Supporting Information

Choi et al. 10.1073/pnas.0910672106

SI Text

Variant Filtering and Annotation. The variants were re-evaluated according to the following rules: (*i*) calls with a Phred-like quality score of less than 45 were excluded; (*ii*) for heterozygous bases, reads supporting both major and minor alleles should have different start and end points; (*iii*) total reads at variant bases should be equal to or greater than 10; (*iv*) for heterozygous bases, the probability of the observed deviation of the frequencies of the major and minor alleles from the binomial distribution should be at least 10^{-7}. We analyzed data including all reads, excluding all potential PCR duplicates by only analyzing reads with independent start sites, or intermediate models. Including all reads provided the highest sensitivity and specificity to detect known SNPs; consequently, data shown are based on analysis of all reads.

The filtered bases were annotated as follows: (*i*) novelty based on overlap with the SNP database (build 130) and 4 previously reported personal genomes (1–4); (*ii*) effect on the encoded protein; (*iii*) conservation information from phyloP scores calculated from multiple alignments of 44 vertebrate species and invertebrate orthologues in *C. elegans* and *D. melanogaster* [PhyloP scores were collected from the University of California Santa Cruz genome browser (http://genome.ucsc.edu) and the orthologue lists were acquired from InParanoid (http://inparanoid.sbc.su.se)]; (*iv*) expression pattern across 79 tissues available from the expression atlas at the Genomics Institute of the Novartis Research Foundation; and (*v*) substitution in known micro-RNA sequences available from miRBase (release 13.0; http://microrna.sanger.ac.uk/).

GenBank Accession Numbers. GenBank accession numbers are as follows: NP_000102.1 (human SLC26A3), NP_067328.1 (mouse SLC26A3, 80% identity, 89% homology), NP_001075567.1 (rabbit SLC26A3, 78% identity, 88% homology), NP_001077145.1 (cow SLC26A3, 79% identity, 88% homology), XP_415945.2 (chicken SLC26A3, 63% identity, 78% homology), NP_001089015.1 (frog SLC26A3, 43% identity, 62% homology), NP_001129155.1 (Zebrafish SLC26A3, 54% identity, 72% homology), NP_649024.1 (fruit fly prestin, 26% identity, 44% homology), NP_491138.2 (worm SULP6, 30% identity, 49% homology), NP_998778.1 (human SLC26A1), NP_000103.2 (human SLC26A2), NP_000432.1 (human SLC26A4), NP_945350.1 (human SLC26A5), NP_075062.2 (human SLC26A6), NP_439897.1 (human SLC26A7), NP_443193.1 (human SLC26A8), NP_001136072.1 (human SLC26A9), and NP_775897.2 (human SLC26A11).

1. Bentley DR, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
2. Ng PC, et al. (2008) Genetic variation in an individual human exome. *PLoS Genet* 4:e1000160.
3. Wang J, et al. (2008) The diploid genome sequence of an Asian individual. *Nature* 456:60–65.
4. Wheeler DA, et al. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452:872–876.

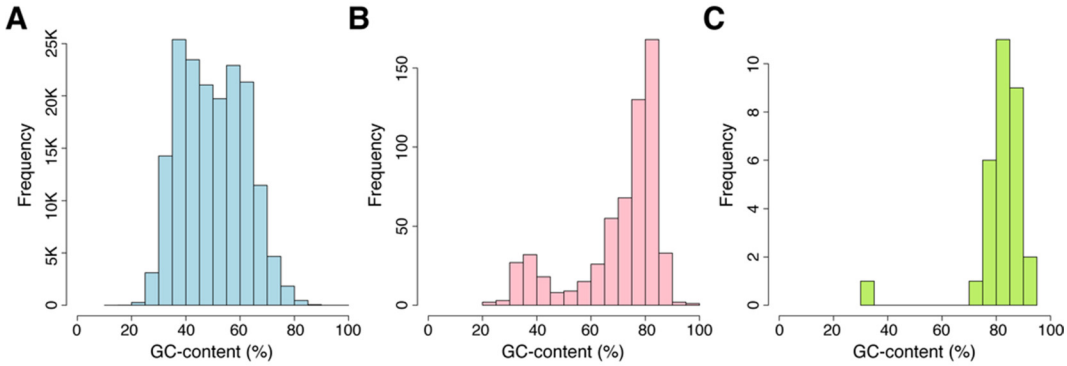

Fig. S1. Distribution of GC content from (*A*) all targeted intervals ($n = 177{,}564$; mean GC content, 49.9%), (*B*) targeted intervals with all bases covered less than 10× ($n = 787$; mean GC content, 68.4%), and (*C*) targeted intervals with no coverage at all ($n = 30$; mean GC content, 81.9%).

Table S1. LOH intervals in GIT 264–1

Chr	Start	End	Length, bp
chr2	20,629,444	26,652,535	6,023,092
chr2	221,496,416	227,234,674	5,738,259
chr4	761,587	7,196,934	6,435,348
chr4	171,974,450	183,281,048	11,306,599
chr5	33,867,772	46,174,214	12,306,443
chr5	49,705,738	68,831,574	19,125,837
chr5	86,987,264	121,560,760	34,573,497
chr5	153,966,237	169,946,389	15,980,153
chr6	6,761,331	10,934,277	4,172,947
chr6	19,569,799	58,786,624	39,216,826
chr6	62,027,492	99,386,170	37,358,679
chr6	132,920,951	139,733,915	6,812,965
chr6	144,089,526	148,970,821	4,881,296
chr7	25,527,037	40,749,097	15,222,061
chr7	66,167,307	123,940,012	57,772,706
chr9	21,673,289	39,088,698	17,415,410
chr9	70,174,192	83,632,247	13,458,056
chr9	92,698,855	124,001,125	31,302,271
chr9	134,867,108	137,127,054	2,259,947
chr10	8,388,335	12,046,364	3,658,030
chr12	29,790,707	34,565,140	4,774,434
chr12	36,330,311	92,059,960	55,729,650
chr14	62,638,241	84,025,011	21,386,771
chr15	59,255,759	84,920,344	25,664,586
chr16	77,542,251	83,428,910	5,886,660
chr19	53,868,842	57,601,319	3,732,478

Table S2. Homozygous deletions in GIT 264–1

Chr:start-end	Size, bp	SNPs	Mean LogR	Overlap with DGV CNVs, %	Overlap with genic region	Description
chr4:173,226,303–173,227,450	1,148	3	−4.874	100	Yes	Intron of GALNT6
chr6:29,959,422–29,962,187	2,766	3	−4.476	100	No	–
chr6:29,967,687–29,969,546	1,860	4	−4.110	100	No	–
chr6:29,976,043–29,976,963	921	3	−3.392	100	Yes	Intron of HLA-A*0226
chr6:29,980,521–29,981,029	509	2	−4.510	100	No	–
chr6:29,982,364–29,983,504	1,141	3	−4.851	100	No	–
chr6:29,985,843–29,987,783	1,941	4	−3.515	100	No	–
chr6:29,992,625–29,992,775	151	2	−4.341	100	No	–
chr6:30,000,415–30,003,207	2,793	4	−4.280	100	Yes	Intron of HLA-A*0226
chr6:30,005,598–30,007,126	1,529	2	−3.167	100	No	–
chr6:30,009,547–30,010,176	630	2	−3.585	100	No	–
chr6:32,081,159–32,082,827	1,669	2	−3.728	100	No	–
chr6:32,648,263–32,650,112	1,850	2	−3.630	100	No	–
chr6:32,672,690–32,672,705	16	2	−3.944	100	No	–
chr6:32,754,210–32,756,221	2,012	2	−2.859	100	No	–
chr7:112,707,316–112,734,558	27,243	4	−5.253	100	No	–
chr9:28,039,518–28,040,777	1,260	2	−6.297	93	Yes	Intron of LINGO2
chr12:33,193,705–33,197,122	3,418	6	−5.148	100	No	–
chr12:58,227,690–58,228,389	700	2	−5.186	100	No	–
chr12:89,012,380–89,014,390	2,011	2	−5.550	100	No	–

The precise endpoints of the clustered deletions in the approximate 3 Mb interval on chromosome 6 have not been defined, and some apparently non-contiguous deletions could represent a single continuous deletion.

Table S3. Coverage of GIT 264–1 at mean 99× per-base coverage

Bases covered at least	Bases	Percent, %
10×	33,448,694	98.39
9×	33,521,237	98.60
8×	33,588,222	98.80
7×	33,650,912	98.98
6×	33,708,677	99.15
5×	33,760,953	99.30
4×	33,808,496	99.44
3×	33,851,516	99.57
2×	33,890,624	99.69
1×	33,925,936	99.79
All bases	33,997,546	100.00

Table S4. Novel homozygous protein-altering variations in GIT 264–1

Chr	Position	Base change	Reads	PhyloP score	Gene	Amino acid change	Worm orthologue
chr16	82,678,492	C>T	148	5.673	MBTPS1	R369H	NA*
chr9	35,707,627	C>A	69	5.513	TLN1	A718S	A
chr7	107,201,654	C>T	173	5.417	SLC26A3	D652N	D
chr12	47,373,772	C>G	129	4.916	CCNT1	D498H	†
chr9	70,288,722	G>A	123	4.900	PGM5	V473M	NA
chr14	72,507,560	G>A	70	4.735	ZFYVE1	A706V	NA
chr9	114,064,663	T>G	39	4.587	ROD1	Q158P	NA
chr12	48,284,364	C>G	76	3.410	FAM186B	W107C	NA
chr12	55,934,911	C>T	40	2.803	R3HDM2	R948H	–
chr6	29,735,201	C>A	107	2.701	MOG	P72H	NA
chr7	92,573,124	G>A	74	1.966	SAMD9	R75W	NA
chr12	62,771,333	G>A	29	1.910	SRGAP1	V483I	Q
chr7	101,900,005	C>T	19	1.886	LRWD1	A545V	NA
chr5	35,991,652	C>A	142	1.859	UGT3A1	V383L	V
chr15	64,644,195	G>A	45	1.762	LCTL	T52M	I
chr6	32,196,828	A>G	18	1.578	ATF6B	S203P	NA
chr12	91,699,921	G>A	120	1.397	EEA1	S1091L	I
chr2	26,393,310	A>G	71	1.347	GPR113	L100P	NA
chr7	97,771,596	C>T	70	1.214	BAIAP2L1	V424M	NA
chr7	29,572,841	C>A	13	1.196	PRR15	P124H	NA
chr6	42,940,693	G>A	35	1.142	KIAA0240	R924Q	NA
chr14	73,467,985	A>T	107	0.561	ZNF410	E475D	NA
chr19	54,908,654	A>G	71	0.514	CPT1C	M787V	NA
chr4	1,335,522	A>C	136	0.442	KIAA1530	N150T	A
chr6	26,080,158	A>G	236	0.340	TRIM38	Y197C	S
chr12	48,236,487	T>C	78	−0.374	KCNH3	S846P	NA
chr12	48,267,781	C>G	108	−0.428	FAM186B	E852D	NA
chr12	51,959,832	G>T	127	−0.469	ESPL1	R805I	NA
chr12	48,280,321	A>C	49	−1.286	FAM186B	F457V	NA

*The protein does not have invertebrate orthologue.
†The amino acid sequence does not align to the orthologue.
NA, not applicable.

Table S5. *SLC26A3* primer sequences

Exon	Forward	Reverse	Product size, bp
2	5′-TGTTCACCACATAGTTGCAAAG-3′	5′-GAAAATTTAGGGGAAAGTGCTTC-3′	253
3	5′-TACCAATTTCCCACCGACAG-3′	5′-GAGCTGTACACATTCAGAGGAAG-3′	280
4	5′-CCAAATGAGAATGTCTTTGATTCC-3′	5′-TCCTGGATTATGTGAAATTTGG-3′	276
5	5′-TGTGGTTGAAAGACAATGTGC-3′	5′-GATGTGTAACAGTCAAGATGAACAG-3′	314
6	5′-ATTTCAGTTGCGGATCCTTG-3′	5′-CATGGCAAAGGCATTAAAGG-3′	331
7	5′-GCGTTGTATCCAAATTGTGC-3′	5′-TGAGCTACAAATGGTGAAGGAC-3′	292
8	5′-GGTACAGAAGTGGGTTTTGGC-3′	5′-TACCTGCAGGCTGAACTGC-3′	278
9–11	5′-CATTGATTTGCCCAATGGTT-3′	5′-AACCGGGGTCCTTGATTTAC-3′	688
12	5′-GCTTTGATTTGCTTTGAGGC-3′	5′-TCTAGACTAGTGTTTCATGGAGCC-3′	460
13	5′-TTTTAAAAGACCCCGTCTCG-3′	5′-CTTCTGATTTGGGCAGGTCC-3′	238
14–15	5′-GAAAAGTGAGCCTGCTTTGG-3′	5′-CACAACCCAGTCTCAACAATG-3′	451
16	5′-CCAATCCCTCAAATCACTCC-3′	5′-GGAATGGAACTACACCTTGAATC-3′	270
17	5′-TCGTTTGGCTCAACTCAGAA-3′	5′-TCATTGATTGGTTGTTTCCA-3′	381
18	5′-CACCCGTAAGGTATAATTCTGATTG-3′	5′-AGTTGGCAGTGAGCCTGAAG-3′	202
19–20	5′-CTCAAGGGGAAACCGGTAG-3′	5′-TGGCCTCACTACTTTGAAGG-3′	467
21	5′-TGGTAAATTCTAGCTCCTAAGATGTG-3′	5′-ATGAACTTATCAATAACTTTCTGGG-3′	189