

SUPPLEMENTARY INFORMATION

Supplementary Figure 1S

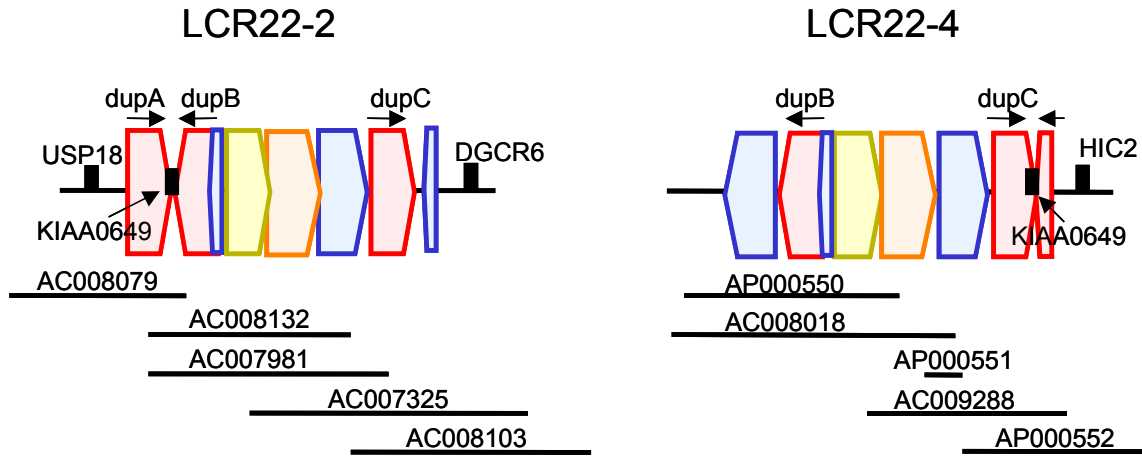


Figure 1S. Block structure of LCR22-2 and LCR22-4. Four functional genes, *USP18*, *GGT*, *GGTLA* and *BCR* map to the LCR22s. Unprocessed pseudogene copies are also present and depicted in different colors (*USP18*, red; *BCR*, blue, *GGT* and *GGTLA*, yellow and *IGSF3*, orange; Babcock et al., 2003). DupA, B and C, correspond to duplications shown in Figure 1. The functional locus of *USP18*, in the forward orientation with respect to the centromere of chromosome 22 is juxtaposed to LCR22-2. The *DGCR6* gene, in the forward orientation, is just telomeric to LCR22-2. A pseudogene, *KIAA0649* is present in the forward orientation in both LCR22-2 and -4. LCR22-2 and -4 have some differences in block structure. For example, the orientation of *KIAA0649* with respect to the orientation of the *GGT*, *IGSF3* and *BCR* blocks are unique to each LCR22. There are two copies of the *BCR* pseudogene in LCR22-4 but only one in LCR22-2. This has made it possible to precisely map the large insert genomic clones as shown below the two LCR22s (GenBank accession numbers provided). In addition, some of the clones contain unique sequences outside the LCR22s, such as for AC008079, AC007325, AP000550, AC008018, AC009288 and AP000552. This made it possible to anchor them simply based upon the sequence of the insert ends of the clones.

Supplementary Figure 2S

A

REGION	IDENTITY	CHR	INTERVAL	GENBANK	LCR
1-383	100.0%	22	17132218-17132600	AC008132	LCR22-2
1-383	99.3%	22	18737251-18737633	AC023490	LCR22-3A
1-383	98.2%	22	19883241-19883623	AP000550	LCR22-4
1-383	95.1%	22	21318066-21318447	D87002	LCR22-5
1-383	94.8%	20	23910156-23910537	AL133466	LCR20
1-383	94.6%	22	23352643-23353026	AP000356	LCR22-8
1-383	90.4%	22	22966012-22966386	AP000354	LCR22-7

LCR22-2 PSV
 TAGTCAGCATCAAGGTGGAGCTGAACAGAGGCAGAAGGAGGAGGACCCAA
 GGTGGTGTCACTCAGGACCCGGGTTCAAGTCCTTATGCTTCTGCAGCCTG 302
 GCCTGGGTCCCCAACCTCCCCAGGGTGACCAAGGGCTTCCAGTCTGCA 331
 CAGAGGACAGGGGGTCTTGACAGCATCAAATGCTAGTGACTAGGAGACAC 406
 TTACGTGGGAAATGCAGACAGACCATGCCTCTAGCCCTTGGCACCAGGCA 418
 CCATCCATCCCTGGGACTTGCTGTCTGCTGAAATGCAGCATGGCTCCTCCAG 506
 GGAGGGGGGCTGTGCCATGTGGGGGCCCCACCCACCTGCAGCTCTTTCC
 CACCCTGGCTGCAGGTCTGCTTCTCTGAATCCA 577

B

REGION	IDENTITY	CHR	INTERVAL	GENBANK	LCR
1-551	100.0%	22	17149097-17149647	AC008132	LCR22-2
1-551	98.4%	22	19900208-19900758	AP000550	LCR22-4
1-551	97.5%	22	23335599-23336149	AP000356	LCR22-8

LCR22-2 551 bp interval: SPSs
 TGCCTGTTGAAAAGGCAGGATATAGACAGCCTGGGTTCAATTTTAGCTTC
 ACCACCTCCCAGCTGTGTGACCTCAGCTGATTTGCATGACCTGTCTGAGC
 CTCAGCATCCCCACCCTGTAAATGGGAATCCACACAGCATCCCCTTGCC
 CAAAGGAGCAGGGAGGGTTGTGAGAGGCTCGTGGGTGAAAAGCACAGAGC
 AGAGCATGGGCCCCAGTGAGCCCTGATCCATGAGGTCTGCTAGCATAATA
 ATTATTCTTCCATGTGCTGCACAGAGTGGCCCCGGAGGCCTTAGCAGAA 296, 301, 313
 ATAACAGAAGCTCCCGGCCCTTTACCATGGTGATGATGGTCCTGACCACT
 CACTGTGGGAGGGTGCTATGGGGACAGGAAGGGATGGGGGGTGCTAGAAC
 TGCCCCCTGAACCCTGACGGGAGCAGGCTCTGTGGGCAAGGCCCTTTCCC 476
 GGTGGCTCAGCCAGCTCTGCACCCACGCCCAAGTCTGCCGCATGGCTTA
 CCCTTCTGGGACTGCTCCGAGCTGTTGAACATGCTGGCAAAGGCCAGCCT
 G

Figure 2S. Experimental confirmation of PSVs and SPSs examples. (A) Paralogous sequence variants in LCR22-2 and LCR22-4 in the *GGT* pseudogene loci. The duplicated copies of the chromosome 22 and 20 regions identified by BLAT are shown. The position of the sequence variant is indicated within the sequence of LCR22-2 and the presumed ancestral *GGT* locus in LCR22-8.

(B) Shared polymorphic sites in LCR22-2 and LCR22-4 in the *GGT* pseudogene loci. The regions of homology as identified by BLAT analysis are shown. The putative ancestral locus in the *GGT* genomic interval is shown, with its sequence at the same sites.

A

3

	500	*	520	*	540	*	560	*	
AC007981	:	GACGAAGTTTCGCTCTTGTGGCCAGGTTGGAGTGCAATGGTGTGATCTCAGCTCACTGCAATCCCCGCCTCAGCAGGAGAG							: 562
AC008132	:	-----							: -
AC007325	:	-----							: -
AC008103	:	-----							: -
LCR22-2	:	-----							: -
AC008018	:	-----							: -
AP000550	:	GACGAAGTTTCGCTCTTGTGGCCAGGTTGGAGTGCAATGGTGTGATCTCAGCTCACTGCAATCCCCGCCTCAACAGGAGAG							: 564
LCR22-4-5	:	GACGAAGTTTCGCTCTTGTGGCCAGGTTGGAGTGCAATGGTGTGATCTCAGCTCACTGCAATCCCCGCCTCAACAGGAGAG							: 564
AC009288	:	-----							: -
AP000552	:	GACGAAGTTTCGCTCTTGTGGCCAGGTTGGAGTGCAATGGTGTGATCTCAGCTCACTGCAATCCCCGCCTCAACAGGAGAG							: 568
LCR22-4	:	GACGAAGTTTCGCTCTTGTGGCCAGGTTGGAGTGCAATGGTGTGATCTCAGCTCACTGCAATCCCCGCCTCAACAGGAGAG							: 568

	580	*	600	*	620	*	640	*	
AC007981	:	CAGGAATCTTCAGTGATCCACTGGCGGATCTGCAGCCATTGTGCGCGCCAGGTTCCCAAGTCTTTTGTGCGCGCGCCTCT							: 644
AC008132	:	-----							: -
AC007325	:	-----							: -
AC008103	:	-----							: -
LCR22-2	:	-----							: -
AC008018	:	-----							: -
AP000550	:	CAGGAATCTTCAGTGATCCACTGGCGGATCTGCAGCCATTGTGCGCGCCAGGTTCCCAAGTCTTTTGTGCGCGCGCCTCT							: 646
LCR22-4-5	:	CAGGAATCTTCAGTGATCCACTGGCGGATCTGCAGCCATTGTGCGCGCCAGGTTCCCAAGTCTTTTGTGCGCGCGCCTCT							: 646
AC009288	:	-----							: -
AP000552	:	CAGGAATCTTCAGTGATCCACTGGCGGATCTGCAGCCATTGTGCGCGCCAGGTTCCCAAGTCTTTTGTGCGCGCGCCTCT							: 650
LCR22-4	:	CAGGAATCTTCAGTGATCCACTGGCGGATCTGCAGCCATTGTGCGCGCCAGGTTCCCAAGTCTTTTGTGCGCGCGCCTCT							: 650

	660	*	680	*	700	*	720	*	7
AC007981	:	CCTTCCAGTACCTATGCCGCCAGCGTCCCTAGCCTCCCGCCATTGCCAGCAGGTGCTGAGATCGCGCCATTGCACTCCAGC							: 726
AC008132	:	-----							: -
AC007325	:	-----							: -
AC008103	:	-----							: -
LCR22-2	:	-----							: -
AC008018	:	-----							: -
AP000550	:	CCTTCCAGTACCTATGCCGCCAGCGTCCCTAGCCTCCCGCCATTGCCAGCAGGTGCTGAGATCGCGCCATTGCACTCCAGC							: 728
LCR22-4-5	:	CCTTCCAGTACCTATGCCGCCAGCGTCCCTAGCCTCCCGCCATTGCCAGCAGGTGCTGAGATCGCGCCATTGCACTCCAGC							: 728
AC009288	:	-----							: -
AP000552	:	CCTTCCAGTACCTATGCCGCCAGCGTCCCTAGCCTCCCGCCATTGCCAGCAGGTGCTGAGATCGCGCCATTGCACTCCAGC							: 732
LCR22-4	:	CCTTCCAGTACCTATGCCGCCAGCGTCCCTAGCCTCCCGCCATTGCCAGCAGGTGCTGAGATCGCGCCATTGCACTCCAGC							: 732

	40	*	760	*	780	*	800	*	820
AC007981	:	CTGGGGGACAAGAGCGAAACTCCATGTCAAAAAAAAAAAAAAAAAAGGATGAAAATTTTGGGAAAAATATGGAAGAAACCAAATG							: 808
AC008132	:	-----							: -
AC007325	:	-----							: -
AC008103	:	-----							: -
LCR22-2	:	-----							: -
AC008018	:	-----							: -
AP000550	:	CTGGGGGACAAGAGCGAAACTCCATGTC - AAAAAAAAAAAAAAAAAAGGATGAAAATTTTGGGAAAAATATGGAAGAAACCAAATG							: 809
LCR22-4-5	:	CTGGGGGACAAGAGCGAAACTCCATGTC - AAAAAAAAAAAAAAAAAAGGATGAAAATTTTGGGAAAAATATGGAAGAAACCAAATG							: 809
AC009288	:	-----							: -
AP000552	:	CTGGGGGACAAGAGCGAAACTCCATGTC - AAAAAAAAAAAAAAAAAAGGATGAAAATTTTGGGAAAAATATGGAAGAAACCAAATG							: 811
LCR22-4	:	CTGGGGGACAAGAGCGAAACTCCATGTC - AAAAAAAAAAAAAAAAAAGGATGAAAATTTTGGGAAAAATATGGAAGAAACCAAATG							: 811

		1240	*	1260	*	1280	*	1300	*	
AC007981	:	GTCTCCAACCTCGTACCTCGTGATCCGCCCGCCTCCGACTCCCAAAGTGCCACAGCCAGGCCTTTTTTTTTTAAGACAGAG								: 1300
AC008132	:	-----								: -
AC007325	:	-----								: -
AC008103	:	-----								: -
LCR22-2	:	-----								: -
AC008018	:	-----								: -
AP000550	:	GTCTCCAACCTCGTGACATCGTGATCCGCCCGCCTCCGACTCCCAAAGTGCCACAGCCAGGCCTTTTTTTTTTAAGACAGAG								: 1295
LCR22-4-5	:	GTCTCCAACCTCGTGACATCGTGATCCGCCCGCCTCCGACTCCCAAAGTGCCACAGCCAGGCCTTTTTTTTTTAAGACAGAG								: 1295
AC009288	:	-----								: -
AP000552	:	GTCTCCAACCTCGTGACATCGTGATCCGCCCGCCTCCGACTCCCAAAGTGCCACAGCCAGGCCTTTTTTTTTTAAGACAGAG								: 1294
LCR22-4	:	GTCTCCAACCTCGTGACATCGTGATCCGCCCGCCTCCGACTCCCAAAGTGCCACAGCCAGGCCTTTTTTTTTTAAGACAGAG								: 1294

		1320	*	1340	*	1360	*	1380	*	
AC007981	:	TCTCGGCCCGGCGCGGTGGCTCAGGCCTTTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCTGATCACGAGGTCAGGAGATC								: 1382
AC008132	:	-----								: -
AC007325	:	-----								: -
AC008103	:	-----								: -
LCR22-2	:	-----								: -
AC008018	:	-----								: -
AP000550	:	TCTCGGCCCGGCGCGGTGGCTCAGGCCTTTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCTGATCACGAGGTCAGGAGGTC								: 1377
LCR22-4-5	:	TCTCGGCCCGGCGCGGTGGCTCAGGCCTTTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCTGATCACGAGGTCAGGAGGTC								: 1377
AC009288	:	-----								: -
AP000552	:	TCTCGGCCCGGCGCGGTGGCTCAGGCCTTTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCTGATCACGAGGTCAGGAGGTC								: 1376
LCR22-4	:	TCTCGGCCCGGCGCGGTGGCTCAGGCCTTTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCTGATCACGAGGTCAGGAGGTC								: 1376

		1400	*	1420	*	1440	*	1460	*	
AC007981	:	GAGACCATCCTGGCTAACACGGTCAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGATGTGGTGGCGGGCGCCTGT								: 1464
AC008132	:	-----								: -
AC007325	:	-----								: -
AC008103	:	-----								: -
LCR22-2	:	-----								: -
AC008018	:	-----								: -
AP000550	:	GAGACCATCCTGGCTAACACGGTCAAACCCCGTCTCTACTAAAAATGCAAAAAAATTAGCCGGGTGTGGTGGCGGGCGCCTGT								: 1459
LCR22-4-5	:	GAGACCATCCTGGCTAACACGGTCAAACCCCGTCTCTACTAAAAATGCAAAAAAATTAGCCGGGTGTGGTGGCGGGCGCCTGT								: 1459
AC009288	:	-----								: -
AP000552	:	GAGACCATCCTGGCTAACACGGTCAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGTGTGGTGGCGGGCGCCTGT								: 1458
LCR22-4	:	GAGACCATCCTGGCTAACACGGTCAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGTGTGGTGGCGGGCGCCTGT								: 1458

		1480	*	1500	*	1520	*	1540	
AC007981	:	AGTCCGAGCTACTCCCGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 1531
AC008132	:	-----CGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 60
AC007325	:	-----CGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 60
AC008103	:	-----CGTGAACCCAGTAGGCGGAGCTTGCGGTGA							: 60
LCR22-2	:	-----CGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 60
AC008018	:	-----CGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 60
AP000550	:	AGTCCGAGCTACTCCCGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 1526
LCR22-4-5	:	AGTCCGAGCTACTCCCGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 1526
AC009288	:	-----CGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 60
AP000552	:	AGTCCGAGCTACTCCCGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 1525
LCR22-4	:	AGTCCGAGCTACTCCCGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGTAGGCGGAGCTTGCGGTGA							: 1525

B

AC007981	:	ACTACTCCAGAGGCTGAGGCAGGAGAATGG	CGTGAACCCGGTAGGCGGAGCTTGCGGTGA
AC008132	:	GCTACTCCAGAGGCTGAGGCAGGAGAATGG	CGTGAACCCGGTAGGCGGAGCTTGCGGTGA
AC007981	:	GCTACTCCCGAGGCTGAGGCAGGAGAATGG	CGTGAACCCGGTAGGCGGAGCTTGCGGTGA

Figure 3S. Shared polymorphism of a 1471 bp long indel between LCR22-2 and LCR22-4.

(A) Alignment of the indel region from LCR22-2/4 clones. This breakpoint corresponds to positions 121622-123105 in the clone alignment (Figure 2). We also included the genomic sequences from the hg17 assembly of the human genome. LCR22-4 has an additional copy of this segment at the 5' end covered by BAC AP000550. This deletion was caused by homologous recombination between two highly identical *AluY* elements in an *Alu*-rich cluster. Five deleted *Alu* repeats constitute nearly 70% of the deleted sequence. **(B)** Alignment of the duplication breakpoints.