Optimal number of cells to use if starting from whole cells. **Due to the stochastic separation of chromosomal fragements after DNA isolation 10 more cells worth of DNA are prefered when starting from isolated DNA. All values are given in double stranded DNA, but for current LFR all DNA is denatured to single stranded prior to aliquoting. For this reason half the listed optimal amount of DNA or cells were used in LFR. All of the above calculations are based on an 80x genome coverage, 60x effective genome coverage library.

Item	Cost
MDA	14.50
CoRE fragmenting	35.40
Barcode Adapter Ligation	32.40
Bead purification after pooling	6.90
Plates, pipette tips, plate seals, etc.	18.80
Total	**108.00**

Supplementary Table 2. LFR reagent cost. Cost per each step was based upon manufacturer list prices. All prices are in US dollars. A complete breakdown of reagents used in LFR can be found in the Supplementary Methods.

Library	NA19240 10 Cell	NA19240 Replicate 1	NA19240 Replicate 2	NA19240, Pipeline 2.0	NA19240 High Coverage	NA12877 Replicate 1	NA12877 Replicate 2	NA12885	NA12886	NA12891	NA12892	NA20431 High Coverage	NA19240 STD	NA12877 STD	NA12878 STD	NA12885 STD	NA12886 STD	NA12891 STD	NA12892 STD	NA20431 STD
dbSNP build	130	130	130	130	132	130	130	130	130	130	130	132	130	130	130	130	130	130	130	130
NCBI genome build	36.3	36.3	36.3	36.3	37.2	36.3	36.3	36.3	36.3	36.3	36.3	37.2	36.3	36.3	36.3	36.3	36.3	36.3	36.3	36.3
Gender	FEMALE	FEMALE	FEMALE	FEMALE	FEMALE	MALE	MALE	FEMALE	MALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	FEMALE	MALE	MALE	FEMALE	MALE
Gross mapping yield (Gb)	307.64	237.42	312.74	237.45	509	258.1	238.39	272.39	293.37	279.83	284.52	513.63	176.45	217.89	216.32	220.86	216	245.92	213.3	192.28
100k normalized coverage variability	0.2059	0.1573	0.1742	0.161	0.159	0.1305	0.1477	0.1567	0.1507	0.1553	0.1149	0.2158	0.0153	0.0123	0.0119	0.015	0.0137	0.0183	0.0131	N/A
Mate distribution mean	326	263	302	263	262	241	250	261	262	298	266	235	301	333	331	311	307	295	339	N/A
Mate distribution range (95% "CI")	(160, 617)	(105, 580)	(145, 598)	(105, 580)	(105, 579)	(110, 530)	(108, 536)	(106, 569)	(113, 558)	(123, 615)	(108, 575)	(94, 530)	(185, 477)	(223, 492)	(219, 488)	(210, 460)	(211, 447)	(175, 444)	(224, 478)	N/A
SNP Transitions/transversions	2.12	2.10	2.13	2.09	2.15	2.13	2.13	2.11	2.13	2.09	2.12	2.14	2.14	2.13	2.15	2.14	2.15	2.13	2.13	2.14
SNP het/hom ratio	1.60	1.67	1.68	1.931	1.95	1.41	1.46	1.44	1.49	1.43	1.55	1.49	1.94	1.56	1.53	1.56	1.60	1.55	1.56	1.56
INS het/hom ratio	1.71	1.47	2.50	1.79	1.95	1.36	1.40	1.47	1.57	1.58	1.49	1.84	1.58	1.38	1.28	1.35	1.39	1.36	1.38	1.35
DEL het/hom ratio	1.99	2.20	2.18	2.435	2.40	1.65	1.78	1.69	1.75	1.78	1.80	1.83	2.31	1.68	1.64	1.66	1.73	1.61	1.64	1.66
SUB het/hom ratio	2.07	2.40	2.24	2.797	2.71	1.78	1.96	2.08	2.05	2.48	2.04	2.28	2.21	1.76	1.69	1.76	1.82	1.71	1.69	1.75
SNP total count	3,515,822	3,391,102	3,636,702	3,720,905	3,751,078	2,891,061	2,876,795	3,016,298	3,104,007	3,003,110	3,130,825	3,041,456	4,053,968	3,347,317	3,237,673	3,341,870	3,301,954	3,370,841	3,335,390	3,261,375
Het SNPs	2,162,154	2,120,945	2,281,748	2,269,965	2,410,575	1,693,868	1,705,994	1,780,295	1,859,283	1,767,838	1,900,711	1,721,862	2,676,177	2,039,517	1,960,235	2,034,951	2,029,760	2,049,456	2,032,007	1,988,716
INS total count	209,174	155,395	238,528	188,553	207,387	135,578	128,991	151,898	161,797	158,847	160,985	185,009	212,648	193,143	166,295	183,735	177,321	205,447	199,403	194,899
DEL total count	206,102	168,752	199,706	202,476	212,367	142,522	137,948	157,602	165,511	161,930	170,391	181,820	236,038	211,211	182,699	201,660	194,632	224,235	217,276	203,206
SUB total count	73,061	67,405	75,372	77,832	79,035	55,438	55,811	63,515	66,387	69,891	65,769	67,642	83,374	73,442	66,170	71,740	70,641	75,698	72,121	67,942
SNP novel rate	0.043	0.051	0.050	0.085	0.030	0.069	0.071	0.057	0.053	0.055	0.034	0.139	0.045	0.067	0.018	0.044	0.043	0.024	0.024	0.066
INS novel rate	0.349	0.303	0.457	0.359	0.338	0.238	0.231	0.263	0.275	0.306	0.243	0.713	0.291	0.215	0.190	0.207	0.204	0.214	0.208	0.218
DEL novel rate	0.326	0.352	0.337	0.382	0.334	0.250	0.251	0.257	0.254	0.296	0.252	0.262	0.340	0.240	0.226	0.234	0.233	0.235	0.233	0.236
SUB novel rate	0.434	0.452	0.448	0.507	0.397	0.346	0.355	0.389	0.383	0.456	0.364	0.408	0.413	0.351	0.302	0.335	0.333	0.335	0.319	0.339
Fully called genome fraction	0.905	0.885	0.913	0.899	0.936	0.910	0.898	0.910	0.931	0.913	0.923	0.938	0.961	0.967	0.952	0.962	0.961	0.974	0.965	0.958
Partially called genome fraction	0.016	0.022	0.015	0.001	0.000	0.017	0.020	0.016	0.013	0.016	0.013	0.000	0.007	0.005	0.010	0.007	0.007	0.003	0.006	0.007
No-called genome fraction	0.078	0.093	0.072	0.1	0.064	0.073	0.082	0.074	0.056	0.071	0.064	0.062	0.032	0.028	0.038	0.031	0.032	0.023	0.029	0.035
Fully called genome fraction, Coding	0.796	0.752	0.845	0.78	0.870	0.807	0.796	0.808	0.870	0.824	0.818	0.884	0.943	0.965	0.950	0.957	0.960	0.920	0.949	0.928
Partially called genome fraction, Coding	0.024	0.030	0.018	0.001	0.000	0.024	0.025	0.024	0.013	0.023	0.023	0.000	0.010	0.007	0.008	0.007	0.006	0.018	0.006	0.014
No-called genome fraction, Coding	0.181	0.218	0.137	0.22	0.130	0.170	0.179	0.168	0.056	0.153	0.160	0.115	0.048	0.037	0.042	0.036	0.033	0.062	0.045	0.058
Synonymous SNP loci	8,266	7,502	9,312	9,360	11,147	6,526	6,496	6,866	7,745	6,899	7,041	9,106	11,713	9,386	9,201	9,475	9,528	9,572	9,199	8,707
Missense SNP loci	8,173	7,726	9,137	10,875	11,106	6,769	6,837	7,449	8,107	7,557	7,309	9,315	11,032	9,383	8,982	9,368	9,467	9,274	8,926	8,502
Nonsense SNP loci	102	112	122	218	144	78	84	107	107	113	96	143	125	96	84	92	104	95	87	96
Nonstop SNP loci	21	20	26	27	19	16	20	18	22	19	21	13	24	21	22	22	24	24	29	23
Frame-shifting INS loci	229	140	542	265	346	145	135	204	220	244	175	349	148	131	119	124	145	138	122	121
Frame-shifting DEL loci	105	151	134	287	250	107	130	156	126	203	136	275	160	119	120	125	117	129	122	105
Frame-shifting SUB loci	26	44	21	43	40	17	26	47	38	52	30	45	26	23	25	29	26	31	21	26
Frame-preserving INS loci	86	72	95	140	164	79	81	76	93	71	78	157	108	121	113	122	106	137	109	106
Frame-preserving DEL loci	94	77	108	136	144	90	84	84	98	82	85	129	144	127	118	135	138	128	121	105
Frame-preserving SUB loci	214	211	274	378	365	165	183	282	246	277	258	360	275	257	307	286	294	279	288	229
Frame-shifting/preserving ratio	0.91	0.93	1.46	0.91	0.95	0.81	0.84	0.92	0.88	1.16	0.81	1.04	0.63	0.54	0.49	0.51	0.54	0.55	0.51	0.57
Nonsyn/syn SNP ratio	0.99	1.03	0.98	1.19	1.01	1.04	1.05	1.08	1.05	1.10	1.04	1.04	0.94	1.00	0.98	0.99	0.99	0.97	0.97	0.98
Insertion/deletions ratio	1.01	0.92	1.19	0.93	0.98	0.95	0.94	0.96	0.98	0.98	0.94	1.02	0.90	0.91	0.91	0.91	0.91	0.92	0.92	0.96
Ins+del/SNP ratio	0.12	0.10	0.12	0.11	0.11	0.10	0.09	0.10	0.11	0.11	0.11	0.12	0.11	0.12	0.11	0.12	0.11	0.13	0.12	0.12
Coding insertion/deletions ratio	1.55	0.92	2.60	0.96	1.31	1.13	1.00	1.16	1.38	1.09	1.12	1.27	0.82	1.01	0.95	0.93	0.97	1.05	0.93	0.90
Coding SNP/all SNP ratio	0.0053	0.0051	0.0057	0.0061	0.0063	0.0053	0.0054	0.0055	0.0059	0.0056	0.0053	0.0065	0.0063	0.0065	0.0065	0.0065	0.0066	0.0065	0.0063	0.0060
Coding (ins+del)/all (ins+del) ratio	0.0012	0.0013	0.0020	0.0021	0.0021	0.0015	0.0016	0.0017	0.0016	0.0019	0.0014	0.0024	0.0012	0.0012	0.0013	0.0013	0.0013	0.0012	0.0011	0.0008
Pct with mates on different contigs or more than 50kb apart	6.68%	2.42%	5.94%	2.42%	2.42%	2.96%	1.87%	1.84%	1.87%	1.99%	2.52%	2.05%	1.91%	1.36%	1.74%	1.66%	1.52%	1.02%	1.09%	1.96%
Pct with mates within 50kb on different strands (order undefined)	3.73%	2.40%	4.50%	2.40%	2.40%	2.09%	2.09%	2.14%	2.37%	2.46%	2.02%	2.22%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%

Supplementary Table 3. Library statistics. The table lists Complete Genomics' summary output metrics for each genome sequence analyzed in this study. In general, LFR libraries produced lower quality genome metrics. All libraries with STD in the title are genomes processed using the Complete Genomics standard sequencing pipeline and are part of a public release of genome data[28].

(a)	(b)	(c)	(d)	e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)	(n)	(o)	(p)	(q)	(r)	(s)
Sample	Ethnicity	NCBI Genome Build	Contig Coverage (Mb)	Total Contigs	Discordant Heterozygous SNPs	Attempted Heterozygous SNPs	LFR phased Heterozygous SNPs	Parental phased Heterozygous SNPs	LFR and Parental Phased Heterozygous SNPs	LFR and Parental shared Heterozygous SNPs	Parental only Heterozgyous SNPs	LFR only Heterozygous SNPs	Parent phasing rate	Phasable, not phased by LFR	Phasable Hets	LFR phasing rate	Parent+ LFR phasing rate	Final Parent phasing rate
NA19240-Replicate 1	Yoruban	36	2471.012	5184	523	2750435	2386741	2063441	2559045	1861621	201820	495604	0.81	250294	2637035	0.91	0.97	0.78
NA19240-Replicate 2	Yoruban	36	2448.412	5976	571	2764157	2433621	2063930	2592567	1878517	185413	528637	0.80	232903	2666524	0.91	0.97	0.77
NA19240-10 cell pipeline 2.0	Yoruban	37	2429.906	6492	498	2763449	2369433	2063763	2584299	1822845	240918	520536	0.80	301684	2671117	0.89	0.97	0.77
NA19240-Replicate 1 High Coverage	Yoruban	37	2519.329	4653	571	2829582	2578903	2066534	2644040	1974361	92173	577506	0.78	117931	2696834	0.96	0.98	0.77
NA19240-Replicates 1&2 combined	Yoruban	36	2536.297	4415	612	2850047	2646352	2063441	2665715	2010040	53401	602274	0.77	68988	2715340	0.97	0.98	0.76
NA19240-Replicate 1 LFR only pipeline 2.0	Yoruban	36	2407.376	5655	301	2292288	2031514	2063957	2480153	1597832	466125	416196	0.83	560119	2242288	0.91*	N/A	N/A
NA19240-Replicate 1 High Coverage LFR only	Yoruban	37	2475.34	4765	356	2433547	2274696	2066534	2559719	1764689	301845	493185	0.81	373881	2383547	0.95*	N/A	N/A
NA19240-Replicate 1 High accuracy parental phasing	Yoruban	36	2476.683	5055	50	2805342	2451098	1698566	2545108	1579454	119112	846542	0.67	178476	2755342	0.89	0.98	0.62
NA19240-Replicate 2 High accuracy parental phasing	Yoruban	36	2422.567	5664	67	2808022	2454182	1698566	2560182	1562866	135700	861616	0.66	204535	2758022	0.89	0.98	0.62
NA12877-Replicate 1	European	36	1952.645	8648	680	2046264	1831032	1455700	1906444	1351382	104318	450744	0.76	136619	1967651	0.93	0.97	0.74
NA12877-Replicate 2	European	36	1973.28	8565	607	2052808	1810540	1455933	1895131	1340238	115695	439198	0.77	150596	1961136	0.92	0.97	0.74
NA12877-Replicates 1&2 combined	European	36	2045.964	8163	693	2104621	1946089	1455700	1955308	1412988	42712	499608	0.74	57371	2003460	0.97	0.98	0.73
NA12885	European	36	2045.509	8945	474	2127176	1850409	1539991	1953981	1409873	130118	413990	0.79	165097	2015506	0.92	0.97	0.76
NA12886	European	36	1993.077	8584	482	2082023	1854360	1494882	1929621	1390698	104184	434739	0.77	134483	1988843	0.93	0.97	0.75
NA12891	European	36	2000.171	8075	N/A	2070884	1825427	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2020884	0.90*	N/A	N/A
NA12892	European	36	2094.884	8318	N/A	2121645	1917442	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2071645	0.93*	N/A	N/A
NA12892 LFR only	European	36	2044.514	8522	N/A	1825812	1720750	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1775812	0.97*	N/A	N/A
NA20431 High Coverage	European	37	1929.948	9611	N/A	2085859	1703047	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2035859	0.84*	N/A	N/A

Supplementary Table 4. Phasing statistics. This table lists output statistics from the LFR phasing algorithm. For genomes NA12891, NA12892, and NA20431 we do not have parental sequence data. Trio phasing results are unavailable for these genomes. Discordantly phased SNPs (f) are predominately due to errors in parent-based phasing exept when high accuracy parental data is uesd (only 50-67 discordances). The following columns were calculated as (n)=(i)/(j), (o)=(l)/(n), (p)=(h)+(o), (q)=(h)/(p), (r)=(j)/(p), and (s)=(i)/(p). *Phasing rate was calculated by dividing the number of LFR phased SNPs by the number of attempted SNPs-50,000 estimated errors, this is used where parents aren't available or in LFR only libraries were less SNPs are detected.

Sample	Ethnicity	# of Clear RLHs	Mb in Clear RLHs	% of Genome Clear	% of Chr. X Clear	Expected % of Chr. X Clear	% Increase in Chr. X vs. Genome
NA19025	Luhya, Kenya	4,490	369.2	13.6			
NA19703	African in USA	4,678	369.0	13.6			
NA19240	Yoruba, Nigeria	5,068	413.2	14.5	17.3	19.3	20
NA21737	Maasai, Kenya	5,149	394.7	14.6			
MP1*	German	9,267	809.3	29.9			
NA20431	European	9,878	833.2	30.8			
NA12877	European	9,008	837.3	30.9			
NA12885	European	9,907	896.0	31.4	51.0	41.8	63
NA12892	European	9,534	897.0	31.4	49.0	41.8	56
NA20847	Gujarati Indian in USA	9,094	900.6	31.5	48.4	42.0	54
NA18555	Han Chinese	9,922	1065.1	39.3			

Supplementary Table 5. Regions of Low Heterozygosity (RLHs) in Diverse Genomes.
RLHs are genomic regions of greater than or equal to 30kb with less than or equal to 1.4 hets per 10kb. Clear RLHs exclude RLHs comprising reference sequence with undetermined bases. The percent of genome and chromosome X calculation is based on the fully sequenced portions of build 36 of the NCBI genome, 2.86 Gb and 151 Mb respectively. For male samples, the sequenced portion of the X chromosome was subtracted from the total genome size (2.71 Gb) because males are haploid for the X chromosome. *Not sequenced by Complete Genomics.

Heterozgyous SNPs/Mb (Adjusted)	NA19025		NA19703		NA19240		NA12877		NA20431		NA18555	
	# of RLHs	% of Total	# of RLHs	% of Total	# of RLHs	% of Total	# of RLHs	% of Total	# of RLHs	% of Total	# of RLHs	% of Total
0-10	52	14%	42	11%	74	17%	49	5%	43	4%	40	3%
11-30	60	17%	68	18%	81	19%	165	18%	205	18%	337	23%
31-50	74	21%	70	19%	76	18%	320	35%	420	37%	555	37%
51-70	83	23%	89	24%	89	21%	219	24%	259	23%	344	23%
71-90	55	15%	65	17%	75	17%	112	12%	155	14%	150	10%
91-130	35	10%	44	12%	36	8%	53	6%	48	4%	56	4%
Total	359	100%	378	100%	431	100%	918	100%	1130	100%	1482	100%

Supplementary Table 6. Distribution of RLHs greater than 200kb by density of heterozygous SNPs. For all samples analyzed 90-95% of RLHs are shorter than 200kb in length. RLHs are defined as genomic regions of greater than or equal to 30kb with less than or equal to 1.4 hets per 10kb. To compensate for "edge" effects and errors, four heterozygous SNPs are removed from the count.

	NA19240		NA20431	
	# of RLHs	% of Total	# of RLHs	% of Total
All RLHs	5069	100	9008	100
30-100kb	4059	80.1	6612	73.4
>100 kb	1010	19.9	2396	26.6
>300 kb	146	3.6	324	4.9
>1000 kb	5	0.5	9	0.4

Supplementary Table 7. Size distribution of RLHs. RLHs are defined as genomic regions of 30 kb with less than 1.4 heterozygous SNPs per 10kb. European genomes have more RLHs due to the out of Africa bottleneck (Supplementary Table 6) but similar length distribution.

PartA	NA12877	NA19240
Comparable heterozygous SNPs	1726638	2209986
Fraction of total phased heterozygous SNPs	95%	93%
Estimated discordant individual heterozygous SNPs per library in contig without flips	96	64
Discordance rate(%)	0.0056	0.0029
Discordance frequency	1 in 17,986 het SNPs	1 in 34,531 het SNPs
Phasing error rate (diploid genome lenght: 2700Mb male and 2850Mb female)	1 in 28Mb	1 in 44Mb
Part B		
All 2+ HETs contigs	8620	6132
Estimated contigs with flips per library	116	135
% contigs affected	1.35	2.20

Supplementary Table 8. Haplotype discordance between replicate LFR libraries. Two replicate libraries from samples NA12877 and NA19240 were compared at all shared phased heterozygous SNP loci. This is a comprehensive comparison because most of phased loci are shared between the two libraries.

Chr.	NA12877 Replicate 1			NA12877 Replicate 2		
	Shared phased SNPs	Discordant	% Discordant	Shared phased SNPs	Discordant	% Discordant
1	121047	6	0.00496	121706	8	0.00657
2	121255	1	0.00082	122056	1	0.00082
3	108892	3	0.00276	109815	5	0.00455
4	117360	3	0.00256	119134	3	0.00252
5	99796	3	0.00301	100801	0	0.00000
6	105194	4	0.00380	105695	5	0.00473
7	63016	7	0.01111	63866	11	0.01722
8	90852	0	0.00000	91338	3	0.00328
9	69480	2	0.00288	69921	6	0.00858
10	81618	5	0.00613	82340	7	0.00850
11	57598	12	0.02083	58067	6	0.01033
12	79176	3	0.00379	79998	4	0.00500
13	54107	1	0.00185	54603	2	0.00366
14	49354	0	0.00000	49511	5	0.01010
15	46996	10	0.02128	47366	23	0.04856
16	43564	8	0.01836	44072	18	0.04084
17	44564	0	0.00000	44975	4	0.00889
18	45319	0	0.00000	45456	2	0.00440
19	28491	3	0.01053	29007	3	0.01034
20	33636	0	0.00000	34035	4	0.01175
21	17421	1	0.00574	17531	1	0.00570
22	21218	1	0.00471	21335	2	0.00937
total	1499954	73	0.00487	1512628	123	0.00813
total all*	1499954	256	0.01707	1512628	398	0.02631

Supplementary Table 9. Comparison of LFR contigs to Roach et al. contigs. LFR contigs were compared to those generated by Roach et al.[2]. Since ISB used the parental data only for validation, and not in phasing, while comparing ISB results with ours, some contigs are expected not to match in parental designation. If not accounted for, this apparent discrepancy can inflate the true discordance measure. In order to have a more realistic estimation of the discordances, only the contigs that were confirmed to align in parental designation were used. This was achieved by discounting the contigs that did not align in parental designation, as measured by having greater than 85% discordance.

Chr.	Position	Reference	Variant	Replicate 1		Replicate 2		Replicates 1 and 2		Result	RLH
				mom	dad	mom	dad	mom	dad		
1	181262791	C	A	A	C	A	C	A	C	Phased	No
1	213849510	A	T	A	T	A	T	A	T	Phased	No
1	243613456	A	G	G	A	G	A	G	A	Phased	No
2	25946767	G	A	-	-	-	-	-	-	Not phased	Yes
2	151377073	A	G	G	A	G	A	G	A	Phased	No
3	11793938	T	C	-	-	T	C	T	C	Phased	No
3	72300288	G	A	A	G	A	G	A	G	Phased	No
3	118683826	C	T	C	T	C	T	C	T	Phased	No
4	41027699	T	C	T	C	T	C	T	C	Phased	No
5	3882079	T	G	G	T	G	T	G	T	Phased	No
5	74064809	C	T	T	C	T	C	T	C	Phased	No
5	160180969	G	A	G	A	G	A	G	A	Phased	No
6	20526350	C	T	T	C	-	-	T	C	Phased	No
6	75923343	G	A	G	A	G	A	G	A	Phased	No
7	125546271	C	G	G	C	G	C	G	C	Phased	No
7	125547785	C	G	G	C	G	C	G	C	Phased	No
8	3720596	C	A	A	C	-	-	A	C	Phased	No
8	76910527	G	A	A	G	-	-	A	G	Phased	No
8	140579460	C	T	C	T	C	T	C	T	Phased	No
9	78623566	G	A	G	A	G	A	G	A	Phased	No
9	98146694	T	A	-	-	-	-	-	-	Not called in standard or LFR libraries	N/A
9	102781329	T	C	T	C	T	C	T	C	Phased	No
10	133424322	A	G	-	-	A	G	A	G	Phased	No
11	24444780	G	A	G	A	G	A	G	A	Phased	No
11	25362086	G	A	A	G	A	G	A	G	Phased	No
11	125708187	G	A	A	G	A	G	A	G	Phased	No
11	130878775	C	T	T	C	T	C	T	C	Phased	No
12	59046995	G	A	A	G	A	G	A	G	Phased	No
13	33576317	C	T	-	-	-	-	-	-	Not phased	Yes
13	40031189	C	T	C	T	C	T	C	T	Phased	No
13	97628143	C	T	C	T	C	T	C	T	Phased	No
14	53878165	G	T	G	T	G	T	G	T	Phased	No
16	45177203	T	C	-	-	T	C	T	C	Phased	Yes
17	1623091	G	A	G	A	G	A	G	A	Phased	No
17	56613332	C	T	T	C	T	C	T	C	Phased	No

Supplementary Table 10. Conrad *et al. de-novo* mutation phasing. 35 de novo mutations in NA19240 identified by Conrad *et al.* [32] were phased by at least one replicate or the commbined replicates fo NA19240 LFR libraries. All coordinates are in NCBI build 36.

	NA12877				NA19240			
	Standard library SNPs/Replicate 1 phasing	Standard library SNPs/Replicate 2 phasing	Replicate 1 (3+ wells)	Replicate 2 (3+ wells)	Standard library SNPs/Replicate 1 phasing	Standard library SNPs/Replicate 2 phasing	Replicate 1 (3+ wells)	Replicate 2 (3+ wells)
de-novo -like heterozygous SNPs	3,813	3,813	26,494	32,931	3,633	3,633	50,137	38,859
Phased by LFR1 and LFR2	1,589	1,589	1,297	1,447	1,057	1,057	998	1,127
Error-like	2,224	2,224	25,197	31,484	2,576	2,576	49,139	37,732
Error-like-phased	243	286	433	463	282	227	556	854
Error reduction factor	9x	8x	58x	68x	9x	11x	88x	44x
Avoided errors	1,981	1,938	24,764	31,021	2,294	2,349	48,583	36,878

Supplementary Table 11. LFR enabled error reduction. Standard library heterozygous SNP calls alone and in combination with LFR calls were phased independently by replicate LFR libraries. In general LFR introduced approximately 10 fold more false positive variant calls. This most likely occurred as a result of the stochastic incorporation of incorrect bases during phi29 based multiple displacement amplification. Importantly, if heterozygous SNP calls are required to be phased and found in 3 or more independent wells the error reduction is dramatic and better than the standard library without error correction. LFR can remove errors from the standard library as well; improving call accuracy by approximately 10 fold.

Gene Symbol	Chr.	Position	Span (bp)	Gene region or TFBS	TFBS max score	TFBS Orientation	P1 variant	P1 TFBS score	P2 variant	P2 TFBS score	Ref	P1 % frequency	P2 % frequency
PDE4DIP	1	144918957	16358	CDS			A		T		T	86	14
		144930594		CDS			A		C		C	85	15
		144934080		*AP1*	10.7	+	T	6.7	A	10.7	A		
		144934761		*Nkx2-5*	9.4	-	G	5.1	T	9.4	T		
		144934860		*Gata1*	14.5	-	T	12.0	C	4.0	C		
		144935315		*NKx3-1*	11.5	-	G	11.5	A	8.2	A		
DNAJC10	2	183578218	64913	*TLX1::NFIC*	20.4	-	C	7.1	T	11.2	C		
		183580882		*RREB1*	24.4	+	G	12.0	A	9.5	A		
		183593697		CDS			T		C		C	43	56
		183642878		UTR3			A		G		G	34	64
		183643131		UTR3			T		C		C	35	64
EDEM1	3	5226415	32570	*Foxa2*	17.4	+	T	13.4	A	6.9	A		
		5258078		UTR3			A		G		G	74	26
		5258431		UTR3			A		G		G	74	26
		5258985		UTR3			C		T		T	59	39
VPS13A	9	79792920	193137	*PPARG::RXRA*	20.3	+	T	8.5	G	11.4	G		
		79792920		*STAT1*	19.6	-	T	10.6	G	6.5	G		
		79986057		CDS			G		A		A	31	69
ALDH1L2	12	105415948	66390	UTR3			G		A		A	9	91
		105417406		UTR3			T		C		C	4	96
		105479562		*Lhx3*	18.0		A	13.4	G	9.5	G		
		105482184		*En1*	9.4		G	8.0	T	4.7	T		
		105482338		*NFKB1*	16.6		G	6.5	C	10.7	C		
UBE2G2	21	46189094	36302	UTR3			C		T		C	33	66
		46191270		UTR3			T		C		C	35	65
		46221494		*PLAG1*	21.2	+	C	9.7	G	13.8	G		
		46225396		*HNF1A*	19.0	-	T	13.1	C	11.4	C		

Supplementary Figure 12. Genes with allelic expression differences and TFBS altering SNPs in NA20431. Out of a nonexhaustive list of genes that demonstrated significant allelic differences in expression, 6 genes were found with SNPs that altered TFBSs and correlated with the differences in expression seen between alleles. All position are given in reference to NCBI build 37. CDS stands for conding sequence and UTR3 for 3' untranslated region.

Sample	Ethnicity	Number of 10kb regions with >0.33% divergence (>1.5 M years)	% genome	Number of 10kb regions with >0.4% divergence (>2.0 M years)	% genome	Number of 10kb regions with >0.48% divergence (>2.5 M years)	% genome	Number of 10kb regions with >0.63% divergence (>3.5 M years)	% genome
NA19240	Yoruban	5020	1.76	2634	0.92	1279	0.45	400	0.14
NA12877	European	3967	1.47	2167	0.80	1086	0.40	391	0.14
NA12892	European	3714	1.30	1957	0.69	964	0.34	317	0.11
African adjusted	Yoruban		2.05		1.07		0.52		0.16
European adjusted	European		2.13		1.16		0.58		0.21

Supplementary Table 13. Highly diverged segments in the human genomes. Highly diverged regions were identified by comparing haplotypes within each individual genome. The two Europeans used in this table are unrelated, yet have a very similiar number of phased variants. Time of divergence is based upon 0.15% divergence per 1 million years (Myr) and a standing variation of 0.1%. This rate of 0.15% per Myr is based upon an aproximation of 1% human versus chimp divergence over a 6 Myr time of independent evolution since a common ancestor. Adjusted percentages reflect that 14% of African and 31% of European genomes are in RLHs, and indicate that the divergent regions were present before the exit of humans from Africa.