American Journal of Human Genetics, Volume *93*

Supplemental Data

Long Runs of Homozygosity Are Enriched for Deleterious Variation

Zachary A. Szpiech, Jishu Xu, Trevor J. Pemberton, Weiping Peng, Sebastian Zöllner, Noah A. Rosenberg, and Jun Z. Li

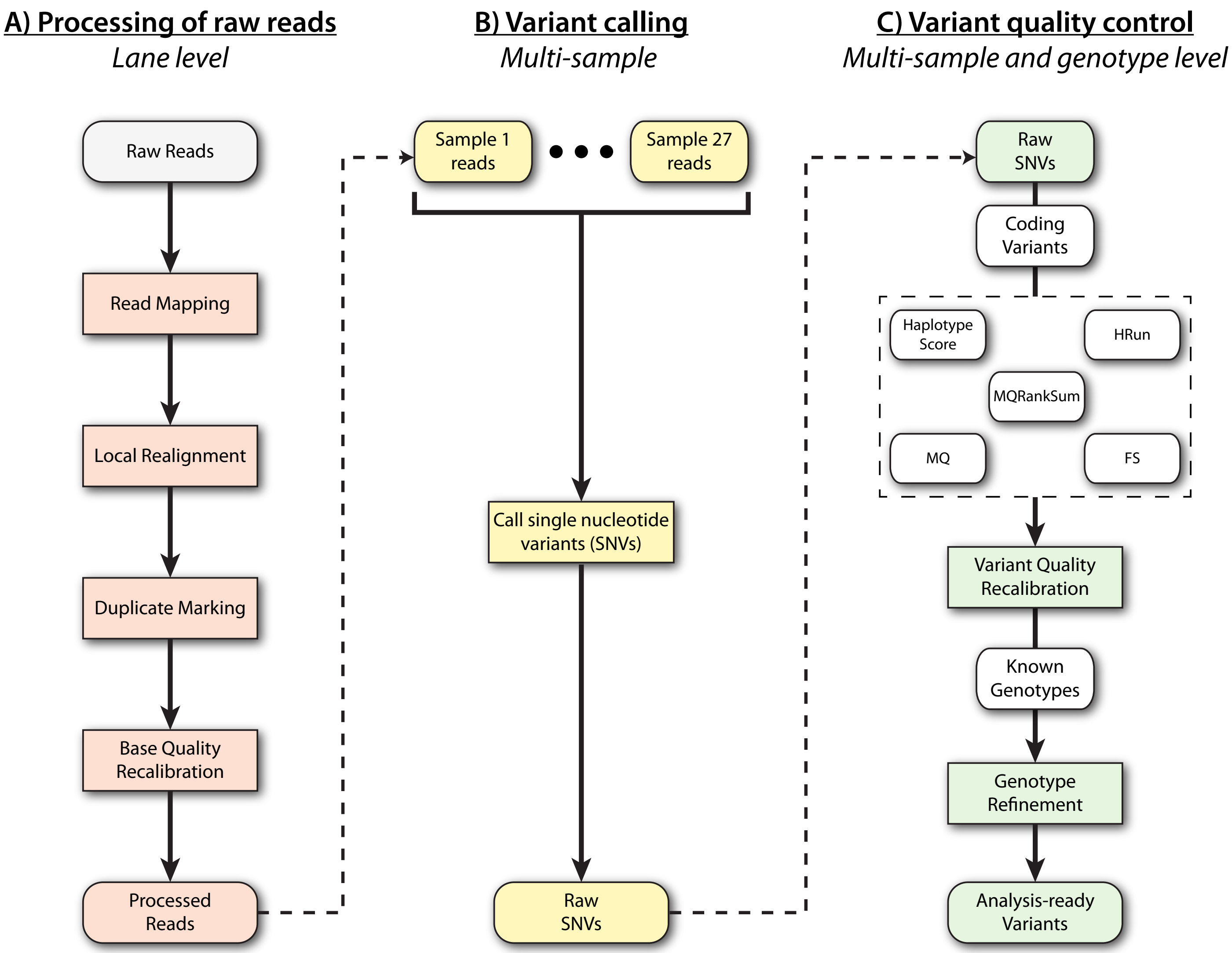

Figure S1 – The exome-sequencing calling and quality-control pipeline.

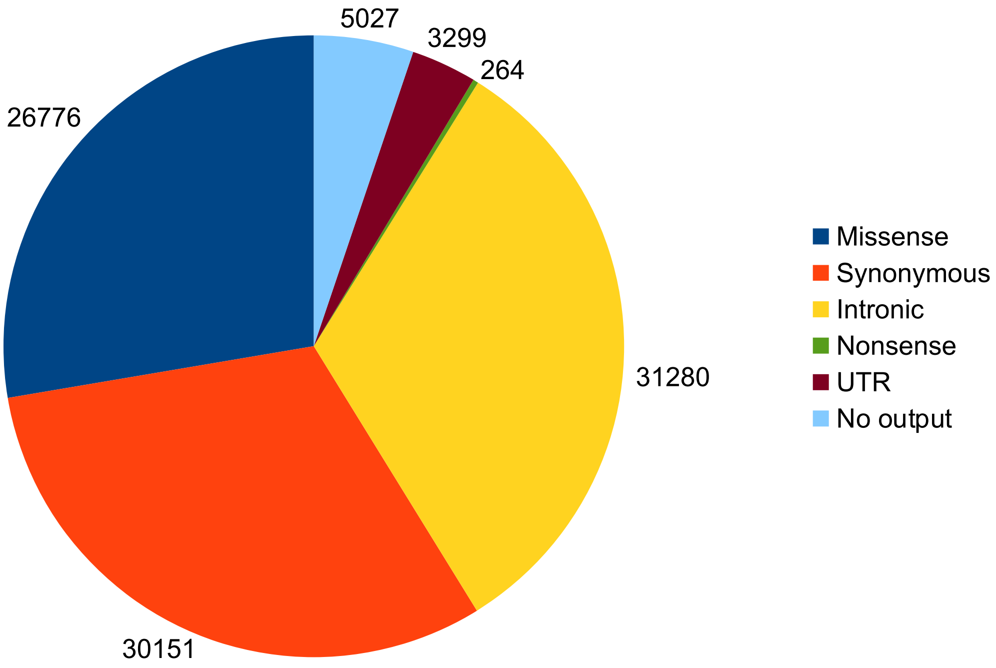

Figure S2 – Classification of 96,797 variant sites by genomic location, as determined by the MapSNPs program included with PolyPhen2. The 96,797 sites represent 96,096 unique sites, as 701 sites were mapped to two different CCDS sequences and are counted twice in this figure.

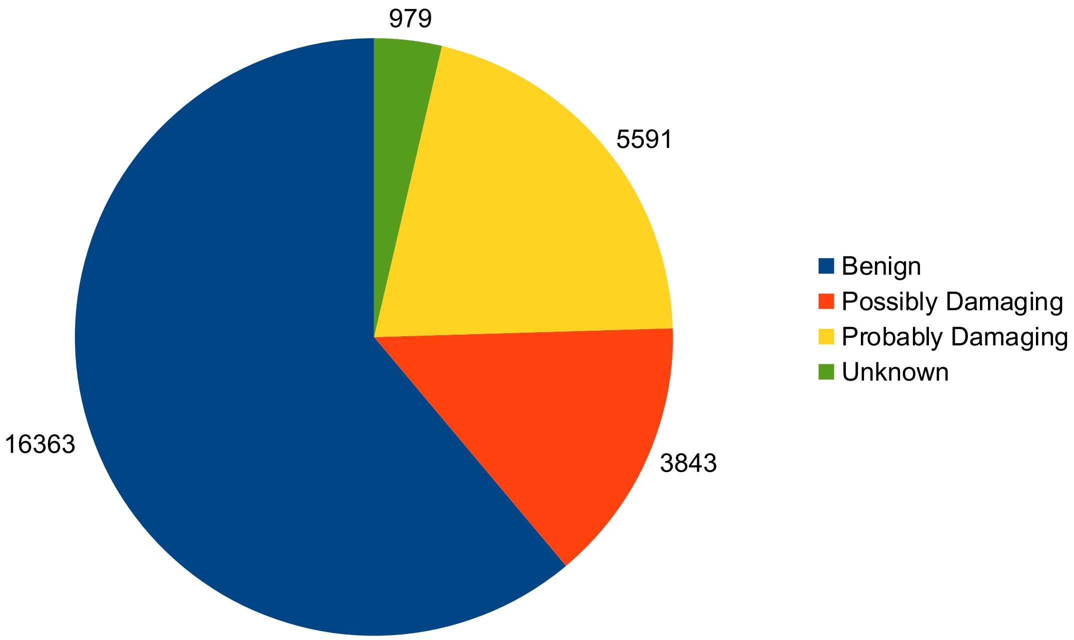

Figure S3 – Classification of missense variant sites by PolyPhen2.

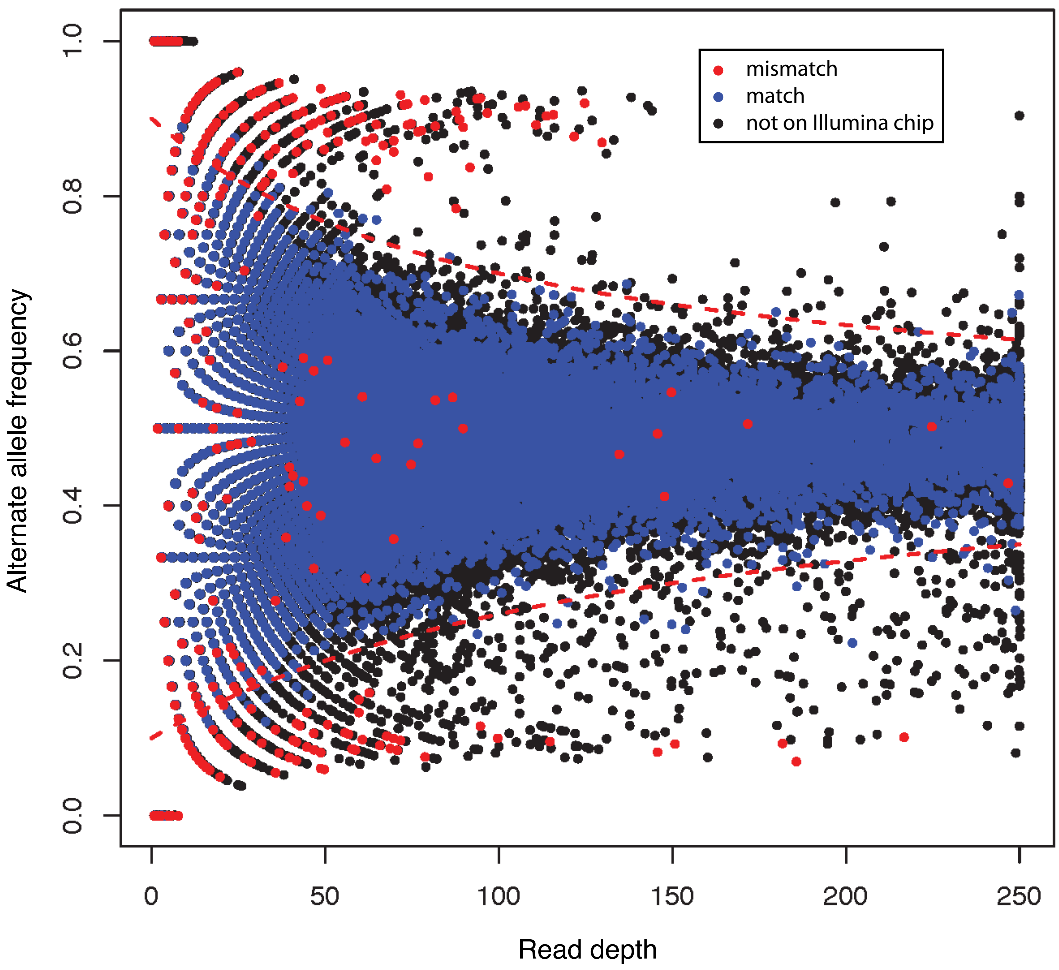

Figure S4 – Alternate allele frequency versus read depth for heterozygous genotypes. Red and blue points represent called exome heterozygotes that are homozygotes and heterozygotes in the Illumina genotype data, respectively. Black points represent called exome heterozygotes that do not occur in the Illumina genotype data. Dotted red lines represent the quality control cutoff for heterozygous genotypes. Any heterozygotes that fall between the two red dotted lines are retained. Others are removed from the data set.

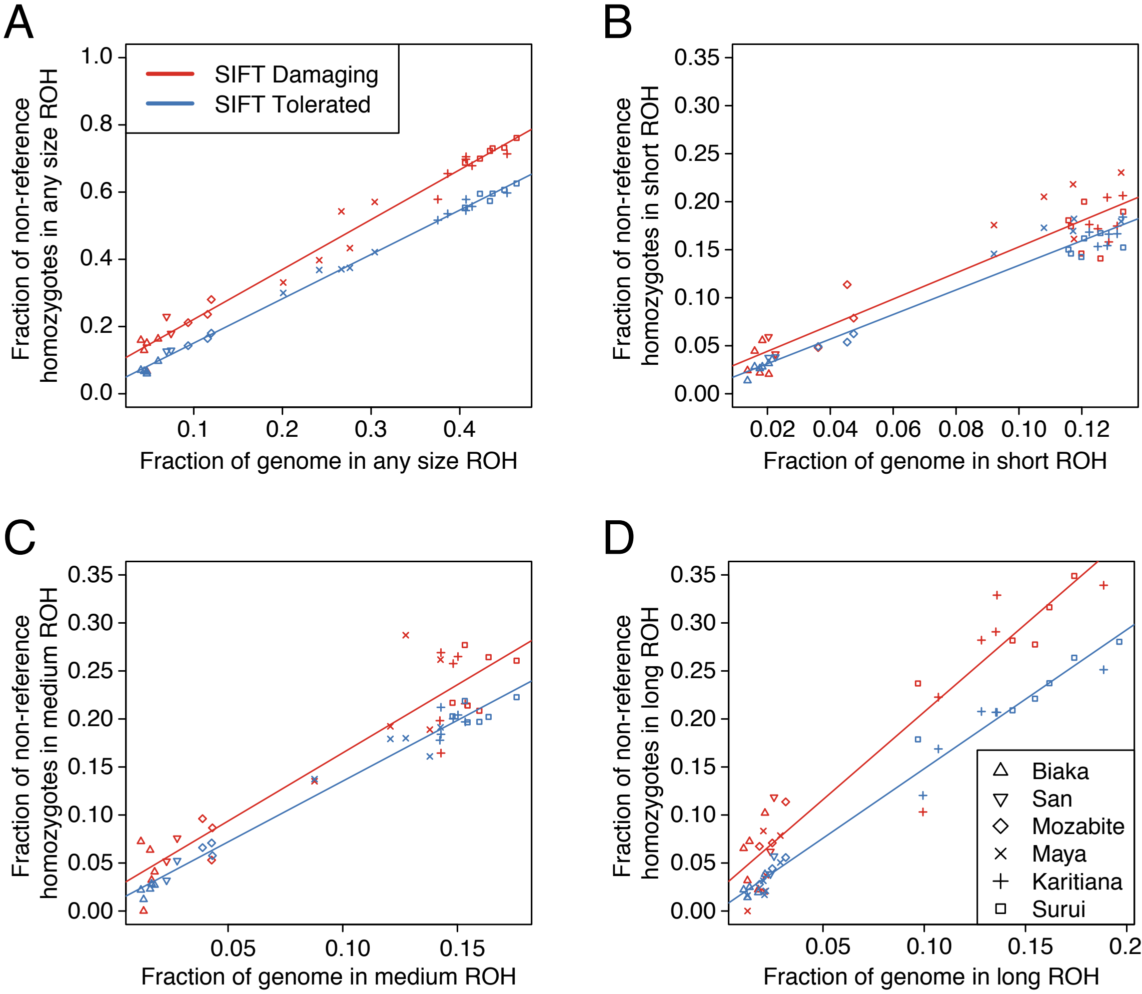

Figure S5 – The fraction of all genome-wide non-reference homozygotes falling in ROH regions versus the fraction of the genome covered by ROH for each individual. (A) All ROH (intercept: $\beta_2 = 0.05413$, $p = 1.875\times10^{-4}$; slope: $\beta_3 = 0.1650$, $p = 4.340\times10^{-4}$) (B) short (class A) ROH (intercept: $\beta_2 = 0.01133$, $p = 0.3070$; slope: $\beta_3 = 0.07911$, $p = 0.4816$) (C) medium (class B) ROH (intercept: $\beta_2 = 0.01404$, $p = 0.2863$; slope: $\beta_3 = 0.1536$, $p = 0.1753$) and (D) long (class C) ROH (intercept: $\beta_2 = 0.02116$, $p = 0.06029$; slope: $\beta_3 = 0.3798$, $p = 1.243\times10^{-3}$). Red points represent damaging homozygotes as predicted by SIFT, and blue points represent non-damaging homozygotes as predicted by SIFT.

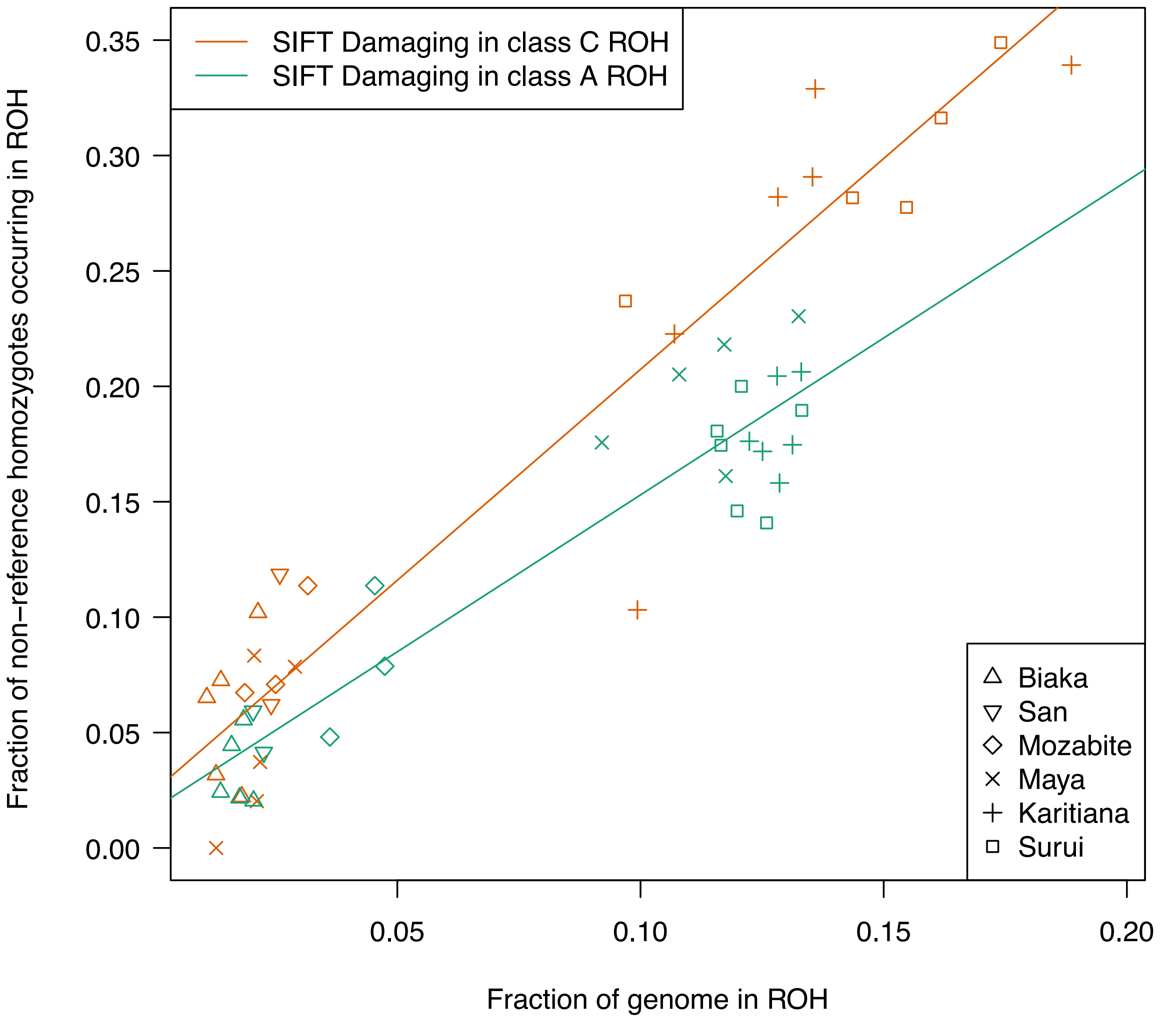

Figure S6 – The fraction of all genome-wide non-reference homozygotes falling in long (class C) and short (class A) ROH regions versus the fraction of the genome covered by ROH, for each individual. Orange points represent damaging homozygotes in class C (long) ROH regions, and green points represent damaging homozygotes in class A (short) ROH regions (intercept: $\beta_2 = 7.622 \times 10^{-3}$, p = 0.6124; slope: $\beta_3 = 0.4676$, p = 3.299×10^{-3}). Damaging alleles are predicted by SIFT.

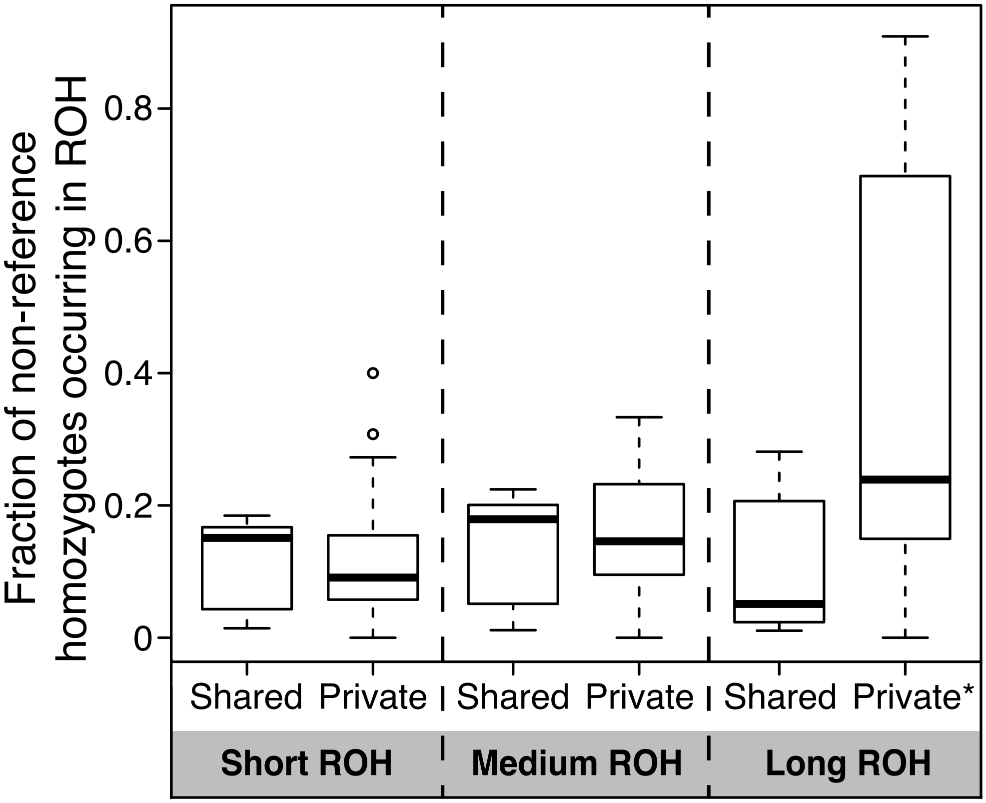

Figure S7 – The fraction of all genome-wide non-reference homozygotes falling in short (class A), medium (class B), and long (class C) ROH regions for shared and private variants. A significant difference in the mean across individuals for private homozygotes and shared homozygotes is found only for long ROH ($p = 2.036 \times 10^{-5}$, marked with an asterisk).

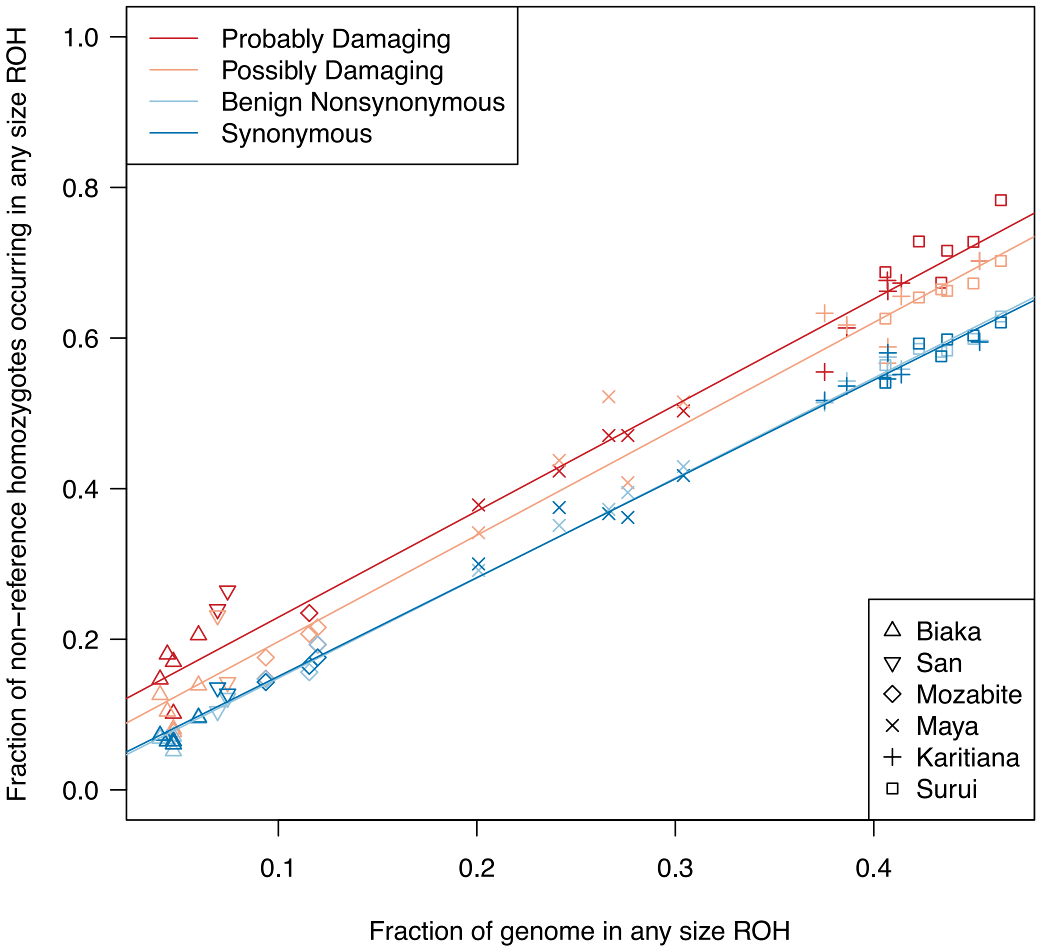

Figure S8 – The fraction of all genome-wide non-reference homozygotes falling in ROH regions versus the fraction of the genome covered by any size ROH, for each individual. Dark red points represent probably damaging homozygotes, light red points represent possibly damaging homozygotes, light blue points represent benign nonsynonymous homozygotes, and dark blue points represent synonymous homozygotes.

Population	Individual ID	Reference homozygote concordance (%)	Heterozygote concordance (%)	Non-reference homozygote concordance (%)
San	991	99.7	99.6	99.9
	992	99.7	99.2	99.5
Biaka	454	99.6	99.5	99.4
	457	99.7	99.4	99.6
	458	99.5	98.7	98.9
	459	99.6	99.4	99.4
	460	98.6	86.0	97.5
Mozabite	1264	99.7	99.3	99.5
	1267	99.8	99.3	99.8
	1274	99.9	99.7	99.8
Maya	854	99.4	99.8	99.4
	855	99.6	99.7	99.5
	856	99.8	99.8	99.5
	860	99.0	99.5	98.1
	868	99.4	99.6	98.6
Karitiana	999	99.6	99.4	99.7
	1012	99.9	99.6	99.7
	1014	99.6	99.1	99.5
	1015	99.6	99.4	99.7
	1018	99.8	99.9	99.7
	1019	99.6	99.2	99.6
Surui	837	99.3	98.5	99.0
	838	99.4	98.9	99.3
	843	98.7	97.1	98.9
	845	99.3	98.6	99.6
	846	99.4	99.1	99.0
	849	99.4	99.6	99.5

Table S1 – Called exome genotype concordance with known Illumina genotypes.

Population	Individual ID	Mean coverage (number of copies)	Percentage of sites with at least 20× coverage
San	991	44	75
	992	38	70
Biaka	454	70	81
	457	58	82
	458	56	75
	459	42	75
	460	49	70
Mozabite	1264	36	81
	1267	38	84
	1274	44	90
Maya	854	62	79
	855	68	82
	856	72	83
	860	55	72
	868	48	71
Karitiana	999	57	77
	1012	74	83
	1014	66	78
	1015	53	73
	1018	75	84
	1019	60	76
Surui	837	47	62
	838	81	79
	843	66	69
	845	83	78
	846	64	72
	849	74	79

Table S2 – Mean coverage and percentage of sites with at least 20× coverage for the 96797 quality-controlled sites.

Population	Individual ID	Class A ROH	Class B ROH	Class C ROH	Total ROH
San	991	0.0225	0.0278	0.0241	0.0744
	992	0.0204	0.0232	0.0258	0.0694
Biaka	454	0.0184	0.0160	0.0127	0.0471
	457	0.0137	0.0167	0.0137	0.0440
	458	0.0204	0.0180	0.0214	0.0598
	459	0.0176	0.0120	0.0108	0.0404
	460	0.0159	0.0132	0.0180	0.0471
Mozabite	1264	0.0474	0.0432	0.0250	0.1156
	1267	0.0454	0.0428	0.0316	0.1198
	1274	0.0361	0.0389	0.0186	0.0937
Maya	854	0.1172	0.1275	0.0218	0.2665
	855	0.1325	0.1426	0.0289	0.3041
	856	0.1175	0.1380	0.0206	0.2761
	860	0.0921	0.0877	0.0211	0.2009
	868	0.1080	0.1208	0.0128	0.2415
Karitiana	999	0.1330	0.1428	0.0994	0.3752
	1012	0.1251	0.1534	0.1354	0.4139
	1014	0.1286	0.1502	0.1283	0.4071
	1015	0.1312	0.1482	0.1070	0.3864
	1018	0.1281	0.1428	0.1359	0.4068
	1019	0.1224	0.1424	0.1886	0.4533
Surui	837	0.1199	0.1479	0.1963	0.4640
	838	0.1332	0.1758	0.0969	0.4058
	843	0.1259	0.1532	0.1436	0.4227
	845	0.1165	0.1596	0.1741	0.4502
	846	0.1207	0.1544	0.1618	0.4370
	849	0.1157	0.1636	0.1547	0.4340

Table S3 – Fraction of individual genomes covered by ROH classes.

Population	Individual ID	Class A ROH			Class B ROH			Class C ROH			Non-ROH		
		1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0
San	991	12	12	194	21	13	231	15	1	136	185	1237	7137
	992	12	8	179	16	3	139	26	2	266	175	1168	7189
Biaka	454	10	4	124	9	9	118	5	1	68	169	1225	7384
	457	2	4	83	10	1	143	16	2	107	168	1224	7396
	458	12	5	144	8	1	89	17	1	230	178	1100	7216
	459	7	5	129	10	1	142	12	0	119	183	1171	7379
	460	8	10	134	2	4	68	3	1	97	134	1032	7217
Mozabite	1264	18	8	373	22	8	280	10	2	206	176	923	7211
	1267	12	13	260	14	9	355	18	8	309	172	946	7128
	1274	9	19	299	12	5	331	9	0	169	156	1005	7246
Maya	854	58	30	935	78	33	1137	7	4	159	146	715	5853
	855	69	31	1049	58	27	1268	20	7	374	142	678	5415
	856	61	23	965	57	32	1021	16	9	242	171	690	5885
	860	43	30	674	31	20	688	10	5	118	150	746	6256
	868	52	26	876	56	37	1040	6	3	76	151	676	6018
Karitiana	999	84	40	1096	80	30	1249	42	4	692	143	583	5111
	1012	67	34	938	81	50	1358	110	4	116	130	549	4687
	1014	70	46	1067	77	28	1412	98	6	1202	144	495	4462
	1015	64	36	1079	82	40	1180	73	4	1000	137	547	4842
	1018	61	36	955	79	33	1219	100	9	1398	144	568	4671
	1019	68	28	1043	84	25	1161	117	3	1665	114	480	4272
Surui	837	53	29	789	91	35	1365	122	5	1782	91	461	3873
	838	59	44	1072	92	40	1358	72	7	1028	116	518	4528
	843	55	30	910	86	18	1254	95	6	1180	104	493	4247
	845	55	27	912	77	41	1308	122	18	1492	108	452	4299
	846	51	36	932	92	29	1238	93	5	1360	106	521	4453
	849	57	34	944	92	29	1391	94	12	1364	120	492	4264

Table S4 – Genotype counts per individual for sites classified as damaging.

Population	Individual ID	Class A ROH			Class B ROH			Class C ROH			Non-ROH		
		1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0
San	991	181	74	876	240	64	1056	174	5	674	4149	8110	28986
	992	179	59	720	139	24	631	268	5	1203	4134	8025	28997
Biaka	454	114	27	547	104	12	533	58	0	306	3831	8212	30253
	457	65	25	348	119	19	466	98	12	389	4008	8344	30416
	458	127	39	600	113	11	433	156	1	816	3739	7877	29494
	459	115	25	592	98	7	449	92	0	403	4009	8359	30172
	460	89	42	635	39	27	269	63	5	430	3151	7633	28678
Mozabite	1264	214	49	1536	199	45	1282	151	15	993	2927	6812	30447
	1267	201	59	1228	251	37	1392	207	7	1262	2952	6863	30355
	1274	184	40	1189	247	45	1357	100	1	555	3142	7344	30739
Maya	854	676	133	4211	713	156	5033	85	4	549	2522	5009	25091
	855	766	160	4843	815	126	5177	214	14	1332	2461	4701	23549
	856	747	154	4190	670	88	4589	130	13	868	2591	4956	25279
	860	499	117	2963	470	73	2951	59	8	489	2435	5408	26106
	868	671	134	3877	682	114	4523	56	4	326	2438	5264	25259
Karitiana	999	814	172	4867	960	151	5510	563	14	2899	2192	4269	21757
	1012	730	150	4488	914	175	5989	937	21	5367	2077	3877	19656
	1014	781	159	4645	980	162	6065	972	20	5186	1993	3772	19201
	1015	776	138	4646	918	133	5386	772	22	4182	2112	4068	20704
	1018	740	130	4388	844	126	5488	962	36	5183	2105	3960	20398
	1019	789	159	4445	839	111	5134	1211	25	7417	1929	3417	18197
Surui	837	654	129	3895	885	135	5423	1260	23	7396	1690	3385	16831
	838	654	195	4537	948	157	5779	768	38	4290	1945	3978	19707
	843	691	152	4000	900	106	5184	863	36	5074	1705	3371	18235
	845	649	161	4035	869	144	5609	1163	51	6437	1774	3598	18242
	846	732	170	4322	847	139	5364	1049	30	6072	1805	3792	18424
	849	673	148	4358	901	119	5664	1022	33	5909	1896	3681	18358

Table S5 – Genotype counts per individual for sites classified as non-damaging.

Population	Individual ID	Class A ROH			Class B ROH			Class C ROH			Non-ROH		
		1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0
San	991	0	0	11	0	0	15	1	0	5	4	32	185
	992	0	0	10	0	0	4	0	0	10	4	28	192
Biaka	454	0	0	1	0	0	10	0	0	2	4	31	202
	457	0	0	3	0	0	4	0	0	5	7	33	200
	458	0	0	2	0	0	3	0	0	7	4	27	203
	459	1	0	4	0	0	5	0	0	3	2	30	206
	460	0	0	4	0	0	1	0	0	3	3	20	203
Mozabite	1264	0	0	19	0	0	6	0	0	6	4	22	196
	1267	0	2	4	0	2	5	0	0	11	2	29	197
	1274	1	0	10	0	0	8	0	0	4	2	27	204
Maya	854	1	1	29	2	2	28	0	0	7	0	17	165
	855	4	1	33	2	1	29	1	0	13	1	16	150
	856	3	1	27	0	0	25	3	0	6	2	16	170
	860	1	0	17	2	0	14	1	0	2	2	15	181
	868	0	2	29	0	0	31	0	0	5	3	24	152
Karitiana	999	2	0	33	5	1	51	1	2	16	4	19	117
	1012	0	1	22	1	0	34	3	2	36	3	15	137
	1014	2	2	26	3	1	35	5	0	34	5	12	124
	1015	2	1	36	2	2	30	6	0	36	2	17	115
	1018	2	0	31	1	5	28	7	0	33	4	9	133
	1019	1	6	33	1	0	28	4	0	48	5	15	111
Surui	837	0	0	21	6	0	35	4	0	39	2	12	116
	838	0	0	27	4	1	36	4	0	36	2	13	124
	843	2	2	29	2	1	27	2	0	35	1	11	117
	845	0	0	28	2	1	40	2	2	43	5	12	110
	846	2	0	33	2	0	38	2	1	37	1	17	114
	849	1	0	28	4	1	46	2	0	46	1	10	106

Table S6 – Genotype counts per individual for nonsense sites.

Population	Individual ID	Class A ROH			Class B ROH			Class C ROH			Non-ROH		
		1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0	1/1	0/1	0/0
San	991	0	0	4	0	0	1	1	0	1	0	9	50
	992	0	0	1	0	0	2	0	0	2	2	12	47
Biaka	454	0	0	0	0	0	3	0	0	0	2	12	47
	457	0	0	0	0	0	1	0	0	1	4	15	44
	458	0	0	1	0	0	0	0	0	3	2	11	48
	459	0	0	0	0	0	2	0	0	1	1	14	48
	460	0	0	0	0	0	1	0	0	0	1	15	47
Mozabite	1264	0	0	3	0	0	0	0	0	0	2	11	50
	1267	0	1	0	0	0	1	0	0	1	2	14	47
	1274	1	0	5	0	0	2	0	0	0	0	10	48
Maya	854	0	1	7	2	0	11	0	0	0	0	8	35
	855	1	0	6	2	0	5	1	0	4	0	5	40
	856	1	0	6	0	0	4	3	0	1	1	8	41
	860	1	0	4	2	0	6	1	0	1	1	5	41
	868	0	2	4	0	0	5	0	0	1	2	11	38
Karitiana	999	2	0	7	3	0	8	0	0	3	2	10	31
	1012	0	0	6	1	0	8	2	0	4	2	5	37
	1014	1	2	4	2	0	8	4	0	7	1	5	31
	1015	0	0	7	2	0	5	3	0	12	2	7	28
	1018	1	0	5	1	0	3	6	0	12	1	3	34
	1019	0	0	7	1	0	4	3	0	15	3	8	25
Surui	837	0	0	6	3	0	5	2	0	9	2	6	31
	838	0	0	8	3	0	5	3	0	4	1	8	31
	843	2	0	6	2	0	5	1	0	11	0	9	28
	845	0	0	10	2	1	8	2	1	9	2	6	23
	846	2	0	6	1	0	8	1	0	12	1	7	26
	849	0	0	3	4	0	9	2	0	9	0	5	32

Table S7 – Genotype counts per individual for loss-of-function nonsense sites.