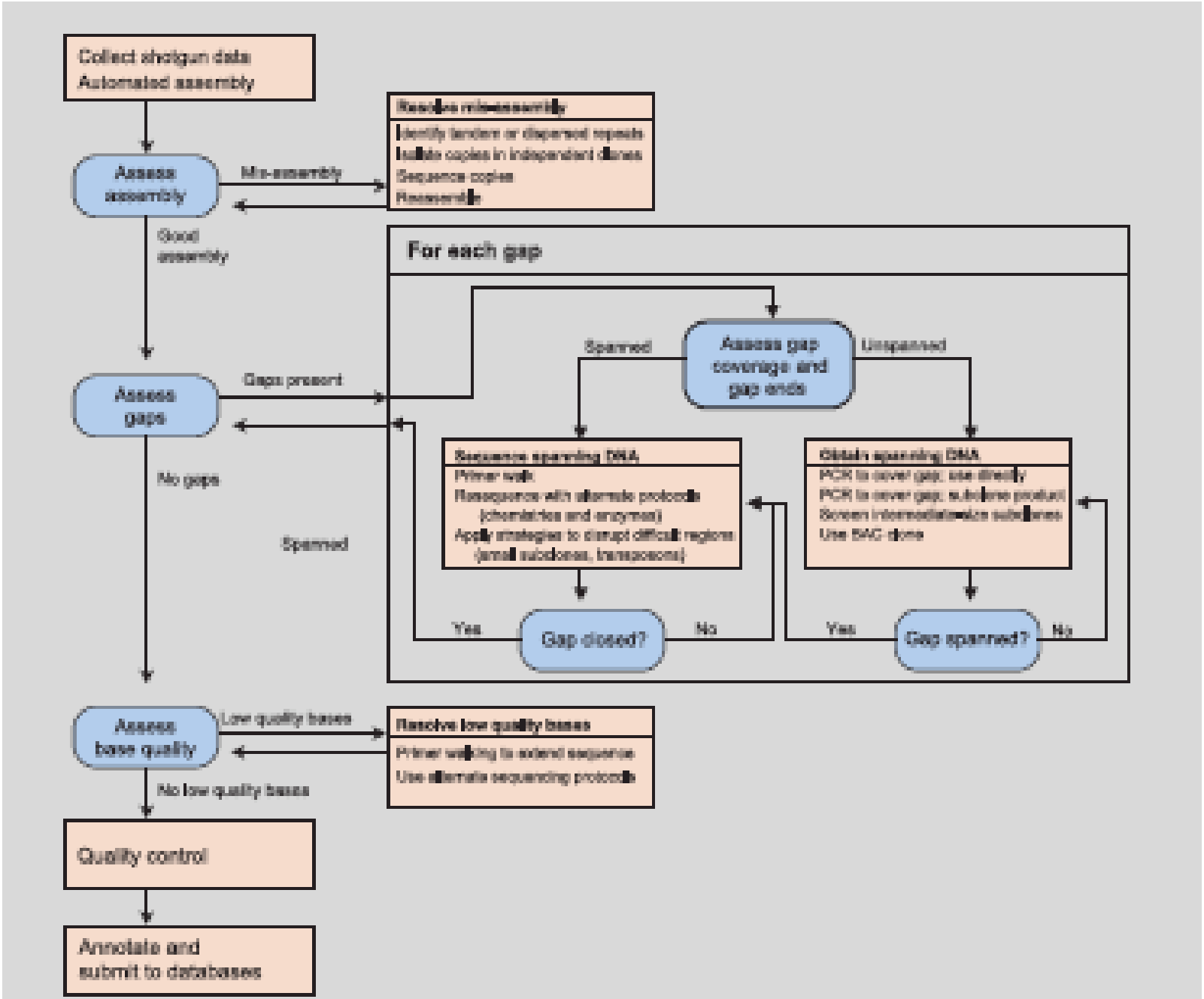
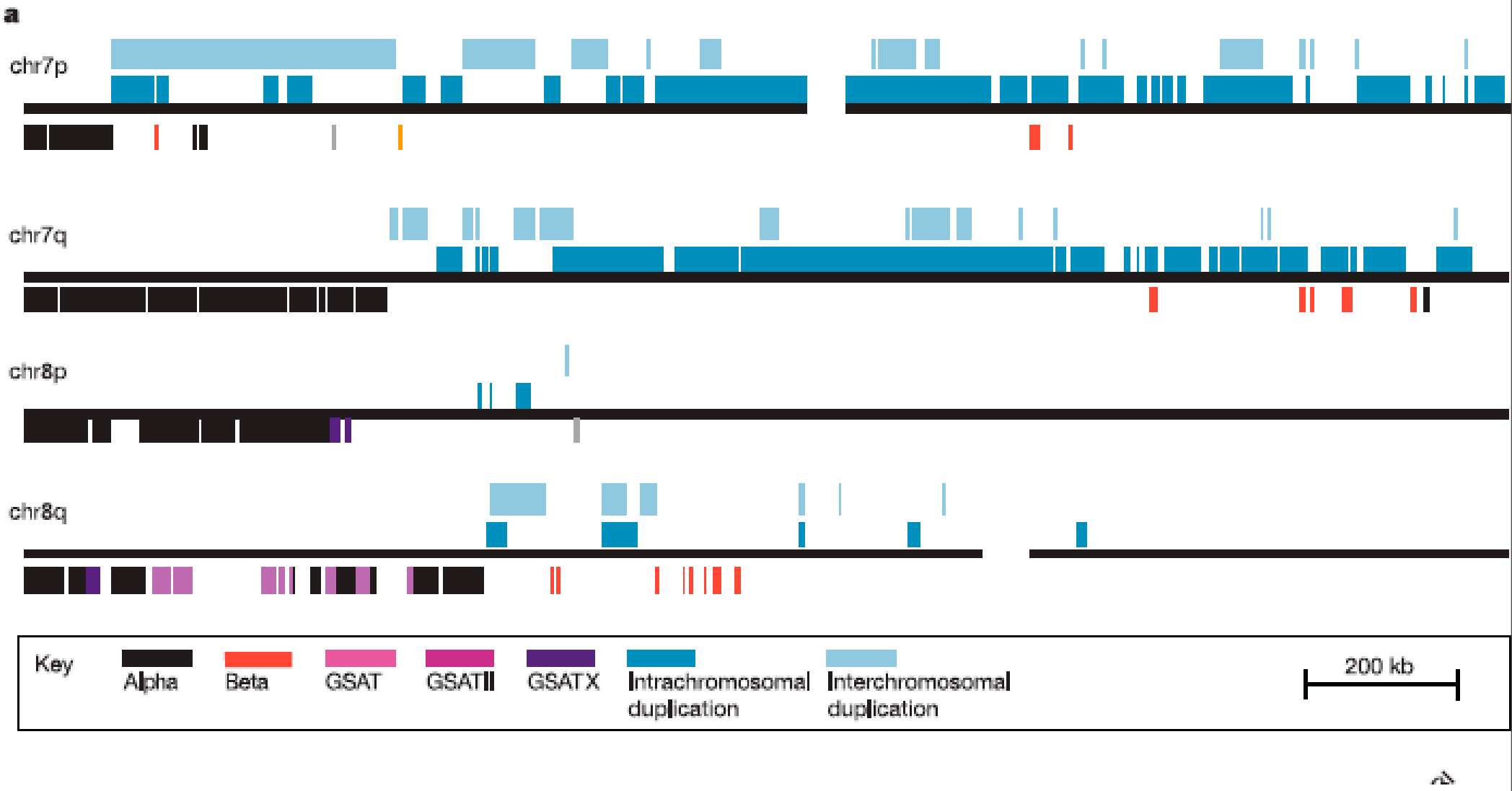


Box 2 Figure 1 Simplified flowchart for finishing of clones.

Table 3 **Chromosome arm length and contiguity in draft and reference sequence**

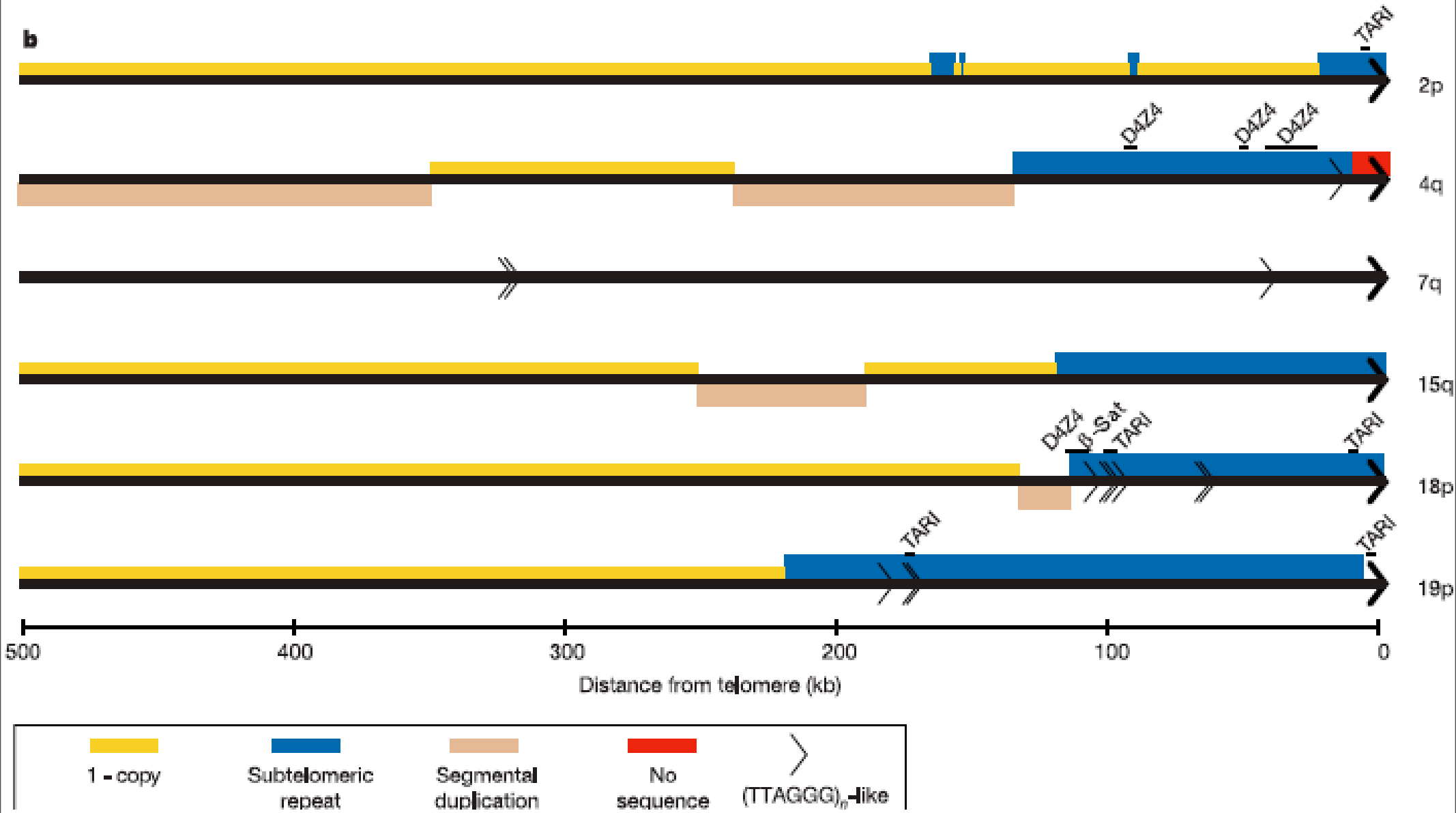
Chromosome	Euch. length* (bp)	N50† draft§ (bp)	Build 35 N50 ref‖ (bp)	Build 35 N-average ref§ (bp)
1p	121,147,476	81,895	16,783,271	33,566,574
1q	104,135,370	45,843	56,331,646	36,675,159
2p	91,748,045	68,853	68,373,980	53,478,029
2q	148,270,183	50,481	84,213,156	54,482,973
3p	90,587,544	39,322	66,080,833	54,853,737
3q	106,018,194	35,734	100,530,261	96,935,077
4p	49,501,045	36,494	9,040,907	13,797,821
4q	138,910,172	31,876	92,070,735	66,386,026
5p	46,441,398	59,470	46,378,398	46,378,398
5q	131,416,467	81,416	41,199,371	33,564,217
6p	58,938,125	251,648	48,945,890	42,200,138
6q	109,037,573	150,424	61,695,806	46,408,435
7p	57,864,988	399,235	47,497,097	40,050,874
7q	97,763,150	298,612	64,426,257	46,810,648
8p	43,958,052	40,151	9,464,880	9,872,060
8q	99,316,773	37,528	57,155,273	47,945,192
9p	46,035,928	87,767	39,435,726	34,619,306
9q	74,393,339	43,983	40,394,264	29,078,785
10p	39,244,941	48,121	20,794,160	15,791,760
10q	93,788,686	47,401	30,112,613	31,833,318
11p	51,450,781	34,383	49,571,094	48,044,101
11q	80,001,602	42,527	17,911,127	26,070,918
12p	34,747,961	197,985	27,615,668	23,435,010
12q	96,306,849	47,272	32,815,934	29,605,325
13p	acro arm	n/a	n/a	n/a
13q	96,274,979	70,497	67,740,325	54,830,719
14p	acro arm	n/a	n/a	n/a
14q	88,298,584	1,370,997	88,290,585	88,290,585
15p	acro arm	n/a	n/a	n/a
15q	82,078,915	30,303	53,619,965	38,049,097
16p	35,143,302	160,390	25,336,229	20,462,803
16q	43,883,952	86,933	42,003,582	40,305,188
17p	22,187,133	114,901	21,163,833	20,341,190
17q	56,487,608	82,866	11,472,733	15,591,618
18p	15,400,898	59,951	15,400,898	15,400,898
18q	59,352,257	50,087	33,548,238	26,073,241
19p	26,923,622	82,369	15,825,424	12,506,733
19q	33,888,028	167,408	31,383,029	31,383,029
20p	26,267,569	1,436,102	26,259,569	26,259,569
20q	34,402,734	1,301,134	26,144,333	21,428,992
21p¶	490,223	n/a	490,223	490,223

Near centromeres repeat satellites; duplication factory

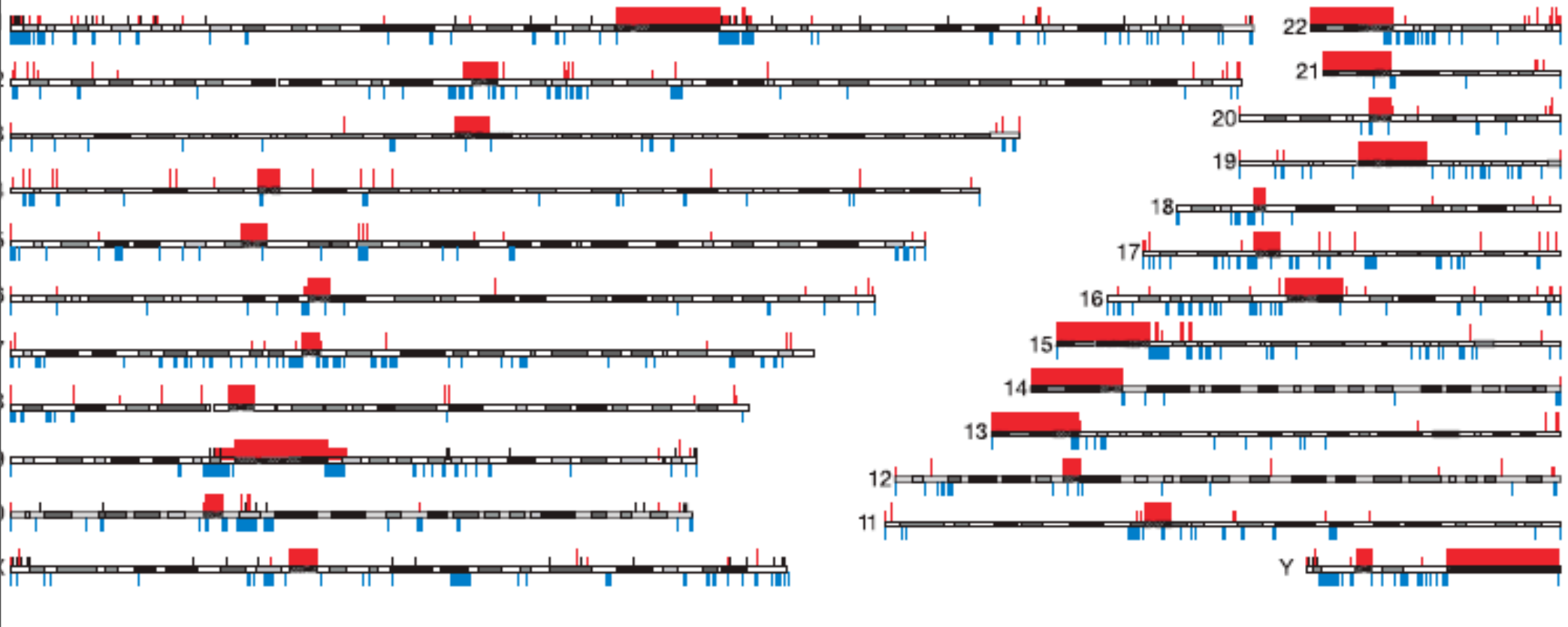


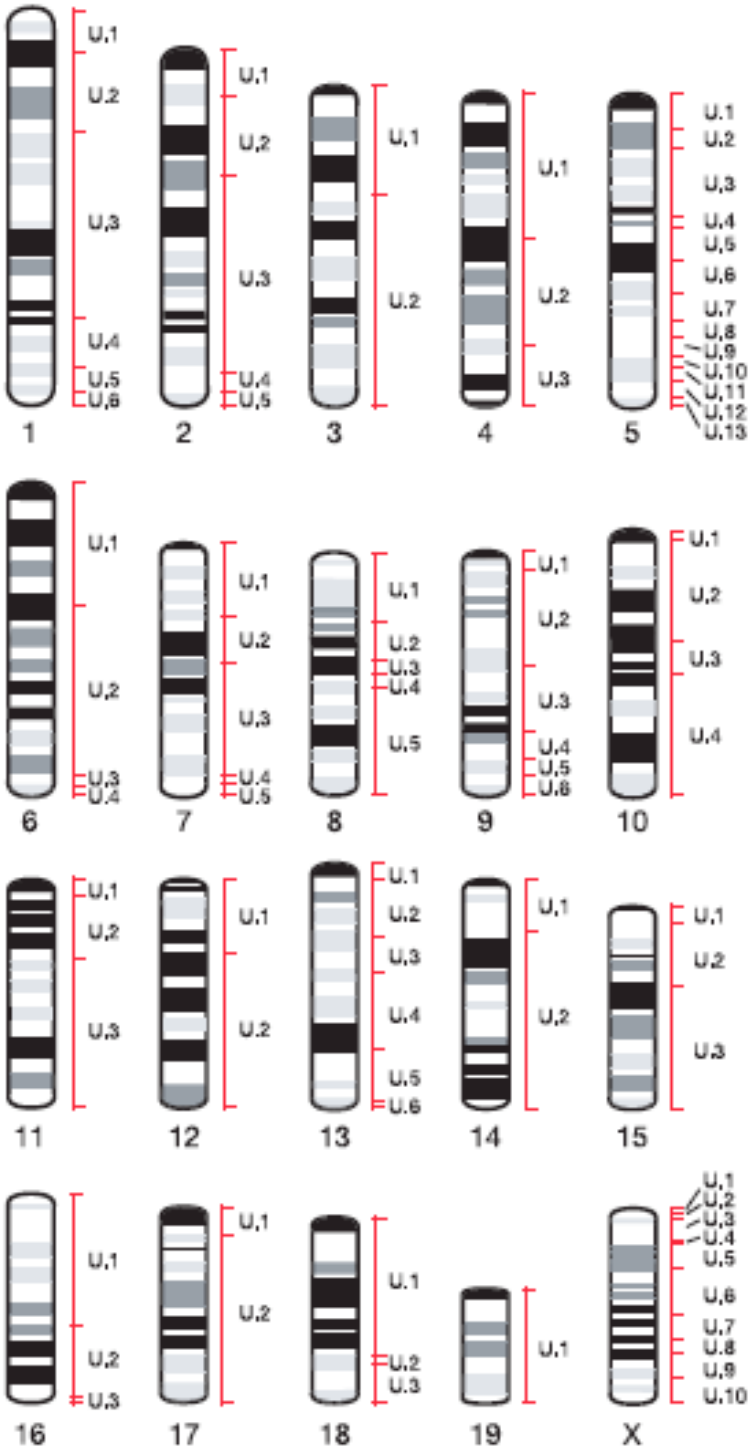
Near telomeres again repeats; seg duplications; ??

J



segmental duplications --wow - unexpected and difficult
To identify correctly (or at all)





Ultra contigs of the mouse Genome

How its sequence was determined

WGS: whole genome shotgun)
HGS: BAC based hierarchical shotgun
BAC Finishing
(BAC definition, origins)

Magabase scale mouse/human similarity
Reciprocal unique shared sequences

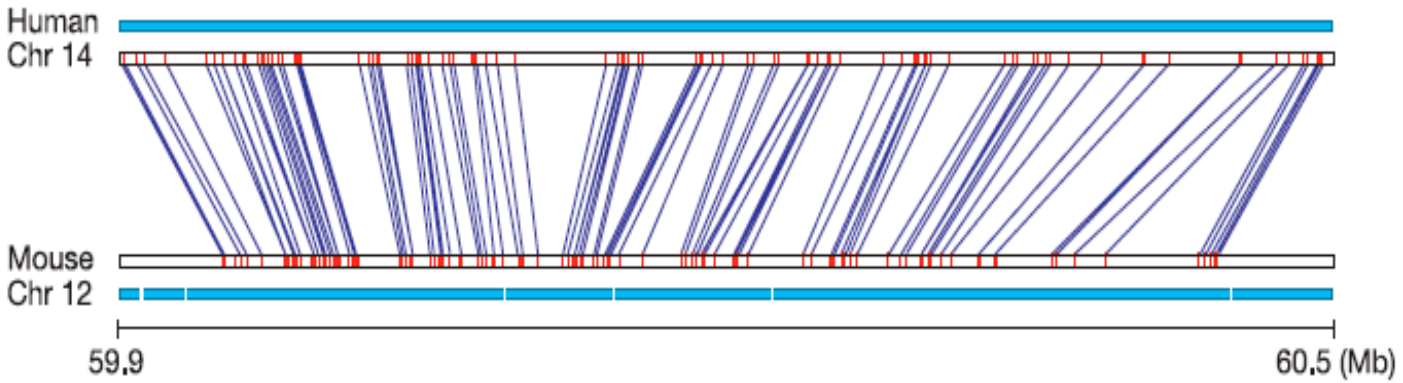
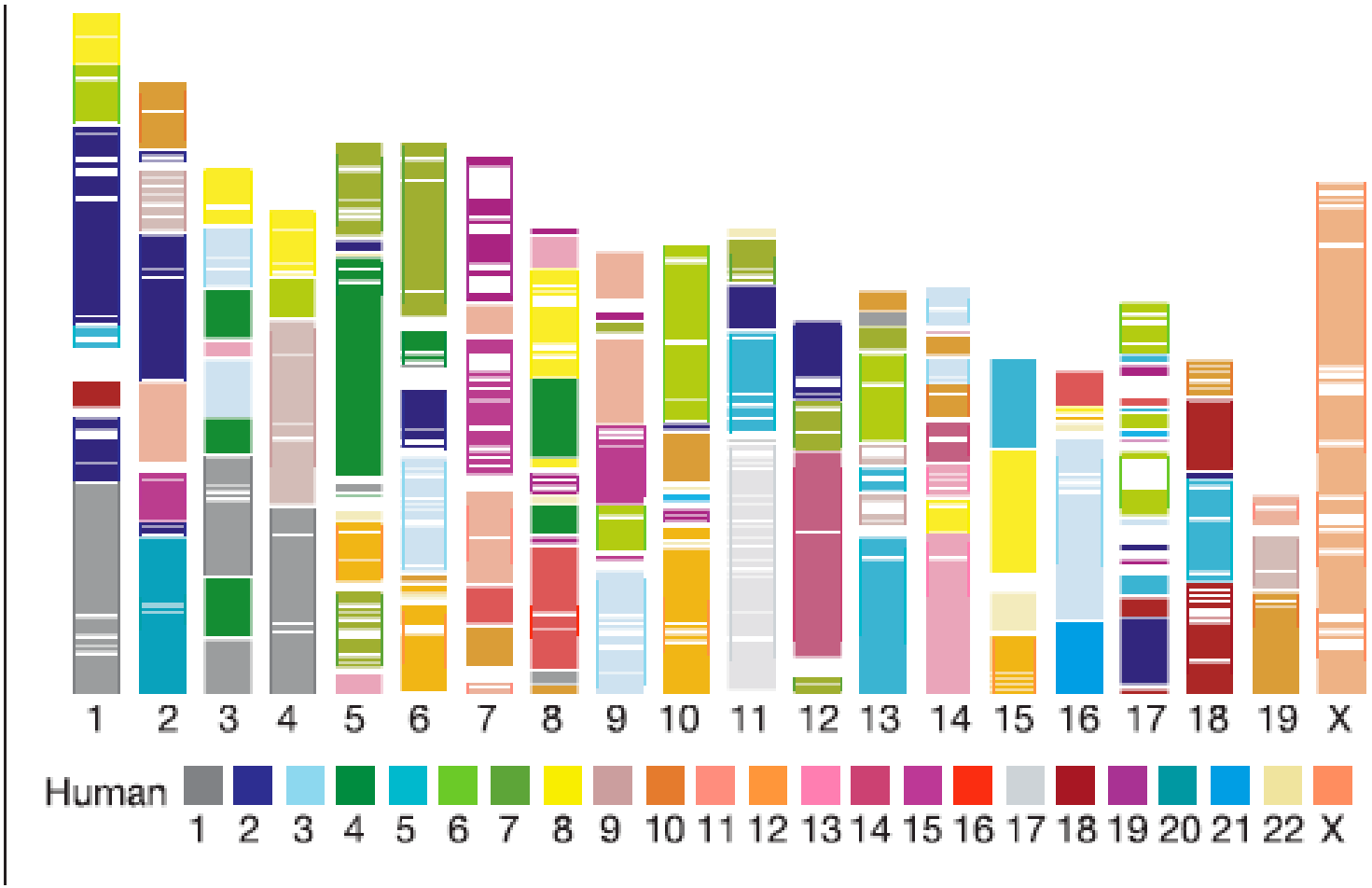


Figure 2 Conservation of synteny between human and mouse. We detected 558,000 highly conserved, reciprocally unique landmarks within the mouse and human genomes, which can be joined into conserved syntenic segments and blocks (defined in text). A typical 510-kb segment of mouse chromosome 12 that shares common ancestry with a 600-kb section of human chromosome 14 is shown. Blue lines connect the reciprocal unique matches in the two genomes. The cyan bars represent sequence coverage in each of the two genomes for the regions. In general, the landmarks in the mouse genome are more closely spaced, reflecting the 14% smaller overall genome size.

Chromosome scale similarity Mouse to Human

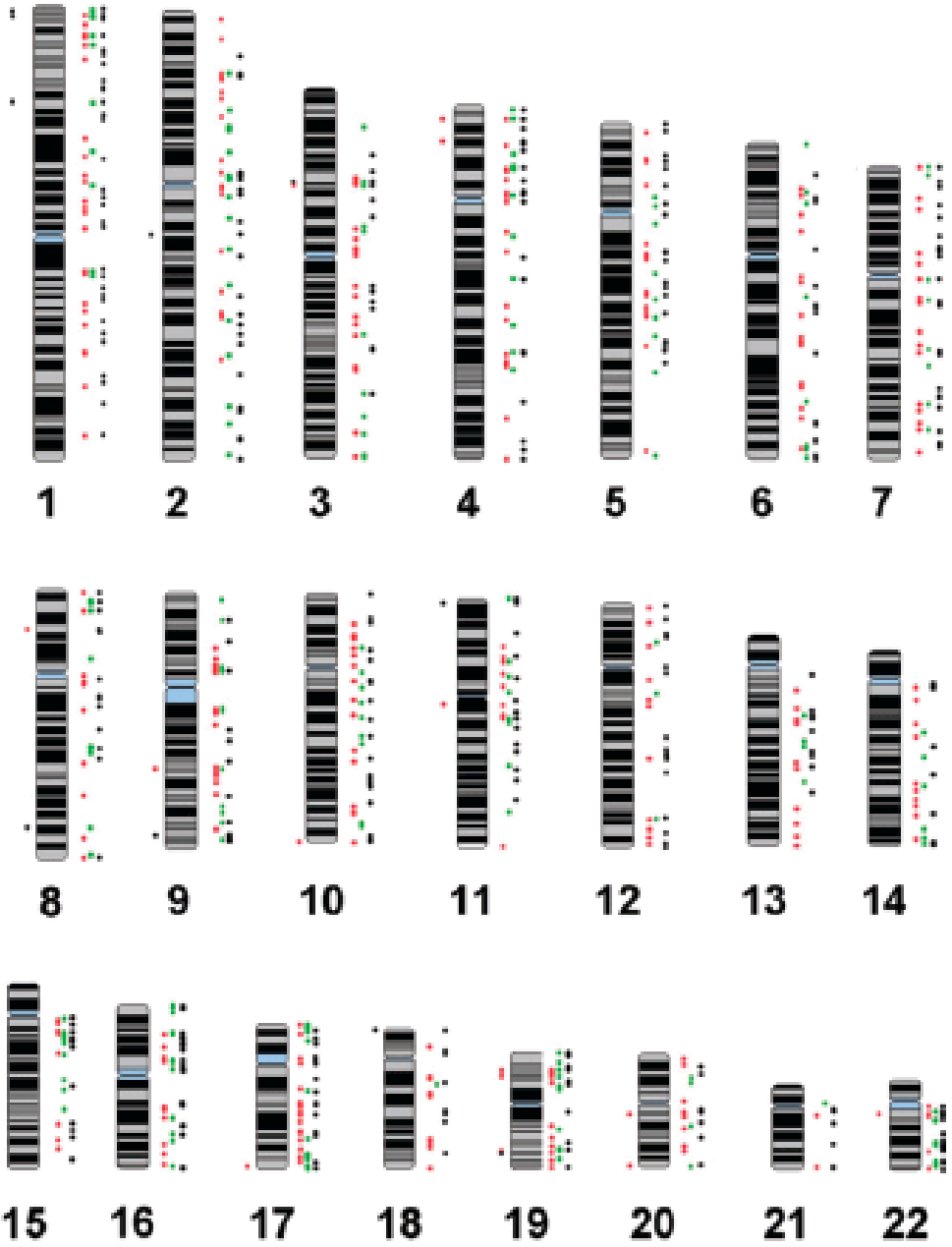
Big blocks, but considerable, apparently random reorganization

295 events define most parsimonious path separating mouse from human



Repeat architecture of mouse & human

	Mouse				Human	
	Thousands of copies	Length occupied (Mb)	Fraction of genome (%)	Lineage specific (%)	Fraction of genome (%)	Lineage specific (%)
LINEs	660	475.3	19.20	16.46	20.99	7.94
LINE1	599	464.8	18.78	16.46	17.37	7.94
LINE2	53	9.4	0.38	–	3.30	–
L3/CR1	8	1.2	0.05	–	0.32	–
SINEs	1,498	202.9	8.22	7.63	13.64	10.74
B1 (Alu)	564	67.3	2.66	2.66	10.74	10.74
B2	348	59.6	2.39	2.39	–	–
B4/RSINE	391	57.1	2.36	2.36	–	–
ID	79	5.3	0.25	0.25	–	–
MIR/MIR3	115	14.1	0.57	–	2.90	–
LTR elements	631	244.3	9.87	8.72	8.55	4.09
ERV_classI	34	16.8	0.68	0.58	2.92	2.02
ERV_classII	127	79.1	3.14	3.14	0.30	0.30
ERV_classIII	37	14.0	0.58	0.32	1.55	0.19
MaLRs (III)	388	112.2	4.82	4.02	3.78	1.58
DNA elements	112	21.8	0.88	0.36	3.03	1.00
Charlie	82	15.2	0.62	0.35	1.41	0.14
Other hATs	8	1.6	0.06	–	0.31	–
Tigger	24	4.4	0.17	–	1.06	0.76
Mariner	1	0.2	0.01	0.01	0.10	0.07
Unclassified	26	9.2	0.38	0.37	0.15	0.14
Total	2,926	953.6	38.55	33.53	46.36	24.05
Small RNAs	19	1.5	0.06	0.04	0.04	0.02
Satellites	7	0.7	0.30	NA	0.34	NA
Simple repeats	960	56.1	2.27	NA	0.87	NA



igure 6. Distribution of CNV clones. High-frequency CNV clones are shown as dots to the right of each chromosome; red, green, and lack dots represent presence in three, four or five, and six or more individuals, respectively. Dots to the left of the chromosomes epresent locations of CNVs that overlap microRNAs (*red dots*) and select cancer genes (*black dots*).

Human
Segmental dups

1kb to 500kb
>90% similar

2 - 6 copies

>5% of genome

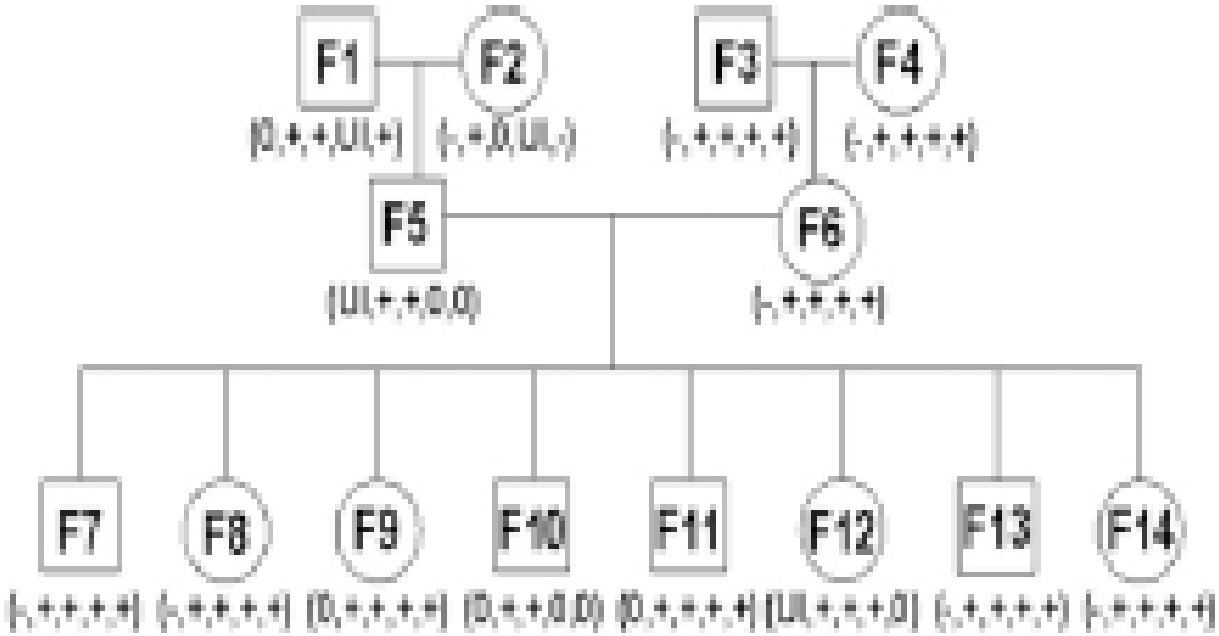


Figure 8. Inheritance of CNVs at five olfactory receptor loci in 14 members of a CEPH pedigree. The five loci (and clones), in the order shown, are *OR2A1* (RP11-466J6), *OR2Z1* (RP11-367L15 and RP11-282G19), *OR4K1* (RP11-449I24 and CTD-2024K23), *OR4M1* (RP11-597A11), and *OR4Q3* (RP11-490A23). − = Copy-number loss; + = copy-number gain; 0 = no copy-number change; UI = uninformative. Male and female family members are shown as squares and circles, respectively.

Table 3. Sensory-Related Genes Associated with CNVs

Chromosome Band	Gains and Losses[a]	Gene(s)[b]	Product[c]	Disease[c]	Clone(s) in Locus[d]
1p36.31	25	*TAS1R1*	Sweet taste receptor T1r isoform a,b,c,d	...	RP11-58A11, RP11-719E21
3p21.31	18	*GNAT1*	Guanine nucleotide binding protein, alpha	Night blindness, congenital stationary	RP11-787O14
7q32.1	5	*IMPDH1*	Inosine monophosphate dehydrogenase 1 isoform a,b	Retinitis pigmentosa-10	RP11-636E12
7q32.1	3	*OPN1SW*	Opsin 1 (cone pigments), short-wave-sensitive	Colorblindness, tritan	RP11-638M14
7q35	54	*OR2A12, OR2A14, OR2A2, OR2A25, OR2A5, OR2A1, OR2A42, OR2A7*	Olfactory receptor, family 2, subfamily A	...	RP11-703N5, RP11-466J6
8p23.3	5	*OR4F21, OR4F29*	Olfactory receptor, family 4, subfamily F	...	RP11-418D21
11q11	8	*OR4C6, OR4P4, OR4S2, OR5D13*	Olfactory receptor, family 4, subfamily C,P,S,D	...	RP11-626N6
11q12.3	3	*ROM1*	Retinal outer segment membrane protein 1	Retinitis pigmentosa, digenic	RP11-484M5
12p13.2	3	*TAS2R14, TAS2R44, TAS2R48, TAS2R49, TAS2R50*	Taste receptor, type 2, member 14,44,48,49,50	...	RP11-202N1
12q13.2	3	*OR6C2, OR6C4, OR6C68, OR6C70*	Olfactory receptor, family 6, subfamily C	...	RP11-222A15
14q11.2	61	*OR4M1, OR4Q3, OR4K1, OR4K2, OR4K5, OR4N2, OR4K13, OR4K14, OR4K15*	Olfactory receptor, family 4, subfamily M,Q,K,N	...	RP11-597A11, RP11-490A23, RP11-449I24, CTD-2024K23
15q11.2	26	*OR4M2, OR4N4*	Olfactory receptor, family 4, subfamily M,N	...	RP11-281J20
16p13.3	7	*OR1F1*	Olfactory receptor, family 1, subfamily F	...	RP11-680M24
17q25.3	18	*ACTG1, FSCN2*	Actin, gamma 1 propeptide; fascin 2	Deafness, autosomal dominant 20/26; retinitis pigmentosa-30	RP11-730A9, RP13-550B21
19p13.2	62	*OR2Z1*	Olfactory receptor, family 2, subfamily Z	...	RP11-282G19, RP11-367L15
22q11.1	15	*OR11H1*	Olfactory receptor, family 11, subfamily H	...	RP11-561P7
22q12.3	5	*MYH9*	Myosin, heavy polypeptide 9, nonmuscle	Deafness, autosomal dominant 17	RP11-108P21

[a] Total number of copy-number gains and losses observed for a CNV locus.

Table 4. Select Examples of CNVs Associated with Cancer-Related Genes

Chromosome Band	Gains and Losses[a]	Gene(s)[b]	Product[c]	Clone(s) in Locus[d]
1p36.33	40	*SKI*	V-ski sarcoma viral oncogene homolog	RP11-83K22, RP11-181G12
1p36.32	12	*TP73*	Tumor protein p73	RP11-631K6
1p36.31	16	*TNFRSF25*	Tumor necrosis factor receptor superfamily,	RP11-58A11
1p32.3	32	*RAB3B*	RAB3B, member RAS oncogene family	RP11-469M21, RP11-91A18
1p13.3	6	*VAV3*	Vav 3 oncogene	RP11-480L11
2q14.2	18	*RALB*	V-ral simian leukemia viral oncogene homolog B	RP11-818M2
2q37.3	6	*BOK*	BCL2-related ovarian killer	RP11-343P10
3p21.31	20	*NAT6, TUSC2, TUSC4*	Putative tumor suppressor FUS2, tumor suppressor candidates 2 & 4	RP11-787O14, RP13-487A19
4q31.1	3	*RAB33B*	RAB33B, member RAS oncogene family	RP11-124P22
6q21	3	*C6orf210*	Candidate tumor suppressor protein	RP11-601O12
6q25.1	20	*ESR1*	Estrogen receptor 1	RP11-655H19
7p22.3	19	*MAFK*	V-maf musculoaponeurotic fibrosarcoma oncogene	RP11-16P10
7p22.3	6	*MAD1L1*	MAD1-like 1	RP11-32509
8q24.21	4	*MYC*	V-myc myelocytomatosis viral oncogene homolog	CTD-2034C18
9q34.2	22	*VAV2*	Vav 2 oncogene	RP11-352K12, RP11-651E2
10p11.23	11	*MAP3K8*	Mitogen-activated protein kinase kinase kinase	RP11-350D11
11p15.4	15	*CDKN1C*	Cyclin-dependent kinase inhibitor 1C	RP11-494F4
11p13	3	*WT1, WIT-1*	Wilms tumor 1 isoform A/B/C/D, Wilms tumor associated protein	RP11-710L2
11p11.2	3	*C1QTNF4*	C1q and tumor necrosis factor related protein 4	RP11-425G10
11q13.1	3	*MEN1*	Menin isoform 1	RP11-48509
11q13.3	6	*CCND1, ORAOV1*	Cyclin D1, oral cancer overexpressed 1	RP11-124K14
12q13.12	4	*MLL2*	Myeloid/lymphoid or mixed-lineage leukemia 2	RP11-66M13
13q31.1	4	*C13orf10*	Cutaneous T-cell lymphoma tumor antigen se70-2	RP11-86D5
14q32.32	3	*TNFAIP2*	Tumor necrosis factor, alpha-induced protein 2	RP11-455L5
16p13.3	19	*AXIN1*	Axin 1 isoform a/b	RP11-598I20
16q22.3	3	*BCAR1*	Breast cancer anti-estrogen resistance 1	RP11-109K6
17p13.2	6	*TAX1BP3*	Tax1 (human T-cell leukemia virus type I)	RP11-753P16
17q11.2	6	*NF1*	Neurofibromin	RP11-518B17
17q21.32	3	*PHB*	Prohibitin	RP11-472H5
17q25.3	17	*MAFG*	V-maf musculoaponeurotic fibrosarcoma oncogene	RP11-634L10, RP11-712H22
17q25.3	6	*C1QTNF1*	C1q and tumor necrosis factor related protein 1	RP11-167M2
18p11.32	15	*YES1*	Viral oncogene yes-1 homolog 1	RP11-806L2
18q21.1	8	*DCC*	Deleted in colorectal carcinoma	RP11-346H17
19p13.3	6	*SH3GL1*	SH3-domain GRB2-like 1	RP11-406I1
19p13.3	4	*TNFSF9, TNFSF7, TNFSF14*	Tumor necrosis factor (ligand) superfamily, members	RP11-526C20
19p13.3	4	*VAV1*	Vav 1 oncogene	CTD-2200O16
19p13.11	16	*RAB3A*	RAB3A, member RAS oncogene family	RP11-512B16
19q13.33	15	*PTOV1*	Prostate tumor overexpressed gene 1	RP11-507G9
19q13.33	7	*BAX*	BCL2-associated X protein isoform sigma/gamma/epsilon/delta/beta/alpha	CTD-2017J20
19q13.33	8	*RRAS*	Related RAS viral (r-ras) oncogene homolog	RP11-264M8, RP11-808J4
20q13.13	3	*BCAS4*	Breast carcinoma amplified sequence 4 isoform a/b	RP11-124P7
22q11.21	3	*HIC2*	Hypermethylated in cancer 2	CTD-2245I11

[a] Total number of copy number gains and losses observed for a CNV locus.

Select CNVs Overlapping Genes Associated with Diseases or Disease Susceptibility

me	Gains and Losses[a]	Gene(s)[b]	Product(s)[c]	Disease[d]	Clone(s) in Locus[e]
	7	*NROB2*	Short heterodimer partner	Obesity, mild, early-onset	RP11-402E20
	7	*TTN*	Titin isoform N2-A, N2-B; isoform novex-1,2,3	Muscular dystrophy, limb-girdle, type 2J	RP11-95I17
	3	*SGCB*	Sarcoglycan, beta (43kDa dystrophin-associated)	Muscular dystrophy, limb-girdle, type 2E	RP11-61F5
	60	*SMA3, SMA4*	SMA3, SMA4	Spinal muscular atrophy-2,-1	RP11-313J5, RP11-155O16
	6	*SMN1*	Survival of motor neuron 1, telomeric isoform d	Spinal muscular atrophy-4	RP11-195E2
	34	*LPA*	Lipoprotein, Lp(a)	Coronary artery disease, susceptibility to	CTD-2310B5
	5	*PARK2*	Parkin isoform 1, 2, 3	Parkinson disease, juvenile, type 2	CTD-2019O18
	10	*GCK*	Glucokinase isoform 2,3	Diabetes mellitus, neonatal-onset	RP11-808H7
	4	*GPR51*	G protein-coupled receptor 51	Nicotine dependence, susceptibility to	RP11-786E15
	3	*BSCL2*	Seipin	Spinal muscular atrophy, distal, type V	RP11-484M5
	79	*A2M*	Alpha-2-macroglobulin precursor	Alzheimer disease, susceptibility to	RP11-536M6
	29	*TBXA2R*	Thromboxane A2 receptor isoform 2	Bleeding disorder due to defective thromboxane A2 receptor	RP11-584K12
	3	*FKRP*	Fukutin-related protein	Muscular dystrophy, limb-girdle, type 2I	RP11-422M7
	6	*COMT*	Catechol-O-methyltransferase isoform S-COMT	Schizophrenia, susceptibility to	RP11-651A4

number of gains and losses observed for a CNV locus.

Regulatory gene family evolution: Stubbs group