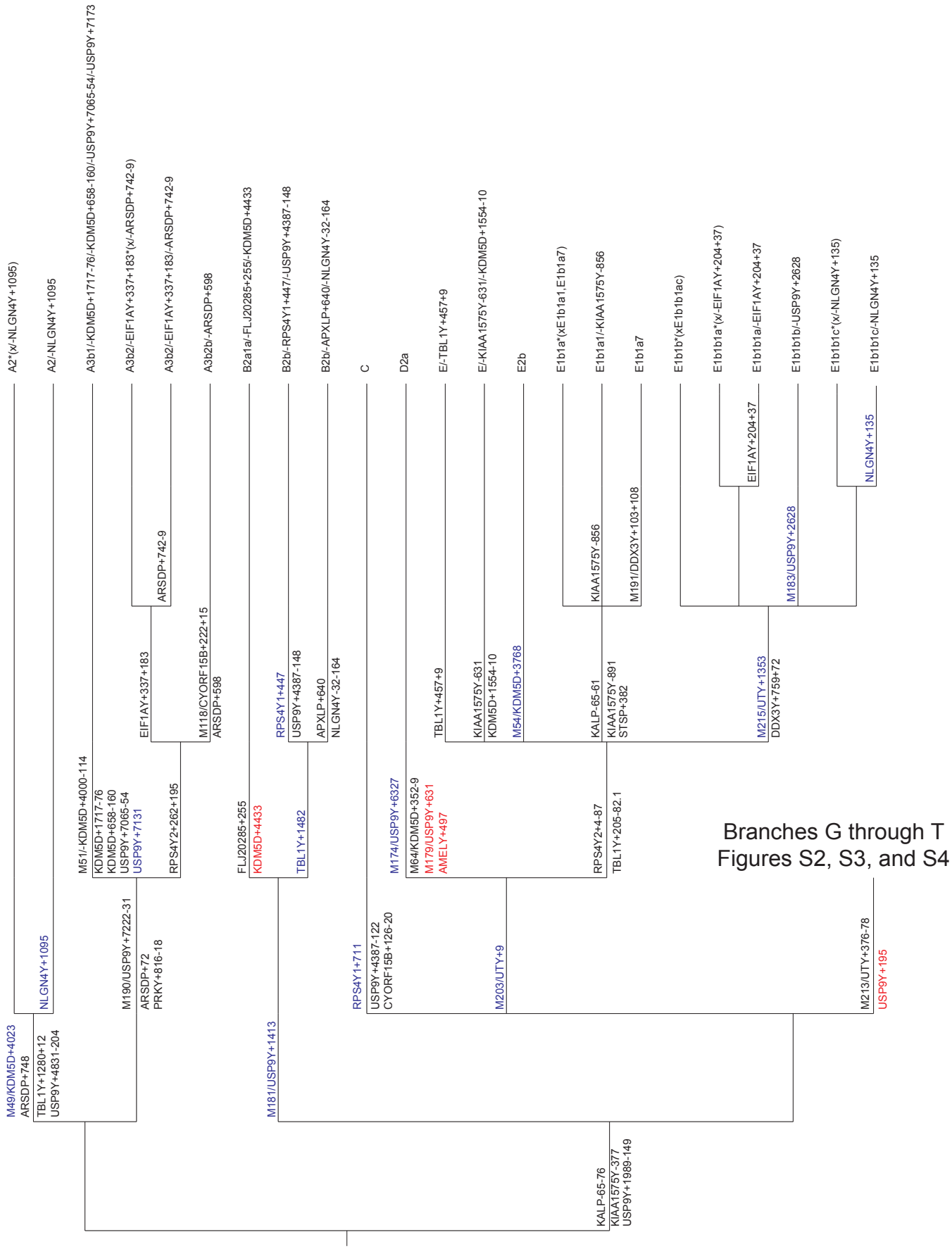
American Journal of Human Genetics, Volume *85*

Supplemental Data

Remarkably Little Variation in Proteins Encoded by the Y Chromosome's Single-Copy Genes, Implying Effective Purifying Selection

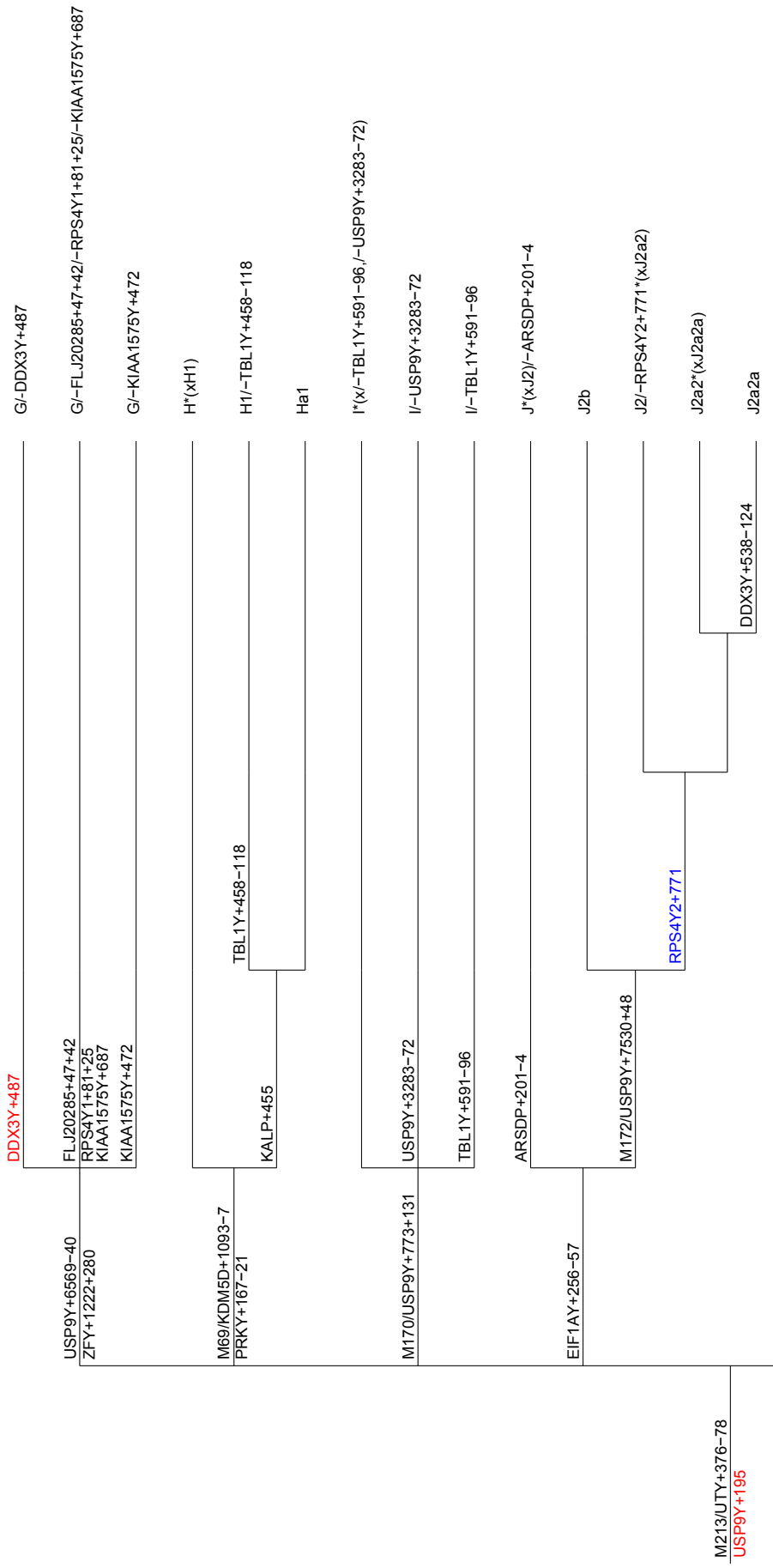
Steve Rozen, Janet D. Marszalek, Raaji K. Alagappan, Helen Skaletsky, and David C. Page

Figure S1. Genealogical Tree of Human Y Chromosomes.



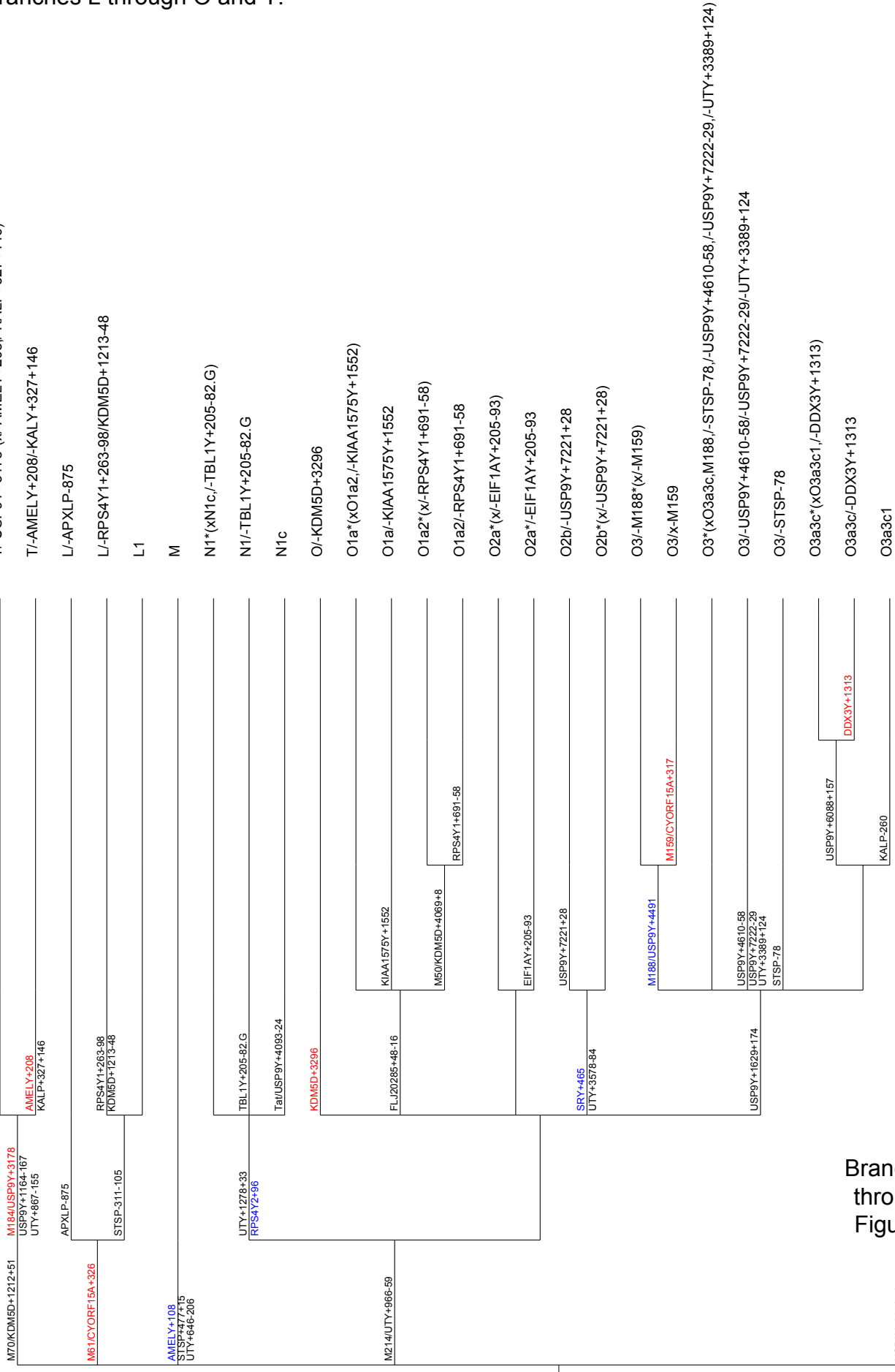
Genealogical tree of human Y chromosomes as in main text Figure 2, with addition of mnemonic variant IDs (see Table S4) and detailed haplogroup (SNP-based haplotype) designations (Karafet et al., 2008, see main text).

Figure S2. Genealogical Tree of Human Y Chromosomes Continued, Branches G through J.



Branches L through T
Figures S3 and S4

Figure S3. Genealogical Tree of Human Y Chromosomes Continued, Branches L through O and T.



Branches P through R Figure S4

Figure S4. Genealogical Tree of Human Y Chromosomes Continued, Branches P, Q, and R.

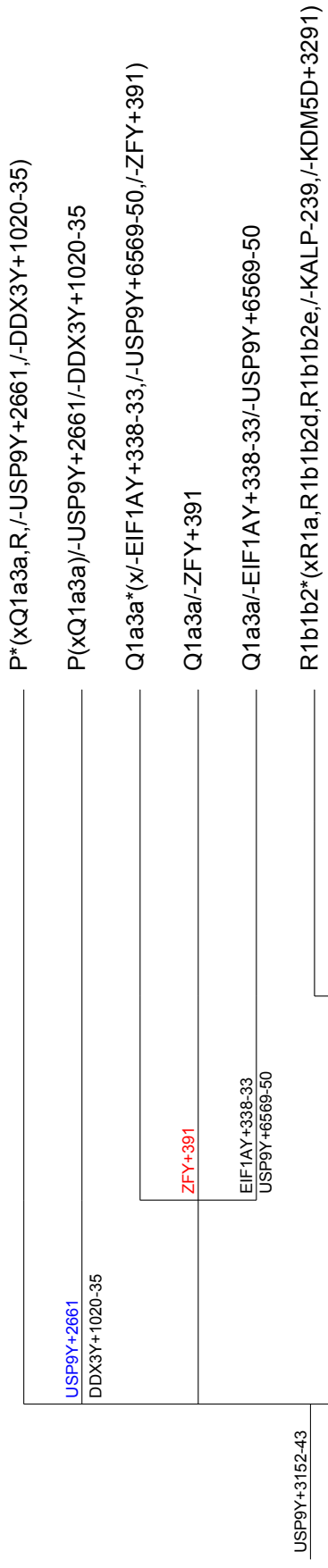


Figure S5. Alignments of 12 *USP9* Genes to Human *USP9Y* near the D65E Mutation.

```
Usp9x-isof2-Rattus-norvegicus      EQGQGDAPPQIEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
Usp9x-Mus-musculus                 EQGQGDAPPQIEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9X-isof4-Homo-sapiens           EQGQGDAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9X-isof2-Macaca-mulatta         EQGQGDAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9X-isof3-Equus-caballus         EQGQGDAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9-isof3-Monodelphis-domesti     EQGQGDAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9-isof1-Ornithorhynchus-ana     EQGQGDAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9-isof1-Gallus-gallus           EQGQGDAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9-Taeniopygia-guttata           EQGQADAPPQLEDEEPAFPHTDLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9-Danio-rerio                   DQGQGDSPTPLEEEEPAFPHTELAKLDDMINRPRWVVPVLPKGELEVLLE
USP9Y-Bos-taurus                   EQGQGDAPPQHEDEEPAFPHADLAKLDDMINRPRWVVPVLPKGELEVLLE
USP9Y-Homo-sapiens                 EQGQGDAPPQHEDEDPAFPHTELANLDDMINRPRWVVPVLPKGELEVLLE
Usp9y-Mus-musculus                 EQGQGDAPPQHEEEDPSFPHTDLAKLEDMINRSRWVVPVLPKGELEVLLE
```

These alignments show the preponderance of glutamic acid at sites corresponding to the human D65E mutation. Region shown corresponds to residues 51 through 100 in the human *USP9Y* sequence. Vertical rectangle indicates residue 65 in human *USP9Y*.

We selected sequences for this multiple alignment by first querying the NCBI refseq_protein database for best matches to human *USP9Y*. We did this using the blast server (BLASTP 2.2.21+ at http://blast.ncbi.nlm.nih.gov/Blast.cgi) using default parameters. In instances where multiple isoforms for a given species provided good matches to human *USP9Y*, we retained only the best match. We then selected the 12 best matches to human *USP9Y* and computed a multiple alignment with clustalw2 2.0.10 (ftp://ftp.ebi.ac.uk/pub/software/clustalw2/) with default parameters.

Rozen et al., Table S1. The105 samples that were resequenced and the MSY haplogroups [a] (SNP-based haplotypes) used in selecting them.

Sample identifier	Haplogroup [a] for sample selection
NA03043	A2
YCC034	A2
YCC022	A2
YCC038	A3b1
Tel-Aviv-4566	A3b2*(xA3b2b)
Tel-Aviv-4626	A3b2*(xA3b2b)
NA08688A	A3b2*(xA3b2b)
GM06342	A3b2b
PD061	B2a1a
NA10470	B2b
NA10494	B2b
WHT3552	C
WHT3474	C
NA04535	D2a
PD178	D2a
PD339	E*(xE2,E3ab)
PD383	E*(xE2,E3ab)
PD347	E2b
YCC037	E2b
NA02090	E1b1a*(xE1b1a1)
PD091	E1b1a*(xE1b1a1)
NA02064	E1b1a*(xE1b1a1)
PD399	E1b1a1
GM02091	E1b1a1
PD123	E1b1b(xE1b1b1ac)
PD196	E1b1b(xE1b1b1ac)
PD111	E1b1b1a
PD144	E1b1b1a
WHT3159	E1b1b1b
WHT3204	E1b1b1b
WHT2736	E1b1b1c
WHT3027	E1b1b1c
PD126	F*(xHT)
PD136	F*(xHT)
WHT2611	F*(xHT)
PD073	H1*(xH1a)
PD276	H*(xH1)
PD146	H1a
PD211	H1a
WHT3449	I
PD384	I
WHT1659	I
PD054	J*(xJ2)
WHT3635	J*(xJ2)
OXEN	J*(xJ2)
PD416	J2*(xJ2a2,J2b)
WHT3255	J2*(xJ2a2,J2b)

PD335	J2b
PD388	J2b
PD430	J2a2*(J2a2a)
PD437	J2a2*(J2a2a)
WHT2426	J2a2a
WHT3560	J2a2a
PD448	T
WHT3257	T
WHT3299	L*(xL1)
WHT3543	L*(xL1)
PD116	L1
PD378	L1
NA10541A	M
PD321	N1*(xN1c)
PD403	N1*(xN1c)
PD427	N1c
WHT716	N1c
PD016	O*(xO1a,O2ab,O3)
PD189	O1a*(xO1b)
PD192	O1a*(xO1b)
PD274	O1a2
PD349	O1a2
PD143	O2a
PD151	O2a
PD197	O2b
NA14819	O2b
NA11587	O3*(xO3a3c)
PD037	O3*(xO3a3c)
PD125	O3*(xO3a3c)
PD098	O3*(xO3a3c)
PD227	O3*(xO3a3c)
PD264	O3*(xO3a3c)
PD122	O3*(xO3a3c)
PD131	O3a3c*(xO3a3c1)
PD377	O3a3c*(xO3a3c1)
PD170	O3a3c1
PD284	O3a3c1
PD421	P*(xQ1a3a,R)
PD070	P*(xQ1a3a,R)
PD222	P*(xQ1a3a,R)
NA11200	Q1a3a
PD024	Q1a3a
PD386	Q1a3a
PD444	Q1a3a
WHT3242	R1a1
WHT3586	R1a1
PD041	R1*(xR1a,R1b1b2de)
PD118	R1*(xR1a,R1b1b2de)
PD329	R1*(xR1a,R1b1b2de)
GM02294	R1*(xR1a,R1b1b2de)
PD223	R1*(xR1a,R1b1b2de)
PD217	R1b1b2d
Boleth	R1b1b2d

WHT2630	R1b1b2e
PD295	R1b1b2e
PD306	R2
WHT3707	R2
NA10667	R2

[a] Haplogroup designations from Karafet et al. (2008), New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. *18* , 830-838.

Rozen et al., Table S2. Pseudogenes surveyed for DNA sequence variation.

Pseudogene	Numbers of nucleotide sites surveyed
ARSDP	2,472
APXLP	2,133
KALP	3,777
STSP	2,160
BCORP	4,709
Total	**15,251**

NOTE: Table S3 provides details of STSs (primer pairs) used to amplify pseudoexons and surrounding sequence.

Rozen et al., Table S3. GenBank accession numbers for STSs (PCR products) used for resequencing genes and pseudogenes. The GenBank entries provide primer pairs and reaction conditions.

STS Name [a]	GenBank Accession	Notes
AMELY-2	BV678971	
AMELY-3	BV678972	
AMELY-4	BV678973	
AMELY-5	BV678974	
AMELY-6-1	BV678975	b
AMELY-6-2	BV678976	b
APXLP-3	BV678977	
APXLP-4	BV678978	
APXLP-4-2	BV679205	b
APXLP-4-3	BV679206	b
APXLP-4-4	BV679207	b
APXLP-5	BV678979	
ARSDP-2	BV678980	
ARSDP-3	BV678981	
ARSDP-4	BV678982	
ARSDP-7	BV678983	
ARSDP-8	BV678984	
ARSDP-9	BV678985	
ARSDP-10	BV678986	
CYORF15A-1	BV678987	
CYORF15A-2	BV678988	
CYORF15A-3	BV678989	
CYORF15A-4	BV678990	
CYORF15B-1	BV678991	
CYORF15B-2	BV678992	
CYORF15B-3	BV678993	
CYORF15B-4	BV678994	
DDX3Y-1	BV678995	
DDX3Y-2	BV678996	
DDX3Y-3	BV678997	
DDX3Y-4	BV678998	
DDX3Y-5	BV678999	
DDX3Y-6	BV679000	
DDX3Y-7	BV679001	
DDX3Y-8	BV679002	
DDX3Y-9	BV679003	
DDX3Y-10	BV679004	
DDX3Y-11	BV679005	
DDX3Y-12	BV679006	
DDX3Y-13	BV679007	
DDX3Y-14	BV679008	
DDX3Y-15	BV679009	
DDX3Y-16	BV679010	
DDX3Y-17	BV679011	
EIF1AY-1	BV679012	
EIF1AY-2	BV679013	
EIF1AY-3	BV679014	
EIF1AY-4	BV679015	
EIF1AY-5	BV679016	

STS Name [a]	GenBank Accession	Notes
EIF1AY-6	BV679017	
EIF1AY-7	BV679018	
BCORP-9	BV679019	
BCORP-10	BV679020	
BCORP-12	BV679021	
BCORP-13	BV679022	
BCORP-14	BV679023	
KALP-2	BV679024	
KALP-4	BV679025	
KALP-5	BV679026	
KALP-6	BV679027	
KALP-7	BV679028	
KALP-10	BV679029	
KALP-11	BV679030	
KALP-12	BV679031	
KALP-13	BV679032	
KALP-14	BV679033	
BCORP-4-1	BV679034	b
BCORP-4-2	BV679035	b
BCORP-4-3	BV679036	b
BCORP-4-4	BV679037	b
BCORP-4-5	BV679038	b
BCORP-4-6	BV679039	b
NLGN4Y-1	BV679040	c
NLGN4Y-2	BV679041	
NLGN4Y-3	BV679042	
NLGN4Y-4	BV679043	
NLGN4Y-5-1	BV679044	b
NLGN4Y-5-2	BV679045	b
PRKY-2	BV679046	
PRKY-3	BV679047	
PRKY-4	BV679048	
PRKY-5	BV679049	
PRKY-7	BV679050	
RPS4Y-1	BV679051	a
RPS4Y-2	BV679052	a
RPS4Y-3	BV679053	a
RPS4Y-4	BV679054	a
RPS4Y-5	BV679055	a
RPS4Y-6	BV679056	a
RPS4Y-7	BV679057	a
RPS4Y2-1	BV679058	
RPS4Y2-2	BV679059	
RPS4Y2-3	BV679060	
RPS4Y2-4	BV679061	
RPS4Y2-5	BV679062	
RPS4Y2-6	BV679063	
RPS4Y2-7	BV679064	
SMCY-1	BV679065	a
SMCY-2	BV679066	a
SMCY-3	BV679067	a
SMCY-4	BV679068	a

STS Name [a]	GenBank Accession	Notes
SMCY-5	BV679069	a
SMCY-6	BV679070	a
SMCY-7	BV679071	a
SMCY-8	BV679072	a
SMCY-10	BV679073	a
SMCY-11	BV679074	a
SMCY-12	BV679075	a
SMCY-13	BV679076	a
SMCY-14	BV679077	a
SMCY-15	BV679078	a
SMCY-16	BV679079	a
SMCY-17	BV679080	a
SMCY-18	BV679081	a
SMCY-19	BV679082	a
SMCY-20	BV679083	a
SMCY-21	BV679084	a
SMCY-22	BV679085	a
SMCY-23	BV679086	a
SMCY-24	BV679087	a
SMCY-25	BV679088	a
SMCY-26	BV679089	a
SMCY-27	BV679090	a
SRY-flank-1	BV679091	d
SRY-flank-2	BV679092	d
SRY-flank-3	BV679093	d
SRY-flank-4	BV679094	d
SRY-1-1	BV679095	b
SRY-1-2	BV679096	b
SRY-1-3	BV679097	b
SRY-1-4	BV679098	b
SRY-1-5	BV679099	b
SRY-1-6	BV679100	b
STSP-2	BV679101	
STSP-3	BV679102	
STSP-4	BV679103	
STSP-5	BV679104	
STSP-6	BV679105	
STSP-9	BV679106	
TBL1Y-2	BV679107	c
TBL1Y-3	BV679108	c
TBL1Y-6	BV679109	
TBL1Y-7	BV679110	
TBL1Y-8	BV679111	
TBL1Y-9	BV679112	
TBL1Y-10	BV679113	
TBL1Y-11	BV679114	
TBL1Y-12	BV679115	
TBL1Y-13	BV679116	
TBL1Y-14	BV679117	
TBL1Y-15	BV679118	
TBL1Y-16	BV679119	
TBL1Y-17	BV679120	

STS Name [a]	GenBank Accession	Notes
TBL1Y-18	BV679121	
TMSB4Y-1	BV679122	
TMSB4Y-2	BV679123	
USP9Y-2	BV679124	c
USP9Y-3	BV679125	
USP9Y-4	BV679126	
USP9Y-5	BV679127	
USP9Y-6	BV679128	
USP9Y-7	BV679129	
USP9Y-8	BV679130	
USP9Y-9	BV679131	
USP9Y-10	BV679132	
USP9Y-11	BV679133	
USP9Y-12	BV679134	
USP9Y-13	BV679135	
USP9Y-14	BV679136	
USP9Y-15	BV679137	
USP9Y-16	BV679138	
USP9Y-17	BV679139	
USP9Y-18	BV679140	
USP9Y-19	BV679141	
USP9Y-20	BV679142	
USP9Y-21	BV679143	
USP9Y-22	BV679144	
USP9Y-23	BV679145	
USP9Y-24	BV679146	
USP9Y-25	BV679147	
USP9Y-26	BV679148	
USP9Y-27	BV679149	
USP9Y-28	BV679150	
USP9Y-29	BV679151	
USP9Y-30	BV679152	
USP9Y-31	BV679153	
USP9Y-32	BV679154	
USP9Y-33	BV679155	
USP9Y-34	BV679156	
USP9Y-35	BV679157	
USP9Y-36	BV679158	
USP9Y-37	BV679159	
USP9Y-38	BV679160	
USP9Y-39	BV679161	
USP9Y-40	BV679162	
USP9Y-41	BV679163	
USP9Y-42	BV679164	
USP9Y-43	BV679165	
USP9Y-44	BV679166	
USP9Y-45	BV679167	
USP9Y-46	BV679168	
UTY-1	BV679169	
UTY-2	BV679170	
UTY-3	BV679171	
UTY-4	BV679172	

STS Name [a]	GenBank Accession	Notes
UTY-5	BV679173	
UTY-6	BV679174	
UTY-7	BV679175	
UTY-8	BV679176	
UTY-9	BV679177	
UTY-10	BV679178	
UTY-11	BV679179	
UTY-12	BV679180	
UTY-13	BV679181	
UTY-14	BV679182	
UTY-15	BV679183	
UTY-16-1	BV679184	b
UTY-16-2	BV679185	b
UTY-17	BV679186	
UTY-18	BV679187	
UTY-19	BV679188	
UTY-20	BV679189	
UTY-21	BV679190	
UTY-22	BV679191	
UTY-23	BV679192	
UTY-24	BV679193	
UTY-25	BV679194	
UTY-26	BV679195	
UTY-27	BV679196	
UTY-28	BV679197	
ZFY-4	BV679198	
ZFY-5	BV679199	
ZFY-6	BV679200	
ZFY-7	BV679201	
ZFY-8	BV679202	
ZFY-9	BV679203	
ZFY-10	BV679204	

[a] For mnemonic purposes, we constructed STS names from (1) the gene or pseudogene name, (2) the exon or pseudoexon number, and (3) when multiple STSs were needed to cover an exon, the STS number within the exon. The official name of RPS4Y was subsequently changed to RPS4Y1, and the official name of SMCY was subsequently changed to JARID1D and then KDM5D.

[b] Multiple STSs needed to cover exon.

[c] Not coding.

[d] Genomic flank of SRY.

Rozen et al., Table S4. Detected sequence variants in Y-chromosomal single-copy genes and pseudogenes.

Variant ID [a]	Gene or pseudogene name	Exon number	Sequence class [b]	Sequence to left of variant	Variants	Seqence to right of variant	Mnemonic variant ID [c,d]
PAGE|S00004	AMELY	2	I	GGCACTTTATGGTGCACATG	T C	TTGAGAAGAGATGAGAAAAG	AMELY-12-43
PAGE|S00001	AMELY	4	4	TATCTTTCTCTTAAGGTGCT	C A	ACCCCTTTGAAGTGGTACCA	AMELY+108
PAGE|S00002	AMELY	4	0	ACCACCAAATCATCCCCGTG	G A	TGTCCCAACAGCACCCCCTG	AMELY+208
PAGE|S00003	AMELY	4	0	GCCTCCAATGTTCCCCCTGC	G A	GCCCCTGCCCCCCATACTTC	AMELY+497
PAGE|S00005	APXLP	4	P	GGTGGTCCCACCCGTCCTGC	A C	TCAGAGCCTGAAGACAGTTC	APXLY+640
PAGE|S00006	APXLP	4	P	ACAGCCTGGACCCCTCTGGG	C T	GCCTCTTGGTGGCCAAGTCC	APXLY-875
PAGE|S00010	ARSDP	2	P	GAATCTTTGCCGGTGCTACT	G C	TTTTTATGCTTGCTTCTGAA	ARSDP+72
PAGE|S00008	ARSDP	3	P	ATGACATTTGTGTGCTTTGC	T C	CAGAATGCCAAATATTGACC	ARSDP+201-4
PAGE|S00009	ARSDP	8	P	GGTGGGAAGGGCATAGGAGG	A G	TGGGAAGATGGGATCCATGT	ARSDP+598
PAGE|S00011	ARSDP	9	P	TCTGCTCTGTGTTCTCTTCA	C T	GGAATCAGGGTGATCGACAA	ARSDP+742-9
PAGE|S00095	ARSDP	9	P	CTTCACGGAATCAGGGTGAT	C T	GACAACCGCAGCTTGGTGCC	ARSDP+748
PAGE|S00007	ARSDP	10	P	AAGGGTAGGTGCCGCAGTGT	C T	GAAGCATTGGCAGACCCTGA	ARSDP+1089
PAGE|S00025	BCORP	4	P	CAAGTGGGCCAACCTCCAAC	C T	GTGCCTGTCAAGCAAAGCAG	KIAA1575Y+687
PAGE|S00026	BCORP	4	P	TTTCCTACCTGCCATCACCT	C T	GCTATTTCAGTCCCCACATC	KIAA1575Y-377
PAGE|S00027	BCORP	4	P	GGCTCCTGTTCCCCAGAAAC	A T	CTGTCTTTCTAGTTTGAGCA	KIAA1575Y-856
PAGE|S00106	BCORP	4	P	ACTGATCATGGACCACACAG	G A	CCTGAAACAGGAAGGGCTCC	KIAA1575Y-891
PAGE|S00107	BCORP	4	P	AAATGCTTTCAGCACAGCAG	A G	AACACTTGGCTTGGACAGGC	KIAA1575Y-631
PAGE|S00108	BCORP	4	P	AGGTGATGGACTGGATGATA	A G	TACAACAGCAAAGTTCTTTT	KIAA1575Y+472
PAGE|S00109	BCORP	4	P	CAGAGAAAGATCCCATTCCC	A C	TGAAAGAGCCTGCCACAAAG	KIAA1575Y+1552
PAGE|S00019	BCORP	10	P	CCTAGAGCAGCTGCTGCTTC	C T	AGCTGGAAAGCAGGGCTTGC	FLJ20285+47+42
PAGE|S00020	BCORP	12	P	TGCATGAGTTTAATCCCACC	G A	CACATCTCTTACTAGGCCTC	FLJ20285+48-16
PAGE|S00018	BCORP	13	P	AATAGCTCTTGGGAGTTCTA	C T	GGCAGCTCTGTCTGTGGTGA	FLJ20285+255
PAGE|S00043	CYORF15A	3	0	GCAAGCCCTAAACACTCTTT	C T	GACTCCAGAGGAGAAGCTGG	M61/CYORF15A+326
PAGE|S00096	CYORF15A	3	0	ATTATTGATGCAAGCCCTAA	A C	CACTCTTTCGACTCCAGAGG	M159/CYORF15A+317
PAGE|S00012	CYORF15B	2	I	GGTATTTACATATTTTTAAT	A T	GAATATTATGTAAAAAGTAA	CYORF15B+126+20
PAGE|S00098	CYORF15B	3	I	GAAGAGGTAAGAAGGTTTTA	C T	GTGACTAAATAATGAGTACT	M118/CYORF15B+222+15
M191	DDX3Y	2	I	TTTTCTTTACAACTTGACTA	G T	ATGAAAATATGAGATATTTT	M191/DDX3Y+103+108
PAGE|S00099	DDX3Y	5	0	TTAACTTTGAGAAATATGAT	G A	ATATACCAGTAGAGGCAACC	DDX3Y+487
PAGE|S00014	DDX3Y	7	I	AAACTAAATCAGACAGTTTA	G A	TTGGTTACTTCCATTAATAT	DDX3Y+538-124
PAGE|S00015	DDX3Y	8	I	AGCCCTGTTGACTTTTCTAA	C T	GGATGCCAGATACACCTTAT	DDX3Y+759+72
PAGE|S00029	DDX3Y	8	I	AATTCAAGGGCATTTAGAAC	A C	CTTTGTCATCTGTTAATATT	M173/DDX3Y+674-52
PAGE|S00100	DDX3Y	11	I	GTAATTAGTTTCTCAGATCT	A G	ATAATCCAGTATCAACTGAG	DDX3Y+1020-35
PAGE|S00101	DDX3Y	11	0	GCCTTATAATTTTTCAGGGA	G C	TGATTCACTTACTTTAGTGT	DDX3Y+1313
PAGE|S00102	EIF1AY	3	I	CTTTTCAATAAAAATTTGCC	G A	CAAAAAATGTCTCTGCTTTA	EIF1AY+204+37
PAGE|S00103	EIF1AY	4	I	TTTTTTTTTTCACACTTGGT	G A	AATAGTATACCTGTCTTGTG	EIF1AY+205-93
PAGE|S00016	EIF1AY	5	I	AATTTGAAAGTAACTTGTGA	A C	ACAACTGGTGATATTTTGGT	EIF1AY+256-57
PAGE|S00017	EIF1AY	5	I	TTTGGAAAAGATATTAATAA	C T	TGGAAATCTCTAATAAAAAC	EIF1AY+337+183
PAGE|S00104	EIF1AY	6	I	AATTGGTGGGACTACAGGTT	A G	TATAATCAGTATGTCTGTTA	EIF1AY+338-33
PAGE|S00044	KDM5D	4	I	TTCATCCATATTGTTTCTTA	T C	ACTTTCAGATTGTGATTGAG	M64/KDM5D+352-9
PAGE|S00061	KDM5D	7	I	CCATTTGCTCCTACCTGGCC	A G	TACCACACTCACAGTAAAAA	KDM5D+658-160
PAGE|S00045	KDM5D	10	I	GTCTTGCTGAAATATATTTT	A G	TTTCAGGAGTGTAAACAGCC	M69/KDM5D+1093-7
PAGE|S00046	KDM5D	10	I	TTCTGTTGTGGTAGTCTTAG	T G	TCTCATGGAGACATGAGTCC	M70/KDM5D+1212+51
PAGE|S00116	KDM5D	10	I	TAGTTCTCATGGAGACATGA	G A	TCCAAAGTATAGTGGGCTAT	KDM5D+1213-48
PAGE|S00117	KDM5D	12	I	ACTTGACGTGTCTTCTGTCT	A G	CGTGAGCAGGGGTGAGCCGA	KDM5D+1554-10
PAGE|S00058	KDM5D	14	I	TCCTACATCTGTTACAGCAT	C G	TAGAGACATCTTTGAGTTCA	KDM5D+1717-76
PAGE|S00059	KDM5D	23	0	TGCAGACGCTGGCTCAGACA	G A	CACCAAGCGTAGCCGGTGGA	KDM5D+3296
PAGE|S00119	KDM5D	23	4	CCGTGTGCAGACGCTGGCTC	A G	GACAGCACCAAGCGTAGCCG	KDM5D+3291
M50	KDM5D	24	I	GGAAAGGGCTCTGGTAAGAC	A G	GGTGTGGTTTGGGTAGGCTG	M50/KDM5D+4069+8
M54	KDM5D	24	4	GTTGCCCTGCAGAGGCTGCC	C T	GTGCGGCTGCCTGAGGGTGA	M54/KDM5D+3768
PAGE|S00041	KDM5D	24	4	CAAGGGCTGCTGGAGAATGG	A G	GACAGTGTGACCAGTCCTGA	M49/KDM5D+4023
PAGE|S00042	KDM5D	24	I	CTCTGATCCCTGTTGGAAGC	C T	TGTGTCTACTCTGCTTCAGG	M51/KDM5D+4000-114
PAGE|S00060	KDM5D	24	0	GATTATGTCTCAGGTGGGCC	G A	AGAAGAAGAACATTATCAGG	KDM5D+4433
PAGE|S00023	KALP	7	P	TGGGGTCTCCCCGAGGAGCC	G A	GACATCCTTGTGCATCACTA	KALP-260
PAGE|S00105	KALP	7	P	GACATCCTTGTGCATCACTA	C G	AATGTCTTTTGGAGCTGGAC	KALP-239
PAGE|S00024	KALP	11	P	ATCCTTATTGGATGGAATAT	G T	ATTTCAATTCTTGCCAGTGT	KALP-65-76
rs56217212	KALP	11	P	AATATTATTTCAATTCTTGC	C A	AGTGTTTTTTAATCCCTAAT	KALP-65-61
PAGE|S00021	KALP	12	P	TTCTTCTGAGTGTATCTTTT	G C	CAAGTATTAACACAACACAA	KALP+327+146
PAGE|S00022	KALP	13	P	CCAGATGTGGAAGCTCCCAC	C T	CTCTTCAGCACACAGTGAGT	KALP+455
PAGE|S00048	NLGN4Y	3	I	ATGTAATAGCATTAGTCAAG	T A	TACTTAGAAACTCTTATCAA	NLGN4Y-32-164
PAGE|S00047	NLGN4Y	4	4	TCTGTAGGGTTTTTAAGTAC	C T	GGTGACCAGGCAGCAAAAGG	NLGN4Y+135
PAGE|S00110	NLGN4Y	5	4	TGGACGAACTTCGCCAAAAC	T C	GGGTACGTTCTTCTTCATGT	NLGN4Y+1095
PAGE|S00049	PRKY	2	I	CAGGAGACTTAATGACCTCT	G A	TCGATTTCTTCTCTCTCCAG	PRKY+167-21
PAGE|S00050	PRKY	7	I	ATTCTGTACACTGACGACAC	G A	TGTTGTTTCTTTTGAAGAAC	PRKY+816-18
PAGE|S00052	RPS4Y1	5	I	TATAACTTTGTTTTTTGTGG	T G	TTTTTTTTGTTTTTTTTAGT	RPS4Y1+81+26
PAGE|S00113	RPS4Y1	7	I	GACTCCTAGGCTGCTAAAAC	G A	TTCTCAGGTTCAACCTGGCA	RPS4Y1+263-98
PAGE|S00114	RPS4Y1	8	2	GATGCTCGAACCATCCGCTA	C T	CCAGATCCTGTCATCAAGGT	RPS4Y1+447
PAGE|S00051	RPS4Y1	10	4	GGCAATAAACCTTGGATTTC	C T	CTGCCCAGGGGAAAGGGCAT	RPS4Y1+711
PAGE|S00115	RPS4Y1	10	I	ATGTGTGTTTTGGTGGGATG	T C	TGTTTTTCTCTCCTCCCTTT	RPS4Y1+691-58
PAGE|S00054	RPS4Y2	2	I	CTCCTCCGATCCGGTATTAC	C T	CGTTAGCAGTTGTTACAGCT	RPS4Y2+4-87
PAGE|S00053	RPS4Y2	3	I	AAGTGAAAAGCTGGAAAGGG	G A	TTGGTAATGTAATGGGATGA	RPS4Y2+262+195
PAGE|S00056	RPS4Y2	3	4	CTACAGGCACCTCGTCCATC	G A	ACAGGTCCTCACAAGCTGAG	RPS4Y2+96
PAGE|S00055	RPS4Y2	7	4	GAGAGAGACAAGAGGCTGGC	T C	GCCAAACAGAGCAGTGGCTA	RPS4Y2+771

Variant ID [a]	Gene or pseudogene name	Exon number	Sequence class [b]	Sequence to left of variant	Variants	Seqence to right of variant	Mnemonic variant ID [c,d]
PAGE|S00063	SRY	1	2	CCCGCTTCGGTACTCTGCAG	C T	GAAGTGCAACTGGACAACAG	SRY+465
PAGE|S00068	STSP	3	P	ACGTTGGGGGTTGCAGTGAG	C T	GGAGATCGCACTACTGCACT	STSP-311-113
PAGE|S00121	STSP	3	P	GGTTGCAGTGAGCGGAGATC	G A	CACTACTGCACTCCATCCTG	STSP-311-105
PAGE|S00069	STSP	5	P	CCCCCATTTGTCTTCTCAGA	G T	AAATGGCACATTGGGATAAG	STSP-78
PAGE|S00066	STSP	6	P	GCTTGTCTTGTCCTACCTCC	A G	CGTGCACATGGCCTTGTTCT	STSP+382
PAGE|S00067	STSP	6	P	GTGTGGGTATGTCTCTCCTC	G A	GTGAATACTTAGAAAGCTGC	STSP+477+15
PAGE|S00122	TBL1Y	3	I	GGCTAATCATTTTTTTTCCT	C T	TTTATAGGCATTGATGTCCT	TBL1Y-265-8
PAGE|S00073	TBL1Y	8	I	TATGCTCCACTTCCCCGTGC	C G T	TCTCCTGCCCAGAGAGTAGA	TBL1Y+205-82
PAGE|S00074	TBL1Y	8	I	CACATGAAATCAGTGAGTGC	G A	CAGGCTCTGGAAGTTTGGTG	TBL1Y+457+9
PAGE|S00075	TBL1Y	9	I	GCCACAATTTAGTTAGACTC	G A	TTAGAGGTAGTGTTACTTAG	TBL1Y+458-118
PAGE|S00076	TBL1Y	10	I	TGATAAAGACAGAAGAGCAC	A G	TGCCTACAATGTCACCTGGC	TBL1Y+591-96
PAGE|S00071	TBL1Y	16	I	GTTGGCAAGGTAAGGGCCGG	C A	AGCACAACTGGTACAGCTCC	TBL1Y+1280+12
PAGE|S00072	TBL1Y	18	4	CACAGCTACCAAGGCACTGG	C A	GGTATCTTCGAGGTGTGCTG	TBL1Y+1482
PAGE|S00080	USP9Y	4	2	CCACAGCATGAAGATGAAGA	T G	CCTGCATTTCCACATACTGA	USP9Y+195
PAGE|S00031	USP9Y	7	0	CTGAAGATGAATTATTTGCT	C T	GTTCTTCAGATCCTCGATCA	M179/USP9Y+631
M70	USP9Y	8	I	TTACTTAAAAATCATTGTTC	A C	TTTTTTTCAGTGTGGGTTGT	M170/USP9Y+773+131
PAGE|S00078	USP9Y	8	I	GGTTTAATAAATGATTAGTT	G C	TTTGAAACTGTTGAGGTAAC	USP9Y+1164+167
PAGE|S00032	USP9Y	12	2	CAACTTGATCATCTTTTTGA	T C	TGCTTTAAGGTAGTAGCTTG	M181/USP9Y+1413
PAGE|S00079	USP9Y	13	I	TCTCGGCTTACTGCAAGCTC	C T	GCCTCTCGGGTTCATGCCAT	USP9Y+1629+174
PAGE|S00081	USP9Y	17	I	CACATTTTAAATTCCAGATA	C T	GTGTTTAACAAATTATTTTG	USP9Y+1989-149
PAGE|S00033	USP9Y	19	4	CACAAGGAAAGAATGATTCT	A C	CCTATGTCGAGGTTTGTGTG	M183/USP9Y+2628
PAGE|S00082	USP9Y	20	4	GCATTTCGTGGCAAACACCT	C T	TCTCTTATAGTTCGGTTTCC	USP9Y+2661
PAGE|S00034	USP9Y	23	0	CAGCTGTAGAAAAATTACGA	G A	CTGTTTGTTTGGACCATGCA	M184/USP9Y+3178
PAGE|S00083	USP9Y	23	I	ATGTGAAACATTAATAACAC	A G	TGAGTTACACTTTATTTTAG	USP9Y+3152-43
PAGE|S00123	USP9Y	23	I	ATTTCTTAGCAATGATCAGA	G A	GAGAAATAGATGTTACTAAG	USP9Y+3283-72
PAGE|S00084	USP9Y	25	2	ATTCCTAATCCCTCATCCGA	G A	TGCGTACTTAGAAATGAGTC	USP9Y+3636
PAGE|S00070	USP9Y	29	I	AGTGTAGACTTGTGAATTCA	T C	GTTGTTTTAATTTAATATTT	Tat/USP9Y+4093-24
M188	USP9Y	31	4	ATTCCAGTCTGTAGTTCACC	C T	GTTACCATCAATGCCGGTTT	M188/USP9Y+4491
PAGE|S00085	USP9Y	31	I	TTTTTGTGGGGTTTTGTTTT	G T	AGATATAATGAGATATTTAA	USP9Y+4387-122
PAGE|S00086	USP9Y	31	I	TTACATTTATTAATTGTACT	C T	GGGTTTTTTTGTGGGGTTTT	USP9Y+4387-148
PAGE|S00124	USP9Y	32	I	TATTAATTTGTTAGCCTGAA	A C	ATGGTGAACAGAATCACTGT	USP9Y+4610-58
PAGE|S00087	USP9Y	33	I	AATGTATTTTTTTAACTATT	C T	TTTTATGATAACTTATCTGT	USP9Y+4831-204
PAGE|S00125	USP9Y	36	I	TATATGAATGTGTTGGCTTT	T C	TTGGTATAGTTATTTTAAAA	USP9Y+6088+157
PAGE|S00030	USP9Y	38	4	GAATACCTTCTGGAGTGCCC	T C	AGTGCAGAAGTGAGGGGTGC	M174/USP9Y+6327
PAGE|S00126	USP9Y	40	I	TTTTAATTATTGCAGTACTT	A C	CCCTCTTAGTTTTTTTTTCT	USP9Y+6569-50
rs13304344	USP9Y	40	I	TGCAGTACTTACCCTCTTAG	T C	TTTTTTTTCTACATATTATT	USP9Y+6569-40
PAGE|S00035	USP9Y	43	I	CTGATACTGAAAATCATTCT	A G	AATTCTAAATAGTTTTATTT	M190/USP9Y+7222-31
PAGE|S00088	USP9Y	43	I	ATGGCACATAATTAGGAACT	C G	AAATGTTAGCTACTATTGGA	USP9Y+7065-54
PAGE|S00089	USP9Y	43	4	TTCGATACAATACAGCGCTC	G A	AAGAATCACTATCAAAAACG	USP9Y+7131
PAGE|S00090	USP9Y	43	I	TTTTTCTCTTATAAATTTGT	A G	GAAACCTCTGTCACAAGTAA	USP9Y+7221+28
PAGE|S00127	USP9Y	43	I	GATACTGAAAATCATTCTAA	A G	TTCTAAATAGTTTTATTTTT	USP9Y+7222-29
PAGE|S00028	USP9Y	45	I	CCAAACCCATTTTGATGCTT	T G	ACTTAAAAGGTCTTCAATTA	M172/USP9Y+7530+48
PAGE|S00036	UTY	1	4	TTTGTTGTTTCCATGAAATC	C G	TGCGCAGTGTCGCTCACTAC	M203/UTY+9
PAGE|S00037	UTY	3	I	AAGGTATTGTTATTCTCTTT	T C	TAAATTTCTTGCTTGACTTA	M207/UTY+325+18
PAGE|S00038	UTY	5	I	AACTTAAAACATCTCGTTAC	A G	TGACTTCTTATTAATATATG	M213/UTY+376-78
PAGE|S00093	UTY	9	I	AACATTATTAGTATGTAAAT	C A	TTCATTGCTAGCTATTTTAA	UTY+646-206
PAGE|S00129	UTY	11	I	ATTAAAATAAATCATATAAT	G A	TACATCGACAATGTAAAATC	UTY+867-155
PAGE|S00039	UTY	12	I	AGACACTGTCTGAAAACAAC	A G	GAAAAATAAACGAACGAAAG	M214/UTY+966-59
PAGE|S00091	UTY	13	I	TGACGATCTTTCCCCCAATT	C T	GAAAGCAGTAGAAACAGTAA	UTY+1278+33
PAGE|S00040	UTY	14	2	CAGCTGGAACAGTTAGAAAG	T C	CAGTTTGTCTTAATGCAGCA	M215/UTY+1353
PAGE|S00130	UTY	23	I	ACTGCTCAGCAGAATTTTTT	T A	AAAATGATTTGATTTCAAAT	UTY+3389+124
PAGE|S00092	UTY	25	I	AGAATTGGTATTGCAGTTGT	C T	ACCTGAGAGGTGATAAATGA	UTY+3578-84
PAGE|S00131	ZFY	5	0	TTACTTCAACCTCAATGTCT	A G	TGCCAGAACATGTTTTAACG	ZFY+391
PAGE|S00094	ZFY	9	I	ATAAAGCAGGTATAATTTAC	C T	GAGAAGTGGAAGAAGTACCT	ZFY+1222+280

[a] dbSNP submitter ID, dbSNP rs#, or YCC "M" ID.

[b] 0=coding non-degenerate; 2=coding twofold degenerate; 4=coding fourfold degenerate; I=intron; P=pseudogene.

[c] Mnemonic identifier, which is also used in Figures S1 through S4.

The purpose of these identifiers is to make it somewhat easier to determine at a glance the gene in which a particular variant occurs and to determine whether it is coding or intronic. However, the information submitted to dbSNP, rather than the positions represented in the mnemonic identifiers, provide the definitive references for the variants.

The mnemonic identifiers have three possible forms: [M*m*/]*G*+*P*, [M*m*/]*G* +*P*+I_{after}, or [M*m*/]*G*+*P*-I_{before}, where

[M*m*/] is an optional YCC “M" ID for previously named variants (main text refs. 2 and 3) that we detected in our survey, followed by the character “/”,

G is a gene or pseudogene name,

P is a nucleotide position in *G*’s coding sequence,

+I_{after} indicates a position in the intron following the exon that ends at *P*, with position +1 assigned to the first nucleotide after the end of the exon, and

-I_{before} indicates a position in the intron preceding the exon that starts at *P*, with the position of the variant reckoned backward (5’) from the exon starting at *P* and with position -1 assigned to the last intronic nucleotide before the exon.

[d] FLJ20285 and KIAA1575Y represent parts of BCORP, which we did not initially recognize as homologous to a single X-chromosome gene.

Rozen et al., Table S5. The human Y-chromosome's X-degenerate genes show much lower variability at non-degenerate sites than at four-fold degenerate sites, in introns and in pseudogenes. This table is similar to Table 3 in the main text, except that rather than examining non-synonymous and synonymous sites, it examines non-degenerate sites—coding nucleotide sites at which any substitution would lead to a change in the encoded amino acid—and four-fold degenerate sites—coding sites at which any substitution would leave the amino acid unchanged. This analysis excludes two-fold degenerate sites, at which two of the three possible nucleotide substitutions would change the encoded amino acid while the third would leave it unchanged. For rows other than those with *P* values, the columns for introns and pseudogenes contain the same data as in main text Table 3.

	Non-degenerate	**Four-fold degenerate**	**Intron**	**Pseudogene**	**Four-fold degenerate, intron, & pseudogene**
Variant nucleotides	11	16	64	29	109
Invariant nucleotides	17,472	3,760	40,265	15,222	59,247
Total nucleotides	17,483	3,776	40,329	15,251	59,356
Proportion of variant sites	6.29×10^{-4}	4.24×10^{-3}	1.59×10^{-3}	1.90×10^{-3}	1.84×10^{-3}
P-values of proportions of variant versus invariant sites (Fisher's exact test, two sided)					
Non-degenerate vs. …		1.7×10^{-6}	2.4×10^{-3}	1.2×10^{-3}	1.8×10^{-4}
Four-fold degenerate vs. …			1.1×10^{-3}	1.4×10^{-2}	
Intron vs. …				0.42	
Nucleotide diversity	2.62×10^{-5}	2.34×10^{-4}	1.22×10^{-4}	9.8×10^{-5}	1.23×10^{-4}
P-values of differences in diversities (Wilcoxon rank sum test, two sided)					
Non-degenerate vs. …		1.6×10^{-8}	3.3×10^{-3}	1.0×10^{-3}	3.8×10^{-4}
Four-fold degenerate vs. …			2.5×10^{-4}	8.2×10^{-2}	
Intron vs. …				0.42	