Supp. Files S1-S4 can be found under Supporting Information as individual .bed files.

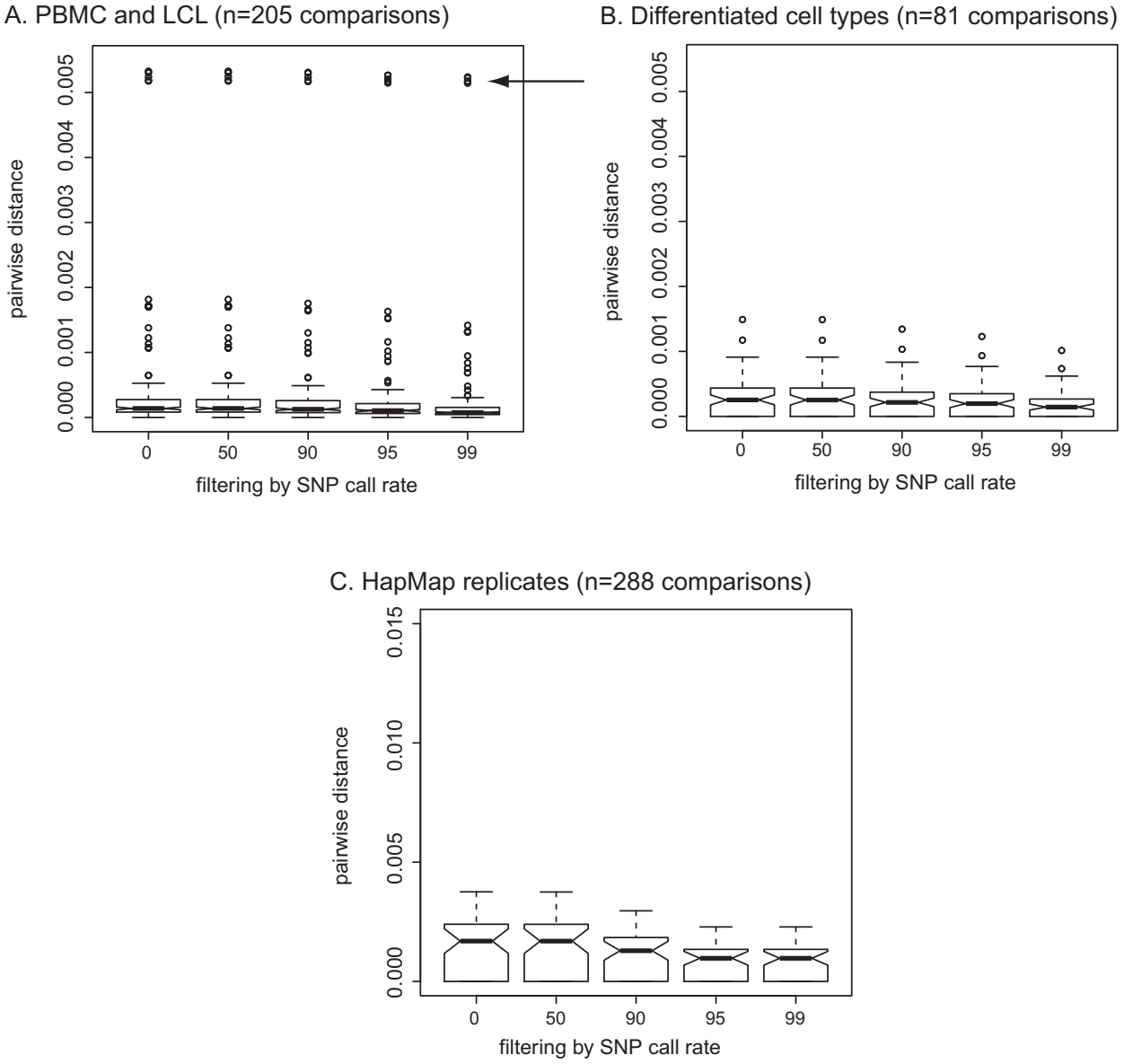
Supp. File S1. A .bed file that describes putative LCL associated loci in GENEVA SAGE, identified using CNVineta. This file may be viewed as a custom track in UCSC Genome Browser.

Supp. File S2. A .bed file that describes PBMC and LCL CNV calls. This file may be viewed as a custom track in UCSC Genome Browser.

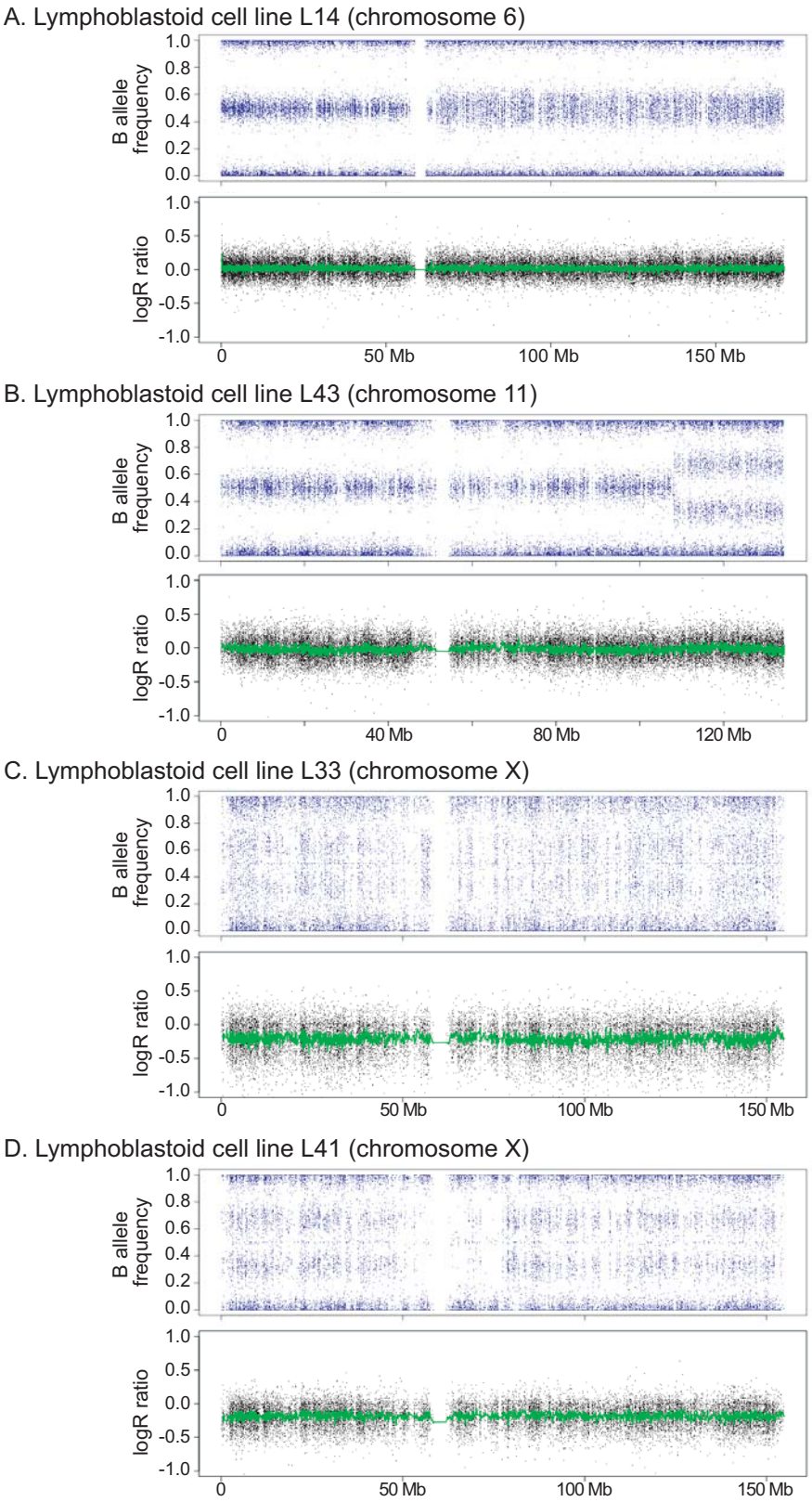
Supp. File S3. A .bed file that describes DCL CNV calls. This file may be viewed as a custom track in UCSC Genome Browser.

Supp. File S4. A .bed file that describes HapMap Replicate CNV calls. This file may be viewed as a custom track in UCSC Genome Browser.

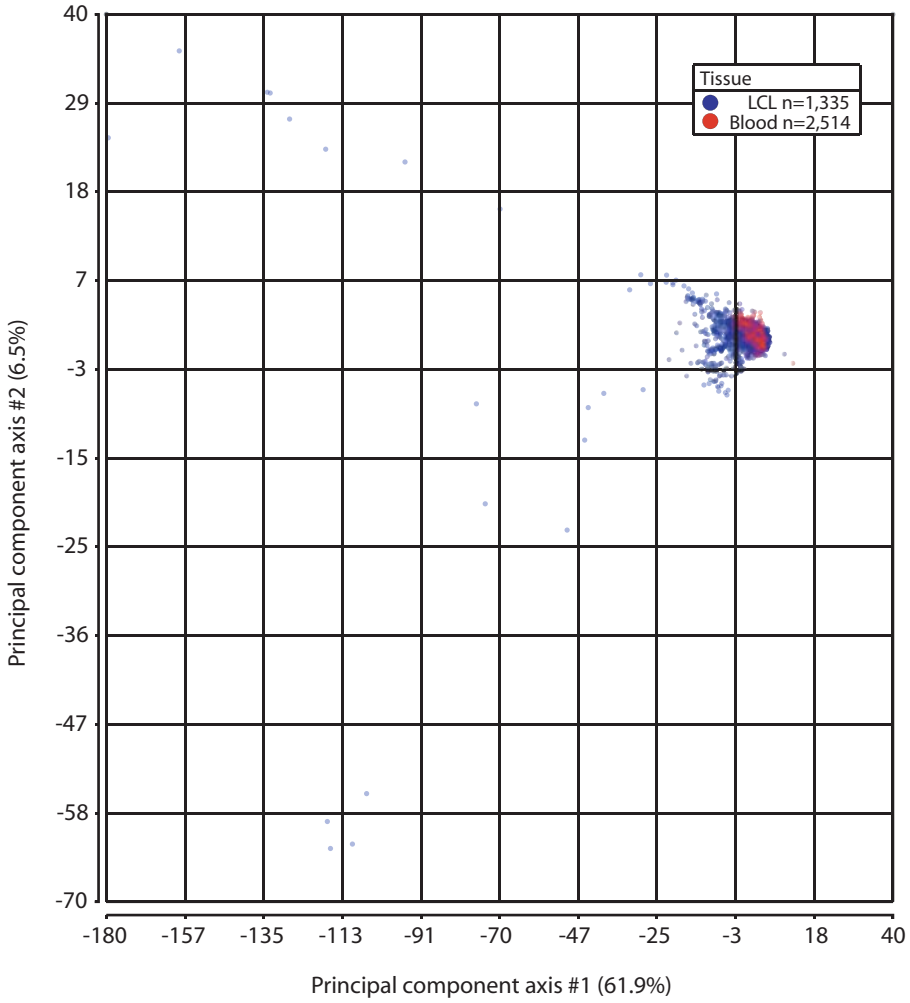
This file contains Supp. Figures S1-S3, and Supp. Tables S1-S5.



Supp. Figure S1. Boxplots of SNP concordance using PLINK software. Pairwise distances between genotypes from the same individual (y-axis) are plotted for unfiltered SNPs (left side) and SNPs for which there were <50%, <90%, <95%, and <99% NoCalls across all samples. (A) Data for PBMC/LCLs (n=205 comparisons). Note that outlier data points (arrow) reflect genotype discrepancies introduced by the comparison of samples B5 and L51 (described in Figures 2 and 3). (B) Data for differentiated cells (n=81 comparisons). Maximum values in the distribution ranged from 0.0008 (no filtering) to 0.0006 (99% filtering). (C) HapMap replicate samples from 6 individuals (n=334 comparisons).



Supp. Figure S2. Mosaic abnormalities in multiple LCLs. Plots of B allele frequency and logR ratio are shown for samples (A) L14; 12% segmental mosaic UPD on chromosome 6, (B) L43;38% segmental mosaic UPD on chromosome 11, (C) L33; mosaic loss of X chromosome and (D) L41; mosaic loss of X chromosome.



Supp. Figure S3. PCA of LogR intensity data for 106 SNPs (n=4032 samples) in a chromosome 22 immunoglobulin locus (hg18 coordinates chr22:21,028,552-21,443,164). While blood-derived DNA samples were discretely localized on the PCA plot, indicating similarity among those samples, the LCL-derived samples displayed much greater heterogeneity. The first principal component (PC #1) accounted for 61.9% of the variance, while PC #2 accounted for 6.5% of the variance. Note that all the outlier data points along PCs #1 and #2 consisted of LCL samples.

Supp. Table S1. Multiple samples from the same subject utilized in intra-subject comparisons

Sample ID	Coriell ID	Cell Type	Gender	Age (Years)	Race
B1	--	PBMC	Male	55	Caucasian
L11	GM22641	LCL	Male	55	Caucasian
L12	GM22642	LCL	Male	55	Caucasian
L13	GM22643	LCL	Male	55	Caucasian
L14	GM22644	LCL	Male	55	Caucasian
L15	GM22645	LCL	Male	55	Caucasian
B2	--	PBMC	Male	26	Caucasian
L21	GM22647	LCL	Male	26	Caucasian
L22	GM22648	LCL	Male	26	Caucasian
L23	GM22649	LCL	Male	26	Caucasian
L24	GM22650	LCL	Male	26	Caucasian
L25	GM22651	LCL	Male	26	Caucasian
B3	--	PBMC	Female	28	Caucasian
L31	GM22671	LCL	Female	28	Caucasian
L32	GM22672	LCL	Female	28	Caucasian
L33	GM22673	LCL	Female	28	Caucasian
L34	GM22674	LCL	Female	28	Caucasian
L35	GM22675	LCL	Female	28	Caucasian
B4	--	PBMC	Female	28	Caucasian
L41	GM22677	LCL	Female	64	Caucasian
L42	GM22678	LCL	Female	64	Caucasian
L43	GM22679	LCL	Female	64	Caucasian
L44	GM22680	LCL	Female	64	Caucasian
L45	GM22681	LCL	Female	64	Caucasian
B5	--	PBMC	Male	32	African-American
L51	GM22731	LCL	Male	32	African-American
L52	GM22732	LCL	Male	32	African-American
L53	GM22733	LCL	Male	32	African-American
L54	GM22734	LCL	Male	32	African-American
L55	GM22735	LCL	Male	32	African-American
B6	--	PBMC	Female	26	African-American
L61	GM22737	LCL	Female	26	African-American
L62	GM22738	LCL	Female	26	African-American
L63	GM22739	LCL	Female	26	African-American
L64	GM22741	LCL	Female	26	African-American

Abbreviations: PBMC, peripheral blood mononuclear cells; LCL, Lymphoblastoid Cell Line

Supp. Table S2. Multiple differentiated cell types from the same individual utilized in intra-subject comparisons

Sample ID	Coriell ID	Cell Type	Gender
1F	GM22133	Fibroblast	male
1M	GM22134	Melanocyte	male
1K	GM22135	Keratinocyte	male
2M	--	Melanocyte*	male
2K	GM22185	Keratinocyte	male
2F	GM22186	Fibroblast	male
3M	GM22189	Melanocyte	male
3K	--	Keratinocyte**	male
3F	GM22191	Fibroblast	male
4M	GM22250	Melanocyte	male
4K	GM22251	Keratinocyte	male
4F	GM22252	Fibroblast	male
5M	GM22253	Melanocyte	male
5K	GM22254	Keratinocyte	male
5F	GM22255	Fibroblast	male
6M	GM22258	Melanocyte	male
6K	GM22259	Keratinocyte	male
6F	GM22260	Fibroblast	male
7M	GM22261	Melanocyte	male
7K	GM22262	Keratinocyte	male
7F	GM22263	Fibroblast	male
8M	GM22266	Melanocyte	male
8K	GM22267	Keratinocyte	male
8F	GM22268	Fibroblast	male
9M	GM22275	Melanocyte	male
9K	GM22276	Keratinocyte	male
9F	GM22277	Fibroblast	male

-- indicates cell lines that are not available from Coriell.

***** Culture only 30% melanocytes (remaining 70% likely fibroblasts)

****** Culture contained 62% keratinocytes (remaining 38% melanocytes)

Supp. Table S3. Concordance of genotype between DCLs consisting of fibroblast (F), keratinocyte (K), and melanocyte (M) samples derived from same individual (n=9 individuals)

ID1	ID2	IBS0	IBS1	IBS2	Concordance
1F	1K	0	288	868142	99.97
1F	1M	0	534	866807	99.94
1K	1M	0	464	867289	99.95
2F	2K	0	724	867107	99.92
2F	2M	0	478	867332	99.94
2K	2M	0	1047	865926	99.88
3F	3K	0	439	866992	99.95
3F	3M	0	355	867434	99.96
3K	3M	0	526	866679	99.94
4F	4K	0	925	865560	99.89
4F	4M	0	665	865814	99.92
4K	4M	0	683	866670	99.92
5F	5K	0	1317	864372	99.85
5F	5M	0	762	865431	99.91
5K	5M	1	1496	862878	99.83
6F	6K	0	309	867247	99.96
6F	6M	0	756	865823	99.91
6K	6M	0	688	866028	99.92
7F	7K	0	1522	863870	99.82
7F	7M	0	1575	862561	99.82
7K	7M	0	728	865674	99.92
8F	8K	1	365	867759	99.96
8F	8M	0	393	867672	99.95
8K	8M	0	200	868258	99.98
9F	9K	0	1135	864617	99.87
9F	9M	0	2571	859794	99.70
9K	9M	0	2030	862348	99.77
				Mean	99.90

Supp. Table S4. Concordance of genotype between HapMap replicates

ID1	ID2	IBS0	IBS1	IBS2	Concordance
NA10851_rep1	NA10851_rep2	87	2469	880903	99.71
NA10851_rep1	NA10851_rep3	95	3178	878618	99.63
NA10851_rep2	NA10851_rep3	90	3375	878025	99.61
NA12239_rep1	NA12239_rep2	3	6000	871691	99.32
NA12239_rep1	NA12239_rep3	0	6606	873272	99.25
NA12239_rep2	NA12239_rep3	2	4519	879352	99.49
NA15510_rep1	NA15510_rep2	1	2663	882424	99.70
NA15510_rep1	NA15510_rep3	3	3797	882406	99.57
NA15510_rep2	NA15510_rep3	1	2574	886880	99.71
NA18517_rep1	NA18517_rep2	0	5492	869112	99.37
NA18517_rep1	NA18517_rep3	1	5664	872654	99.36
NA18517_rep2	NA18517_rep3	0	2808	881534	99.68
NA18576_rep1	NA18576_rep2	2	4205	876359	99.52
NA18576_rep1	NA18576_rep3	1	2993	883551	99.66
NA18576_rep2	NA18576_rep3	1	5139	878427	99.42
NA18980_rep1	NA18980_rep2	1	4052	874831	99.54
NA18980_rep1	NA18980_rep3	1	3982	876386	99.55
NA18980_rep2	NA18980_rep3	0	4022	877319	99.54
				Mean	99.54

Supp. Table S5. Common CNV regions in DCLs (variant in > 50% of samples & size > 25 kb)

Index	Chr	Region start	Region end	Size (bp)	Number of genes	Ig overlap	# CNV	# CNV DGV	#CNV GS/B n=2514	#CNV GS/LCL n=1335	p-value
1	1	72541006	72584492	43486	0		24	32	0	0	NA
2	1	194994473	195083353	88880	4		13	54	105	56	0.047
3	2	88914239	89282353	368114	5	13 IgK	26	114	0	0	NA
4	3	163995351	164109297	113946	1		21	39	0	5	0.013
5	4	34455255	34506399	51144	0		18	27	6	10	0.13
6	6	103841338	103868754	27416	0		13	29	0	1	0.42
7	8	39349340	39506122	156782	13		15	38	8	8	0.61
8	9	44173590	46324418	2150828	12		17	192	25	16	0.75
9	11	55130608	55210152	79544	4		15	75	705	395	2.6e-9
10	12	9525137	9626237	101100	1		15	25	0	0	NA
11	14	105413362	105892781	479419	10		23	183	1	9	0.0026
12	15	32466069	32645302	179233	9		13	106	107	113	0.0050
13	15	18463975	20232614	1768639	34		25	306	250	200	0.31
14	17	41521621	42120174	598553	20		15	110	175	132	0.80
15	20	1505190	1541893	36703	3		17	43	0	0	NA
16	22	22674422	22726814	52392	10		15	47	110	75	0.64

Abbreviations: Chr, chromosome; Number of genes refers to UCSC Genes track (hg18) from the UCSC Genome Browser; Ig overlap, number of immunoglobulin elements in the region based on BioMart at Ensembl (build GRCh37.p3); IgK, immunoglobulin kappa; TCR, T-cell receptor; IgH, immunoglobulin heavy chain; # CNV, number of copy number variants observed in this study; # CNV DGV, number of copy number variants present in the Database of Genomic Variants; #CNV GS/B, number of autosomal copy number variants in GENEVA SAGE samples derived from blood; #CNV GS/LCL, number of autosomal copy number variants in GENEVA SAGE samples derived from LCL; p-value, result of two-sided Fisher's Exact test on GENEVA SAGE samples (case=LCL derived samples, control=blood derived samples).