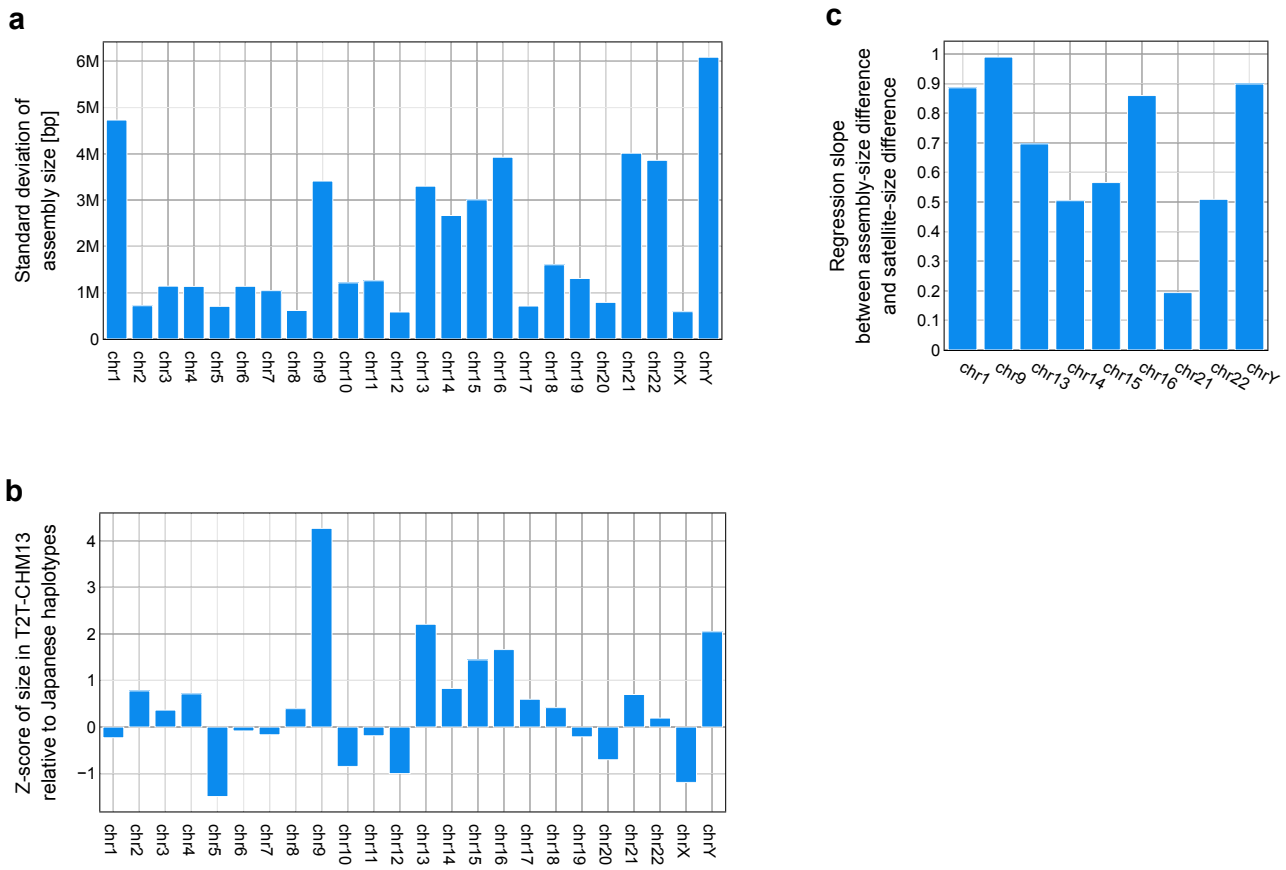
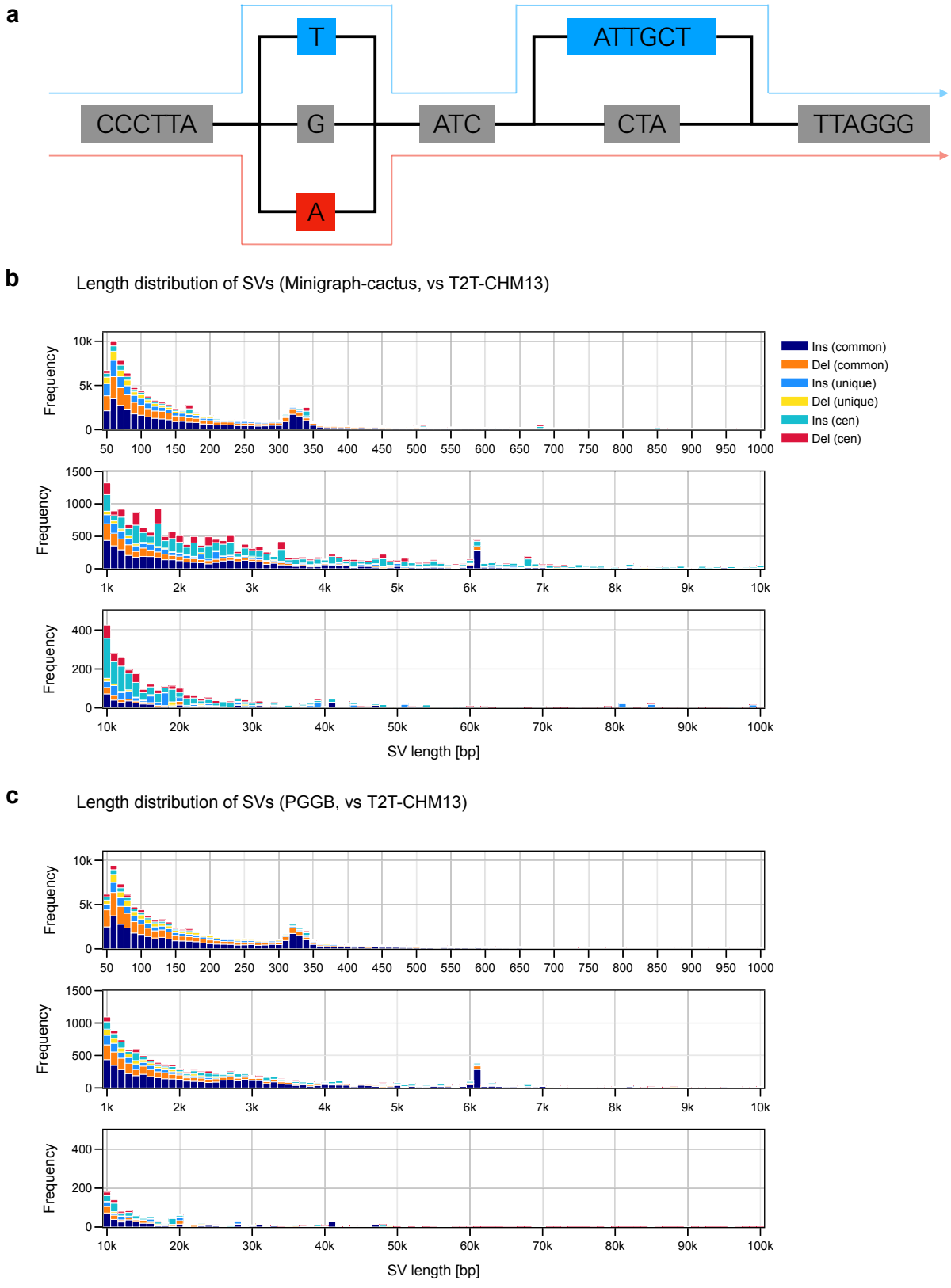


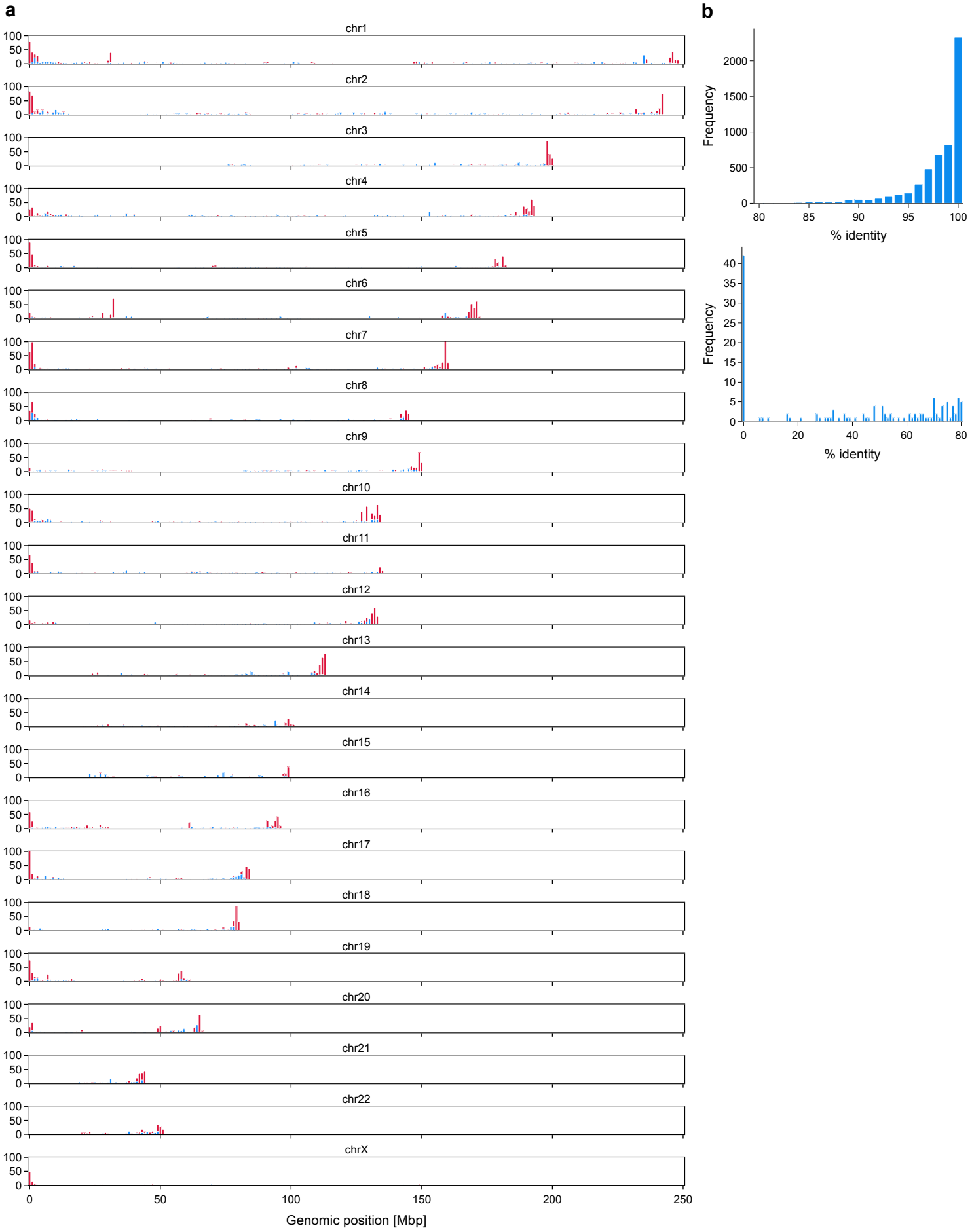
Supplementary Fig. 1 | Assembly size stats of the 20 Japanese haplotypes.

a, Standard deviations of chromosome lengths across the 20 Japanese haplotypes. **b**, Z-scores for each chromosome of the T2T-CHM13 reference, calculated relative to the Japanese haplotype distribution. Positive values indicate that T2T-CHM13 is longer than the Japanese mean, and negative values indicate it is shorter. **c**, Regression slopes from a zero-intercept linear model between the total assembly-size difference and the difference in satellite array size (sum of alpha satellite and HSat1/2/3) for chromosomes with large size variability (standard deviation >2 Mbp), quantifying how much of the assembly-size difference is explained by the difference in satellite size.



Supplementary Fig. 2 | SVs found from our Japanese haplotypes.

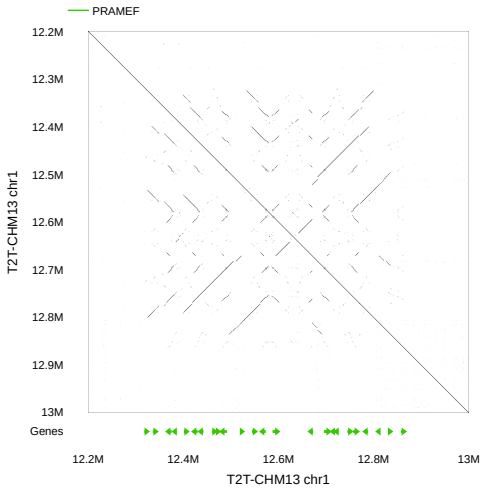
a, A schematic figure illustrating the identification of population-specific variations from a pangenome graph constructed with a reference genome and one haplotype from each of two populations. Gray nodes represent reference genome sequences mostly shared by the two individual haplotypes. Colored nodes represent population-specific sequences that diverge from the reference genome path. The first branch in the graph depicts a polymorphic SNV site where both two individual haplotypes carry distinct variants. The second branch is a longer variant unique to one population. Colored thin lines trace the path of each individual haplotype. **b–c**, The length distributions of the SVs found from the Japanese pangenome graph constructed with two different methods, Minigraph-cactus (**b**) and pggb (**c**). SVs were called based on the path of T2T-CHM13 as the reference genome. Maximum insertion and deletion SV sizes detected were 97 kbp and 99 kbp (minigraph-cactus) and 97 kbp and 59 kbp (pggb), respectively. SVs were stratified with three categories: 1) those whose counterparts were found by the other method based on Truvari (common), 2) those whose counterparts were not found (unique), and 3) those within centromeres.



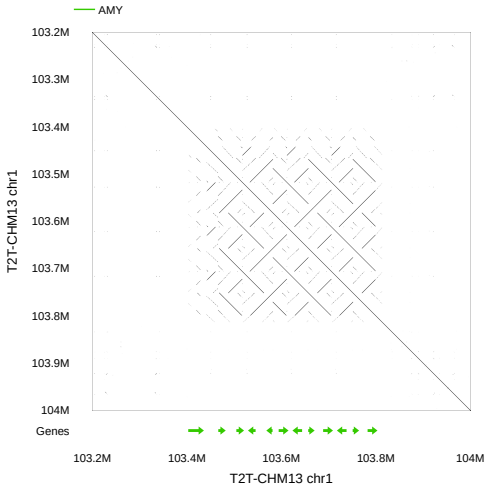
Supplementary Fig. 3 | SVs found only from our Japanese haplotypes.

a, Positional distribution of the counts of Japanese-specific SVs per 1 Mbp bin. Centromeres were excluded. Red are SVs in either subtelomeric regions (1 Mbp from each chromosome end) or segmental duplications (with 100 kbp flanking regions), and blue are SVs in other regions. **b**, Sequence identities of SVs unique to our Japanese haplotypes aligned to the pangenome graph of the HPRC Year-1 and CPC haplotypes.

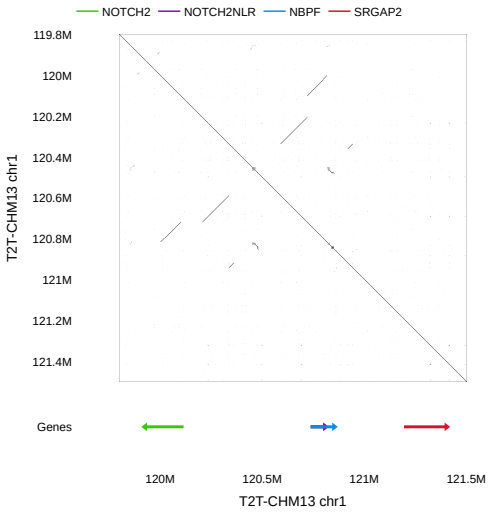
PRAMEF, T2T-CHM13 @ chr1:12,200,001-13,000,000



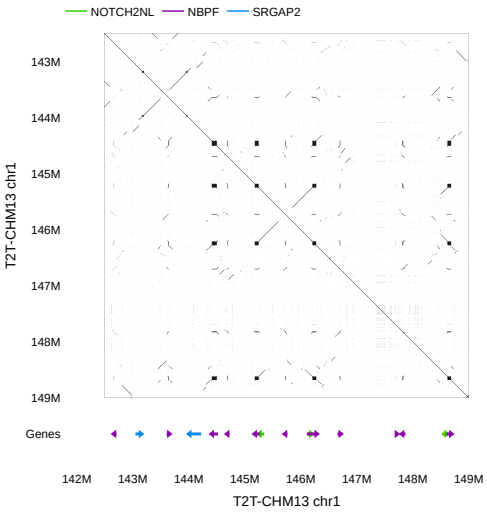
AMY, T2T-CHM13 @ chr1:103,200,001-104,000,000



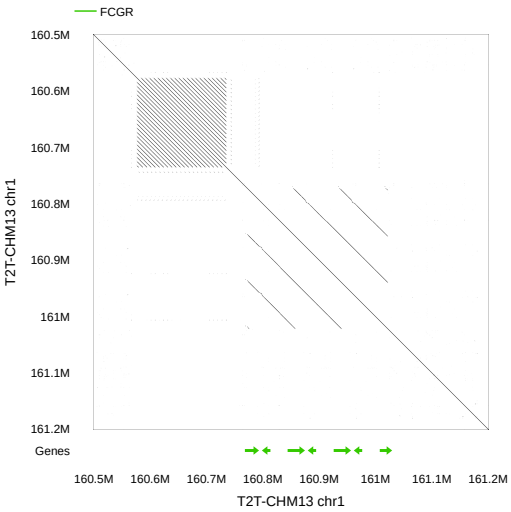
NOTCH2-NOTCH2NLR-SRGAP2, T2T-CHM13 @ chr1:119,800,001



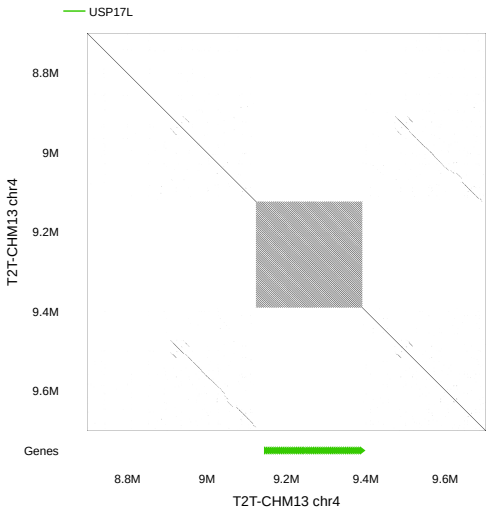
NOTCH2NL-NBPF-SRGAP2, T2T-CHM13 @ chr1:142,500,001-149,



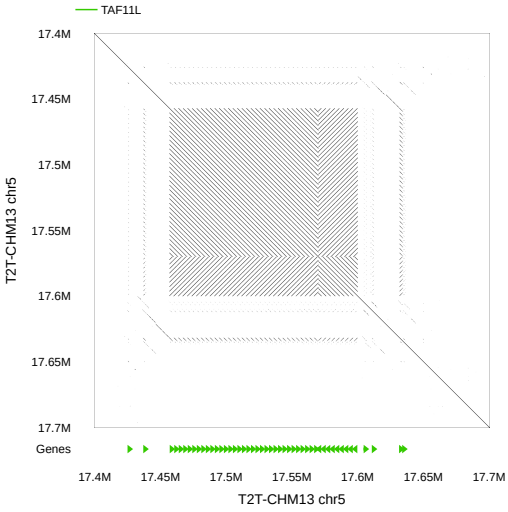
FCGR, T2T-CHM13 @ chr1:160,500,001-161,200,000



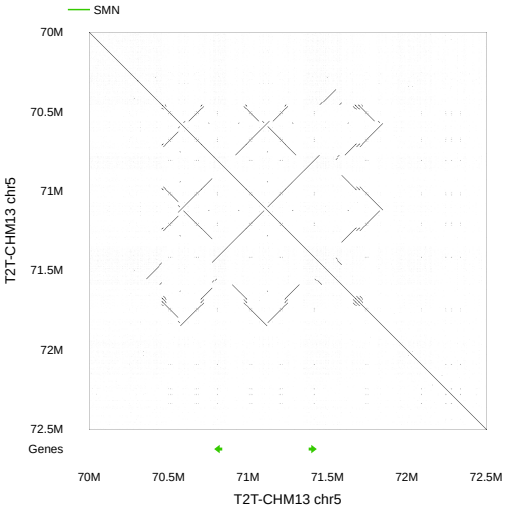
USP17L, T2T-CHM13 @ chr4:8,700,001-9,700,000



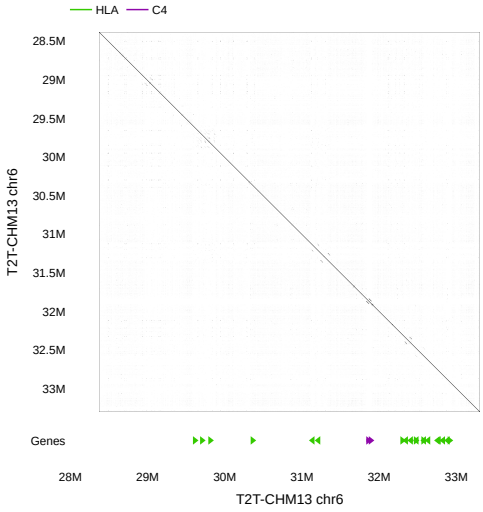
TAF11L, T2T-CHM13 @ chr5:17,400,001-17,700,000



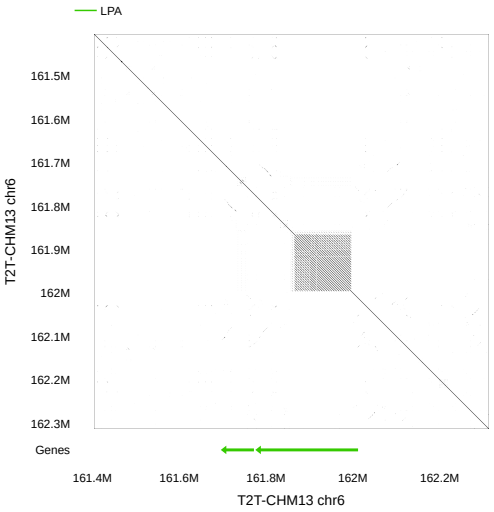
SMN, T2T-CHM13 @ chr5:70,000,001-72,500,000



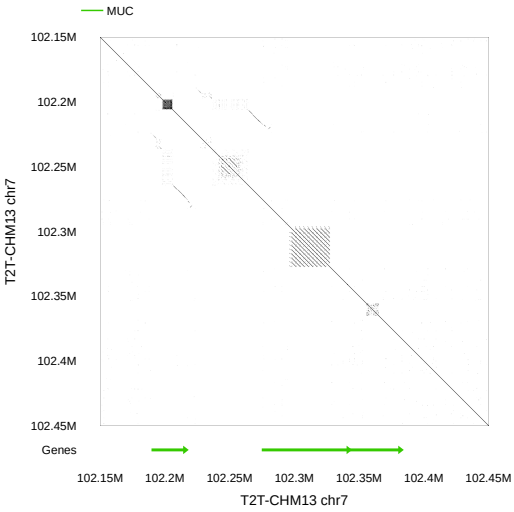
HLA, T2T-CHM13 @ chr6:28,381,551-33,301,940



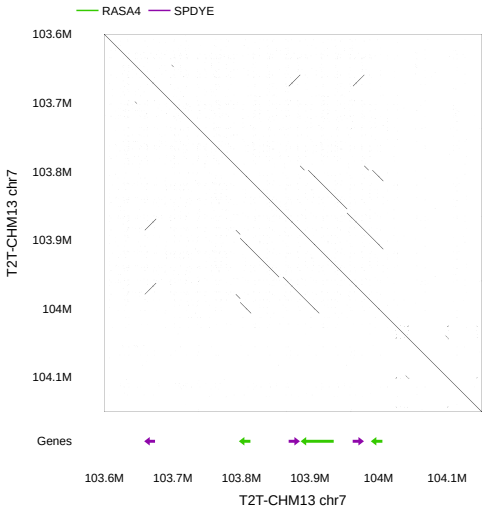
LPA, T2T-CHM13 @ chr6:161,403,623-162,311,762



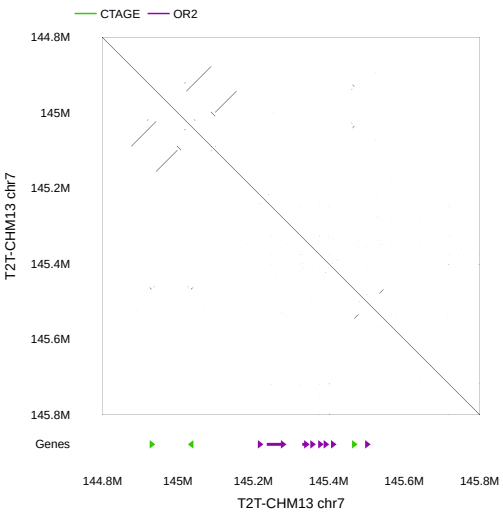
MUC3A, T2T-CHM13 @ chr7:102,150,001-102,450,000



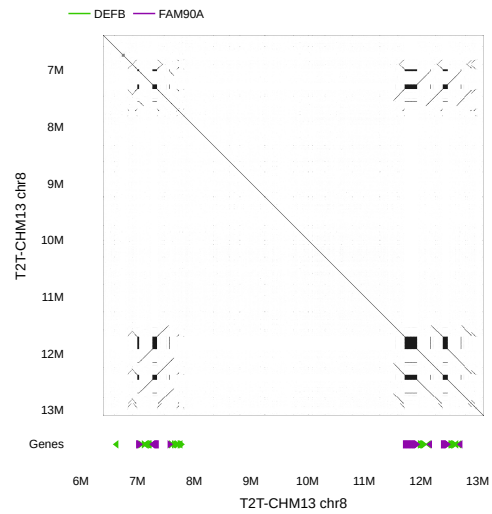
RASA4-SPDYE, T2T-CHM13 @ chr7:103,600,001-104,150,000



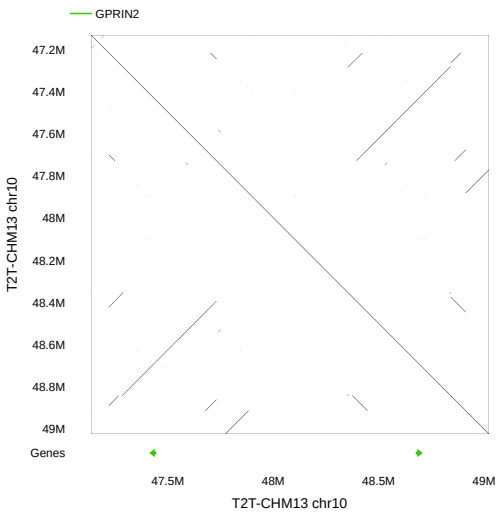
CTAGE-OR2, T2T-CHM13 @ chr7:144,800,001-145,800,000



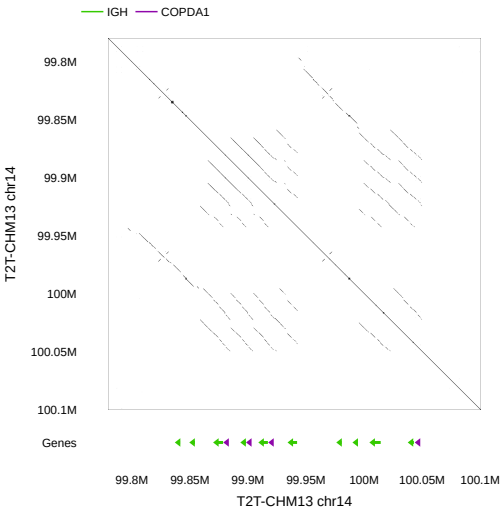
DEFB-FAM90A, T2T-CHM13 @ chr8:6,400,001-13,100,000



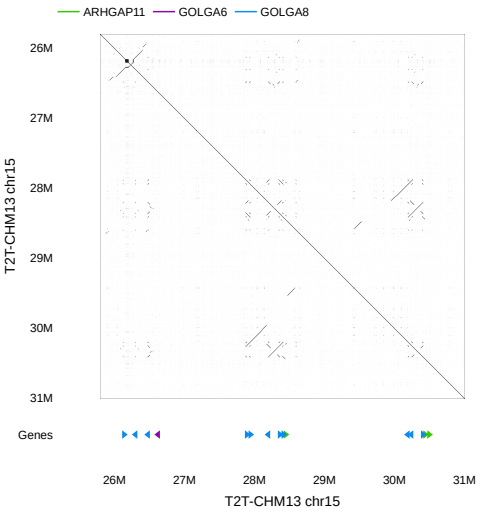
GPRIN2, T2T-CHM13 @ chr10:47,135,400-49,019,220



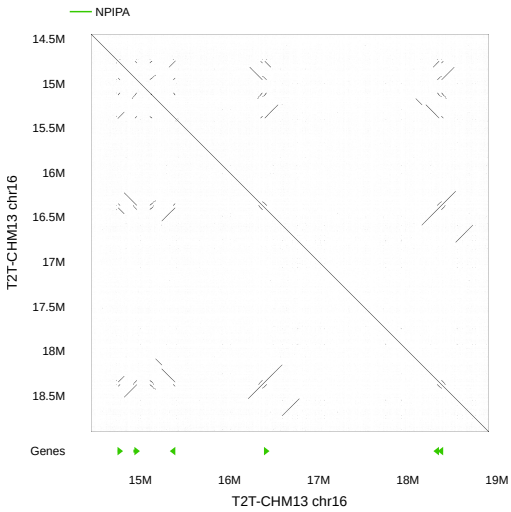
IGH-COPDA1, T2T-CHM13 @ chr14:99,780,001-100,100,000



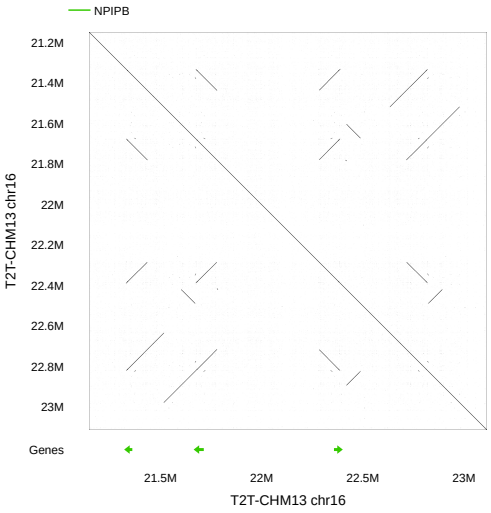
ARHGAP11-GOLGA6-GOLGA8, T2T-CHM13 @ chr15:25,800,001-3



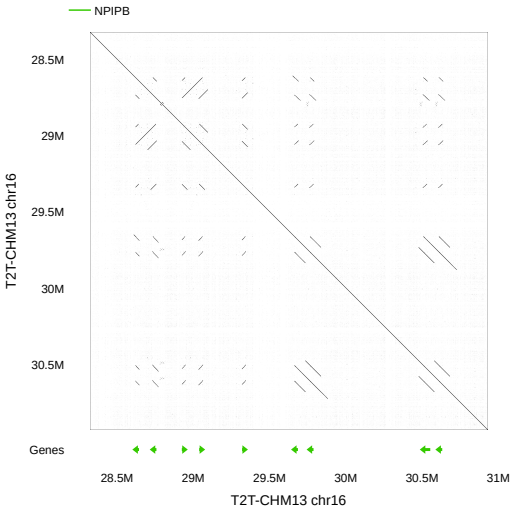
NPIPA, T2T-CHM13 @ chr16:14,446,468-18,900,000



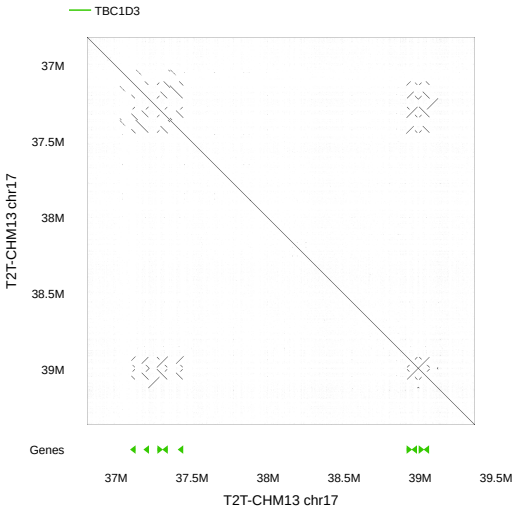
NPIPB3-5, T2T-CHM13 @ chr16:21,150,001-23,112,084



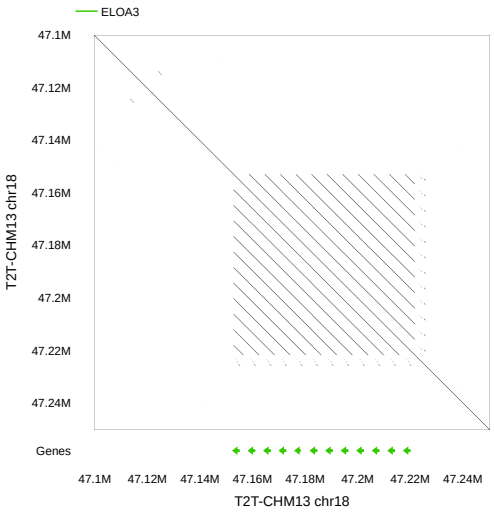
NPIPB6-13, T2T-CHM13 @ chr16:28,323,072-30,926,372



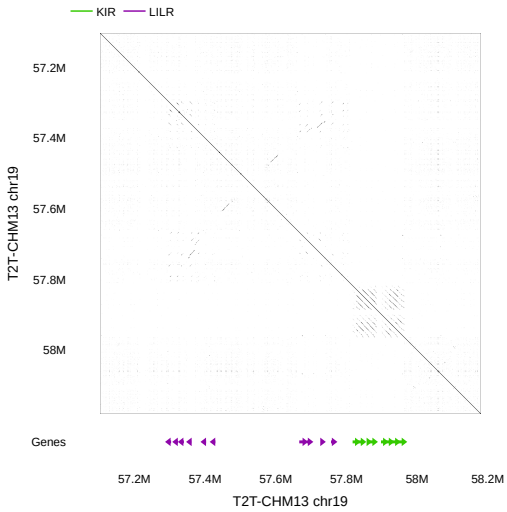
TBC1D3, T2T-CHM13 @ chr17:36,813,574-39,355,611



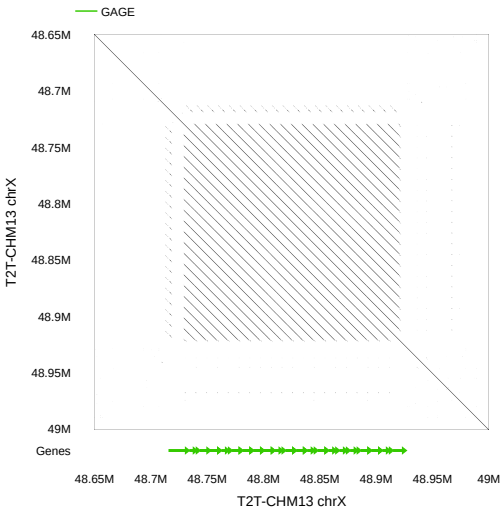
ELOA3, T2T-CHM13 @ chr18:47,100,001-47,250,000



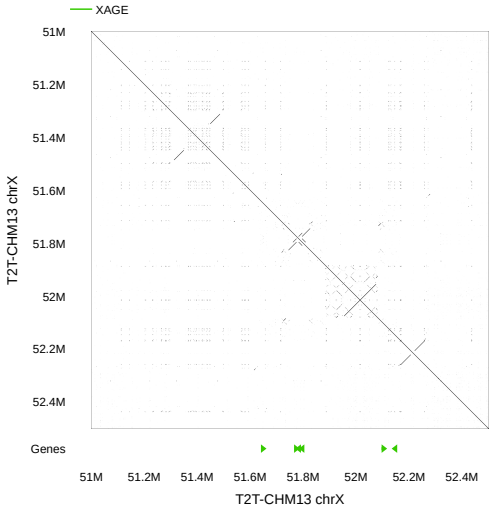
KIR-LILR, T2T-CHM13 @ chr19:57,104,136-58,178,808



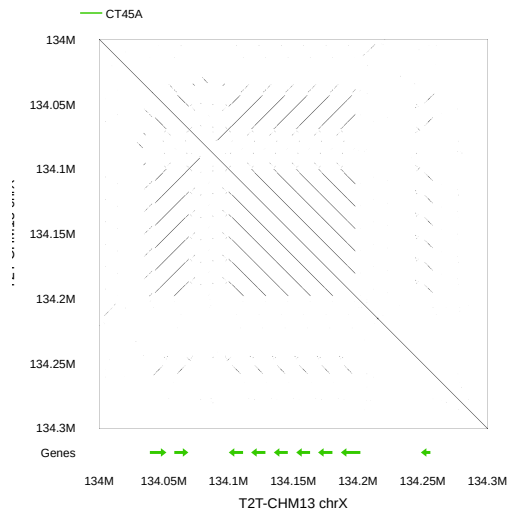
GAGE, T2T-CHM13 @ chrX:48,650,001-49,000,000



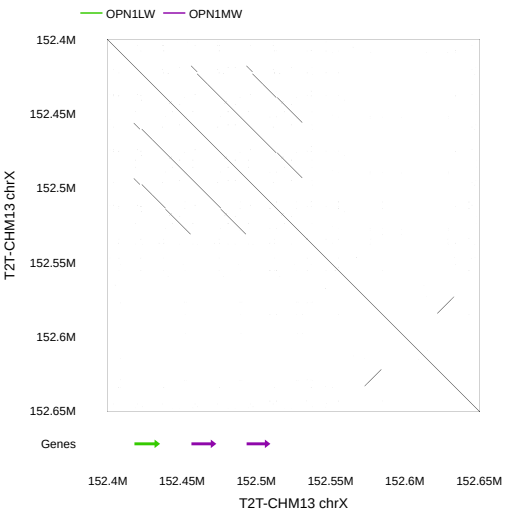
XAGE, T2T-CHM13 @ chrX:51,000,001-52,500,000

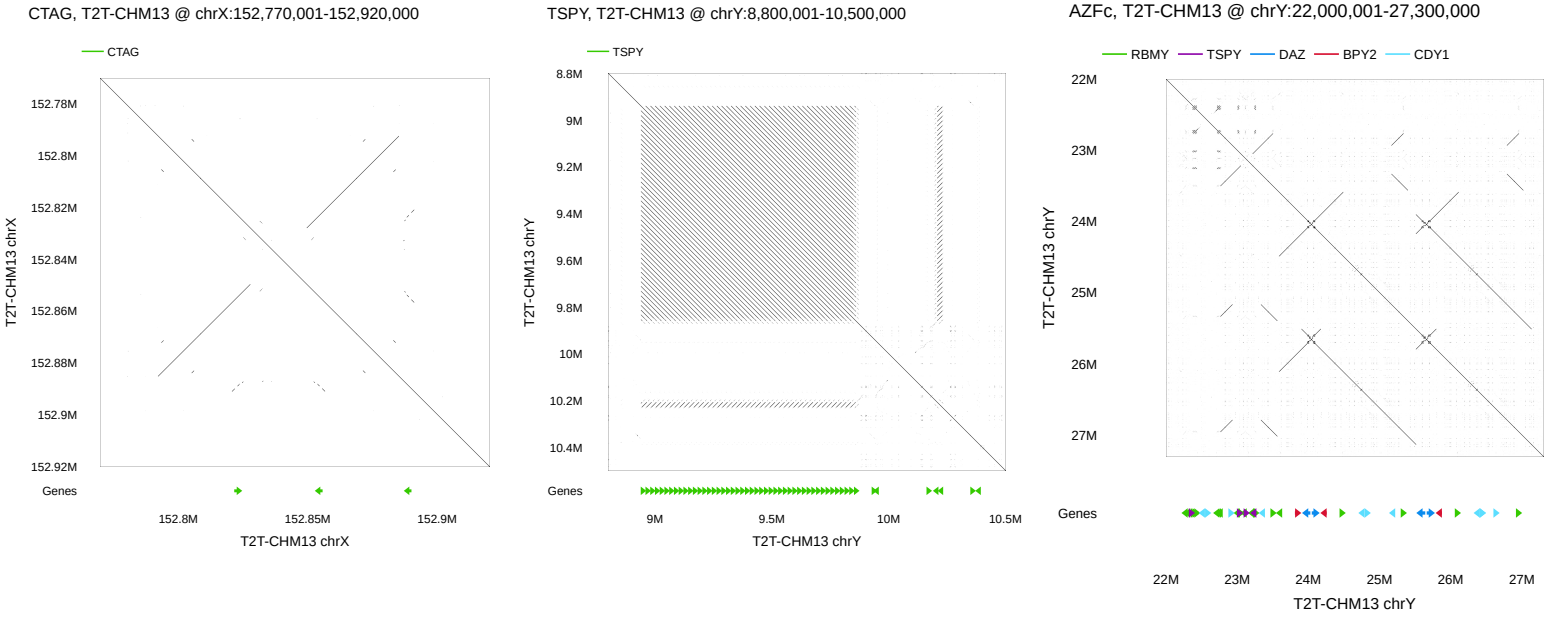


CT45A, T2T-CHM13 @ chrX:134,000,001-134,300,000



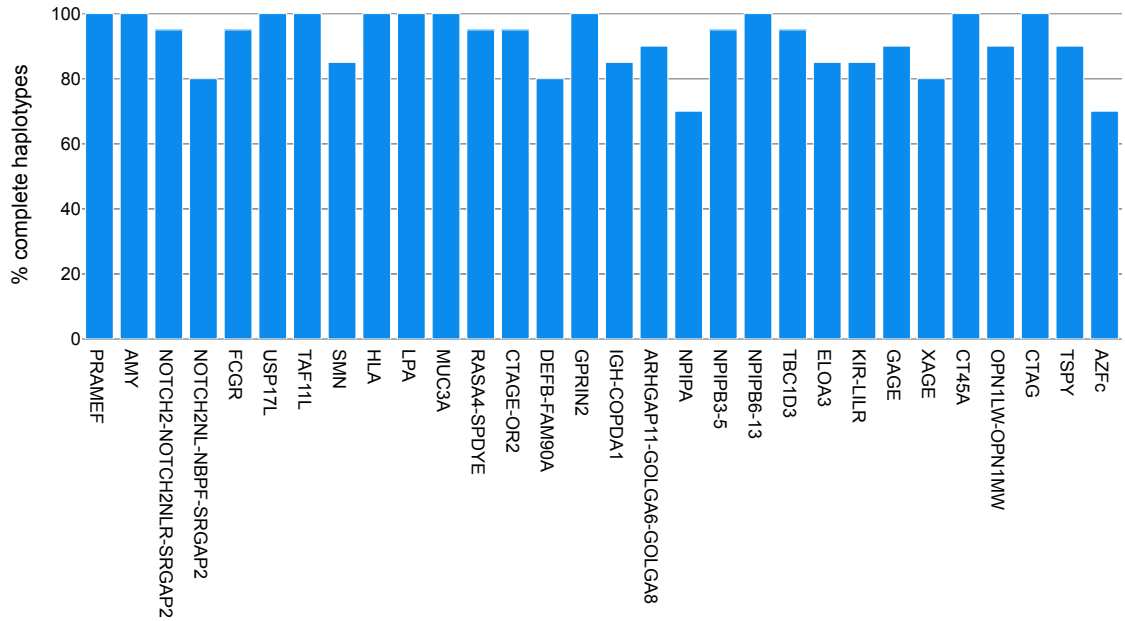
OPN1LW-OPN1MW, T2T-CHM13 @ chrX:152,400,001-152,650,000



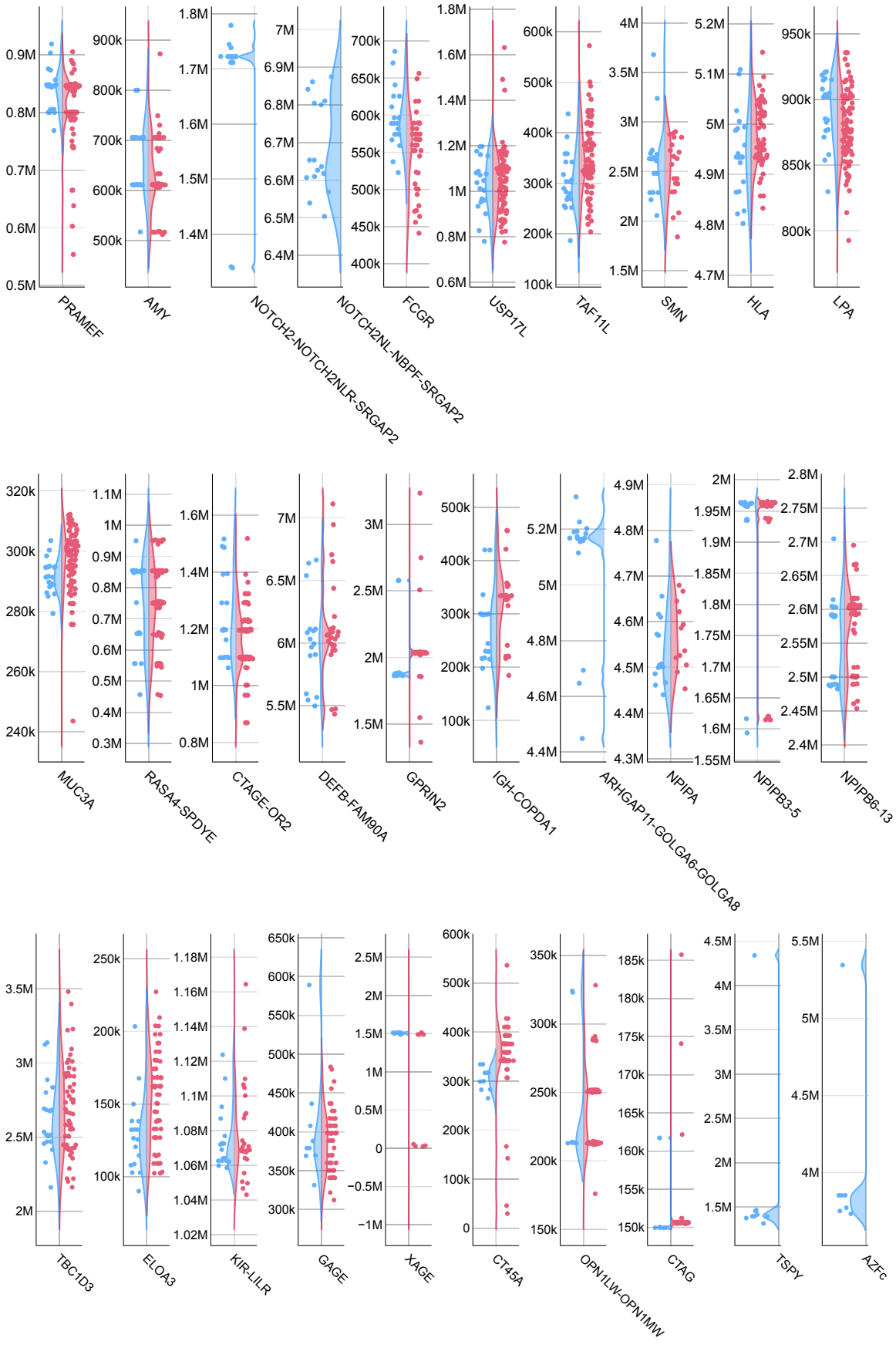


Supplementary Fig. 4 | Sequence structure of each of the 30 medically important complex regions.

For each of the 30 regions, the self dot plot of the T2T-CHM13 in the region is drawn with Gepard. Key genes are annotated at the bottom of each dot plot. The word size is 100bp.



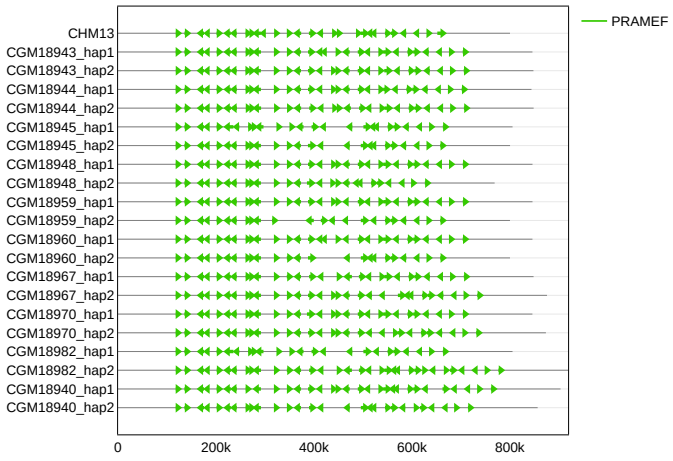
Supplementary Fig. 5 | Complete haplotype reconstruction rate for 30 medically important complex regions.
For each of the 30 regions, the percentage of complete haplotypes that meet all the three criteria (sequence contiguity, structural quality, and base quality) across Japanese haplotypes are shown.



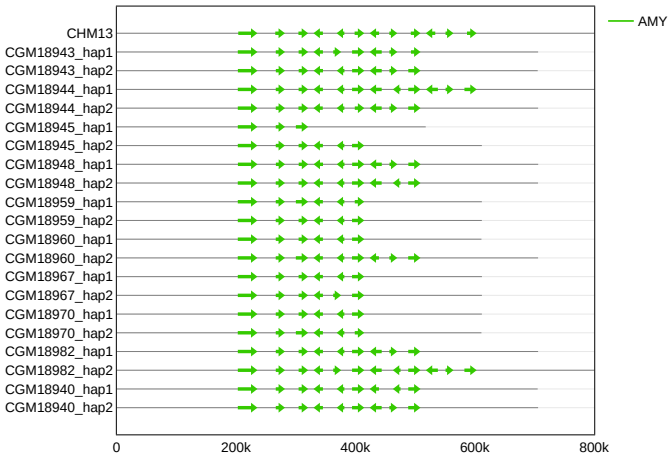
Supplementary Fig. 6 | Size distributions of HPRC Year-1 and our assemblies for 30 medically important complex regions.

For our Japanese assemblies (blue) only complete haplotypes were used for the plot, and for HPRC Year-1 assemblies (red) only haplotypes assembled without sequence gaps were used.

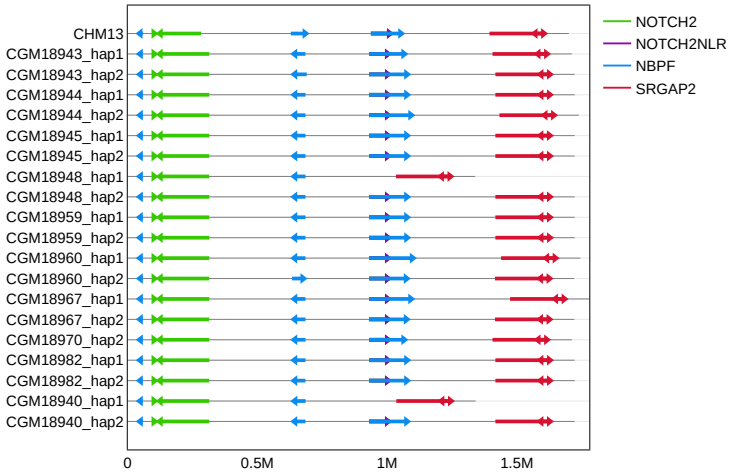
PRAMEF, complete haplotypes



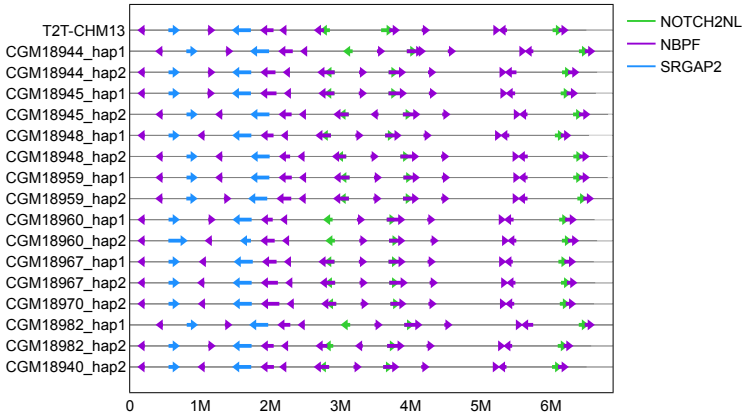
AMY, complete haplotypes



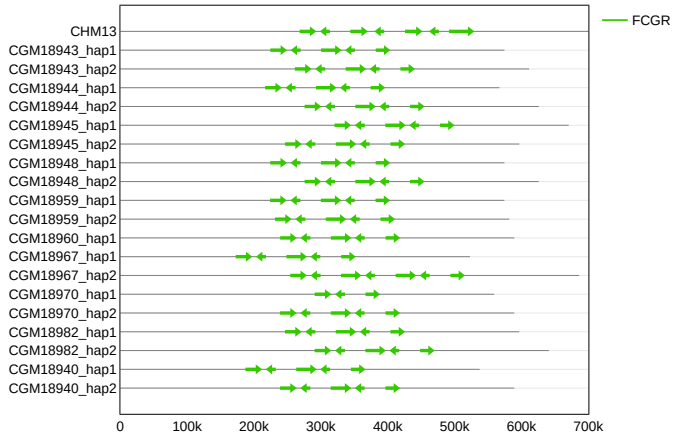
NOTCH2-NOTCH2NLR-SRGAP2, complete haplotypes



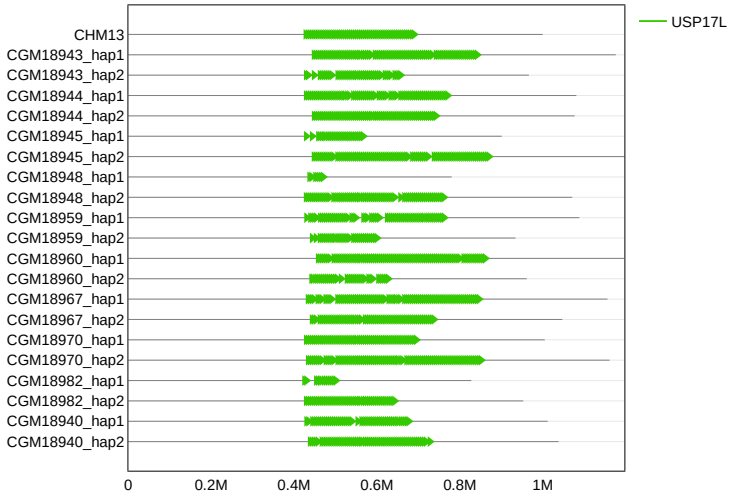
NOTCH2NL-NBPF-SRGAP2, complete haplotypes



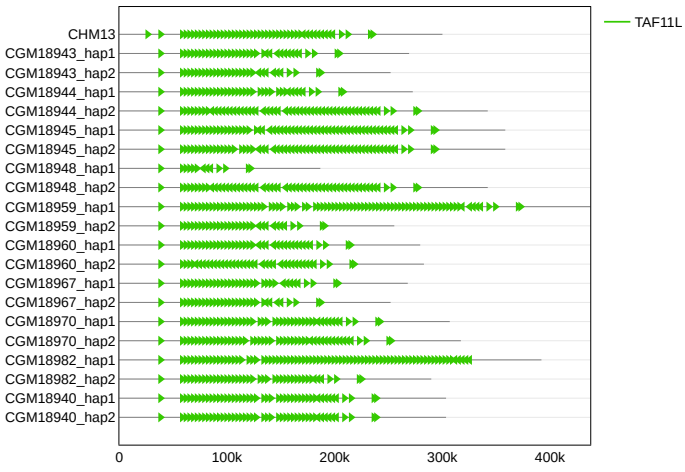
FCGR, complete haplotypes



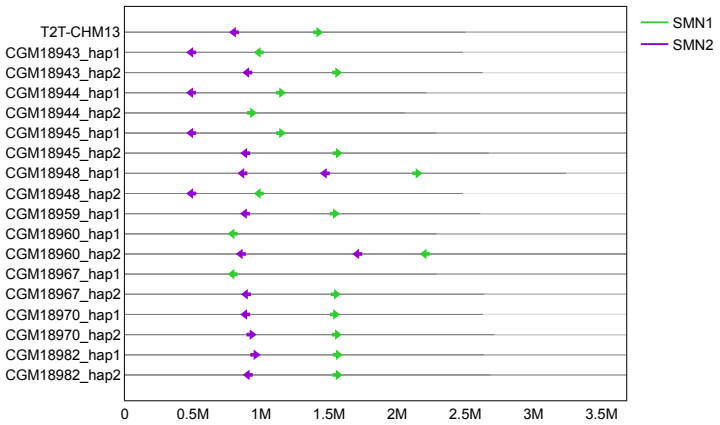
USP17L, complete haplotypes



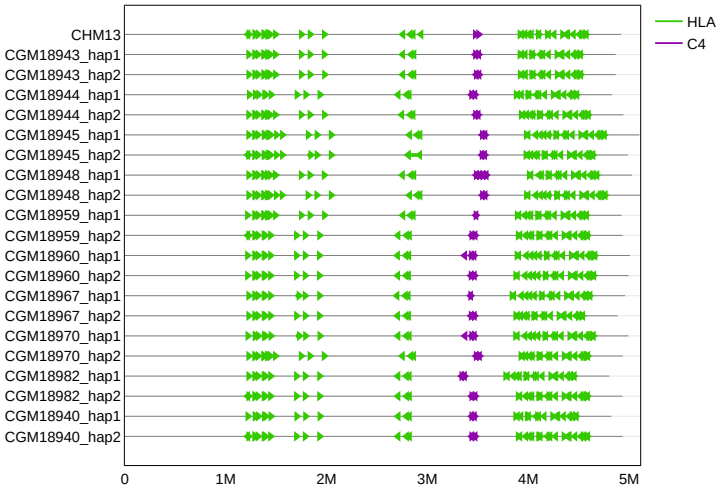
TAF11L, complete haplotypes



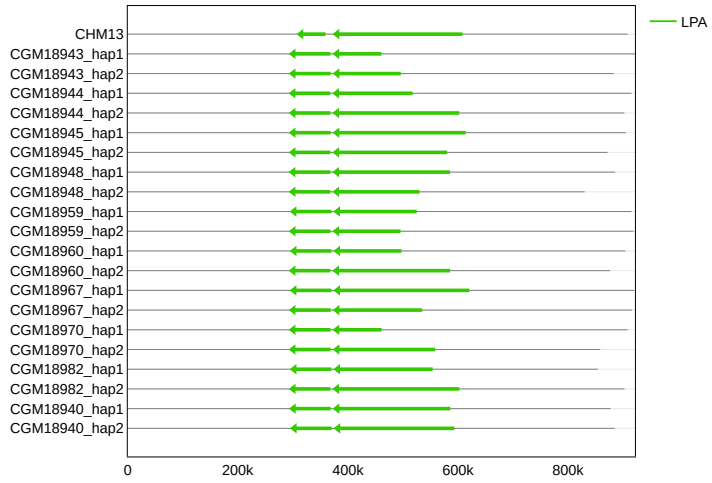
SMN, complete haplotypes



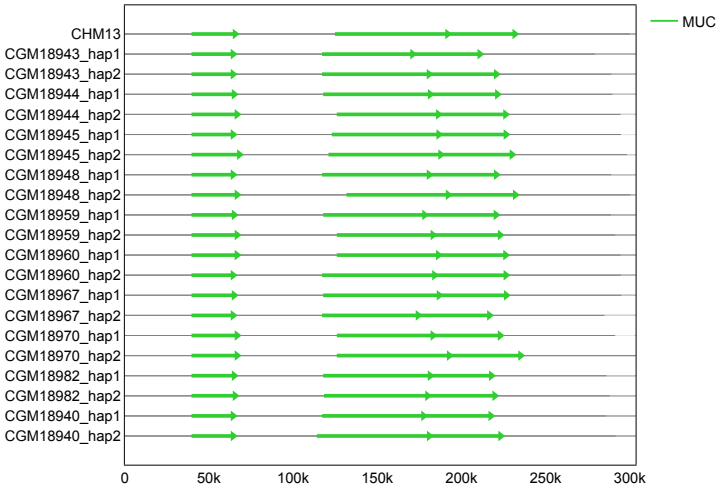
HLA, complete haplotypes



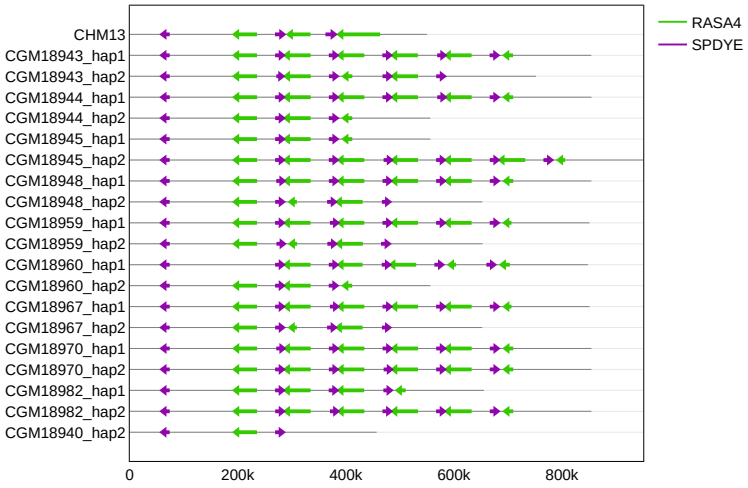
LPA, complete haplotypes



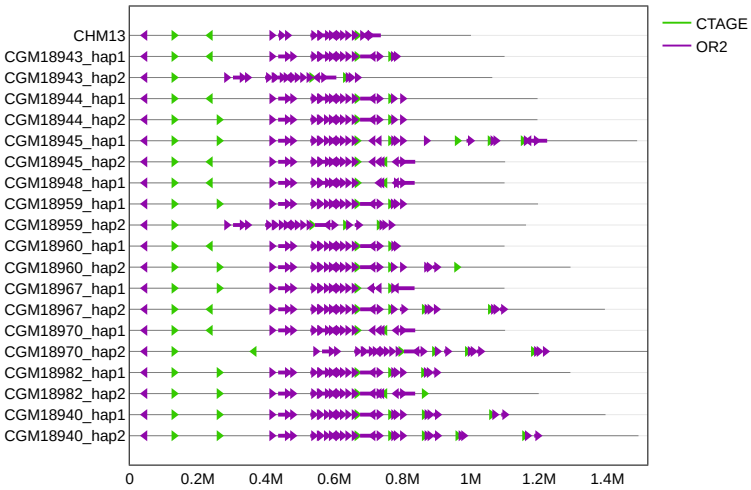
MUC3A, complete haplotypes



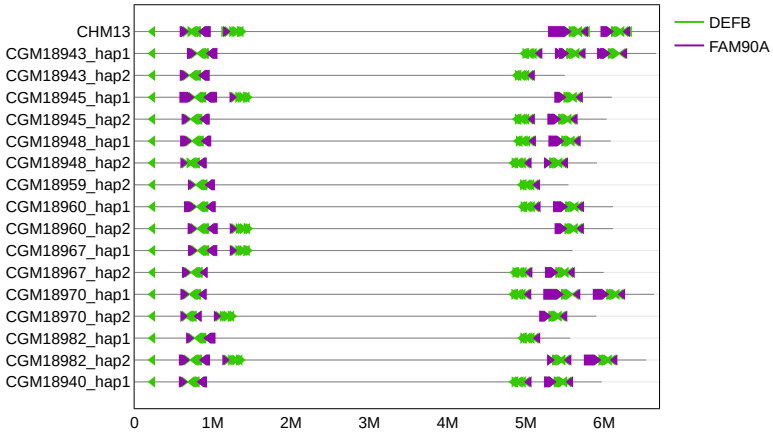
RASA4-SPDYE, complete haplotypes



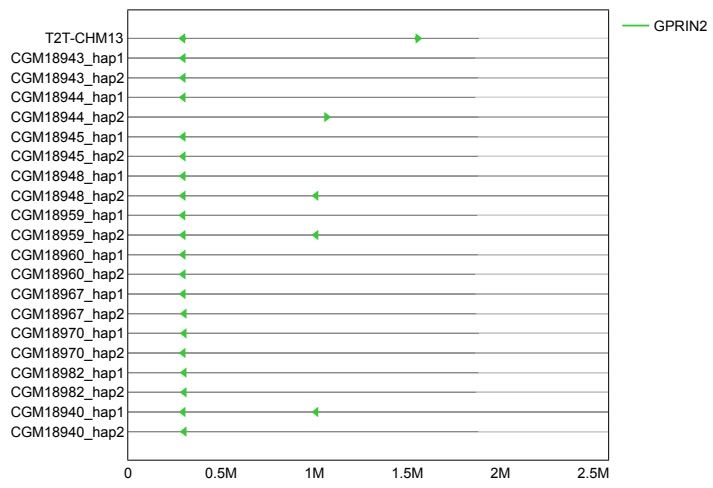
CTAGE-OR2, complete haplotypes



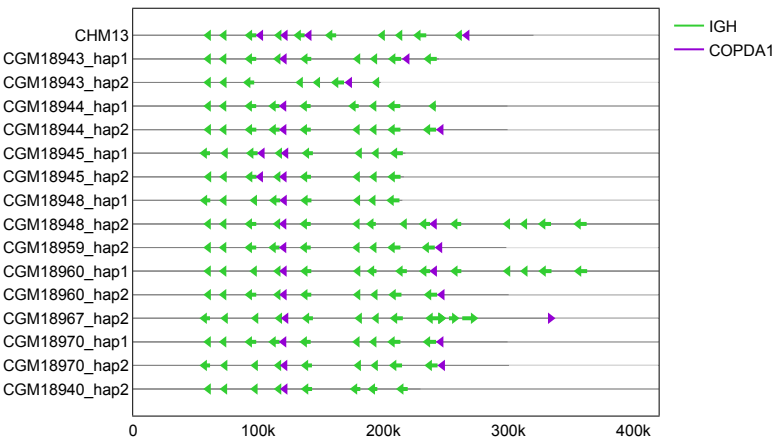
DEFB-FAM90A, complete haplotypes



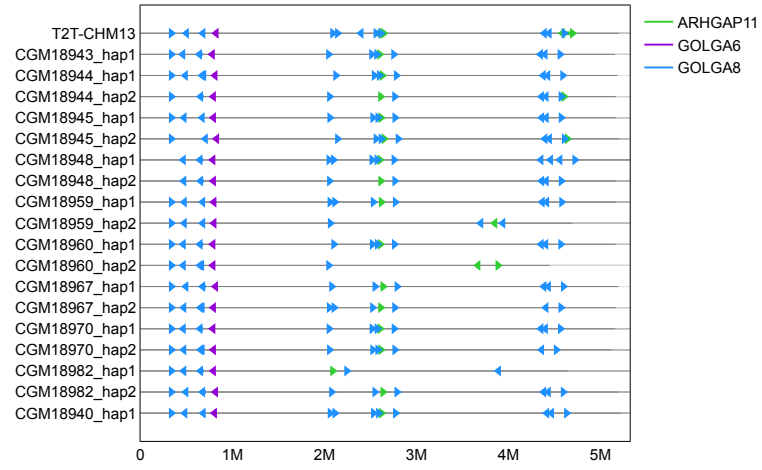
GPRIN2, complete haplotypes



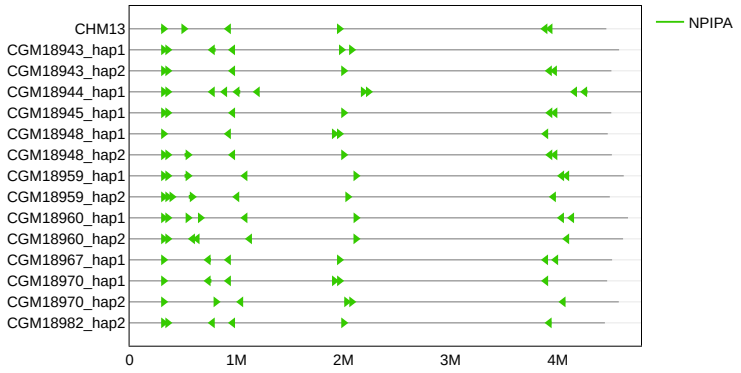
IGH-COPDA1, complete haplotypes



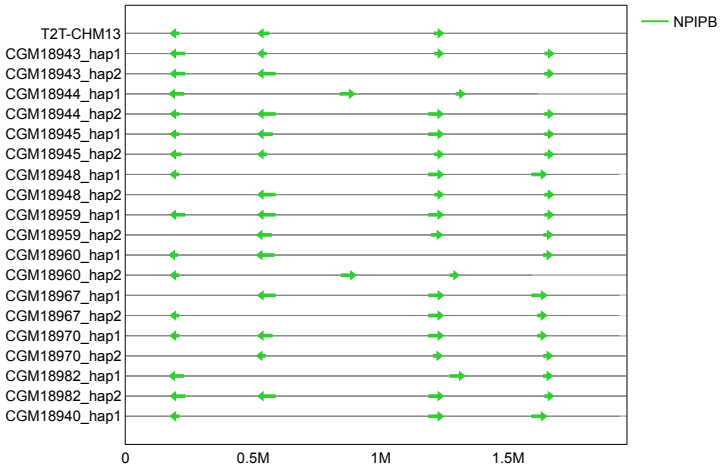
ARHGAP11-GOLGA6-GOLGA8, complete haplotypes



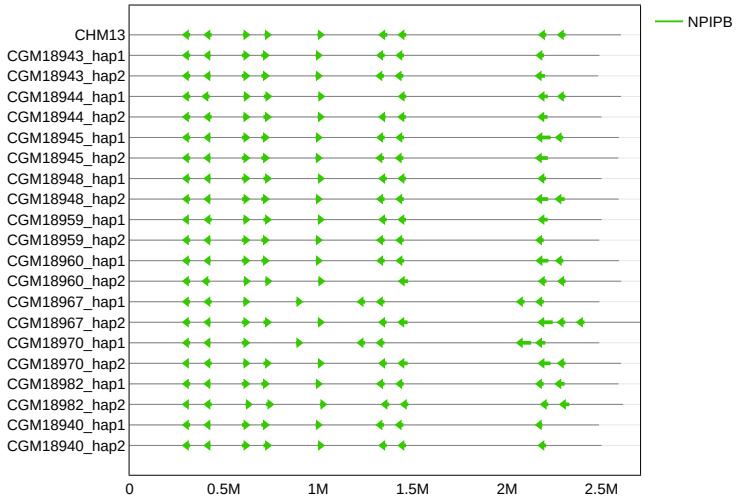
NPIPA, complete haplotypes



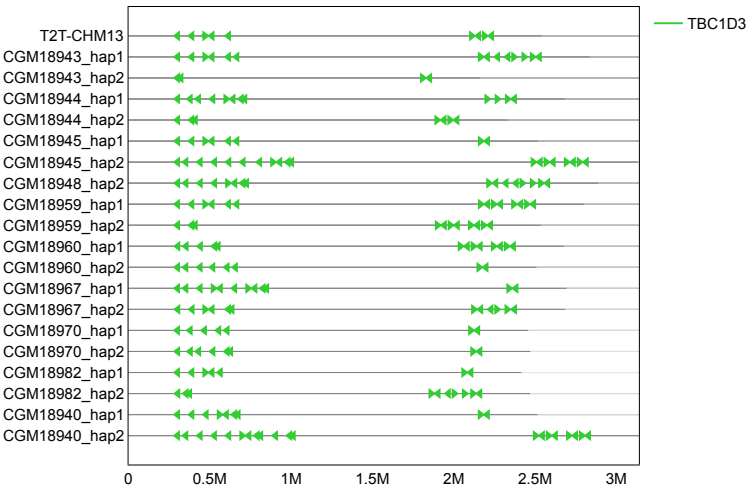
NPIPB3-5, complete haplotypes



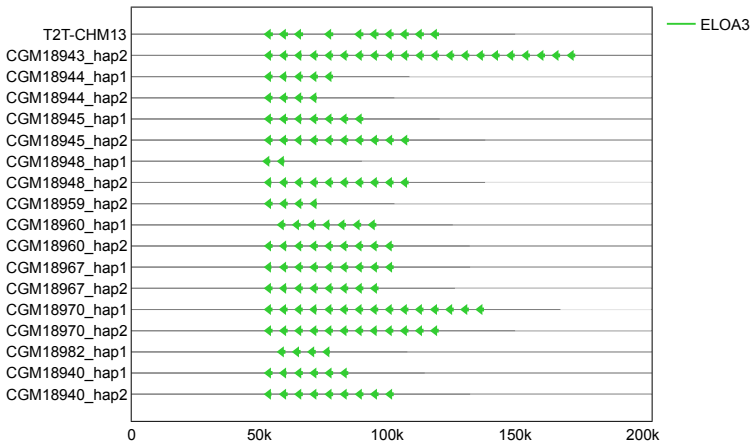
NPIPB6-13, complete haplotypes



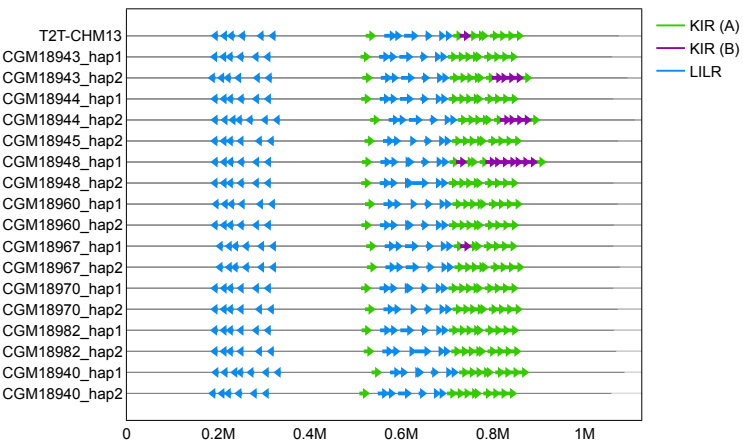
TBC1D3, complete haplotypes



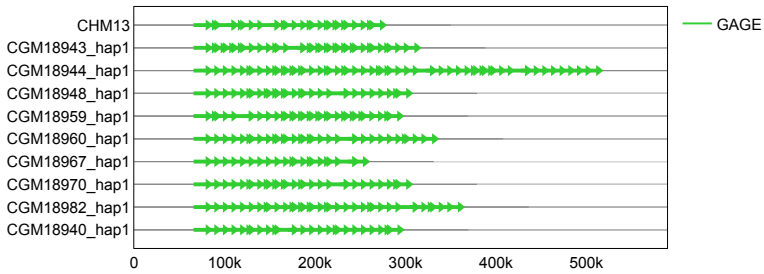
ELOA3, complete haplotypes

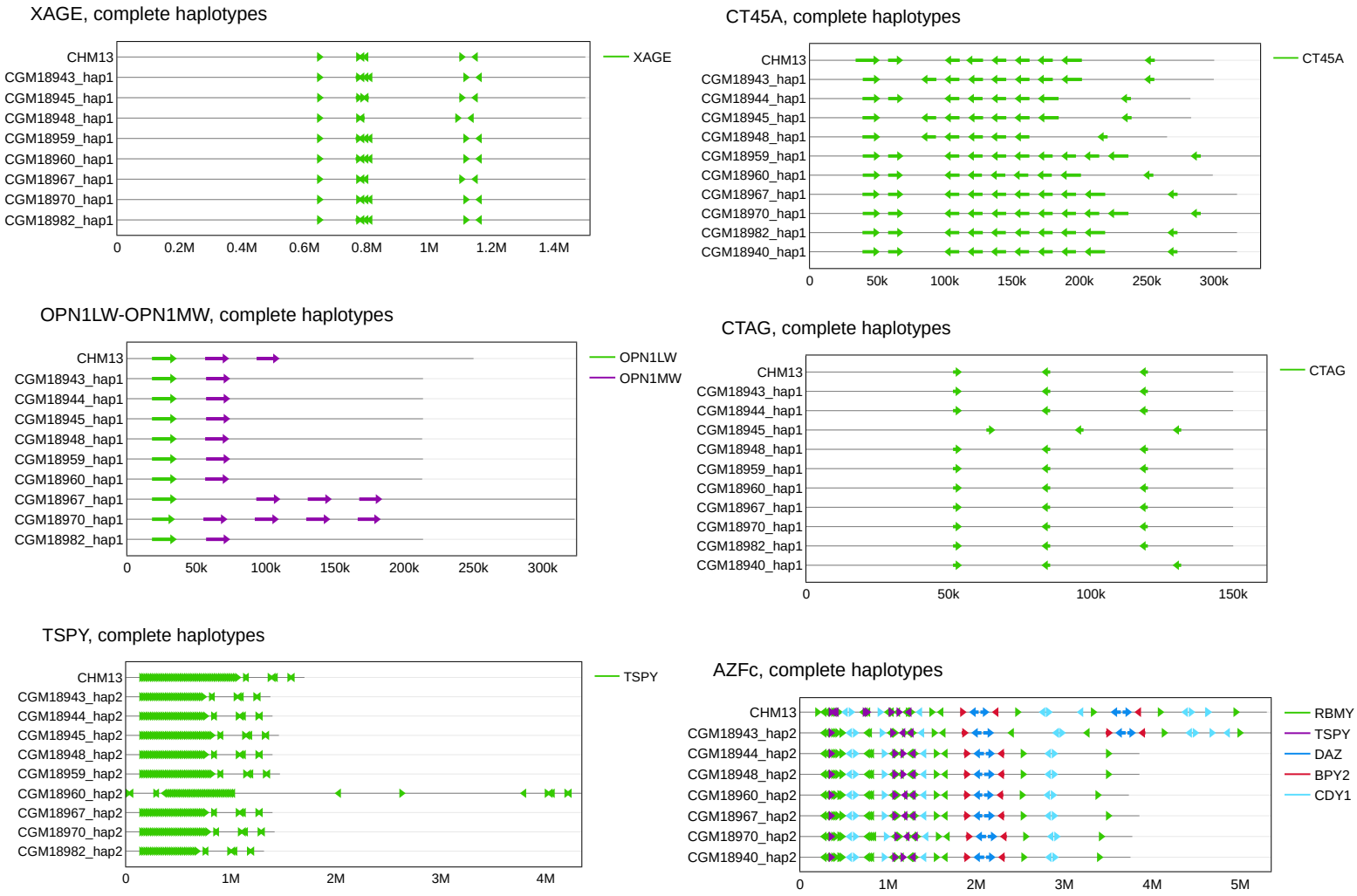


KIR-LILR, complete haplotypes



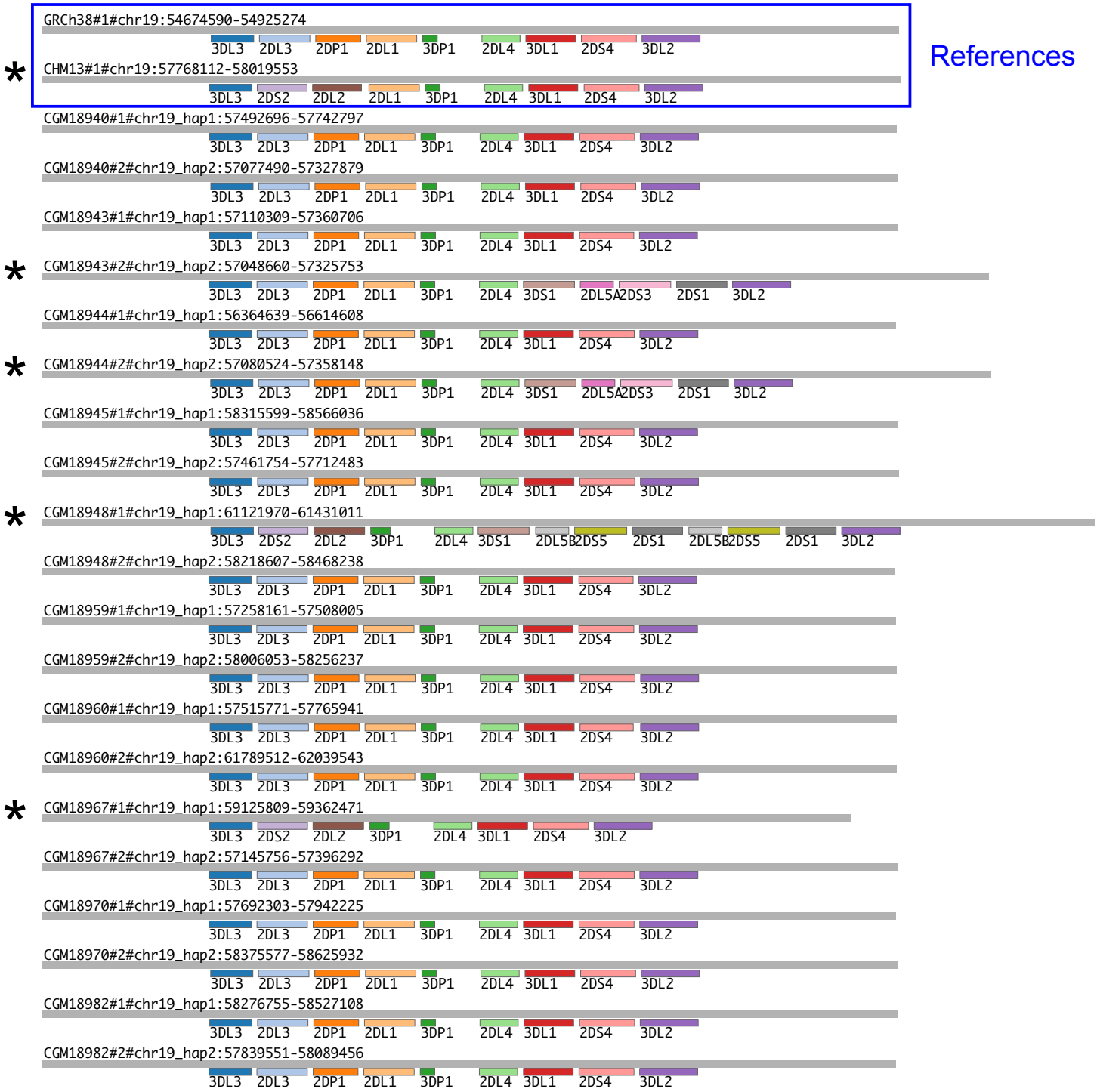
GAGE, complete haplotypes





Supplementary Fig. 7 | Key gene annotations in the complete haplotypes of the 30 medically important complex regions.

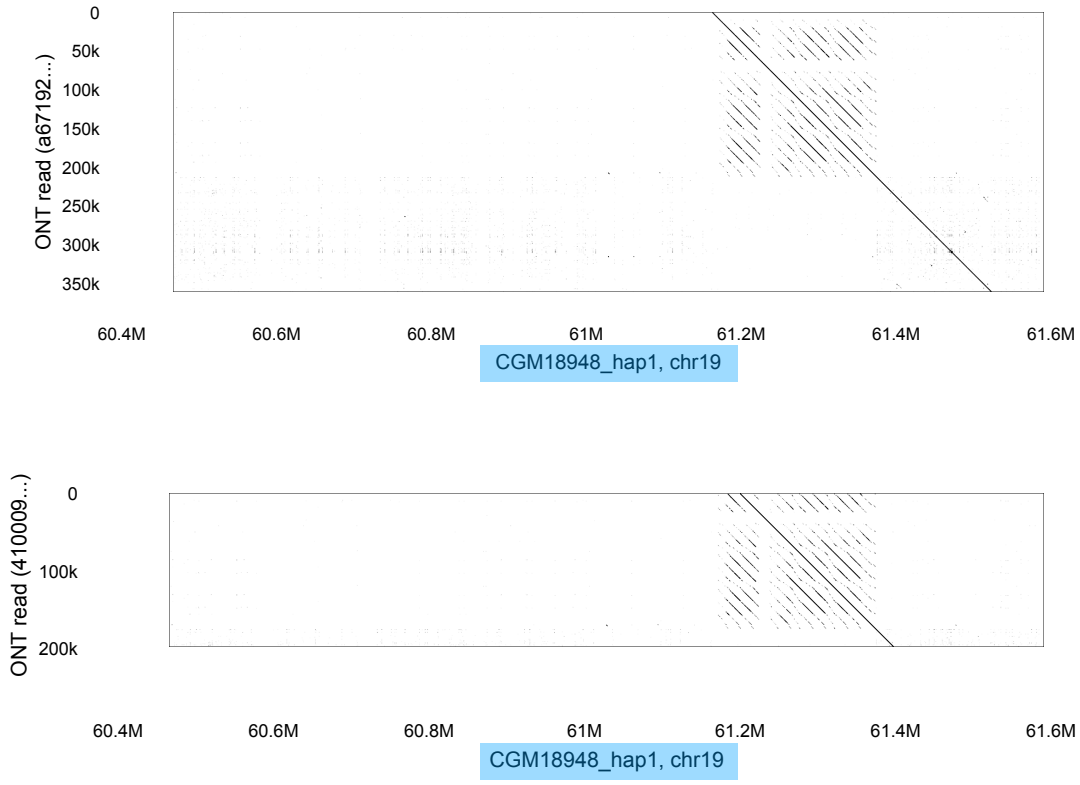
For each of the 30 regions, lift-overed annotations of key genes in the region are shown for each complete Japanese haplotype and T2T-CHM13. The gray bar represents the sequence of each Japanese haplotype corresponding to the region in T2T-CHM13.



(* = B-haplotypes; Others = A-haplotypes)

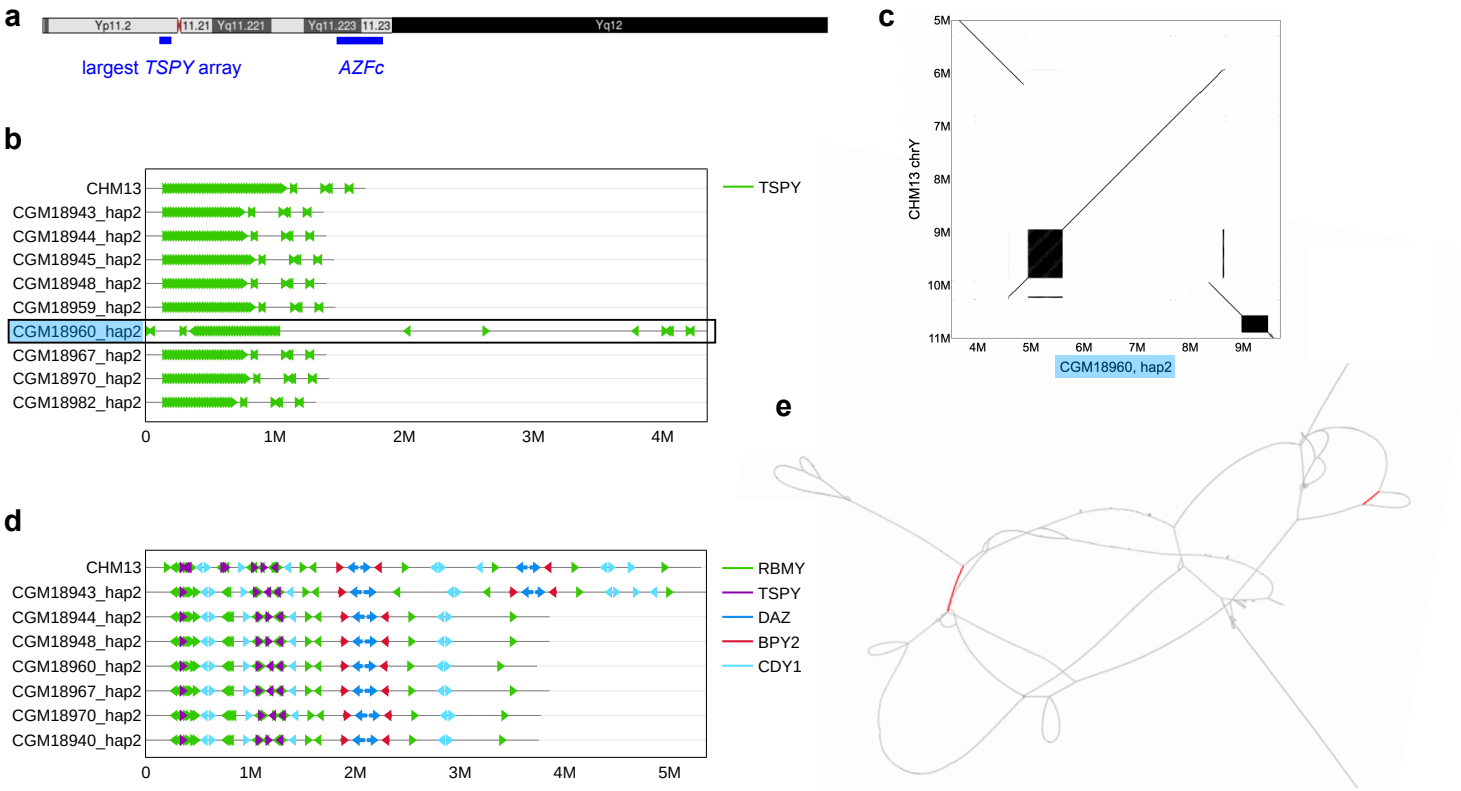
Supplementary Fig. 8 | *KIR* haplotypes annotated with Immuannot.

From top to bottom, annotations of *KIR* genes of the GRCh38 and T2T-CHM13 reference genomes and the 20 Japanese haplotypes are listed. Small bars represent *KIR* genes shown with their names, and the same color indicates the same allele. Haplotypes with asterisks are B-haplotypes.



Supplementary Fig. 9 | Two spanning ONT ultra-long reads of the previously unidentified duplication in a *KIR* gene array.

Comparison between the assembled sequence of the blue haplotype (in Fig. 3a,b) and two ONT ultra-long reads collected for validation that span the expanded part of the *KIR* gene array. Both reads show a clear colinearlity with the assembled sequence and one read fully spans the entire *KIR* gene array.



Supplementary Fig. 10 | Japanese haplotype diversity in highly complex regions in chromosome Y

a, Locations of the largest *TSPY* array and the *AZFc* region on chromosome Y in T2T-CHM13. The illustration of cytobands was taken from the UCSC genome browser. **b**, T2T-CHM13 and Japanese haplotypes around the largest *TSPY* array. **c**, Dot plot between T2T-CHM13 and one Japanese haplotype (highlighted in blue) that harbors a large inversion involving the largest *TSPY* gene array (the black rectangle in the dot plot). **d**, T2T-CHM13 and Japanese haplotypes in the *AZFc* region. **e**, The pangenome graph of the *AZFc* region constructed with pggb from the HPRC Year-1 and Japanese haplotypes. Paths highlighted in red are newly added by Japanese haplotypes.

Supplementary Table 1. Sequencing stats

NOTE: "Coverage" is calculated based on a genome size of 3.1Gbp

PacBio HiFi:

Sample name	# Reads	Total bases [bp]	Coverage [x]	Mean length [bp]	N50 length [bp]	Max length [bp]
CGM18940	9,493,011	124,138,448,652	39.8	13,077	20,574	68,047
CGM18943	9,110,323	126,097,442,051	40.5	13,841	17,080	72,769
CGM18944	7,396,042	109,579,331,059	35.2	14,816	18,500	60,747
CGM18945	8,677,999	123,151,384,768	39.5	14,191	18,175	69,867
CGM18948	8,910,842	118,904,229,526	38.1	13,344	19,348	70,807
CGM18959	9,539,341	110,930,909,119	35.6	11,629	15,502	66,560
CGM18960	8,208,681	153,628,903,202	49.3	18,715	21,581	67,382
CGM18967	9,969,732	146,977,062,417	47.1	14,742	17,882	68,817
CGM18970	9,830,561	170,135,408,766	54.6	17,307	21,162	70,548
CGM18982	7,731,531	147,073,884,789	47.2	19,023	22,876	71,267

Omni-C:

Sample name	# Reads	Total bases [bp]	Coverage [x]	Mean length [bp]	N50 length [bp]	Max length [bp]
CGM18940	821,407,287 x2	123,211,093,050 x2	79.1	150	150	150
CGM18943	918,302,656 x2	137,745,398,400 x2	88.4	150	150	150
CGM18944	959,714,642 x2	143,957,196,300 x2	92.4	150	150	150
CGM18945	690,287,654 x2	103,543,148,100 x2	66.4	150	150	150
CGM18948	661,310,800 x2	99,196,620,000 x2	63.6	150	150	150
CGM18959	556,457,526 x2	83,468,628,900 x2	53.6	150	150	150
CGM18960	564,040,159 x2	84,606,023,850 x2	54.3	150	150	150
CGM18967	615,692,210 x2	92,353,831,500 x2	59.3	150	150	150
CGM18970	619,696,872 x2	92,954,530,800 x2	59.6	150	150	150
CGM18982	690,794,288 x2	103,619,143,200 x2	66.5	150	150	150

ONT ultra-long (UTokyo; all):

Sample name	# Reads	Total bases [bp]	Coverage [x]	Mean length [bp]	N50 length [bp]	Max length [bp]
CGM18940	16,296,087	550,698,206,308	177.6	33,793	71,074	1,844,468
CGM18943	18,702,733	481,727,287,858	155.4	25,757	63,791	1,634,926
CGM18944	18,203,845	530,382,547,597	171.1	29,136	66,157	1,311,980
CGM18945	14,360,293	452,524,652,160	146.0	31,512	69,889	1,491,002
CGM18948	16,958,006	455,316,983,813	146.9	26,850	67,997	1,147,791
CGM18959	18,624,651	484,458,574,907	156.3	26,012	67,219	1,221,094
CGM18960	22,306,045	612,973,304,553	197.7	27,480	65,639	1,736,510
CGM18967	11,534,991	438,084,999,976	141.3	37,979	80,599	1,102,543
CGM18970	12,298,745	447,302,690,133	144.3	36,370	73,789	1,688,809
CGM18982	16,066,227	574,434,794,992	185.3	35,754	73,590	1,501,234

ONT ultra-long (UTokyo; >100kbp):

Sample name	# Reads	Total bases [bp]	Coverage [x]	Mean length [bp]	N50 length [bp]	Max length [bp]
CGM18940	1,257,918	180,062,788,923	58.1	143,144	141,397	1,844,468
CGM18943	983,381	146,435,533,701	47.2	148,910	147,046	1,634,926
CGM18944	1,110,767	165,688,873,833	53.4	149,166	147,113	1,311,980
CGM18945	1,014,621	149,881,043,038	48.3	147,721	146,143	1,491,002
CGM18948	996,922	148,018,414,907	47.7	148,475	147,223	1,147,791
CGM18959	1,049,703	156,594,033,953	50.5	149,179	147,875	1,221,094
CGM18960	1,283,694	187,210,856,459	60.4	145,838	143,786	1,736,510
CGM18967	1,139,034	172,980,694,009	55.8	151,866	151,306	1,102,543
CGM18970	1,061,341	156,006,094,689	50.3	146,990	145,206	1,688,809
CGM18982	1,368,493	195,682,263,966	63.1	142,991	141,135	1,501,234

Supplementary Table 2. Sequencing stats of Nanopore ultra-long reads sequenced at UCSC

Sample_ID	read_N50	Gb	coverage	100kb+	200kb+	300kb+	400kb+	500kb+	1Mb+	whales #	q5+	q10+	q15+	q20+	q25+
GM18940	65,838	310.5	94.1	27.08	3.78	0.55	0.17	0.08	0	17	93.33	82.99	72.4	52.48	9.44
GM18943	68,986	268.53	81.38	29.55	7.87	1.86	0.37	0.07	0	1	80.83	72.98	65.7	51.46	13.75
GM18944	60,965	328.32	99.49	24.89	2.99	0.39	0.1	0.04	0	8	98.93	89.9	80.87	62.89	14.39
GM18945	79,778	214.58	65.01	26.3	7.61	1.99	0.44	0.09	0	2	64.54	58.16	51.9	40.09	9.9
GM18948	80,196	310.58	94.12	37.5	9.44	2.06	0.41	0.09	0	7	93.22	81.49	69.45	48.07	8.78
GM18959	91,018	227.63	68.98	31.16	8.5	1.88	0.39	0.09	0	8	68.45	61.79	55.31	43.38	9.25
GM18960	86,478	255.6	77.46	32.93	8.53	1.96	0.43	0.09	0	1	76.95	69.34	62.23	48.9	10.31
GM18967	88,165	217.31	65.86	28.73	8.36	2.05	0.45	0.12	0	7	65.35	59.19	53.06	40.64	8.08
GM18970	82,209	220.41	66.79	27.98	7.82	1.8	0.38	0.08	0	4	66.29	60.48	54.79	42.76	9.3
GM18982	77,497	204.2	61.88	24.95	7.29	1.86	0.41	0.08	0	1	61.15	54.3	48.1	37.13	9.81

Supplementary Table 3. List of operations of manual curation

Curation of haplotype assignment of contigs:

Sample name	Contig name	Chromosome	Haplotype (before curation)	Haplotype (after curation)
CGM18940	h1tg000026l	chrY	hap1	hap2
CGM18940	h1tg000060l	chr15	hap1	hap2
CGM18940	h1tg000062l	chrY	hap1	hap2
CGM18940	h1tg000063l	chr22	hap1	hap2
CGM18940	h1tg000107l	chr6	hap1	hap2
CGM18948	h1tg000043l	chrY	hap1	hap2
CGM18960	h1tg000032l	chrY	hap1	hap2
CGM18960	h1tg000042l	chrY	hap1	hap2
CGM18967	h1tg000034l	chrY	hap1	hap2
CGM18970	h2tg000069l	chr1	hap2	hap1
CGM18982	h1tg000030l	chrY	hap1	hap2
CGM18982	h1tg000033l	chrY	hap1	hap2
CGM18982	h1tg000036l	chrY	hap1	hap2
CGM18982	h1tg000049l	chrY	hap1	hap2

Manual scaffolding of contigs:

Sample name	Haplotype	Contig name	Chromosome assigned
CGM18943	hap2	h1tg000020l	chrY

Manual curation for CGM18940's chromosomal scaffolds:

Sample name	Haplotype	Chromosome	Description of operation
CGM18940	hap1	chr2	Purged chr2:130032194-150580001
CGM18940	hap1	chr9	Purged chr9:1-17077554
CGM18940	hap2	chr6	Concatenated the forward sequence of h1tg000107l to the prefix of the chr6 scaffold with a sequence gap
CGM18940	hap2	chr15	Concatenated the reverse complemented sequence of h1tg000060l to the suffix of the chr15 scaffold with a sequence gap
CGM18940	hap2	chr22	Concatenated the reverse complemented sequence of h1tg000063l to the prefix of the chr22 scaffold with a sequence gap
CGM18940	hap2	chrY	Replaced with the concatenation of the reverse complemented sequences of h1tg000026l and h1tg000062l with a sequence gap

Supplementary Table 4. Assembly stats.

		Contigs					Scaffolds					Chromosomal Scaffolds (w/o chrM)				QV		Single-copy genes (asmgene)			Telomere-to-telomere assembly status for each chromosome ("C": T2T contigs, "S": T2T scaffolds)																															
Sample	Hap	# Contigs	Total bases	Max length	N50 length	auN value	# Scaffolds	Total bases	Max length	N50 length	auN value	# Scaffolds	Total bases	Max length	N50 length	Merqury	Yak	full_single	full_dup	perc_single	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
CGM18940	hap1	66	3,065,789,104	245,189,110	104,879,181	110,489,056	25	2,989,360,231	245,189,110	156,195,506	154,774,653	23	2,989,263,263	245,189,110	156,195,506	60.79	53.92	33474	73	0.982	C	S	S		S	S	S	C	S	S		C				S	S		S	S	S	S		-
	hap2	42	2,978,019,940	245,102,220	107,812,723	115,550,725	23	2,904,760,418	245,102,220	147,151,709	153,210,017	23	2,904,760,418	245,102,220	147,151,709	61.47	55.03	33406	77	0.9801	C	S	S			S	S	S	C	S		C				S	S	S	S				-	
CGM18943	hap1	69	3,039,273,529	242,826,137	138,568,228	123,285,845	40	3,030,799,884	247,520,969	154,181,322	155,433,740	23	3,021,420,735	247,520,969	154,181,322	63.42	54.45	33514	109	0.9842	S	C		S	C	C		C	C	C	C	C					C	S	C	C		S	C	-
	hap2	125	2,929,911,867	193,485,417	136,080,202	133,975,587	51	2,928,042,134	246,869,894	145,305,005	153,925,924	23	2,919,677,719	246,869,894	145,305,005	64.10	55.1	33497	113	0.9838		S			C	C		C	C	C	C						S	C	C	C	S	S	-	
CGM18944	hap1	164	3,052,330,762	242,265,134	129,483,041	119,987,444	76	3,046,592,783	249,155,589	155,318,995	155,988,712	23	3,029,204,114	249,155,589	155,318,995	63.53	53.64	33489	87	0.9828	S	C			S	C	S	S	S		C	C	S					C	C	C			S	-
	hap2	99	2,930,934,763	242,612,675	135,991,310	124,871,347	54	2,930,010,287	258,196,639	146,218,127	155,191,451	23	2,916,992,232	258,196,639	146,218,127	60.23	53.16	33433	96	0.9815	S	C	C		C	C	S	S	S	C	C					S		S	C	S	C	S	-	S
CGM18945	hap1	120	3,020,679,697	249,293,442	134,223,936	136,655,233	53	3,013,096,913	249,293,442	155,122,962	155,550,010	23	3,008,293,430	249,293,442	155,122,962	60.82	54.02	33525	88	0.9839	C	S	S		C	S		C		C	S	C					C	C	C	C	S		C	-
	hap2	82	2,917,908,637	242,224,868	127,819,649	128,861,197	43	2,917,234,232	257,477,501	146,622,731	154,887,963	23	2,912,847,863	257,477,501	146,622,731	61.74	54.46	33475	97	0.9827	S				C	S	C	C	S	S	S	S				C	C	C	C	C	S		-	
CGM18948	hap1	176	3,075,353,932	253,222,128	135,633,951	130,613,512	62	3,051,108,770	253,305,853	154,131,922	155,992,074	23	3,038,845,072	253,305,853	154,131,922	64.28	53.26	33519	84	0.9836		S		S		C	C	S				C				C	S	C		C			-	
	hap2	84	2,893,339,006	251,424,744	112,864,796	126,547,315	50	2,911,939,063	252,300,762	146,198,965	153,876,911	23	2,897,654,475	252,300,762	146,198,965	60.30	52.91	33470	117	0.9832	S	S		C	C	C	S	C		C	S	C				C	S	C	C	C			-	
CGM18959	hap1	232	3,091,721,898	257,533,225	135,059,907	129,515,609	71	3,071,177,540	257,592,629	154,332,372	157,257,706	23	3,051,734,352	257,592,629	154,332,372	61.38	54.55	33441	190	0.9845			S		C	S		S	C	C	S	C					S	S		S	S	S	S	-
	hap2	131	2,919,106,969	251,714,032	120,427,261	126,437,487	62	2,914,544,910	251,714,032	145,777,493	153,776,412	23	2,886,987,576	251,714,032	145,777,493	63.73	54.74	33405	91	0.9805	C	C			S	C	S		S		C	S					S	S	S	S		S	-	
CGM18960	hap1	161	3,041,643,193	242,270,071	111,328,456	128,176,211	72	3,018,903,344	242,327,658	155,233,135	155,040,344	23	3,006,640,870	242,327,658	155,233,135	62.94	54.15	33529	103	0.9845			S	S	C	S	S	C	S		S	C	C				S	C	S	C		S	S	-	
	hap2	87	2,923,951,872	247,934,492	135,253,169	126,072,276	56	2,941,018,753	247,934,492	145,421,388	153,920,887	23	2,919,782,049	247,934,492	145,421,388	64.06	54.76	33473	120	0.9833	C	S	C	S	C	S	S	C			C	S	C	S		C	C	C	C	C			-	
CGM18967	hap1	211	3,031,233,499	242,795,645	119,971,359	121,347,366	76	3,017,786,863	243,690,463	155,042,577	155,179,917	23	2,996,872,982	243,690,463	155,042,577	63.75	54.25	33508	91	0.9835	S	C	C				C	C	S	S		C	S		S	S	S	S	S	C			S	-
	hap2	120	2,960,150,362	246,696,106	140,401,979	139,019,666	69	2,968,374,987	246,788,055	145,841,060	154,319,441	23	2,935,587,455	246,788,055	145,841,060	63.46	54.84	33442	145	0.9832		C		C	C	S	C	C			C	S					S	S	C	C	S	S	-	S
CGM18970	hap1	118	3,024,779,430	242,248,412	121,324,348	125,356,332	60	3,039,068,900	247,013,048	155,068,577	155,731,408	23	3,024,628,791	247,013,048	155,068,577	65.72	55.45	33480	117	0.9835	S		S			C		S		S	C	C		C	S	C	S	S		C	S	S	S	-
	hap2	70	2,941,895,411	244,699,028	137,101,261	137,664,539	48	2,923,921,428	244,699,028	145,431,912	153,536,682	23	2,917,645,007	244,699,028	145,431,912	65.53	55.44	33523	81	0.9837	C	C	S		S	C			C	C	C	C	C			C	S				S	-	
CGM18982	hap1	144	3,017,357,387	242,295,251	133,646,597	136,075,716	69	3,015,329,384	251,185,660	154,963,437	155,717,606	23	3,003,849,740	251,185,660	154,963,437	62.98	53.68	33498	86	0.9831				C	C		C	C		S	S	C			S		S	S	C	C		S		-
	hap2	85	2,965,009,485	253,550,596	134,404,795	134,535,427	49	2,964,553,196	253,550,596	146,930,931	154,684,839	23	2,948,427,246	253,550,596	146,930,931	65.29	54.87	33496	114	0.9838			S	C	S	S		C		S		C	C		S		S				S		-	
																				65%	70%	55%	40%	80%	90%	65%	90%	60%	70%	75%	90%	30%	10%	20%	50%	90%	85%	75%	90%	45%	55%	70%	20%

NOTE: Hap1 contains chrX and chrM. Hap2 contains chrY.

Supplementary Table 5. List of protein coding genes with copy number differences (two-sided Wilcoxon rank sum test, P < 0.05) between JPT and EAS haplotypes

Gene name	Copy number in T2T-CHM13	Average copy number in JPT haplotypes	Average copy number in EAS haplotypes	P value
PRH1-PRR4	1	1	0.2	0.00043
SYNE2	1	1	0.2	0.00043
RBFOX2	1	1	0.2	0.00043
BCL2L14	1	0.8	0.1	0.00207
RP11-468E2.4	1	1	0.3	0.00207
RP11-468E2.2	1	1	0.3	0.00207
UNC79	1	1	0.3	0.00207
VKORC1L1	1	1	0.3	0.00207
ENSG00000278817	1	0.05	0.7	0.00424
AP1B1	1	0.75	0.1	0.00424
RBMY1E	1	2.5	0.6	0.00705
NDUFS3	1	1	0.4	0.0083
NUP160	1	1	0.4	0.0083
FMN1	1	1	0.4	0.0083
AATF	1	1	0.4	0.0083
EEF1B2	1	1	0.4	0.0083
ATG16L1	1	1	0.4	0.0083
DGAT1	1	1	0.4	0.0083
CPSF1	1	1	0.4	0.0083
CDY2B	1	1	0.2	0.01431
OR4C11	1	0.85	0.3	0.01553
OR4P4	1	0.85	0.3	0.01553
OR4S2	1	0.85	0.3	0.01553
EIF3CL	1	0.35	0.9	0.01553
ENSG00000283830	1	0.85	0.3	0.01553
NPIPB12	1	1.45	0.9	0.0263
TSPY3	17	10.8	5.4	0.02749
RP11-542C16.2	1	1	0.5	0.02783
NANOS3	1	1	0.5	0.02783
RPS3A	1	0.5	1	0.02783
ENSG00000284475	1	0.7	0.2	0.02783
ENSG00000258002	1	1	0.5	0.02783
CTAGE4	1	1.8	1.3	0.02783
SPANXC	1	0.9	0.2	0.03209
SPANXA2	1	1.1	1.8	0.03209
CDY1B	1	0.9	0.2	0.03209
TSPY1	2	1.8	0.8	0.03734
MBD3L5	1	1.1	0.6	0.043
DDTL	2	1.4	0.9	0.043
ENSG00000283955	1	0.75	0.3	0.04773
RP11-152F13.10	1	0.95	0.5	0.04773
XXbac-BPG32J3.20	1	0.55	0.1	0.04773
FAM90A8P	1	0.55	1	0.04773

Supplementary Table 6. List of the 30 medically important complex regions and their coordinates on T2T-CHM13

Chrom	Start	End	Name	Focal gene list	Reference number
chr1	12,200,000	13,000,000	PRAMEF	PRAMEF	8
chr1	103,200,000	104,000,000	AMY	AMY	1,4
chr1	119,800,000	121,500,000	NOTCH2-NOTCH2NLR-SRGAP2	NOTCH2,NOTCH2NLR,NBPF,SRGAP2	1
chr1	142,500,000	149,000,000	NOTCH2NL-NBPF-SRGAP2	NOTCH2NL,NBPF,SRGAP2	1
chr1	160,500,000	161,200,000	FCGR	FCGR	3,4
chr4	8,700,000	9,700,000	USP17L	USP17L	8
chr5	17,400,000	17,700,000	TAF11L	TAF11L	8
chr5	70,000,000	72,500,000	SMN	SMN	1
chr6	28,381,550	33,301,940	HLA	HLA,C4	3,4,6
chr6	161,403,622	162,311,762	LPA	LPA	1,4,6
chr7	102,150,000	102,450,000	MUC3A	MUC	1
chr7	103,600,000	104,150,000	RASA4-SPDYE	RASA4,SPDYE	6,7
chr7	144,800,000	145,800,000	CTAGE-OR2	CTAGE,OR2	4
chr8	6,400,000	13,100,000	DEFB-FAM90A	DEFB,FAM90A	6
chr10	47,135,399	49,019,220	GPRIN2	GPRIN2	1,6
chr14	99,780,000	100,100,000	IGH-COPDA1	IGH,COPDA1	5
chr15	25,800,000	31,000,000	ARHGAP11-GOLGA6-GOLGA8	ARHGAP11,GOLGA6,GOLGA8	1,4
chr16	14,446,467	18,900,000	NPIPA	NPIPA	1
chr16	21,150,000	23,112,084	NPIPB3-5	NPIPB	1
chr16	28,323,071	30,926,372	NPIPB6-13	NPIPB	1
chr17	36,813,573	39,355,611	TBC1D3	TBC1D3	1,4
chr18	47,100,000	47,250,000	ELOA3	ELOA3	9
chr19	57,104,135	58,178,808	KIR-LILR	KIR,LILR	4
chrX	48,650,000	49,000,000	GAGE	GAGE	1
chrX	51,000,000	52,500,000	XAGE	XAGE	1
chrX	134,000,000	134,300,000	CT45A	CT45A	8
chrX	152,400,000	152,650,000	OPN1LW-OPN1MW	OPN1LW,OPN1MW	4
chrX	152,770,000	152,920,000	CTAG	CTAG	7
chrY	8,800,000	10,500,000	TSPY	TSPY	2
chrY	22,000,000	27,300,000	AZFc	RBMY,TSPY,DAZ,BPY2,CDY1	2

References:

[1] Vollger, M. R. *et al.* Segmental duplications and their variation in a complete human genome. *Science* **376**, eabj6965 (2022).

[2] Rhie, A. *et al.* The complete sequence of a human Y chromosome. *Nature* **621**, 344–354 (2023).

[3] Vollger, M. R. *et al.* Increased mutation and gene conversion within human segmental duplications. *Nature* **617**, 325–334 (2023).

[4] Jeong, H. *et al.* Structural polymorphism and diversity of human segmental duplications. *Nat Genet* **57**, 390–401 (2025).

[5] Nie, J. *et al*. T2T-CHM13 versus hg38: accurate identification of immunoglobulin isotypes from scRNA-seq requires a genome reference matched for ancestry, NAR Genomics and Bioinformatics, 7, lqaf074 (2025).

[6] Liao, W.-W. *et al.* A draft human pangenome reference. *Nature* **617**, 312–324 (2023).

[7] Gao, Y. *et al.* A pangenome reference of 36 Chinese populations. *Nature* **619**, 112–121 (2023).

[8] Nassir, N. *et al.* A draft Arab pangenome reference. *bioRxiv* 2024.07.09.602638 (2024) doi:10.1101/2024.07.09.602638.

[9] Morgan, M. A. J. *et al.* ELOA3: A primate-specific RNA polymerase II elongation factor encoded by a tandem repeat gene cluster. *Sci. Adv.* **9**, eadj1261 (2023).