Published in final edited form as:

Pharmacogenomics. 2009 April ; 10(4): 549–563. doi:10.2217/PGS.09.8.

Expression and alternative splicing of folate pathway genes in HapMap lymphoblastoid cell lines

Shiwei Duan^{*}, R Stephanie Huang^{*}, Wei Zhang, Shuangli Mi, Wasim K Bleibel, Emily O Kistner, Nancy J Cox, and M Eileen Dolan

Abstract

Aim—Folate is vital for cell growth and development through its important role in one-carbon metabolism - an essential process in the synthesis of amino acids and nucleic acids. Folate pathway genes have been considered as therapeutic targets of drugs for the treatment of cancer and other diseases. Racial and ethnic disparities of folate metabolism and outcome of antifolate therapies have been reported. In this study, we evaluate the genetic regulation for expression and alternative splicing of folate related genes in HapMap lymphoblastoid cell lines (LCLs) of individuals of European and African descent.

Materials & methods—Gene and exon level expression and alternative splicing of folate pathway genes were compared in LCLs derived from the Centre d'Etude du Polymorphisme Humain (CEPH) from Utah (CEU) and the Yoruba from Ibadan (YRI) using a permutation-based test. A genome-wide association study was performed to search for SNPs associated with folate pathway gene expressions and alternative splicing in the combined population samples.

Results—A total of 52 folate pathway genes were evaluated in the analysis of which 46 were expressed in the LCLs. There were 12 genes (26%) with differential gene-level expression and 23 genes (50%) with differential alternative splicing for exons or UTRs between the CEU and the YRI (permutation $p \leq 0.05$). The expression level of *FPGS* and the splicing indices of eight genes (*ATP13A2*, *ASCC3L1*, *IFIH1*, *SMARCA5*, *SMARCA2*, *SETX*, *DDX52* and *RUVBL2*) were found to be associated with SNP genotypes in the combined populations ($p < 3.2 \times 10^{-8}$, Bonferroni corrected $p < 0.05$).

Conclusion—Our study suggests that LCLs are an *in vitro* system suitable to evaluate the expression levels of folate pathway genes. The differential transcript-level expressions and the differentially alternative splicing events of exons or UTRs and associated SNP markers in 2 populations will enhance our understanding of the folate pathway and, thus, facilitate research in the areas of nutrition and folate metabolism.

†Author for correspondence: 5841 S. Maryland Avenue, Box MC2115, The University of Chicago, Chicago, IL 60637, USA, Tel.: +1 773 702 4441; Fax: +1 773 702 0963; E-mail: edolan@medicine.bsd.uchicago.edu.

*Authors contributed to this work equally

Financial & competing interests disclosure: This Pharmacogenetics of Anticancer Agents Research (PAAR) Group study was supported by NIH/NIGMS grant GM61393 and data deposits are supported by GM61374. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research: The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved. The authors have obtained appropriate institutional review board approval for the studies using de-identified human lymphoblastoid cell lines.

Keywords

alternative splicing; folate; gene expression; HapMap

Folates are vitamins essential to the development of the CNS and therefore deficiencies in folate can result in adverse effects in humans [1]. Prenatal folate supplementation reduces the risk of neural tube defects [2] and probably oral-facial clefts [3]. In adult life, folate deficiency has been known for decades to produce a characteristic form of anemia [4]. As a cofactor in the transfer of single carbon unit, folate is involved in the synthesis of both amino acids and nucleic acids [5].

The understanding that folate is essential in cell growth and development [2] has led to the development of antifolates for use in inflammatory diseases (e.g., rheumatoid arthritis) [6,7] and in the treatment of cancer [8]. Antifolate drugs such as methotrexate, lometrexol, pemetrexed, edatrexate, raltrexed, nolatrexed and PT523 are folate antagonists that competitively inhibit the activity of some folate pathway enzymes including those encoded by the genes *TYMS*, *DHFR*, *GART* and *ATIC* [9,10]. Methotrexate and pemetrexed are broad-spectrum agents with large interindividual variations in their pharmacokinetic profiles and side effects that include suppression of bone marrow function, manifested by neutropenia, thrombocytopenia and anemia. Researchers have developed clinical pharmacogenetic models to predict the outcome of patients treated with some antifolate drugs using the combined information of polymorphic variations of specific folate pathway genes and other clinical phenotypes [11].

Novel gene-expression signatures have been identified to predict sensitivity to antifolate drug treatment [12-14]. Gene-expression signatures predicting the subtypes of acute lymphoblastic leukemia and the sensitivity to antifolate therapy have been reported [13,14]. Gene expression can be evaluated as a quantitative phenotype and used in association studies to identify expression quantitative trait loci (eQTL), including SNPs [15-17] and copy-number variations [18]. Researchers have found an association between the methotrexate pathway gene SNPs and methotrexate-induced toxicity that is race-specific in patients with rheumatoid arthritis [6]. Furthermore, ethnic differences in serum and red blood cell folate concentration have been observed between Caucasian and African-Americans or Mexican-Americans [19] and between Northern and Southern Chinese populations [20]. The focus of this paper is to better understand interethnic differences in expression of folate related genes.

The International HapMap lymphoblastoid cell lines (LCLs) offer a unique resource for studying genetic variants associated with expression because extensive genotypic data is publicly available [21,101]. In addition, we recently reported on expression differences in LCL samples derived from Europeans and Africans [17]. In the present study, we evaluated the expression of folate pathway genes in LCLs to ascertain their utility as a cellular model to evaluate sensitivity to antifolates. Specifically, we evaluated both transcript-level expression and alternative splicing in both the Centre d'Etude du Polymorphisme Humain (CEPH) from Utah (CEU) and the Yoruba from Ibadan (YRI) samples and carried out a genome-wide scan for the eQTLs of the genes whose expression or alternative splicing differed between the two populations.

Materials & methods

Selection of folate pathway genes

A list of the complete set of genes for folate biosynthesis and one-carbon pool by folate was obtained by querying the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database for these two pathways [102].

Cell lines

HapMap LCLs (30 CEU trios and 30 YRI trios) were purchased from Coriell Institute for Medical Research (NJ, USA). The details of maintenance are described elsewhere [16,17,103].

Gene expression evaluation

The Affymetrix GeneChip® Human Exon 1.0 sense target (ST) array (Affymetrix, Inc., CA, USA) was used to evaluate the base line expression of over 17,000 genes in 176 HapMap CEU and YRI LCLs [16,17,22,23]. Probes potentially affected by SNPs were removed before the data summarization procedures [17,24]. Folate pathway genes with an average expression level greater than 25th quantile of the average expression values of 17,000 transcript clusters, which translates to mean $\log_2$ transformed mRNA expression signals greater than 5.34, were defined as expressed in LCLs.

Splicing index calculation

The splicing index (SI) represents the log-transformed normalized exon-level probeset intensities by the gene-level transcript cluster intensities in each sample.

$$SI_{i,j} = \log \left(\frac{e_{i,j}}{g_i} \right)$$

where $e_{i,j}$ is the intensity of the j th probeset of the i th transcript cluster, g_i is the intensity of the i th transcript cluster and $SI_{i,j}$ is the splicing index of the j th probeset of the i th transcript cluster [25]. A probeset with higher SI value means that the probeset-representing exon or UTR has higher occurrence in the various transcript isoforms of a gene, implying a better chance of being spliced for the exon or UTR in the gene transcripts and vice versa.

Comparisons of gene expression & SI values in the CEU & the YRI

Using a t-test-based Westfall–Young (WY) approach that adjusts for the correlations among trio members, we identified the differentially expressed genes and SIs between the CEU and the YRI samples. Because of the relatedness among family members, trios were permuted between the CEU and YRI samples. The permutation-adjusted p-value (permutation $p < 0.05$) was considered statistically significant. The adjusted p-values were calculated from 10,000 permutations using Permax 2.2, which was provided as a contributory library by Robert Gray in the R statistical software.

SNPs associated with race-specific genes & probesets

The SNP genotypes for CEU and YRI populations were downloaded from the online HapMap database (release 22, nonredundant and rs_strand version) [104]. Altogether, 1.57 million common SNPs with minor allele frequencies of greater than 5% in each population and with no Mendelian inheritance transmission errors for the trios were used in the present study. Using the total association option in the Linkage Disequilibrium Analyses for Quantitative and

Discrete Traits (QTDT) software [26], the genome-wide association study was carried out in the combined CEU and YRI samples using race and sex as covariates. The tested phenotypes contained 12 differentially expressed genes and 60 differentially spliced probesets. To distinguish between local and distant association between a SNP and its associated genes or transcript isoforms, we chose to define a local SNP if its associated gene was within 4 Mb on either side of the gene; otherwise it was defined as distant. All genomic positions used in the present study were based on National Center for Biotechnology Information (NCBI) Build 36.1.

Validation of differential expression of *SMARCA2* gene between populations

We randomly chose 30 unrelated CEU and 30 unrelated YRI cell lines

- CEU: GM06985, GM06993, GM07022, GM07034, GM07055, GM07056, GM07345, GM07357, GM11830, GM11832, GM11992, GM11993, GM12003, GM12005, GM12006, GM12044, GM12056, GM12155, GM12717, GM12750, GM12751, GM12760, GM12761, GM12812, GM12813, GM12814, GM12873, GM12875, GM12891, GM12892;
- YRI: GM18501, GM18502, GM18505, GM18507, GM18508, GM18517, GM18853, GM18856, GM18859, GM18861, GM18870, GM18912, GM19092, GM19101, GM19102, GM19131, GM19137, GM19140, GM19141, GM19153, GM19171, GM19172, GM19193, GM19203, GM19204, GM19206, GM19207, GM19209, GM19210, GM19239.

Total RNA was extracted using the RNeasy® Mini kit (Qiagen Inc., CA, USA) following the manufacturer's protocol. RNA quality assessment and quantification were conducted using the optical spectrometry 260:280 nm ratio. Subsequently, mRNA was reverse transcribed to cDNA using Applied Biosystems High Capacity Reverse Transcription kit (Applied Biosystems, CA, USA). TaqMan® Hs01030861_m1 probe was used for the evaluation of the differential expression of *SMARCA2* gene. Expression measurements were performed on the Applied Biosystems 7500 real-time PCR system. Total reaction was carried out in 25 µl volume consisting 12.5 µl ABI SYBR® Universal mix (Applied Biosystems), 1.5 µl primers along with 10 µl diluted cDNA. The thermocycler parameters were: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s/60°C for 1 min. Each cycle threshold (C_t) value obtained for each probeset of interest was quantified into relative expression levels of the control (human β 2-microglobulin) using the relative standard curve method. Each standard curve was created using a mixture of cDNA of known concentration from all samples being tested. A ratio comparing the relative quantity of the probeset of interest (PS315993) relative to the quantity of the constitutive exon (exon 5) was compared with the splicing index values from the expression arrays to determine replication of findings.

Results

Folate pathway genes involved in the current study

Our laboratory has profiled the expression levels of over 17,000 transcript clusters in 176 LCLs from the International HapMap CEU and YRI populations using the Affymetrix Human Exon 1.0 ST Array [16,17,22,23]. In this study, we have specifically evaluated 52 folate pathway genes from the KEGG Pathway Database (Figure 1, Supplementary table 1) [27]. Of the 52 genes, 46 are expressed in LCLs; those not expressed include *ALDH1L1*, *ALPI*, *ALPL*, *ALPP*, *DDX4* and *MOV10L1* (Figure 1).

Differential genes & alternative splicing events between the CEU & the YRI

Evidence suggest racial disparities in folate metabolism [28] and the outcome of the antifolate treatment [6], providing us rationale for evaluating differences in gene expression and alternative splicing of exons and UTRs of folate pathway genes in LCLs derived from individuals of Caucasian and African descent. We found that two and ten genes were significantly higher in the CEU and the YRI samples, respectively (permutation-adjusted $p < 0.05$, Table 1). Since many genes undergo alternative splicing and thus may encode proteins with distinct functions [29], we probed deeper into the folate related genes to determine whether particular transcript isoforms were differentially expressed by comparing the SI using the exon-level expression of the folate pathway genes (a higher SI indicates a higher exon expression). Of the 1210 core probesets (exon-level) that represent the 46 expressed folate pathway genes, 60 core probesets within 23 folate pathway genes showed differential SI values between CEU and YRI samples (permutation-adjusted $p < 0.05$, Table 2). Altogether, there are 18 and 12 genes with at least one probeset with a differential splicing index in the CEU or the YRI samples, respectively (Table 2). Interestingly, six genes (*ATIC*, *IFIH1*, *GCH1*, *SETX*, *SMARCA2* and *DDX54*) contain multiple differentially expressed splicing patterns in these two populations (Figure 1).

The *SMARCA2* gene was shown with differential expression between the CEU and YRI (permuted-adjusted $p = 0.0019$, Figure 2 and Table 1). Using samples randomly selected from the two populations (30 CEU and 30 YRI samples), we performed real-time PCR to validate this observation. Gene-expression levels of *SMARCA2* in CEU and YRI correlated between the exon expression array for probeset 3159993 and real-time PCR for exon 4–5 ($p < 0.0001$, Figure 2). Furthermore, consistent with exon expression array results, exon 4–5 of *SMARCA2* was expressed higher in the CEU than in the YRI samples ($p < 0.001$, $T = 3.78$, Figure 2).

In addition, PS2966721 (10th exon) of *ASCC3* has significantly higher SI in the CEU than in YRI, implying the shortest transcript isoforms of *ASCC3* may be less common in the YRI (Figure 3A). Figure 3B illustrates PS3929740 (3'-UTR) for the *GART* gene with significantly higher SIs in the YRI than in the CEU, suggesting the shortest transcript isoform of *GART* is expressed more frequently in the YRI than in the CEU. The alternative splicing patterns of the exons and UTRs can be complex. As shown in Figure 3C, *IFIH1* gene probesets have a race-specific SI pattern. The longer transcript isoform of *IFIH1* gene is likely to be more common in the YRI, while the shorter isoform is more common in the CEU. Note that, 39 probesets within 11 genes with known alternative splicing events in the probeset-targeting segments have differential splicing indices between the CEU and the YRI (Table 2).

eQTL of the differential expressed genes & differential spliced probesets

To identify potential genetic regulators of population specific gene-expression and alternative splicing, we performed genome-wide association studies on 12 genes and 60 probesets in the combined CEU and YRI population samples. As shown in Table 3, *FPGS* gene and ten probesets of seven genes were significantly associated with one or more SNPs ($p < 3.2 \times 10^{-8}$, Bonferroni corrected $p < 0.05$). Among them, we identified two local associations for the probesets of *IFIH1* and *RUVBL2*. As an example illustrated in Figure 4, the proportion of *SMARCA2* transcript isoforms containing the 28th exon to the total gene transcript is higher in the YRI than in the CEU samples. And two ethnic SNPs (rs6544781 and rs405945; $F_{st} = 0.14$ and 0.21) are shown to be distantly-associated with the differentially spliced exons in the two population samples (Figure 4), suggesting that alternative splicing events are under genetic control and likely heritable [30].

Discussion

The majority of folate pathway genes (46 of 52) are expressed in the LCLs, suggesting that this *in vitro* model may be useful in studying folate pathway genes or cellular sensitivity to antifolates. Of the 46 folate pathway genes in LCLs, two and ten genes are expressed to a greater extent in the CEU and YRI as measured by exon array, respectively. Seven of these 12 genes have at least one probeset (exon-level) with a higher splicing index in the samples from one population compared with the other. When evaluating alternative splicing, we found that approximately half of the total expressed genes were found to have at least one higher spliced probeset that was more common in one of the populations, suggesting that the different contents of the transcript isoforms may help explain the differential expression levels of the folate pathway genes between these two populations. Besides the 52 folate pathway genes, there are other folate related genes evaluated in this study, such as folate transporter genes (*SLC19A1*, *SLC19A2* and *SLC19A3*), folate receptor genes (*FOLR1*, *FOLR2* and *FOLR1P*), folate hydrolase genes (*FOLH1* and *FOLH2*). *FOLR1P* is not probed on the array, and *FOLH1*, *FOLR2* and *SLC19A3* are not expressed. However, we did not find the remaining genes to be differentially expressed nor did we find spliced probesets within these genes (data not shown).

Folate pathway genes are often studied in the context of pharmacogenetics of antifolate drugs, such as methotrexate [31]. In this study, we evaluated both transcript-level expression and exon-level alternative splicing events between different ethnic populations. We identified 60 probesets with differential splicing indices in 23 genes in the two distinct populations. According to the University California, Santa Cruz (UCSC) Genome Browser [32], 39 of these 60 probesets were shown to be alternatively spliced (Table 2). However, 21 probesets residing in 13 genes demonstrate potential alternative splicing events that have not been reported yet. For example, *DHFR*, a key gene in the folate pathway, is represented by ten core probesets. Although *DHFR* gene expression is not different between the two populations, one probeset (PS2864625) in its 3'-UTR is more highly expressed (sevenfold) in the YRI population samples (Supplementary Figure 1) than its neighbor probeset (PS2864626). Given that two known *DHFR* transcript isoforms share the same 3'-UTR, it is plausible that one or more undetected *DHFR* transcript isoforms exist by selectively splicing out the tiny 3'-UTR targeted by PS2864626. Another possible explanation is that the hybridization of the PS2864626 is affected by unknown SNPs within the target mRNA sequences; although we have filtered out one potentially affected probe (probe 1864244) of PS2864626 using the latest SNP database (NCBI dbSNP129) [17]. More evidence needs to be collected to confirm whether there is a cryptic splicing site in the 3'-UTR of *DHFR* and whether this affects the genetic regulation mediated by microRNA [33].

It is noteworthy that the majority of the regulatory SNPs identified in this study are defined as distant eQTL, suggesting that the regulatory effect on the target genes is in *trans*. This may involve distant-acting molecules, such as non-coding RNAs [34]. In the process of maturation of pre-mRNA, changes in alternative splicing events has been implicated in several human genetic diseases [35–37]. Although we identified potential differentially spliced exons and UTRs in folate-related genes, the functional significance of these differential alternative splicing patterns are unknown. Population differences in blood folate level may be a result of differences in alternative splicing.

Severe folate deficiency has been reported to be significantly associated with the susceptibility of multiple diseases including cancer [38], cardiovascular diseases [38] and neurodegenerative disorders [39]. Recent studies of the folate pathway genes have focused on expression [13] or SNPs within candidate folate pathway genes [40]. The exon level alternative splicing patterns have not been fully investigated. We performed a genome-wide analysis to identify SNPs associated with genes found to be differentially expressed or with differential splicing indices

between the populations. Since racial disparities of folate nutrition status, metabolism and incidence of folate responsive diseases exists [27,41,42], our results are likely to provide clues on the genetic contribution to interethnic differences observed.

In summary, we have compared the gene expression and alternative splicing of folate pathway genes between European and African descendants. The knowledge gained from the present study provides evidence for interethnic differences in expression and alternative splicing of folate related genes. These results may propel research in the fields of genetic epidemiology of folate-related diseases and the pharmacogenetics of antifolates.

Future perspective

In the present study, we identified folate pathway genes with both transcript-level expression differences and potential alternative splicing differences within samples from two distinct populations. This work provides guidance for further research evaluating the role of differential expressed genes in disease risk and in sensitivity of cells to antifolate drugs, such as pemetrexed and methotrexate.

Executive summary

Folate pathway genes involved in the current study

- A total of 52 folate pathway genes are investigated in an *in vitro* lymphoblastoid cell lines (LCL) system and 46 of these genes are expressed in the LCLs.

Differential genes & alternative splicing events between the CEU & the YRI

- A total of 12 genes (26%) with differential gene-level expression and 23 genes (50%) with differential splicing indices indicative of potential alternative splicing events in exons or UTRs between the centre d'Etude du Polymorphisme Humain from Utah (CEU) and the Yoruba from Ibadan (YRI).
- Among the 60 probesets residing in 11 genes with differential splicing index values, 39 are known to have alternative splicing events in the probesets-targeting exons or UTRs.

Expression quantitative trait loci of the differential expressed genes & differential spliced probesets

- The *FPGS* gene expression level and the splicing index of eight genes (*ATP13A2*, *ASCC3L1*, *IFIH1*, *SMARCA5*, *SMARCA2*, *SETX*, *DDX52* and *RUVBL2*) are found to be associated with SNP genotypes in the combined populations, implying a genetic explanation for the racial disparity of folate metabolism and incidence of folate-related diseases.
- This is a systematic study providing important information on gene expression and alternative splicing for the folate pathway genes using the International HapMap samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to TA Clark, TX Chen, AC Schweitzer and JE Blume (Expression Research, Affymetrix Laboratory, Affymetrix Inc., Santa Clara, CA 94706, USA) for their contribution in generating Exon Array data.

Bibliography

Papers of special note have been highlighted as:

- of interest
- of considerable interest

1. Djukic A. Folate-responsive neurologic diseases. *Pediatr Neurol* 2007;37:387–397. [PubMed: 18021918]
2. Beaudin AE, Stover PJ. Folate-mediated one-carbon metabolism and neural tube defects: balancing genome synthesis and gene expression. *Birth Defects Res C Embryo Today* 2007;81:183–203. [PubMed: 17963270]
3. Boyles AL, Wilcox AJ, Taylor JA, et al. Folate and one-carbon metabolism gene polymorphisms and their associations with oral facial clefts. *Am J Med Genet A* 2008;146A:440–449. [PubMed: 18203168]
4. Jamieson JA, Kuhnlein HV. The paradox of anemia with high meat intake: a review of the multifactorial etiology of anemia in the Inuit of North America. *Nutr Rev* 2008;66:256–271. [PubMed: 18454812]
5. Grzelakowska-Sztabert B, Zielinska ZM. The transfer of one-carbon units in insect metabolism. Pathways of folate coenzyme synthesis. *J Insect Physiol* 1967;13:1207–1219. [PubMed: 6075987]
6. Ranganathan P, Culverhouse R, Marsh S, et al. Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis. *J Rheumatol* 2008;35:572–579. [PubMed: 18381794]
7. Sokka T, Envalds M, Pincus T. Treatment of rheumatoid arthritis: a global perspective on the use of antirheumatic drugs. *Mod Rheumatol* 2008;18:228–239. [PubMed: 18437286]
8. Stankova J, Lawrance AK, Rozen R. Methylenetetrahydrofolate reductase (MTHFR): a novel target for cancer therapy. *Curr Pharm Des* 2008;14:1143–1150. [PubMed: 18473861]
9. Robien K, Boynton A, Ulrich CM. Pharmacogenetics of folate-related drug targets in cancer treatment. *Pharmacogenomics* 2005;6:673–689. [PubMed: 16207145]▪ Review summarizing the drugs targeting the folate pathway and common polymorphisms in folate-metabolizing enzymes and transport proteins.
10. Assaraf YG. Molecular basis of antifolate resistance. *Cancer Metastasis Rev* 2007;26:153–181. [PubMed: 17333344]
11. Wessels JA, van der Kooij SM, le Cessie S, et al. A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis. *Arthritis Rheum* 2007;56:1765–1775. [PubMed: 17530705]▪ Study describes a model for predicting the efficacy of methotrexate in patients with recent-onset rheumatoid arthritis.
12. Hsu DS, Balakumaran BS, Acharya CR, et al. Pharmacogenomic strategies provide a rational approach to the treatment of cisplatin-resistant patients with advanced cancer. *J Clin Oncol* 2007;25:4350–4357. [PubMed: 17906199]▪ Describes genomic-derived signatures of cisplatin and pemetrexed sensitivity that predicted sensitivity *in vitro* and, in the case of cisplatin, also predicted treatment response in patients treated with cisplatin.
13. Kager L, Cheok M, Yang W, et al. Folate pathway gene expression differs in subtypes of acute lymphoblastic leukemia and influences methotrexate pharmacodynamics. *J Clin Invest* 2005;115:110–117. [PubMed: 15630450]▪ Establishes that acute lymphoblastic leukemia (ALL) from certain genetic lineages accumulates significantly lower methotrexate polyglutamate compared with B-lineage ALL without these genetic abnormalities or compared with hyperdiploid ALL. The results provide new knowledge regarding the potential causes of treatment failure in childhood ALL.
14. Soric MJ, Pottier N, Pei D, et al. *In vivo* response to methotrexate forecasts outcome of acute lymphoblastic leukemia and has a distinct gene expression profile. *PLoS Med* 2008;5:E83. [PubMed: 18416598]▪ Describes a set of genes whose expression was significantly related to the reduction of circulating leukemia cells after initial *in vivo* treatment with methotrexate and validated the gene signature in an independent cohort of children with ALL.
15. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. *Nat Genet* 2007;39:1217–1224. [PubMed: 17873874]▪ Uses the International HapMap samples to demonstrate

that gene expression is heritable and to identify gene expressions associated with common SNPs in multiple populations.

16. Duan S, Huang RS, Zhang W, et al. Genetic architecture of transcript-level variation in humans. *Am J Hum Genet* 2008;82:1101–1113. [PubMed: 18439551]• Authors performed a genome-wide study for both local and distant expression quantitative trait loci (eQTL) in the Caucasian and African descendants.
17. Zhang W, Duan S, Kistner EO, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet* 2008;82:631–640. [PubMed: 18313023]•• Identifies genes differentially expressed between the Caucasian and African descendants using the International HapMap lymphoblastoid cell lines.
18. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;315:848–853. [PubMed: 17289997]•• One of the initial studies that demonstrates the contribution of SNPs and copy-number variations to gene expression.
19. Ford ES, Bowman BA. Serum and red blood cell folate concentrations, race, and education: findings from the third National Health and Nutrition Examination Survey. *Am J Clin Nutr* 1999;69:476–481. [PubMed: 10075333]
20. Hao L, Ma J, Stampfer MJ, et al. Geographical, seasonal and gender differences in folate status among Chinese adults. *J Nutr* 2003;133:3630–3635. [PubMed: 14608086]
21. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861. [PubMed: 17943122]•• Comprehensive report on the second phase of the International HapMap Project.
22. Huang RS, Duan S, Shukla SJ, et al. Identification of genetic variants contributing to cisplatin-induced cytotoxicity by use of a genomewide approach. *Am J Hum Genet* 2007;81:427–437. [PubMed: 17701890]
23. Huang RS, Duan S, Bleibel WK, et al. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *Proc Natl Acad Sci USA* 2007;104:9758–9763. [PubMed: 17537913]•• Describe a cell-based model that identifies a small group of genetic variants that explain variation in drug induced cytotoxicity.
24. Duan S, Zhang W, Bleibel WK, Cox NJ, Dolan ME. SNPInProbe_1.0: a database for filtering out probes in the Affymetrix GeneChip(R) Human Exon 1.0 ST array potentially affected by SNPs. *Bioinformatics* 2008;2(10):469–470. [PubMed: 18841244]
25. Affymetrix Inc. Identifying and validating alternative splicing events. Affymetrix Technical Note. 2006
26. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000;66:279–292. [PubMed: 10631157]
27. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 2004;32:D277–D280. [PubMed: 14681412]
28. Lawrence JM, Watkins ML, Chiu V, Erickson JD, Petitti DB. Do racial and ethnic differences in serum folate values exist after food fortification with folic acid? *Am J Obstet Gynecol* 2006;194:520–526. [PubMed: 16458656]
29. Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 2002;418:236–243. [PubMed: 12110900]•• Describes alternative pre-mRNA splicing that encode proteins with distinct functions.
30. Kwan T, Benovoy D, Dias C, et al. Heritability of alternative splicing in the human genome. *Genome Res* 2007;17:1210–1218. [PubMed: 17671095]• Reports that alternative splicing events in humans is heritable and genetically controlled.
31. Krajcinovic M, Moghrabi A. Pharmacogenetics of methotrexate. *Pharmacogenomics* 2004;5:819–834. [PubMed: 15469405]
32. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res* 2002;12(6):996–1006. [PubMed: 12045153]
33. Mishra PJ, Humeniuk R, Mishra PJ, Longo-Sorbello GS, Banerjee D, Bertino JR. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc Natl Acad Sci USA* 2007;104:13513–13518. [PubMed: 17686970]•• Demonstrates

that a naturally occurring miRNA SNP is associated with dihydrofolate reductase overproduction and resistance to methotrexate.

34. Glinsky GV. Phenotype-defining functions of multiple non-coding RNA pathways. *Cell Cycle* 2008;7:1630–1639. [PubMed: 18469518]
35. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev* 2003;17:419–437. [PubMed: 12600935]
36. Brinkman BM. Splice variants as cancer biomarkers. *Clin Biochem* 2004;37:584–594. [PubMed: 15234240]
37. Novoyatleva T, Tang Y, Rafalska I, Stamm S. Pre-mRNA missplicing as a cause of human disease. *Prog Mol Subcell Biol* 2006;44:27–46. [PubMed: 17076263]
38. van Guelpen B. Folate in colorectal cancer, prostate cancer and cardiovascular disease. *Scand J Clin Lab Invest* 2007;67:459–473. [PubMed: 17763182]
39. Mattson MP, Shea TB. Folate and homocysteine metabolism in neural plasticity and neurodegenerative disorders. *Trends Neurosci* 2003;26:137–146. [PubMed: 12591216]
40. Sharp L, Little J. Polymorphisms in genes involved in folate metabolism and colorectal neoplasia: a HuGE review. *Am J Epidemiol* 2004;159:423–443. [PubMed: 14977639]
41. Liu T, Soong SJ, Wang X, Wilson NP, Craig CB. African–American and white differences in nutritional status among low-income women attending public health clinics. *J Health Care Poor Underserved* 1996;7:323–337. [PubMed: 8908889]
42. Estrada DA, Billett HH. Racial variation in fasting and random homocysteine levels. *Am J Hematol* 2001;66:252–256. [PubMed: 11279635]

Websites

101. International HapMap Project www.hapmap.org
102. KEGG pathway database www.genome.ad.jp/kegg
103. The Coriell Institute for Medical Research www.coriell.org
104. Affymetrix www.affymetrix.com

Figure 1. Ideogram for the folate pathway genes in human genome

Figure 2. Real-time expression of *SMARCA2* gene in CEU and YRI samples

Expression of the *SMARCA2* gene was evaluated using real-time PCR with a probe against exon 4–5. The Affymetrix exon array expression levels for probeset 3159993 are plotted against the real-time PCR results for exon 4–5 of *SMARCA2*. The expression of *SMARCA2* in CEU and YRI are compared. CEU: Centre d'Etude du Polymorphisme Humain from Utah; YRI: Yoruba from Ibadan.

Figure 3. Differentially spliced exons of *ASCC3*, *GART* and *IFIH1* genes in the CEU and the YRI populations
Probesets are asterisked in black and gray to denote higher SI values in the CEU and the YRI. CEU: Centre d'Etude du Polymorphisme Humain from Utah; SI: Splicing index; YRI: Yoruba from Ibadan.

Figure 4. The distant QTLs for the alternative splicing in the *SMARCA2* gene
 Probesets are asterisked in black and gray to denote higher SI values in the CEU and the YRI. The top graphs illustrate the frequency of distribution for probeset 3159979 and 3160044 with their corresponding SNPs rs6544781 and rs405945 that are distantly associated with the differentially spliced exons in CEU and YRI samples. CEU: Centre d'Etude du Polymorphisme Humain from Utah; SI: Splicing index; QTL: Quantitative trait loci; YRI: Yoruba from Ibadan.

Table 1
Differentially expressed folate pathway genes in the CEU and the YRI

TC	Gene name	Gene band	KEGG pathway	Permutation-adjusted p-value	Gene title
Expression higher in CEU					
3159946	<i>SMARCA2</i>	9p24.3	Folate biosynthesis	0.0019	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2
3190061	<i>FPGS</i>	9q34.11	Folate biosynthesis	0.0083	Folylpolylutamate synthase
Expression higher in YRI					
3540007	<i>MTHFD1</i>	14q23.2	One-carbon pool by folate	0.0001	Methylenetetrahydrofolate dehydrogenase 1 (NADP ⁺ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase
2745646	<i>SMARCA5</i>	4q31.21	Folate biosynthesis	0.0005	SWI/SNF related, matrix associated, actin-dependent regulator of chromatin, subfamily a, member 5
3228007	<i>SETX</i>	9q34.13	Folate biosynthesis	0.0006	Senataxin
3144973	<i>RAD54B</i>	8q22.1	Folate biosynthesis	0.0011	RAD54 homolog B (<i>S. cerevisiae</i>)
3405531	<i>DDX47</i>	12p13.1	Folate biosynthesis	0.0046	DEAD (Asp-Glu-Ala-Asp) box polypeptide 47
3629378	<i>MTFMT</i>	15q22.31	One-carbon pool by folate	0.0093	Mitochondrial methionyl-tRNA formyltransferase
3754736	<i>DDX52</i>	17q12	Folate biosynthesis	0.0186	DEAD (Asp-Glu-Ala-Asp) box polypeptide 52
2502300	<i>DDX18</i>	2q14.1	Folate biosynthesis	0.0189	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18
3565524	<i>GCHI</i>	14q22.2-q22.3	Folate biosynthesis	0.0215	GTP cyclohydrolase 1 (dopa-responsive dystonia)
3250019	<i>DDX50</i>	10q21.3	Folate biosynthesis	0.0485	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50

CEU: Centre d'Etude du Polymorphisme Humain from Utah; KEGG: Kyoto Encyclopedia of Genes and Genomes; SNF: Sucrose nonfermentable; SWI: Switch; TC: Transcript cluster; YRI: Yoruba from Ibadan.

Probesets containing differential splicing index in the CEU and the YRI for the folate pathway genes

Table 2

Probeset ID	Transcript cluster ID	Permutation-adjusted p-value	Gene	Genomic regions of probeset	Direction	Alternative splicing*
<i>Higher SIs in CEU</i>						
2387009	2387006	0.004	<i>MTR</i>	Chr1:235025627-235025660	(+)	-
2398743	2398736	0.0032	<i>ATP13A2</i>	Chr1:17185971-17186027	(-)	Known
2398782	2398736	0.0131	<i>ATP13A2</i>	Chr1:17204475-17204634	(-)	Known
2489178	2489172	0.0142	<i>MTHFD2</i>	Chr2:74286425-74286503	(+)	-
2502312	2502300	0.0452	<i>DDX18</i>	Chr2:118295696-118295796	(+)	-
2502314	2502300	0.0228	<i>DDX18</i>	Chr2:118295957-118296011	(+)	-
2526782	2526759	0.0392	<i>ATIC</i>	Chr2:215899790-215899896	(+)	-
2584239	2584207	0.0284	<i>IFIH1</i>	Chr2:162882943-162882969	(-)	Known
2584240	2584207	0.0004	<i>IFIH1</i>	Chr2:162883001-162883063	(-)	Known
2584241	2584207	0.0118	<i>IFIH1</i>	Chr2:162883085-162883184	(-)	Known
2745659	2745646	0.0169	<i>SMARCA5</i>	Chr4:144654238-144654298	(+)	-
2745660	2745646	0.0217	<i>SMARCA5</i>	Chr4:144654300-144654527	(+)	-
2745662	2745646	0.0159	<i>SMARCA5</i>	Chr4:144654577-144654703	(+)	-
2745667	2745646	0.0009	<i>SMARCA5</i>	Chr4:144657963-144658021	(+)	-
2966721	2966636	0.0053	<i>ASCC3</i>	Chr6:101321179-101321300	(-)	Known
3048752	3048749	0.0003	<i>DDX56</i>	Chr7:44572132-44572200	(-)	-
3159979	3159946	<0.0001	<i>SMARCA2</i>	Chr9:2022952-2023081	(+)	Known
3159981	3159946	0.0234	<i>SMARCA2</i>	Chr9:2029490-2029560	(+)	Known
3159982	3159946	0.0152	<i>SMARCA2</i>	Chr9:2029587-2029700	(+)	Known
3159989	3159946	0.0009	<i>SMARCA2</i>	Chr9:2037429-2037483	(+)	Known
3159993	3159946	<0.0001	<i>SMARCA2</i>	Chr9:2044597-2044696	(+)	Known
3159995	3159946	0.0069	<i>SMARCA2</i>	Chr9:2046693-2046838	(+)	Known
3160005	3159946	0.0081	<i>SMARCA2</i>	Chr9:2063243-2063323	(+)	Known
3228013	3228007	0.0004	<i>SETX</i>	Chr9:134129691-134129817	(-)	Known
3228057	3228007	0.0097	<i>SETX</i>	Chr9:134193142-134194763	(-)	Known
3349017	3349014	0.0148	<i>PTS</i>	Chr11:11602314-11602339	(+)	-
3453322	3453319	0.0045	<i>DDX23</i>	Chr12:47511267-47511298	(-)	-
3453323	3453319	0.0057	<i>DDX23</i>	Chr12:47511336-47511366	(-)	-
3453325	3453319	0.0014	<i>DDX23</i>	Chr12:47512138-47512169	(-)	-
3472248	3472225	0.0003	<i>DDX54</i>	Chr12:112094623-112094718	(-)	Known

Probeset ID	Transcript cluster ID	Permutation-adjusted p-value	Gene	Genomic regions of probeset	Direction	Alternative splicing*
3565547	3565524	0.0372	<i>GCHI</i>	Chr14:54401803-54401828	(-)	Known
3754766	3754736	0.004	<i>DDX52</i>	Chr17:33073639-33073710	(-)	Known
3754769	3754736	0.0081	<i>DDX52</i>	Chr17:33076411-33076437	(-)	-
3775853	3775842	0.0009	<i>TYMS</i>	Chr18:649650-649687	(+)	-
3838136	3838118	<0.0001	<i>RUVBL2</i>	Chr19:54205042-54205135	(+)	-
Higher SIs in YRI						
2526803	2526759	0.042	<i>ATIC</i>	Chr2:215922644-215922689	(+)	-
2565320	2565262	0.0474	<i>ASCC3LI</i>	Chr2:96327749-96327879	(-)	Known
2565321	2565262	0.0001	<i>ASCC3LI</i>	Chr2:96328076-96328164	(-)	Known
2565323	2565262	0.0001	<i>ASCC3LI</i>	Chr2:96328302-96328425	(-)	Known
2565325	2565262	0.0072	<i>ASCC3LI</i>	Chr2:96328795-96328884	(-)	Known
2565328	2565262	<0.0001	<i>ASCC3LI</i>	Chr2:96330484-96330515	(-)	Known
2565329	2565262	<0.0001	<i>ASCC3LI</i>	Chr2:96331015-96331136	(-)	Known
2584213	2584207	0.0171	<i>IFIH1</i>	Chr2:162832857-162832975	(-)	Known
2584215	2584207	<0.0001	<i>IFIH1</i>	Chr2:162837044-162837128	(-)	Known
2864625	2864619	0.012	<i>DHFR</i>	Chr5:79960361-79960497	(-)	-
3160042	3159946	<0.0001	<i>SMARCA2</i>	Chr9:2105908-2106013	(+)	Known
3160043	3159946	0.0097	<i>SMARCA2</i>	Chr9:2109458-2109516	(+)	Known
3160044	3159946	<0.0001	<i>SMARCA2</i>	Chr9:2113721-2113831	(+)	Known
3160045	3159946	0.0339	<i>SMARCA2</i>	Chr9:2113896-2113925	(+)	Known
3160062	3159946	<0.0001	<i>SMARCA2</i>	Chr9:2151688-2151886	(+)	Known
3160065	3159946	0.0494	<i>SMARCA2</i>	Chr9:2160430-2160459	(+)	Known
3160075	3159946	0.0397	<i>SMARCA2</i>	Chr9:2172195-2172238	(+)	Known
3228075	3228007	0.0003	<i>SETX</i>	Chr9:134218858-134218947	(-)	Known
3233443	3233431	0.0321	<i>FBXO18</i>	Chr10:5977025-5977085	(+)	Known
3472234	3472225	0.0002	<i>DDX54</i>	Chr12:112084081-112084177	(-)	Known
3472238	3472225	0.0342	<i>DDX54</i>	Chr12:112085129-112085166	(-)	Known
3565528	3565524	0.0049	<i>GCHI</i>	Chr14:54378694-54378975	(-)	Known
3629396	3629378	0.0499	<i>MTFMT</i>	Chr15:63103092-63103131	(-)	-
3629399	3629378	0.0065	<i>MTFMT</i>	Chr15:63106242-63106274	(-)	-
3929740	3929721	0.0242	<i>GART</i>	Chr21:33818403-33818934	(-)	Known

* Alternative splicing events are shown in the UCSC genome browser.

CEU; CEU: Centre d'Etude du Polymorphisme Humain from Utah; Chr: Chromosome; SI: Splicing index; UCSC: University of California, Santa Cruz; YRI: Yoruba from Ibadan.

Table 3
SNPs associated with the differentially expressed genes or spliced probe sets

Probeset ID	Transcript cluster ID	Transcript cluster gene	SNP	SNP gene	Permutation-adjusted p-value	F _{st}	Mode
<i>Gene expression-SNP association</i>							
—	3190061	<i>FPGS</i>	rs9990951	—	3.E-08	0.05	Distant
<i>SL-SNP association</i>							
2398743	2398736	<i>ATPI3A2</i>	rs12912451	—	1.E-08	0.24	Distant
2398743	2398736	<i>ATPI3A2</i>	rs8047395	<i>FTO</i>	2.E-09	0.05	Distant
2565321	2565262	<i>ASCC3LI</i>	rs1919417	—	1.E-08	0.21	Distant
2565321	2565262	<i>ASCC3LI</i>	rs28950772	—	1.E-08	0.11	Distant
2565329	2565262	<i>ASCC3LI</i>	rs1876794	—	1.E-08	0.02	Distant
2584215	2584207	<i>IFIH1</i>	rs3747517	<i>IFIH1</i>	3.E-08	0.03	Local
2584215	2584207	<i>IFIH1</i>	rs6754757	<i>TCF7L1</i>	2.E-08	0.02	Distant
2745667	2745646	<i>SMARCA5</i>	rs17400149	—	1.E-08	0.01	Distant
3159979	3159946	<i>SMARCA2</i>	rs11894159	—	2.E-08	0.17	Distant
3159979	3159946	<i>SMARCA2</i>	rs6544781	—	3.E-08	0.14	Distant
3159979	3159946	<i>SMARCA2</i>	rs9211119	—	2.E-08	0.16	Distant
3160044	3159946	<i>SMARCA2</i>	rs405945	—	6.E-09	0.21	Distant
3228013	3228007	<i>SETX</i>	rs4976053	<i>NRCAM</i>	2.E-08	0.06	Distant
3754766	3754736	<i>DDX52</i>	rs4908473	<i>CAMTA1</i>	2.E-08	0.03	Distant
3838136	3838118	<i>RUVBL2</i>	rs1062708	<i>RUVBL2</i>	6.E-12	0.17	Local

Local mode means that the distance between an SNP and its associated transcript cluster is less than 4 Mb. Otherwise it is distant mode. SI: Splicing index.