Journal of Human Genetics (2013) 58, 16–20

www.nature.com/jhg

ORIGINAL ARTICLE

Genotype instability during long-term subculture of lymphoblastoid cell lines

Ji Hee Oh[1,3,6], Young Jin Kim[1,4,6], Sanghoon Moon[1], Hye-Young Nam[2], Jae-Pil Jeon[2], Jong Ho Lee[3], Jong-Young Lee[1] and Yoon Shin Cho[1,5]

Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines (LCLs) promise to address the challenge posed by the limited availability of primary cells needed as a source of genomic DNA for genetic studies. However, the genetic stability of LCLs following prolonged culture has never been rigorously investigated. To evaluate genotypic errors caused by EBV integration into human chromosomes, we isolated genomic DNA from human peripheral blood mononuclear cells and LCLs collected from 20 individuals and genotyped the DNA samples using the Affymetrix 500K SNP array set. Genotype concordance measurements between two sources of DNA from the same individual indicated that genotypic discordance is negligible in early-passage LCLs (<20 passages) but substantial in late-passage LCLs (>50 passages). Analysis of concordance on a chromosome-by-chromosome basis identified genomic regions with a high frequency of genotypic errors resulting from the loss of heterozygosity observed in late-passage LCLs. Our findings suggest that, although LCLs harvested during early stages of propagation are a reliable source of genomic DNA for genetic studies, investigations that involve genotyping of the entire genome should not use DNA from late-passage LCLs.
Journal of Human Genetics (2013) **58,** 16–20; doi:10.1038/jhg.2012.123; published online 22 November 2012

Keywords: lymphoblastoid cell line; single-nucleotide polymorphism; genome-wide association study

INTRODUCTION

High-throughput microarray-based single nucleotide polymorphism (SNP) genotyping greatly facilitates genome-wide association studies (GWAS) to identify human disease-susceptibility loci. Primary cells or tissue samples are the largest sources of genomic DNA for SNP typing. However, the limited availability of these samples restricts the ease and efficiency with which GWAS can be conducted.

Given that lymphoblastoid cell lines (LCLs), which are human B lymphocytes immortalized by *in vitro* infection with Epstein-Barr Virus (EBV), are a renewable source of DNA, they have emerged as a promising alternative to the use of primary cells or tissue samples as sources of human genomic DNA. Numerous genetic studies, including several GWAS currently underway worldwide, have used LCL samples as a DNA source.[1–4] Although LCLs provide a permanent source of human DNA, the genetic stability of LCLs has not been thoroughly studied in the context of genetic and non-genetic factors.[5]

It has been reported that LCLs were influenced by non-genetic factors such as the amount of individual response to the EBV, the history of passage in cell culture and culture conditions.[5] It is also known that the immortalization process of LCLs by EBV infection has the potential to cause changes in genetics.[6] There are several reports that focused on the genetic changes during the lymphocyte transformation. Specifically, the availability of LCL for GWAS has been primarily evaluated with regard to genotype analysis.[6] Some studies estimated that EBV-transformation process may produce minor artifacts on genomic structure and LCL would be a reliable resource for SNP genotyping and detecting copy number variation under the stringent quality control.[7,8] Furthermore, the recent array comparative genomic hybridization analysis of the B-LCL lines and their parental B cells demonstrated that genomic stability was maintained.[9] LCL stability during the long-term subculture process, however, has been remained unclear.

In this study, we rigorously investigated whether genetic instability of LCLs might cause the accumulation of genetic modifications following their long-term subculture. Substantial genotypic errors were detected mostly in late-passage, but not in early-passage, LCLs. This suggests that LCLs harvested during early propagation stages (<40 passages) are reliable sources of genomic DNA for SNP genotyping.

[1]Division of Structural and Functional Genomics, Center for Genome Science, National Institute of Health, Chungcheongbuk-do, Republic of Korea; [2]National Biobank of Korea, Chungcheongbuk-do, Republic of Korea; [3]Yonsei University Research Institute of Science for Ageing, Yonsei University, Seoul, Republic of Korea; [4]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea and [5]Department of Biomedical Science, Hallym University, Gangwon-do, Republic of Korea
[6]These authors contributed equally to this work.
Correspondence: Professor YS Cho, Department of Biomedical Science, Hallym University, 1 Hallymdaehak-gil, Chuncheon, Gangwon-do 200-702, Republic of Korea.
E-mail: yooncho33@hallym.ac.kr
Received 7 May 2012; revised 27 September 2012; accepted 1 October 2012; published online 22 November 2012

MATERIALS AND METHODS

Samples

The 20 LCL strains used in this study were chosen from the LCL collection of the Korean HapMap project (http://cdc.go.kr). As the first step in generating LCLs, peripheral blood samples from individuals who were 40–69 years old and part of Korean Genome Epidemiologic Study (KoGES) cohorts were subjected to Ficoll–Hypaque gradient centrifugation to obtain peripheral blood mononuclear cells (PBMCs). The PBMCs were prepared according to the protocols suggested for use with Amersham Biosciences (Freiburg, Germany). The subsequent infection of PBMCs with EBV, using procedures described elsewhere,[10] eventually generated LCLs. All LCL strains were cultured in RPMI 1640 medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum at 37 °C in humidified air containing 5% CO_2. Culture medium was replaced with fresh RPMI 1640 at each passage.

Subculture of LCLs

We used continuous subculturing to propagate LCLs until maximal end passage.[10] The maximal passage of each LCL strain was determined the cell number did not increase 4 weeks after subculture.[11] Under our culture conditions, most of the LCLs we studied stopped proliferating after 160 passages. The 17 LCLs that proliferated after this many passages were classified as immortal, and the three LC lines that stopped proliferating at passages 33, 44 and 48 were classified as non-immortal (Table 1). The LCL strains were grown to take about >2 years. The average lifespan of these non-immortal LCL strains was 41 ± 8 passages. We analyzed LCL samples harvested at six designated propagation stages. These samples were named LCL2 (passage 2), LCL4 (passage 4), P1 (between 10–20 passages), P41 (between 50–60 passages), P100 (between 110–120 passages) and P160 (between 170–180 passages).[10,12,13]

Genotyping

We genotyped PBMC and LCL samples using the GeneChip human mapping 500K array set (Affymetrix, Inc., Santa Clara, CA, USA), which comprises 500 568 SNPs on two arrays, named NSP and STY. Genotyping was performed according to the manufacturer's protocol.

Examination of genotype concordance

Genotypes of all samples were called using the Affymetrix BRLMM algorithm.[14] We examined the genotype concordance between PBMCs and LCLs derived from the same individual by using identity-by-state (IBS) analysis.[15]

Table 1 Description of samples

Sample name	*Gender*	*Age*	*Immortalization*	*Used for analysis*
A1	F	65	Immortal	Yes
A2	M	64	Immortal	Yes
A3	M	45	Immortal	Yes
A4	M	57	Immortal	Yes
A5	M	57	Immortal	No
A6	M	47	Immortal	No
A7	F	45	Non-immortal	No
A8	F	53	Immortal	Yes
A9	F	58	Non-immortal	Yes
A10	F	68	Immortal	Yes
K1	F	61	Immortal	Yes
K2	F	46	Immortal	No
K3	M	64	Immortal	Yes
K4	M	41	Non-immortal	Yes
K5	M	65	Immortal	Yes
K6	M	58	Immortal	Yes
K7	F	55	Immortal	Yes
K8	M	47	Immortal	Yes
K9	M	69	Immortal	Yes
K10	F	43	Immortal	Yes

Pairwise IBS distances between PBMC and LCL were calculated for each of the NSP and STY arrays separately, as well as for the combined array set. Briefly, a SNP with perfect genotype matching between two samples (for example, PBMC and LCL2 for A1 line) was assigned for the score of 2. A SNP showing half genotype matching or no matching between two samples was assigned for 1 or 0, respectively. Overall pairwise IBS distance between two samples was determined by dividing the sum of all SNP scores with two times of SNP numbers. We excluded the A5, A6, A7 and K2 strains from further analysis because concordance testing indicated that these LCL strains likely originated from different blood donors. In total, 16 LCL strains were used for further analyses (Table 1).

Examination of large genomic aberration

To ensure possible genomic aberration detected by genotype mismatching and heterozygosity analysis, we calculated the Log R ratio or the B-allele frequency of PBMCs and LCLs. The PennCNV-affy software package was used to obtain the Log R ratio or the B-allele frequency from both NSP and STY data. The Log R ratio or the B-allele frequency was plotted on the chromosomal regions using the R statistics package (http://www.r-project.org).

RESULTS

We estimated genotype instability in LCLs and PBMCs across sample pairs from 16 of the 20 strains generated. Overall, the mean genotype call rates for samples were 98.1%, 98.4% and 98.3% for the STY array, the NSP array and these two arrays combined, respectively. To investigate the concordance of SNP genotypes between PBMCs and LCLs at six different propagation stages from the same line, we calculated the pairwise distance based on IBS analysis using the 500 568 SNPs (hereafter called original SNPs) represented in the Affymetrix 500K array set. The mean pairwise IBS distance of original SNPs between PBMC and LCLs was ~ 0.995 (Table 2), indicating that LCLs are generally a reliable source of DNA for genotyping with microarray-based DNA chips. To estimate within sample variation of genotyping, we randomly selected eight different LCLs and genotyped each sample twice using the separate array chips. Concordance rates between duplicates of the same sample ranged from 0.988 to 0.997. The mean concordance rate for overall test was 0.992 (Supplementary Table 1). These results suggest that within sample variation resulted from genotyping can be disregarded in the estimation of the concordance rate between the LCLs and PBMCs.

Of the 500 568 SNPs, we further tested the concordance of SNPs that are most frequently used in GWAS (hereafter called GWAS SNPs). To select GWAS SNPs, we adapted SNPs that had been analyzed in GWAS for eight quantitative traits as a part of Korea Association Resource (KARE) Project,[16] which involves using the Affymetrix Genome-wide Human SNP Array 5.0 to genotype 500 568 SNPs. We selected 352 228 SNPs identified in KARE GWAS after excluding SNPs owing to a high missing genotype rate ($>5\%$), a low minor allele frequency (MAF) (<0.01) and significant deviation from Hardy–Weinberg equilibrium (HWE) ($P<1\times10^{-6}$). Overall mean genotype concordance (between PBMC and LCLs) was 0.996 for GWAS SNPs (Table 2). Concordance tests involving the GWAS SNP set produced results similar to those obtained using the original SNP set (Figure 1). These results demonstrated that the LCLs are a suitable alternative to PBMCs as a source of DNA for genotyping experiments, such as GWAS.

To understand the source of mismatches of the LCLs, we calculated genotype concordance between PBMCs and LCLs for SNPs arbitrarily grouped according to the HWE *P*-value, MAF and genotype missing rate (Table 3). The underlying assumption of the analysis is based on previous reports that SNPs with lower HWE *P*-values, lower MAF values and higher genotype missing rates tend to be associated with

Table 2 Genotype concordance between PBMCs and LCLs from no filtered or GWAS filtered call rate

		No filtering						*GWAS filtering*					
ID	*IBS*	*LCL2*	*LCL4*	*P1*	*P41*	*P100*	*P160*	*LCL2*	*LCL4*	*P1*	*P41*	*P100*	*P160*
A1		0.996	0.997	0.997	0.996	0.996	0.997	0.996	0.997	0.997	0.997	0.997	0.997
A2		0.995	0.996	0.996	0.995	0.992	0.995	0.996	0.997	0.997	0.996	0.993	0.995
A3		0.997	0.998	0.997	0.997	0.994	0.994	0.998	0.998	0.998	0.998	0.993	0.994
A4		0.997	0.995	0.994	0.992	0.996	0.997	0.997	0.996	0.996	0.994	0.997	0.998
A8		0.994	0.993	0.993	0.992	0.991	0.991	0.995	0.994	0.994	0.993	0.992	0.992
A9		0.997	0.995	0.996	0.993			0.997	0.996	0.997	0.994		
A10		0.996	0.992	0.996	0.995	0.995	0.994	0.997	0.994	0.997	0.996	0.995	0.994
K1		0.996	0.997	0.997	0.996	0.998	0.998	0.997	0.998	0.997	0.997	0.999	0.998
K3		0.998	0.998	0.998	0.996	0.998	0.998	0.998	0.999	0.998	0.996	0.998	0.998
K4		0.995	0.996	0.996	0.995			0.996	0.997	0.997	0.996		
K5		0.997	0.998	0.998	0.996	0.997	0.994	0.998	0.999	0.998	0.996	0.997	0.995
K6		0.995	0.995	0.995	0.994	0.995	0.994	0.996	0.996	0.996	0.995	0.996	0.995
K7		0.997	0.997	0.995	0.996	0.996	0.997	0.997	0.997	0.996	0.996	0.996	0.997
K8		0.994	0.995	0.994	0.993	0.995	0.995	0.995	0.996	0.995	0.994	0.996	0.996
K9		0.988	0.991	0.991	0.994	0.994	0.994	0.991	0.992	0.993	0.994	0.995	0.995
K10		0.998	0.996	0.997	0.994	0.997	0.997	0.998	0.997	0.997	0.995	0.998	0.998
Mean		0.996	0.996	0.996	0.995	0.995	0.995	0.996	0.996	0.996	0.995	0.996	0.996
Total mean		0.995							0.996				

Abbreviations: GWAS, genome-wide association study; IBS, identity-by-state; LCL, lymphoblastoid cell line; PBMC, peripheral blood mononuclear cell.

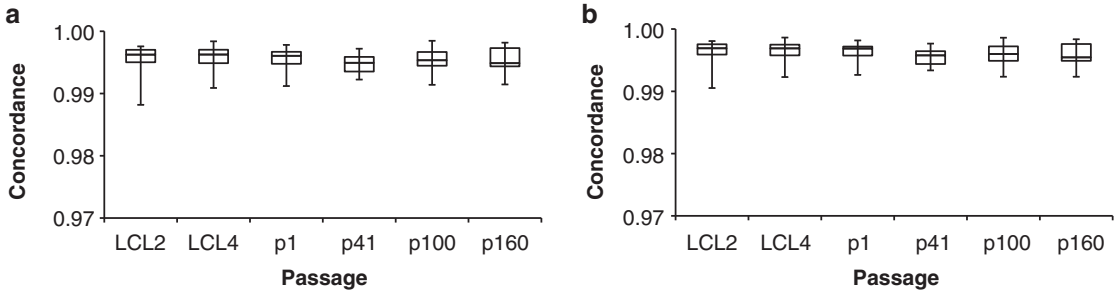

Figure 1 Genotype concordance of original 500 568 SNPs (**a**) and GWAS SNPs (**b**) between PBMC and LCLs at six different propagation stages.

more genotyping errors.[17–19] We generated four groups based on the rates of missing genotypes (<1%, between 1–5%, between 5–10% and >10%), three groups according to HWE *P*-values of SNPs (HWE-$P > 1 \times 10^{-4}$, $1 \times 10^{-4} \geqslant$ HWE-$P > 1 \times 10^{-6}$ and HWE-$P \leqslant 1 \times 10^{-6}$) and five groups according to MAF values of SNPs (MAF<1%, 1%⩽MAF<5%, 5%⩽MAF<10%, 10%⩽MAF<50% and MAF⩾50%). For grouping, we adopted SNP information on HWE *P*-values, MAF values and genotype missing rates that are available from KARE genome-wide scan data.[16] Regardless of LCL passage number, overall results showed no notable difference in concordance among groups in the same category (0.994 average concordance).

We also attempted to identify the chromosomal regions most vulnerable to genotyping errors associated with LCL-derived DNA by scrutinizing genotype concordance across entire chromosomes. A high rate of genotype disconcordance between PBMCs and LCLs was observed on chromosomes 6p, 16q, 18p and 22q in the late-passage LCL strains A3, A10, K3, and A2 (Supplementary Figure S1). In those LCL strains, loss of heterozygosity (LOH) was observed on the loci showing the highest rates of genotype disconcordance with PBMCs, suggesting that LOH might be the major cause of genotype errors for late-passage LCLs (>50 passage) (Table 4 and Supplementary Figure S2). This result indicated that LCLs at late stages of propagation are not reliable source of DNA for genome analysis.

The presence of LOH regions was further proved by detecting either the copy number loss of the large chromosomal region from the analysis of Log R ratio or the heterozygosity loss estimated from the B-allele frequency (about 0.5). Silent LOH showed no change in the Log R ratio but did substantial LOH in the B-allele frequency.[20] In this study, LOH by copy loss was observed on 6p of A3 (p100 and p160) (Supplementary Figure S3A) and on 18p of K3 (in P41, P100 and P160) (Supplementary Figure S3D). Silent LOH was detected on 16q of A3 (in P100 and P160) (Supplementary Figure S3B), 16q of A10 (in P160) (Supplementary Figure S3C) and 22q of A2 (in P100 and P160) (Supplementary Figure S3E).

DISCUSSION

We propagated human LCLs through as many as 160 passages and assessed their stability at selected stages of propagation by SNP genotyping. Overall, we observed no notable differences in genotype concordance between PBMCs and LCLs throughout the course of propagation. However, inspection of each chromosome revealed LOH in four late-stage LCLs. Thus, we recommend against using LCLs at a late stage of propagation for genome analysis, especially SNP

Table 3 Comparison of genotype concordance of LCLs among SNPs grouped by missing rate, HWE *P*-value and MAF

Variables	*Group*	*Concordance*
Missing rate		
	$\leqslant 0.1$	0.989
	$\leqslant 0.05$	0.992
	$\leqslant 0.01$	0.994
	<0.01	0.997
HWE		
	$\leqslant 10^{-6}$	0.989
	$\leqslant 10^{-4}$	0.992
	$>10^{-4}$	0.996
MAF		
	$\geqslant 0.5$	0.996
	$\geqslant 0.1$	0.995
	$\geqslant 0.05$	0.995
	$\geqslant 0.01$	0.996
	<0.01	0.998

Abbreviations: HWE, Hardy–Weinberg equilibrium; LCL, lymphoblastoid cell line; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

Table 4 Loss of heterozygosity with increased numbers of LCL passages through culture

Line	*Chromosome*	*No. of genotype mismatching with PBMC*	*Heterozygosity (%)*[a]	*Passages*
A3	6p	884	1.58	P100, P160
A3	16q	1999	0.25	P100, P160
A10	16q	892	3.02	P160
K3	18p	601	0	P41, P100, P160
A2	22q	1376	0.44	P100, P160

Abbreviations: LCL, lymphoblastoid cell line; PBMC, peripheral blood mononuclear cell.
[a]Heterozygosity = (number of heterozygotes)/(number of homozygotes + number of heterozygotes).
Heterozygosity for LCLs was calculated using LCL samples with lowest concordance relative to blood samples.

genotyping. In addition, karyotype analysis before genotyping is desirable for LCLs subcultured through >50 passages.

The detection of LOH at a specific chromosomal region should not be relied on concordance rates. When a genotype mismatching occurs due to LOH, the between sample IBS score for one SNP will be 1 as usual. As a result, the between sample IBS distance obtained from concordance analysis at a LOH region will be always 0.5. Therefore, LOH showed little effect on concordance rates. Indeed, it was estimated that the genotype mismatching caused by LOH occupies only a very small portion (0.27%) among a total of 1376 mismatches detected in 22q of A2 (Table 4). In this context, additional measures such as Log R ratio and B-allele frequency should be thoroughly examined to analyze LOH in LCLs. Changes in B-allele frequency are specifically important variable to detect silent LOH[20] that cannot be detected by Log R ratio alone (Supplementary Figure S3).

The mechanism underlying genomic aberration observed in LCLs during long-term subculture remains unclear. However, one plausible explanation is a double-strand break induced recombination.[21] Although the frequency of double-strand breaks is strictly regulated by the actions of nonhomologous end-joining proteins and tumor suppressor proteins such as p53, double-strand breaks sometimes produce genomic aberrations.[22,23] Besides genomic changes, phenotypic changes such as activation of the NF-κB pathway and carcinogenesis-related genes have been associated with long-term subculturing of LCLs.[13] Profiles of these differentially expressed genes can be considered as genetic signatures of LCL immortalization or EBV-induced carcinogenesis.[13] Moreover, differential expression of nine microRNAs during long-term subculture of LCLs has provided a signature of terminal immortalization of LCLs that distinguishes this from the initial stage of EBV-mediated B-cell transformation.[12]

Mohyuddin *et al.*[24] studied on microsatellite instability between blood and LCLs by analyzing mutation rate of 20 short tandem repeats on the non-recombining part of the Y chromosome. They reported that mutations were only 0.3% of the analyses. Our study is different from their work in the context of marker type (microsatellite vs SNP) and test region in the genome (Y chromosome vs all autosomes). In addition, Mohyuddin *et al.*[24] did not pay attention to the genomic instability that may be influenced by the propagation stages of LCLs.

Thus, to our knowledge, this is the first study to examine the effect of long-term subculturing on the genomic stability of LCLs. Our findings indicate that EBV transformation does not significantly affect the genotypes of LCLs. However, LCLs subjected to >50 passages through culture are not recommended for SNP genotyping owing to an unacceptable increase in the frequency of genetic artifacts.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by grant from Korea Center for Disease Control and Prevention, Republic of Korea (4845-301, 4851-307) and intramural grant from the Korea National Institute of Health (2011-N73005-00, 2011-N74003-00). YSC acknowledges support from the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (2012R1A2A1A03006155). Lymphoblastoid cell lines and DNAs for genotyping used in this study were provided by the National Biobank of Korea.

1 Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V., Robson, S. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464,** 713–720 (2010).
2 McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40,** 1166–1174 (2008).
3 Simon-Sanchez, J., Scholz, S., Fung, H. C., Matarin, M., Hernandez, D., Gibbs, J. R. *et al.* Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum. Mol. Genet.* **16,** 1–14 (2007).
4 International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789–796 (2003).
5 Choy, E., Yelensky, R., Bonakdar, S., Plenge, R. M., Saxena, R., De Jager, P. L. *et al.* Genetic analysis of human traits *in vitro*: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet.* **4,** e1000287 (2008).
6 Herbeck, J. T., Gottlieb, G. S., Wong, K., Detels, R., Phair, J. P., Rinaldo, C. R. *et al.* Fidelity of SNP array genotyping using Epstein Barr virus-transformed B-lymphocyte cell lines: implications for genome-wide association studies. *PLoS ONE* **4,** e6915 (2009).
7 McElroy, J. P., Nelson, M. R., Caillier, S. J. & Oksenberg, J. R. Copy number variation in African Americans. *BMC Genet.* **10,** 15 (2009).
8 Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).

9 Jeon, J. P., Shim, S. M., Nam, H. Y., Baik, S. Y., Kim, J. W. & Han, B. G. Copy number increase of 1p36.33 and mitochondrial genome amplification in Epstein-Barr virus-transformed lymphoblastoid cell lines. *Cancer Genet. Cytogenet.* **173,** 122–130 (2007).
10 Jeon, J. P., Nam, H. Y., Shim, S. M. & Han, B. G. Sustained viral activity of epstein-Barr virus contributes to cellular immortalization of lymphoblastoid cell lines. *Mol. Cells* **27,** 143–148 (2009).
11 Sugimoto, M., Furuichi, Y., Ide, T. & Goto, M. Incorrect us of 'immortalization' for B-lymphoblastoid cell lines transformed by Epstein-Barr virus. *J. Virol.* **73,** 9690–9691 (1999).
12 Lee, J. E., Hong, E. J., Nam, H. Y., Kim, J. W., Han, B. G. & Jeon, J. P. MicroRNA signatures associated with immortalization of EBV-transformed lymphoblastoid cell lines and their clinical traits. *Cell Prolif.* **44,** 59–66 (2011).
13 Lee, J. E., Nam, H. Y., Shim, S. M., Bae, G. R., Han, B. G. & Jeon, J. P. Expression phenotype changes of EBV-transformed lymphoblastoid cell lines during long-term subculture and its clinical significance. *Cell Prolif.* **43,** 378–384 (2010).
14 Rabbee, N. & Speed, T. P. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* **22,** 7–12 (2006).
15 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
16 Cho, Y. S., Go, M. J., Kim, Y. J., Heo, J. Y., Oh, J. H., Ban, H. J. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41,** 527–534 (2009).
17 Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A. *et al.* Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur. J. Hum. Genet.* **12,** 395–399 (2004).
18 Mitchell, A. A., Zwick, M. E., Chakravarti, A. & Cutler, D. J. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. *Bioinformatics* **20,** 1022–1032 (2004).
19 Fu, W., Wang, Y., Li, R., Lin, R. & Jin, L. Missing call bias in high-throughput genotyping. *BMC Genomics* **10,** 106 (2009).
20 Tan, D. S., Lambros, M. B., Natrajan, R. & Reis-Filho, J. S. Getting it right: designing microarray (and not 'microawry') comparative genomic hybridization studies for cancer research. *Lab. Invest.* **87,** 737–754 (2007).
21 Moynahan, M. E. & Jasin, M. Loss of heterozygosity induced by a chromosomal double-strand break. *Proc. Natl Acad. Sci. USA* **94,** 8988–8993 (1997).
22 Danjoh, I., Saijo, K., Hiroyama, T. & Nakamura, Y. The Sonoda-Tajima Cell Collection: a human genetics research resource with emphasis on South American indigenous populations. *Genome Biol. Evol.* **3,** 272–283 (2011).
23 Sugimoto, M., Tahara, H., Ide, T. & Furuichi, Y. Steps involved in immortalization and tumorigenesis in human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. *Cancer Res.* **64,** 3361–3364 (2004).
24 Mohyuddin, A., Ayub, Q., Siddiqi, S., Carvalho-Silva, D. R., Mazhar, K., Rehman, S. *et al.* Genetic instability in EBV-transformed lymphoblastoid cell lines. *Biochim. Biophys. Acta.* **1670,** 81–83 (2004).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)