Advances in Genomics and Genetics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Differences between the genomes of lymphoblastoid cell lines and blood-derived samples

This article was published in the following Dove Press journal:
Advances in Genomics and Genetics
23 February 2017
Number of times this article has been viewed

Advances in Genomics and Genetics downloaded from https://www.dovepress.com/
For personal use only.

Lena M Joesch-Cohen
Gustavo Glusman

Institute for Systems Biology, Seattle, WA, USA

Abstract: Lymphoblastoid cell lines (LCLs) represent a convenient research tool for expanding the amount of biologic material available from an individual. LCLs are commonly used as reference materials, most notably from the Genome in a Bottle Consortium. However, the question remains how faithfully LCL-derived genome assemblies represent the germline genome of the donor individual as compared to the genome assemblies derived from peripheral blood mononuclear cells. We present an in-depth comparison of a large collection of LCL- and peripheral blood mononuclear cell-derived genomes in terms of distributions of coverage and copy number alterations. We found significant differences in the depth of coverage and copy number calls, which may be driven by differential replication timing. Importantly, these copy number changes preferentially affect regions closer to genes and with higher GC content. This suggests that genomic studies based on LCLs may display locus-specific biases, and that conclusions based on analysis of depth of coverage and copy number variation may require further scrutiny.

Keywords: genomics, whole-genome sequencing, viral transformation, copy number changes, bioinformatics

Introduction

Transformation of peripheral blood mononuclear cells (PBMCs) into lymphoblastoid cell lines (LCLs) through infection by Epstein–Barr virus is a commonly used practice for creating an unlimited supply of cells for use in a variety of studies. Such LCLs are used interchangeably with non-LCLs, often as in vitro model systems[1] or as sources of genomic data.[2,3] Notably, the "benchmark" human genomes used as references by the Genome in a Bottle Consortium are Epstein–Barr virus-transformed LCLs maintained by the Coriell Institute.[4,5]

Although genomic data from LCLs are often used as a bona fide source of genomic data, investigation is still underway to confidently conclude whether LCLs provide a faithful copy of their donor genome. A number of studies have examined differences between LCLs and their donors using a variety of metrics. Studies involving large cohorts of LCLs and controls have investigated mutations in mitochondrial DNA (mtDNA) through the analysis of whole-exome data[6] and gene expression analysis,[7] as well as variation in genotype and copy number variations (CNVs) throughout the genome.[8] Studies performed on single pairs or small groups (n≤20) of blood-derived genomes and their directly derived LCLs have compared CNV calls between LCLs and controls, using both array comparative genomic hybridization and whole genome data.[9,10] Other studies have investigated differences in single nucleotide polymorphisms

Correspondence: Gustavo Glusman
Institute for Systems Biology, 401 Terry Avenue N, Seattle, WA 98109, USA
Tel +1 206 732 1273
Fax +1 206 732 1260
Email Gustavo@SystemsBiology.org

Dovepress

http://dx.doi.org/10.2147/AGG.S128824

© 2017 Joesch-Cohen and Glusman. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v3.0) License (http://creativecommons.org/licenses/by-nc/3.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

(SNPs) and other variations, using both SNP arrays and exome data,[11,12] as well as methylation profiles and gene expression levels.[13,14]

We present an in-depth comparison of a large collection of LCL-derived genomes with matched PBMC-derived genomes based on distributions of coverage and copy number alterations. We found significant locus-specific differences in the depth of coverage and copy number calls, which may be driven by differential replication timing.

Materials and methods

Description of data sets

Drawing from whole-genome assemblies that had been created for previous studies between September 2010 and February 2014, we compiled a set of 126 assemblies (63 males, 63 females) derived from LCLs, sequenced at high quality (>40 × average coverage) by Complete Genomics, Inc. (CGI) and analyzed using human genome freeze GRCh37 (hg19) as reference. We then created two "matched control" sets (MC1 and MC2), composed of nonoverlapping sets of 126 blood-derived whole-genome assemblies. The genomes in these sets were individually matched to the 126 LCLs by metadata. Matching requirements included being of the same sex, having been sequenced on the same platform (CGI), mapped to the same reference genome (GRCh37) and analyzed using the same Complete Genomics Analysis Pipeline software version (Table 1).[15] Most genome assemblies are of European descent; the distribution of populations at continental region resolution shows slightly less diversity in the LCL than control sets (Table 2).

Analysis of normalized coverage profiles

For each genome, we computed its normalized coverage profile as described.[16] This profile reports the normalized coverage level at 1 kb bins along the genome; a value of 100 represents the expected diploid coverage. From these coverage levels, we computed for each genome several summary statistics of normalized coverage, namely, the standard deviation, median and median absolute deviation (MAD). We also computed these statistics for each chromosome, including mtDNA. We combined normalized coverage profiles for each set (LCLs, MC1 and MC2) to calculate the average genome span at each coverage level across all assemblies. We then visualized these distributions, along with their standard deviation around the average coverage, to compare the three sets of genomes.

Table 1 Pipeline software versions used

Pipeline versions	Assemblies per set
1.08.00.30	1
1.08.00.34	1
1.11.00.15	1
1.11.00.18	3
2.0.2.22	27
2.0.2.26	6
2.0.3.2	11
2.0.3.6	1
2.0.4.14	1
2.2.0.26	35
2.4.0.43	1
2.5.0.20	38

Table 2 Population distribution in the three sets of genome assemblies

Region	LCLs	MC1	MC2
EUR	125	117	112
AMR	1	8	12
EAS		1	1
SAS			1

Abbreviations: AMR, admixed American; EAS, East Asian; EUR, European; LCLs, lymphoblastoid cell lines; MC1 and MC2, matched control sets 1 and 2; SAS, South Asian.

Analysis of reference coverage profiles

For each set of genome assemblies, we computed reference coverage profiles (RCPs) as described.[16] This yielded the reference coverage (i.e., the coverage value corresponding to diploid coverage), prenormalized median coverage, the MAD around the prenormalized median coverage, and the distribution of prenormalized coverage levels for every 1 kb bin along the genome. We used RCPs to compare the three sets of assemblies in pairwise fashion (namely, LCL vs MC1, LCL vs MC2, MC1 vs MC2). For each comparison, we computed a two-sample Kolmogorov–Smirnov (KS) statistic on the coverage level distributions in each set of bins. The KS statistic ranges from 0 to 1, which represent an exact match or no overlap between the two distributions, respectively. For each of the three comparisons, we then visualized the reported KS statistic along each chromosome, smoothing over 1 Mb (1000 consecutive bins). To compare the KS values with the GC content, we subdivided the genome into 25 "GC buckets" of similar size and increasing GC percentiles, as described.[16] We computed distances to the nearest exons and segmental duplication level as annotated in the knownGene and genomicSuperDups tracks from the University of California, Santa Cruz database, respectively.[17] We evaluated the relationship with replication timing ratio as observed in the C0202 LCL (GEO: GSM500943).[18,19] We also used RCPs to investigate the coverage level distributions across the three sets in individual bins.

Analysis of CNV calls

For each genome, we computed CNVs as described.[16] Each CNV call is characterized by its observed ploidy level (denoted by integer values between 0 and 4, where 4 represents genomic regions with four or more copies), its location in the genome and its frequency in a reference population. For each genome, we selected rare CNV calls (with population frequency ≤1%) and visualized the distribution of rare CNV counts across all 126 assemblies in each set, grouping the CNVs by ploidy level.

Results

The autosomal coverage distribution of LCLs differs from that of PBMCs

The uncorrected depth of sequencing coverage can fluctuate significantly within each genome, but becomes very uniform by normalization using RCPs.[16] Successfully normalized autosomal coverage follows a narrow distribution, centered on 100% of the expected diploid coverage; the width of this distribution serves as a metric of uniformity of genome coverage.

We compared the averaged distribution of normalized autosomal coverage in three sets of genomes: one consisting of genomes from LCLs and two of matched controls from PBMCs. We observed higher variability of coverage levels in LCLs than in the controls (Figure 1): average coverage counts close to the expected value of 100% (95%–104%) were higher in the controls, whereas average counts farther out from 100% (extending to 80% and 120%) were higher in LCLs. Standard distributions of coverage around average levels show that while there is overlap in average coverage between the LCLs and controls, the two groups are nonetheless distinct (Figure 1, inset).

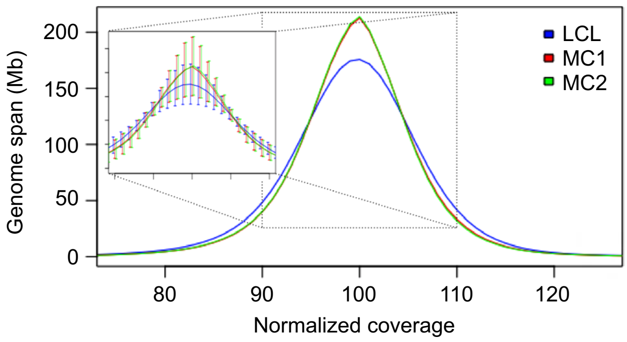

Figure 1 Normalized coverage distributions.

Notes: LCLs display a wider distribution of average normalized coverages than matched controls (MC1 and MC2). Inset: standard deviations around the average coverage in the 90%–110% range.

Abbreviations: LCL, lymphoblastoid cell line; MC1 and MC2, matched control sets 1 and 2.

Inspection of standard deviations of normalized coverage on a per assembly basis showed that the trend of wider normalized coverage distributions is more evident in some pipeline software versions than in others (Figure 2): more recent versions of the pipeline tend to exhibit larger standard deviation (wider distribution of normalized coverage) in LCLs than in PBMCs. This trend is particularly evident starting from version 2.0.3.2: in it and later pipeline versions, most LCLs exhibit less uniform normalized coverage than their matched controls.

The magnitude of this trend is not distributed evenly among the chromosomes. We observed a larger difference in normalized coverage variation between LCLs and controls in chromosomes 12 and 14, but almost no difference in chromosome 19 (Figure 3).

The mitochondrial coverage of LCLs differs from that of PBMCs

We also observed a much larger difference in coverage between LCLs and controls in mtDNA. Most notably, the median average coverage level in mtDNA in LCLs is almost fourfold higher than in the controls, with some increases in mtDNA in LCLs up to 12-fold. This finding is consistent with previous studies, which have reported increases of up to ninefold in mtDNA in LCLs.[7,10] As in autosomes, the distribution of coverage levels in mtDNA is significantly wider in LCLs than in controls. We observed, on average, a standard deviation and MAD of 20.0 and 20.1, 4.1 and 3.5 and 3.7 and 3.1 in LCLs, MC1 and MC2, respectively.

Considering the expanded mtDNA coverage in LCLs, we hypothesized a potential mismapping effect on autosomes, due to the presence of nuclear mtDNA segments.[20] We found a very slight increase in coverage deviation in autosomes with higher nuclear mtDNA segment content, but these are not sufficient to account for the large observed differences between LCLs and controls, and among autosomes.

LCL coverage fluctuates along the chromosomes

To explore the differences in coverage at higher resolution, we performed bin-by-bin comparisons of coverage level distributions using the KS statistic. The results of this analysis show that distributions vary much more between LCLs and controls than between the two sets of controls on a per-bin basis, consistent with our observations on a per-chromosome level. KS statistics comparing LCLs and controls range from 0 to ~0.9, with a median of 0.266 and a MAD of 0.153. KS values comparing the two controls, however, range only from 0 to ~0.3, with a median of 0.079 and a MAD of 0.024.

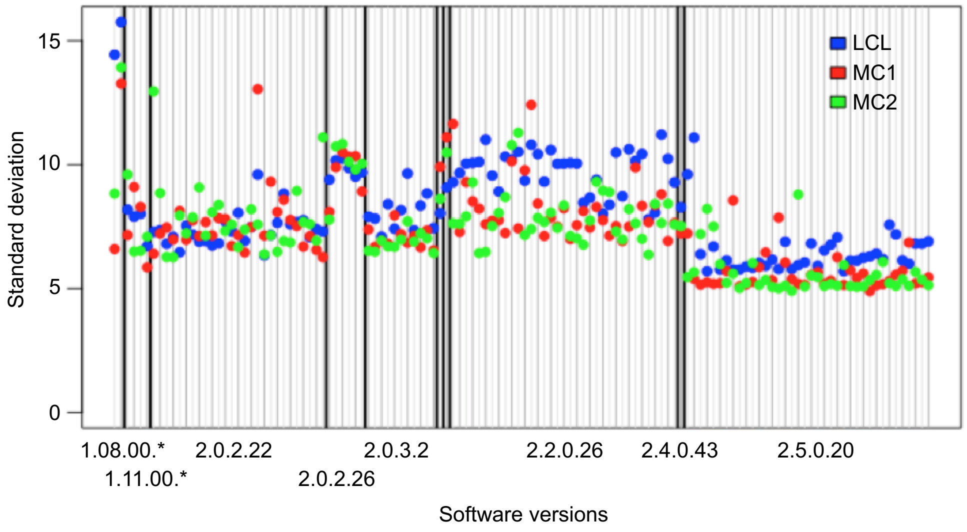

Figure 2 Software version effect.

Notes: Newer software versions have better ability to separate LCLs (blue) from matched controls (MC1 and MC2) by standard deviation of normalized coverage. Genome assemblies are arranged by assembly software version and chronologically.

Abbreviations: LCL, lymphoblastoid cell line; MC1 and MC2, matched control sets 1 and 2.

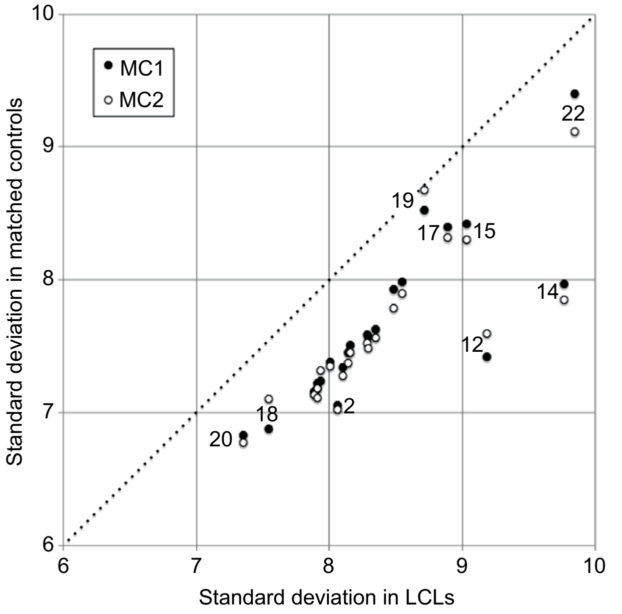

Figure 3 Chromosomal distribution of coverage deviations.

Notes: All autosomes except for chr19 display higher coverage variability (standard deviation of normalized coverage) in LCLs as compared to the control sets.

Abbreviations: LCLs, lymphoblastoid cell lines; MC1 and MC2, matched control sets 1 and 2.

We next visualized the KS statistic along each chromosome and found that the variation between LCLs and controls is present in all autosomes and throughout the length of each chromosome. However, there are strong regional fluctuations in KS values (Figure 4). Examining these regions showed that lower KS values for LCLs vs controls correspond to similar coverage distributions across all three sets, whereas larger KS values show different coverage distributions between LCLs and controls (Figure 4B). This difference typically represents increased coverage in LCLs relative to controls.

The fluctuation of KS values along the chromosomes approximately correlates with chromosomal banding (Figure 4A): LCLs tend to have more distorted coverage in light bands (Giemsa negative, or R bands) relative to controls, but more similar coverage in dark bands (Giemsa positive or G bands). Giemsa banding patterns are related to both GC content and gene density, as well as replication stage, with R bands replicating early.[21] We observed that KS values are weakly correlated with GC content and gene presence, with median KS value rising with increasing GC percentile (Figure 5A) and decreasing with increasing distance from the nearest exon (Figure 5B). We observed a much stronger relationship with replication timing ratio:[18] most of the coverage difference between LCLs and controls is located in early-replicating regions of the genome as previously reported in the C0202 LCL;[19] the earlier the replication timing, the stronger the difference (Figure 5C). The differences between LCLs and controls were not enriched in segmentally duplicated regions of the genome (Figure 5D).

We demonstrate this coverage distortion by example, by comparing one LCL (pipeline version 2.4.0.43) and its matched controls in the early-replicating chromosomal band 2p22.2 (Figure 6). We observe 5%–10% excess coverage in the LCL over a span of almost 1 Mb, with finer-scale excess coverage frequently in the 20%–30% range.

LCLs display modified CNV counts

Distortions in depth of coverage can lead to changes in the inferred ploidy levels, both at known CNV regions and in typically copy-invariant loci. We evaluated the distribution

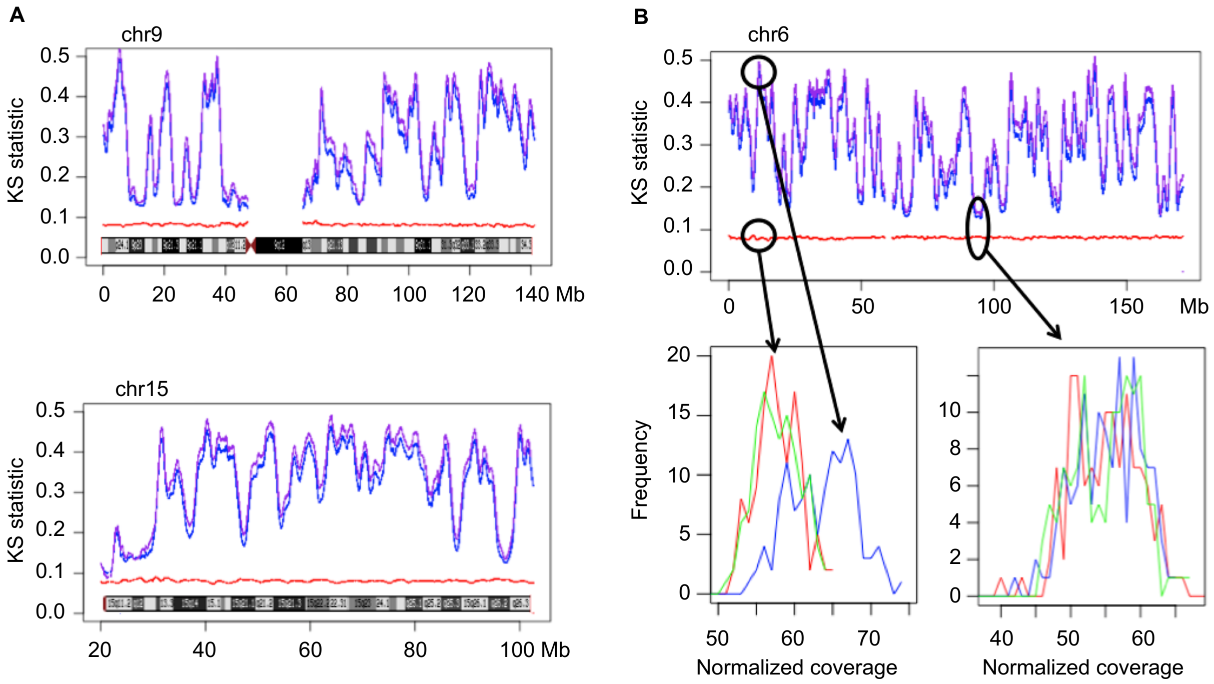

Figure 4 Genome set comparison along chromosomes.

Notes: (**A**) Smoothed KS statistic values (smoothing width=1000 bins=1 Mb) comparing LCLs to MC1 (blue), LCLs to MC2 (purple) and MC1 to MC2 (red) mapped along chromosomes 9 and 15; the inset chromosome ideograms show the increased KS values between LCLs and controls in early-replicating, light Giemsa bands. (**B**) Similar comparisons for chr6, highlighting regions of high dissimilarity between the normalized coverage distributions of LCLs and controls (lower left) and of similarity between all sets (lower right). Blue, red and green denote LCLs, MC1 and MC2, respectively.

Abbreviations: KS, Kolmogorov–Smirnov; LCLs, lymphoblastoid cell lines; MC, matched controls.

of number of rare CNV calls (frequency ≤1% in a reference population) in the genomes of LCLs and their matched controls. We observed in LCLs (relative to controls) an increase in the number of segments with ploidy levels 0, 1 and 4 (null, haploid and tetraploid or greater, respectively); we observed no change for ploidy level 3 (Figure 7).

A previous study reported high similarity of CNV counts in LCL- and PBMC-derived genomes, though results were based only on a small sample.[8] Likewise, another study reported only four regions in which an LCL had a different copy number from its "donor" genome.[10] This second study included only two genome assemblies, constructed using an old version (1.10.0.22) of the CGI pipeline software. We found that early pipeline versions have less ability to observe differences in coverage (Figure 2); most of our comparisons involve genome assemblies constructed using CGI pipeline software versions 2.0.2.22 through 2.5.0.20.

Discussion

We compared genome-wide patterns of depth of coverage (after normalization using RCPs) of LCL- and PBMC-derived genomes. While PBMCs represent the actual somatic genome as derived from direct tissue (blood) samples, LCLs are immortalized using viral transformation; their genomes are expected to be different from the "donor" genomes in a number of ways. We indeed found differences: LCLs have a broader distribution of coverage (after normalization; Figure 1); the differences display a nontrivial pattern along the chromosomes (Figure 4), including higher copy number of the mitochondrial chromosome; LCLs have regions with deeper coverage than PBMCs (Figure 4B); changes in coverage are not enriched in segmentally duplicated regions, but they are somewhat correlated with the GC content and gene density, and are even more strongly correlated with replication timing (Figure 5). These regional differences in coverage (Figure 6) can lead to differences in CNV calls (Figure 7).

Several prior studies have compared LCLs to their donor genomes using techniques such as SNP typing and gene expression analysis, and have found, for the most part, only minor notable differences in coverage. Using SNP arrays, one study reported a high SNP concordance between early-passage LCLs and controls, but suggested that loss of heterozygosity may explain genotype discordance in late-passage LCLs.[12] Another study reported high concordance rates between genotypes and copy number in LCLs and controls.[8] More recent studies have also used whole-exome and whole-genome sequences, but have also reported only minimal differences in coverage between LCL- and PBMC-derived

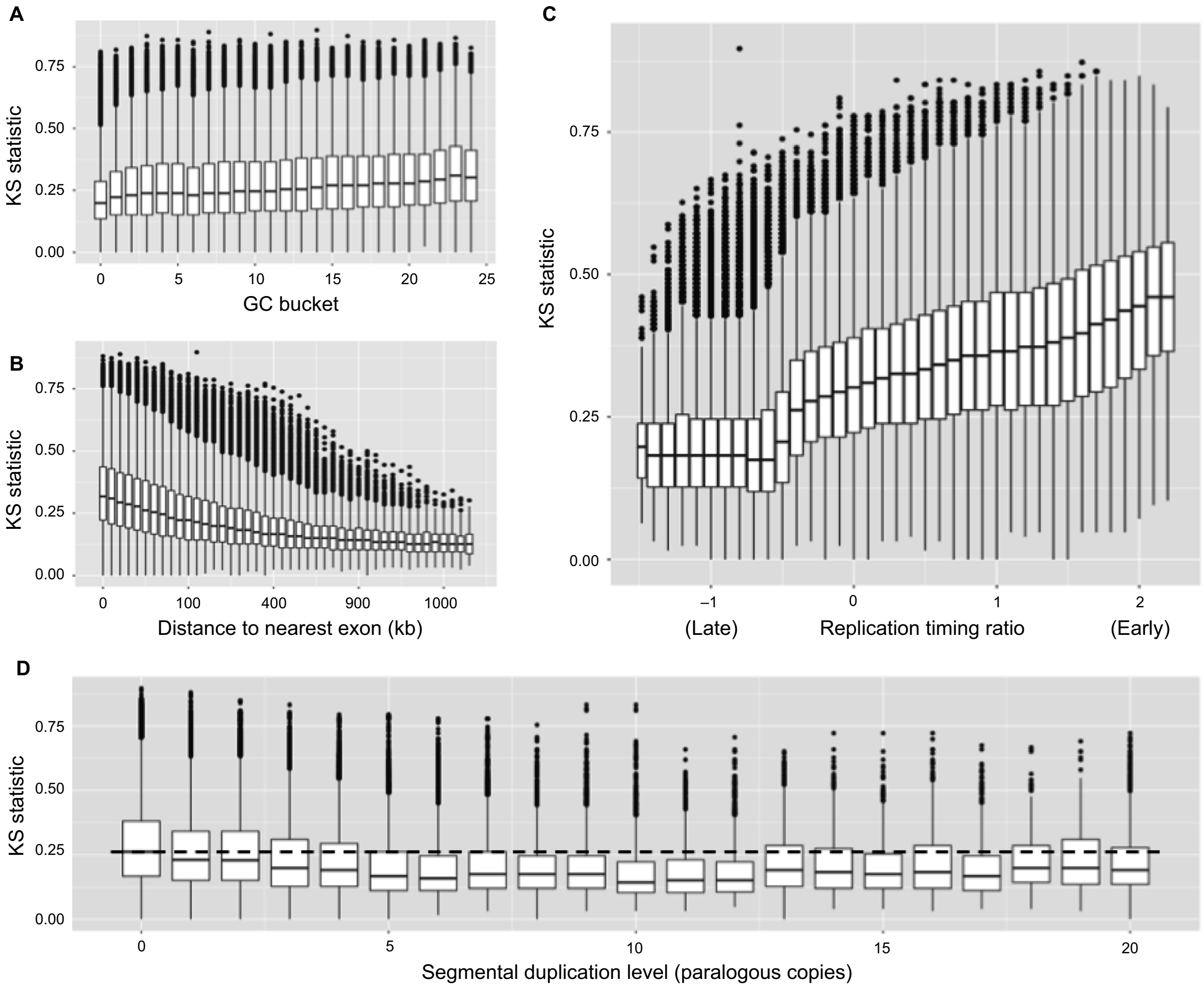

Figure 5 Correlation with various genomic parameters.

Notes: (**A**) Distribution of KS values grouped into 25 "GC buckets" (percentiles) of increasing GC percentage. (**B**) Distribution of KS values grouped by increasing distance to the nearest exon (square root scale). (**C**) Distribution of KS values grouped by replication timing ratio, $\log_2$ (early/late). Lower values = late replication; higher values = early replication. (**D**) Distribution of KS values grouped by segmental duplication level, from 0 (outside segmental duplications, 94% of the genome) to 20 or more paralogous copies. The dashed line highlights the median KS value for the regions outside segmental duplications.

Abbreviations: KS, Kolmogorov–Smirnov.

genomes.[10,11] Differences were found, however, in coverage levels of mtDNA. Researchers reported a copy number increase in LCLs in 1p36.33 and attributed this increase to higher levels of mtDNA.[9] Other findings also suggested a higher level of mtDNA, as well as mitochondria-related gene expression in LCLs.[7] Two features of our study may have contributed to detecting previously unreported differences: the larger sample size and the improved normalization using RCPs. While our study design (driven by sample availability) did not allow us to compare LCL and PBMC genomes from the same individuals, we strove to effectively match our LCLs and controls by metadata to avoid batch effects and inherent population biases. Importantly, the set of LCLs in our study is slightly less diverse than the sets of controls; nevertheless, the coverage comparisons between the two sets of controls showed more similarity than to the set of LCLs. We conclude that the genomic differences displayed by LCLs are not related to population structure.

Since LCLs are actively replicating cells, differential timing of replication can reasonably be expected to lead to different observed coverage levels between early- and late-replicating regions of the genome. Indeed, the differences we observed between LCLs and PBMCs were concentrated in early-replicating regions of the genome. This is consistent with the cell division states of the LCLs, namely, a higher proportion of cells during or after S phase. We also observed an increase in CNV calls (Figure 7), particularly for tetraploid state, which is consistent with regional duplication due to early replication. The increase in CNV calls for haploid and null states may reflect events of DNA loss, many of which

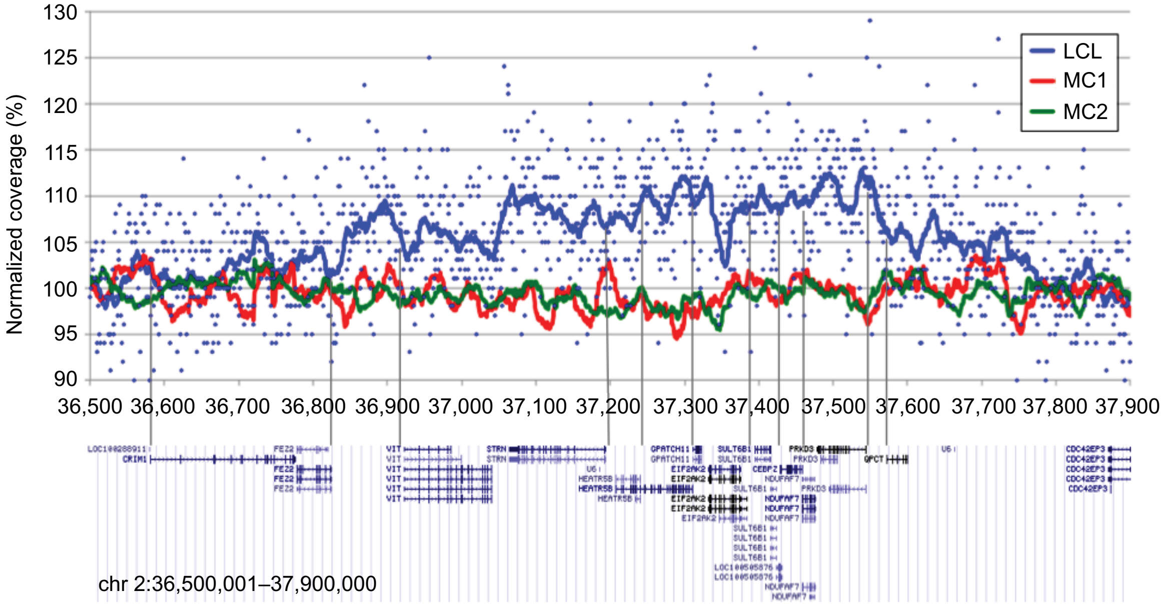

Figure 6 Example of regional coverage distortion.

Notes: Normalized coverage trace for one LCL (blue) vs its matched controls (red and green) in the 2p22.2 early-replicating band, averaged in overlapping 25 kb windows (upper panel). Blue points represent the actual 1 kb resolution normalized coverages for the LCL. Vertical lines connect to the transcription start sites of the known genes in this region (lower panel).

Abbreviations: LCLs, lymphoblastoid cell lines; MC, matched controls.

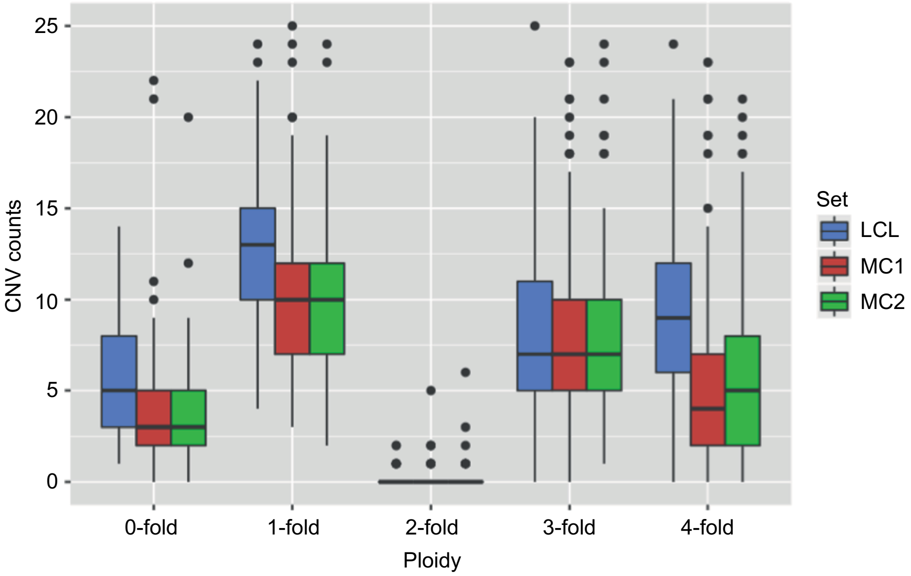

Figure 7 Effect on copy number calls.

Notes: Distribution of rare (frequency <1%) CNV counts in LCLs (blue) and controls (red and green), stratified by ploidy (0 to 4+ fold).

Abbreviations: CNV, copy number variant; LCLs, lymphoblastoid cell lines; MC, matched controls.

would be tolerated in a cell line growing in vitro. This result is consistent with the findings of previous studies, which suggest that earlier replicating regions of the genome are more likely to contain actively transcribed genes than later replicating regions.[22] In turn, these actively transcribed regions have been shown to harbor a higher percentage of indel and substitution mutations,[23] as well as CNVs caused by nonallelic homologous recombination.[24] The process of nonallelic homologous recombination is related to the process of homologous recombination, which has also been shown to preferentially occur in transcriptionally active chromatin, following a double-stranded break.[25] The elevated level of recombination in early-replicating regions of the genome may thus contribute to the differential coverage observed in these regions.

CNVs have been found to play a significant role in causing human disease, especially when CNVs occur in gene-rich areas.[26] Genome-wide association studies, formerly analyzing only SNPs, are now being used to identify de novo and rare CNVs and to associate those CNVs with diseases.[27] Results of such studies have already shown de novo CNVs to be causal factors in both autism and schizophrenia. As such, the accurate identification and analysis of CNVs become vitally important in the clinic, especially in prenatal settings, in which the discovery of disease-causing CNVs can help to shape care and management of such diseases.[26]

In summary, and in contrast with previous reports, we observed significant differences in the depth of coverage between LCL- and PBMC-derived genome assemblies, leading to differential CNV calls. Importantly, these copy number changes preferentially affect regions with higher GC content and closer to genes. This suggests that genomic studies based on LCLs may display locus-specific biases, and that conclusions based on depth of coverage analysis and copy number considerations may require further scrutiny.

Acknowledgments

The authors wish to thank Max Robinson and Vivek Rai for helpful discussions, Denise Mauldin and Chris Witwer for technical support, the Center for Systems Biology (P50GM075647) and the many individuals whose genomes were sequenced. Genome data production was supported by the University of Luxembourg-Institute for Systems Biology Program.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Choy E, Yelensky R, Bonakdar S, et al. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet*. 2008;4(11):e1000287.
2. Sie L, Loong S, Tan EK. Utility of lymphoblastoid cell lines. *J Neurosci Res*. 2009;87(9):1953–1959.
3. Simon-Sanchez J, Scholz S, Fung H-C, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet*. 2007;16(1):1–14.
4. Zook JM, Chapman B, Wang J, et al. Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat Biotechnol*. 2014;32(3):246–251.
5. Zook JM, Catoe D, McDaniel J, et al. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci data*. 2016;3:160025.
6. Diroma MA, Calabrese C, Simone D, et al. Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. *BMC Genomics*. 2014;15(Suppl 3):S2.
7. Chakrabarty S, D'Souza RR, Kabekkodu SP, Gopinath PM, Rossignol R, Satyamoorthy K. Upregulation of TFAM and mitochondria copy number in human lymphoblastoid cells. *Mitochondrion*. 2014;15(1):52–58.
8. Shirley MD, Baugher JD, Stevens EL, et al. Chromosomal variation in lymphoblastoid cell lines. *Hum Mutat*. 2012;33(7):1075–1086.
9. Jeon JP, Shim SM, Nam HY, Baik SY, Kim JW, Han BG. Copy number increase of 1p36.33 and mitochondrial genome amplification in Epstein-Barr virus-transformed lymphoblastoid cell lines. *Cancer Genet Cytogenet*. 2007;173(2):122–130.
10. Nickles D, Madireddy L, Yang S, et al. In depth comparison of an individual's DNA and its lymphoblastoid cell line using whole genome sequencing. *BMC Genomics*. 2012;13:477.
11. Londin ER, Keller MA, D'Andrea MR, et al. Whole-exome sequencing of DNA from peripheral blood mononuclear cells (PBMC) and EBV-transformed lymphocytes from the same donor. *BMC Genomics*. 2011;12:464.
12. Oh JH, Kim YJ, Moon S, et al. Genotype instability during long-term subculture of lymphoblastoid cell lines. *J Hum Genet*. 2013;58(1):16–20.
13. Çalişkan M, Cusanovich DA, Ober C, Gilad Y. The effects of EBV transformation on gene expression levels and methylation profiles. *Hum Mol Genet*. 2011;20(8):1643–1652.
14. Aberg K, Khachane AN, Rudolf G, et al. Methylome-wide comparison of human genomic DNA extracted from whole blood and from EBV-transformed lymphocyte cell lines. *Eur J Hum Genet*. 2012; 20(9):953–955.
15. Carnevali P, Baccash J, Halpern AL, et al. Computational techniques for human genome resequencing using mated gapped reads. *J Comput Biol*. 2011;19(3):279–292.
16. Glusman G, Severson A, Dhankani V, et al. Identification of copy number variants in whole-genome data using reference coverage profiles. *Front Genet*. 2015;6:45.
17. Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC Known Genes. *Bioinformatics*. 2006;22(9):1036–1046.
18. Hiratani I, Ryba T, Itoh M, et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol*. 2008;6(10):2220–2236.
19. Ryba T, Battaglia D, Pope BD, Hiratani I, Gilbert DM. Genome-scale analysis of replication timing: from bench to bioinformatics. *Nat Protoc*. 2011;6(6):870–895.
20. Gaziev AI, Shaikhaev GO. Nuclear mitochondrial pseudogenes. *Mol Biol*. 2010;44(3):358–368.
21. Kim MA, Johannsmann R, Grzeschik KH. Giemsa staining of the sites replicating DNA early in human lymphocyte chromosomes. *Cytogenet Cell Genet*. 1975;15(6):363–371.
22. Schwaiger M, Schübeler D. A question of timing: Emerging links between transcription and replication. *Curr Opin Genet Dev*. 2006;16(2): 177–183.
23. Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. *Nat Rev Genet*. 2015;16(4): 213–223.
24. Koren A, Polak P, Nemesh J, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet*. 2012;91(6):1033–1040.
25. Aymard F, Bugler B, Schmidt CK, et al. Transcriptionally active chromatin recruits homologous recombination at DNA double strand breaks. *Nat Struct Mol Biol*. 2015;21(4):366–374.
26. Martin CL, Kirkpatrick BE, Ledbetter DH. Copy number variants, aneuploidies, and human disease. *Clin Perinatol*. 2015;42(2):227–242.
27. McCarroll SA. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet*. 2008;17(R2):135–142.

Advances in Genomics and Genetics

Dovepress

Publish your work in this journal

Advances in Genomics and Genetics is an international, peer reviewed, open access journal that focuses on new developments in characterizing the human and animal genome and specific gene expressions in health and disease. Particular emphasis will be given to those studies that elucidate genes, biomarkers and targets in the development of new or improved therapeutic interventions. The journal is characterized by the rapid reporting of reviews, original research, methodologies, technologies and analytics in this subject area. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/advances-in-genomics-and-gene-expression-journal