Population Differences in the Rate of Proliferation of International HapMap Cell Lines

Amy L. Stark,[1] Wei Zhang,[2] Tong Zhou,[3] Peter H. O'Donnell,[4,5] Christine M. Beiswanger,[6] R. Stephanie Huang,[4,5] Nancy J. Cox,[7] and M. Eileen Dolan[4,5,*]

The International HapMap Project is a resource for researchers containing genotype, sequencing, and expression information for EBV-transformed lymphoblastoid cell lines derived from populations across the world. The expansion of the HapMap beyond the four initial populations of Phase 2, referred to as Phase 3, has increased the sample number and ethnic diversity available for investigation. However, differences in the rate of cellular proliferation between the populations can serve as confounders in phenotype-genotype studies using these cell lines. Within the Phase 2 populations, the JPT and CHB cell lines grow faster (p < 0.0001) than the CEU or YRI cell lines. Phase 3 YRI cell lines grow significantly slower than Phase 2 YRI lines (p < 0.0001), with no widespread genetic differences based on common SNPs. In addition, we found significant growth differences between the cell lines in the Phase 2 ASN populations and the Han Chinese from the Denver metropolitan area panel in Phase 3 (p < 0.0001). Therefore, studies that separate HapMap panels into discovery and replication sets must take this into consideration.

The International HapMap Project was designed to characterize genetic variation of different individuals from multiple populations, and the samples generated for the project provide the research community with a unique resource for investigating genetics and cellular phenotypes that would be difficult to investigate in humans. Phases 1 and 2 (referred to as Phase 2 for this study) of the project focused on 270 Epstein-Barr virus (EBV)-transformed lymphoblastoid cell lines (LCLs) from four populations: 90 Yoruba (YRI) individuals comprised of 30 trios collected from Ibadan, Nigeria; 45 unrelated Japanese individuals collected from Tokyo (JPT) and 45 unrelated Han Chinese individuals collected from Beijing (CHB), often considered together as 90 Asian (ASN) lines; and 90 Utah residents with ancestry from northern and western Europe (CEU), comprising 30 trios. Phase 3 of the HapMap project expanded the populations and number of cell lines available. The original populations were supplemented, and seven additional populations were added. With the exception of the CEU cell lines, which were collected and transformed approximately thirty years ago, all of the other cell lines were collected within the last ten years and transformed by Coriell Institute for Medical Research (Camden, NJ, USA).

The HapMap cell lines provide extensive publicly available genotyping data,[1,2] allowing for studies of the contribution of genetics to baseline gene expression[3–8] and pharmacologic phenotypes.[9–11] These lines have been utilized for other studies, including evaluation of copy-number variation,[12–14] and are part of the 1000 Genomes Project,[15] in which low-pass sequencing is used to identify most genetic variants that have frequencies of at least 1% in the populations studied. Thus, the cell lines are extraordinarily rich in genetic information, making them valuable for genotype-phenotype studies using the results of any cellular phenotype.

A major advantage to these cell lines is that they offer an alternative to pharmacogenomic studies that would be considered challenging, if not impossible, to perform in human subjects.[16] The fundamental challenge in attempting to identify pharmacogenomic markers from patient trials is that such studies would require a homogenous population of patients treated with the same dosage regimen and minimal confounding variables. In oncology, the standard of care tends to change as new therapies are tested, and the vast majority of patients receive multiple drugs.

In order to facilitate the use of cell lines for pharmacogenetic studies, investigators have undertaken a variety of studies designed to understand potential variability among cell lines, such as the EBV baseline copy-number, rate of cellular proliferation, and ATP levels.[17] Furthermore, the effect of one confounding variable, cellular proliferation rate, on cellular susceptibility to chemotherapeutic-induced cytotoxicity using these LCLs was previously reported.[18]

In this study, we have made a systematic examination of the rate of proliferation within and among populations, with the goal of recognizing and providing additional tools for future study involving these LCLs. The alamarBlue assay described previously was used to measure cellular growth rate in all cell lines.[18] Of the 11 populations included in the HapMap Project, only the CEU LCLs existed as previously established cell lines. The other ten

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA; [2]Institute for Human Genetics and Department of Pediatrics, University of Illinois College of Medicine, Chicago, IL 60612, USA; [3]Department of Medicine, Univerity of Illinois College of Medicine, Chicago, IL 60612, USA; [4]Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL 60637, USA; [5]Committee on Clinical Pharmacology and Pharmacogenomics, University of Chicago, Chicago, IL 60637, USA; [6]Coriell Institute for Medical Research, Camden, NJ 08103, USA; [7]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL 60637, USA
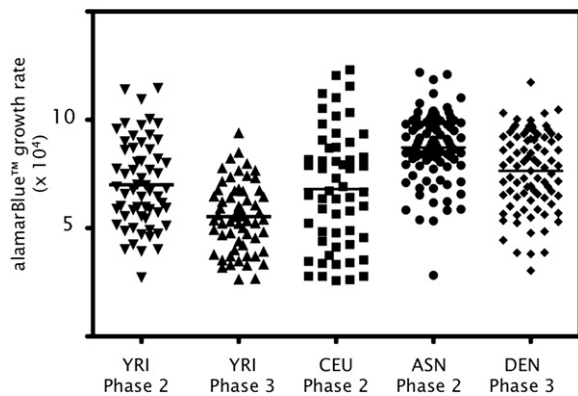*Correspondence: edolan@medicine.bsd.uchicago.edu

Figure 1. Intra- and Interpopulation Growth-Rate Differences among HapMap Panels
The two Yoruba panels' growth rates as measured with alamarBlue were significantly different (p < 0.0001). The original three HapMap populations had similar growth rates for the CEU and YRI, but the samples making up the ASN population grew faster than those of the other two populations (p < 0.0001). The Han Chinese from Denver samples grew significantly slower than those of the ASN (p < 0.0002).

Figure 2. Principal Component Analysis Shows that YRI Phase 2 and 3 Are Overall Genetically Similar
PCA was performed with the use of the NIPALS algorithm in the ade4 library of the R package and the use of ~650,000 common SNPs (MAF > 0.05). The two populations did not separate into distinct clusters. The first two major principal components explain only 1% of the total data variance.

populations were collected and established as cell lines specifically for the HapMap Project over the years 2002 through 2007. Peripheral-blood samples were collected from the participating populations according to strict institutional review board and community-engagement protocols. Mononuclear cells were isolated from the samples and transformed with EBV as described previously.[19] Cell lines were minimally expanded (approximately 100-fold expansion) before inclusion in the HapMap population panels.

In the original Phase 2 populations, there was no significant difference in the cellular proliferation rate between the YRI and CEU lines; however, the ASN (JPT and CHB) cell lines grew faster than the other two populations, as seen in Figure 1 (p < 0.0001). We also observed a significant difference in the rate of cellular proliferation within the cell lines derived from YRI samples: the second set of 90 YRI cell lines (Phase 3) grow significantly slower than the first set of 90 YRI lines (Phase 2), with p < 0.0001 (Figure 1). In contrast, there are no significant growth-rate differences between the two panels of the CEU cell lines. All population-growth comparisons were performed as conservative, nonparametric t tests.

In addition, we evaluated growth differences between the cell lines in the Phase 2 ASN populations with an Asian population released in Phase 3, the Han Chinese from the Denver metropolitan area panel. The requirement of having at least three grandparents of Han Chinese ancestry was the same for both panels. Again, we observe a significantly (p = 0.0002) different rate of cellular proliferation (Figure 1). The significant difference remains even when comparing only the Chinese from Beijing with the Chinese from Denver (p < 0.0001, data not shown). These differences across populations with apparently similar genetic backgrounds provoked a more detailed investiga-
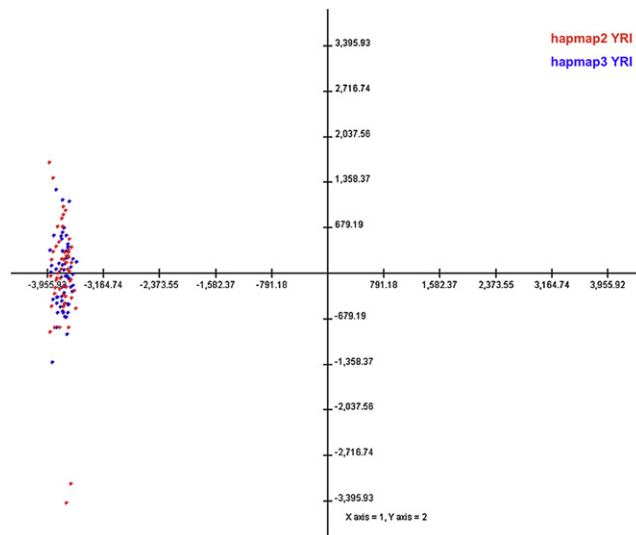
tion into whether the differences may be attributed to genetic or environmental causes.

Principal component analysis (PCA) was performed with the non-linear iterative partial least-squares (NIPALS)[20] algorithm in the ade4 library[21] of the R package and the use of ~650,000 common SNPs (minor allele frequency [MAF] > 0.05) between the two panels; the two populations did not separate into distinct clusters (Figure 2). The first two major principal components explain only 1% of the total data variance. This confirmed the expectation that the two panels of the YRI population are not genetically different. Because the panels did not separate, linked SNPs could not have been driving the separation. However, we repeated the analysis using only unlinked SNPs, with similar results (Figures S1–S4 available online). Similarly, we found no difference between the Han Chinese samples from Beijing in Phase 2 and those from the Denver metropolitan area in Phase 3 by using principal components (data not shown).

We calculated Fst values, comparing the YRI Phase 2 and Phase 3 panels and found no evidence for widespread allele-frequency differences. The Fst values of the YRI populations demonstrated higher values (max 0.336) for males compared to females than for Phase 2 compared to Phase 3 (max 0.206).

The YRI cell lines were collected from the same collection site (Ibadan, Nigeria) and are part of the same HapMap population (Yoruba), the only difference being what appeared to be a random division into two sets (90 cell lines/set) for the public release. To ensure that our observations were not due to the fact that the two panels were ordered, received, and grown at different times within
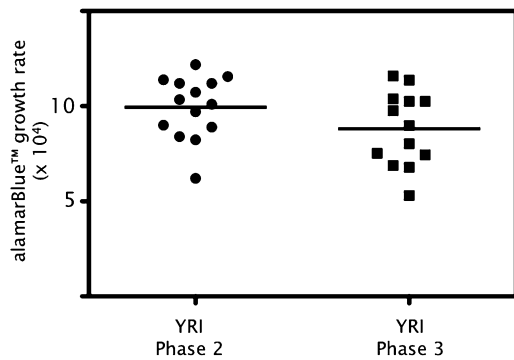
Figure 3. Subsets from Two Yoruba Populations Validate Population-Growth Differences
When controlling for all variables, including shipping time, media and supplements, and personnel, a random subset of cell lines from the two Yoruba panels showed the same growth difference that the entire collection had shown ($p = 0.0599$).

our laboratory, we reevaluated 27 randomly selected, unrelated cell lines from the YRI population (14 YRI Phase 2 and 13 YRI Phase 3). In this subset, we used a one-tailed t test because we had a directional hypothesis. The YRI Phase 3 cell lines grew slower ($p = 0.0599$), suggesting that the difference in growth was not due to differences in the conditions and timing at which the cell lines were purchased and maintained (Figure 3).

We investigated whether this difference may be related to the time required to establish an immortalized cell line after EBV transformation. We compared the time in days for the YRI Phase 2 and YRI Phase 3 cell lines to be frozen down (signifying successful immortalization and cell growth). Phase 3 cells required significantly more time compared to Phase 2 cells ($p = 0.0002$; with a significant difference, $p = 0.0242$, remaining even after removal of the ten outlier cell lines), as seen in Figure 4.

The original cell lines transformed and then maintained at Coriell were selected for release with the first panel of
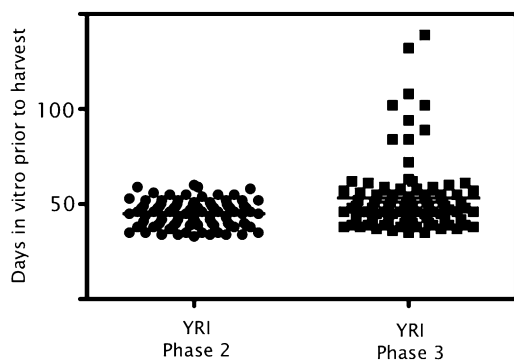

Figure 4. Significant Differences in the Days from Whole-Blood Culture to Initial Freeze at Coriell for the Two Yoruba Populations
Within the two Yoruba populations, the number of days before the whole-blood culture was frozen down was significantly ($p = 0.0002$) higher for the HapMap 3 panel. When the ten cell lines that had the highest number of days until freeze, appearing as outliers on the graph, are removed, the difference remains significant ($p = 0.0242$).

HapMap primarily on the basis of how quickly they reached adequate cell density for harvesting of DNA. Therefore, cell lines that were growing more slowly were likely to become part of the second release of cell lines (Phase 3). Although all cell lines grew adequately to allow for eventual DNA harvest of comparable quality, trios with the least number of slow-growing cell lines were more likely to be included in Phase 2. Therefore, the difference in the cellular growth rate for the YRI panels is largely attributed to the selection and division at Coriell.

The division of the population into two panels with respect to cellular-growth-rate variation has important ramifications for cellular-phenotype studies, particularly for studies that are designed to use one population as discovery and another as replication. Cellular growth rate has been shown to affect some pharmacological phenotypes, including drug sensitivity.[9,18] Studies using non-HapMap, patient-derived LCLs have recognized the importance of looking at the cellular growth rate in pharmacological studies.[22] For example, using the alamarBlue assay of cell-growth inhibition described previously,[9,18] we found that the YRI Phase 2 population was significantly more sensitive to three chemotherapeutic agents—5'deoxyfluorouridine (5'-DFUR), pemetrexed, and carboplatin—compared to the YRI Phase 3 panel, with $p \leq 0.0001$ (Figure 5). The slower-growing HapMap Phase 3 cell lines were more resistant to the drugs, consistent with previous results demonstrating an inverse relationship between slower-growing cells and drug sensitivity.[18] To address this problem when performing genotype-phenotype relationships in this population, the two YRI panels are best analyzed as a single cohort or, alternatively, randomly divided into two sets that are not significantly different in proliferation rate, thus allowing for a discovery and a replication population.

There are ramifications beyond pharmacology for genotype-phenotype studies using panels of cell lines that grow at different rates, including gene expression. Many different studies of gene expression have utilized the HapMap cell lines.[3–8] Baseline gene expression for the Phase 2 CEU and YRI cell lines has been generated with the use of the Affymetrix GeneChip Human Exon 1.0 ST Array.[8] Within the CEU population, we found no gene-expression signatures significantly correlated with growth rate; in contrast, there were 217 transcript clusters that had a Q value less than 0.10 in the YRI panel (Table S1). Of those, 135 genes had differential expression values between the CEU and YRI, with a p value less than 0.05 (Q value less than 0.037).[8] The Phase 2 CEU and YRI panels do not grow at different rates; however, this gene list could be further evaluated in the Phase 3 YRI panel. Interpretation of results from gene-expression studies among populations should consider growth rate, particularly for panels that grow at different rates.

Cellular growth rate was found to be 30% heritable;[18] there could be several rare alleles affecting the growth rate that are not randomly distributed between the two panels.
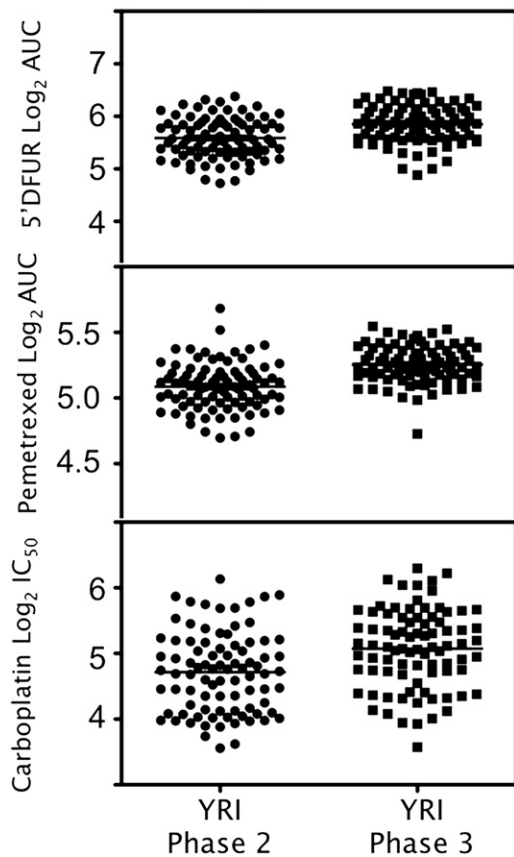
Figure 5. Drug Sensitivity Is Significantly Different between Two Yoruba Populations in a Direction Consistent with the Growth-Rate Differences
Significant differences in pharmacological sensitivity to three independent chemotherapeutic agents exist between the two Yoruba HapMap panels (p ≤ 0.0001 for carboplatin, pemetrexed, and 5'-DFUR).

These nonrandomly distributed alleles could also affect other phenotypes, not just cellular growth rate. The PCA merely confirms a similarity in overall genetic makeup but does not fully address the potential confounding factor of an artificial selection between the two populations. Furthermore, although estimates of heritability across large pedigrees provide a robust measure, it remains that ~70% of cellular-growth-rate variability has environmental, and potentially technical causes that are not well understood or easily identifiable. Regardless of their origins, the importance of this finding is that these differences can have important ramifications for phenotypes of interest and must be considered in experimental design and analysis.

In summary, we found in vitro cellular-growth-rate differences to be confounding and significant within the fully released HapMap YRI population and between the HapMap Han Chinese populations. Overall, we found important growth-rate differences between many of the HapMap populations. On the basis of the data reported here, we believe it is essential to consider growth rate for any studies that utilize different HapMap populations in genotype-phenotype studies. In particular, HapMap popu-lations should be carefully evaluated when incorporated in genetic studies in which replication and validation panels are needed, or when comparisons within and among ethnic groups will be made.

Supplemental Data

Supplemental Data include four figures and one table and can be found with this article online at http://www.cell.com/AJHG/.

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes Project, http://www.1000genomes.org/
Coriell Cell Repositories, http://ccr.coriell.org/
HapMap Project, http://www.HapMap.org/
The R Project for Statistical Computing, http://www.r-project.org/

References

1. International HapMap Consortium. (2003). The International HapMap Project. Nature *426*, 789–796.
2. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al; International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
3. Morley, M., Molony, C.M., Weber, T.M., Devlin, J.L., Ewens, K.G., Spielman, R.S., and Cheung, V.G. (2004). Genetic analysis of genome-wide variation in human gene expression. Nature *430*, 743–747.
4. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M., et al. (2007). A genome-wide association study of global gene expression. Nat. Genet. *39*, 1202–1207.

5. Spielman, R.S., Bastone, L.A., Burdick, J.T., Morley, M., Ewens, W.J., and Cheung, V.G. (2007). Common genetic variants account for differences in gene expression among ethnic groups. Nat. Genet. *39*, 226–231.

6. Storey, J.D., Madeoy, J., Strout, J.L., Wurfel, M., Ronald, J., and Akey, J.M. (2007). Gene-expression variation within and among human populations. Am. J. Hum. Genet. *80*, 502–509.

7. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., et al. (2007). Population genomics of human gene expression. Nat. Genet. *39*, 1217–1224.

8. Zhang, W., Duan, S., Kistner, E.O., Bleibel, W.K., Huang, R.S., Clark, T.A., Chen, T.X., Schweitzer, A.C., Blume, J.E., Cox, N.J., and Dolan, M.E. (2008). Evaluation of genetic variation contributing to differences in gene expression between populations. Am. J. Hum. Genet. *82*, 631–640.

9. Huang, R.S., Kistner, E.O., Bleibel, W.K., Shukla, S.J., and Dolan, M.E. (2007). Effect of population and gender on chemotherapeutic agent-induced cytotoxicity. Mol. Cancer Ther. *6*, 31–36.

10. Jones, T.S., Yang, W., Evans, W.E., and Relling, M.V. (2007). Using HapMap tools in pharmacogenomic discovery: the thiopurine methyltransferase polymorphism. Clin. Pharmacol. Ther. *81*, 729–734.

11. Hartford, C.M., Duan, S., Delaney, S.M., Mi, S., Kistner, E.O., Lamba, J.K., Huang, R.S., and Dolan, M.E. (2009). Population-specific genetic variants important in susceptibility to cytarabine arabinoside cytotoxicity. Blood *113*, 2145–2153.

12. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., and Pritchard, J.K. (2006). A high-resolution survey of deletion polymorphism in the human genome. Nat. Genet. *38*, 75–81.

13. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature *444*, 444–454.

14. Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science *315*, 848–853.

15. Zhang, W., and Dolan, M.E. (2010). Impact of the 1000 genomes project on the next wave of pharmacogenomic discovery. Pharmacogenomics *11*, 249–256.

16. Welsh, M., Mangravite, L., Medina, M.W., Tantisira, K., Zhang, W., Huang, R.S., McLeod, H., and Dolan, M.E. (2009). Pharmacogenomic discovery using cell-based models. Pharmacol. Rev. *61*, 413–429.

17. Choy, E., Yelensky, R., Bonakdar, S., Plenge, R.M., Saxena, R., De Jager, P.L., Shaw, S.Y., Wolfish, C.S., Slavik, J.M., Cotsapas, C., et al. (2008). Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. PLoS Genet. *4*, e1000287.

18. Stark, A.L., Zhang, W., Mi, S., Duan, S., O'Donnell, P.H., Huang, R.S., and Dolan, M.E. (2010). Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. Pharmacogenomics J.

19. Beck, J.C., Beiswanger, C.M., John, E.M., Satariano, E., and West, D. (2001). Successful transformation of cryopreserved lymphocytes: a resource for epidemiological studies. Cancer Epidemiol. Biomarkers Prev. *10*, 551–554.

20. Wold, S., Esbensen, K., and Geladi, P. (1987). Principal components analysis. Chemom. Intell. Lab. Syst. *2*, 37–52.

21. Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. *22*, 1–20.

22. Morag, A., Kirchheiner, J., Rehavi, M., and Gurwitz, D. (2010). Human lymphoblastoid cell line panels: novel tools for assessing shared drug pathways. Pharmacogenomics *11*, 327–340.