TITLE PAGE

Article title:
Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia

Authors and affiliations:
Stephan Busche¹, Bing Ge², Ramon Vidal³, Jean-François Spinella³, Virginie Saillour³, Chantal Richer³, Jasmine Healy³, Shu-Huang Chen¹, Arnaud Droit⁴, Daniel Sinnett³,⁵, Tomi Pastinen¹,²
¹Department of Human Genetics, McGill University, Montréal, Québec, Canada
²McGill University and Génome Québec Innovation Centre, Montréal, Québec, Canada
³Department of Hematology-Oncology, CHU Sainte-Justine Research Center, Université de Montréal, Montréal, Québec, Canada
⁴Department of Molecular Medicine, Centre de Recherche du CHUQ, Université Laval, Québec, Canada
⁵Department of Pediatrics, Université de Montréal, Montréal, Québec, Canada

Running title:
Epigenomic alterations in childhood ALL

Keywords:
Childhood pre-B acute lymphoblastic leukemia, DNA methylation, ETv6-RUNXI, gene expression, biomarker

Financial support:
This work was supported by research funds provided by Genome Quebec and Genome Canada, the Terry Fox Research Institute, the Fonds de la Recherche en Santé du Québec (FRSQ), the Canadian Institutes for Health Research (CIHR), and the Cole foundation. SB is the recipient of
a post-doctoral fellowship from the Réseau de Médecine Génétique Appliquée. RV is the recipient of a post-doctoral research fellowship from the Foreign Affairs and International Trade Canada. JFS is the recipient of a Cole Foundation scholarship. DS holds the François-Karl Viau Chair in Paediatric Oncogenomics and is a scholar of the Fonds de la Recherche en Santé du Québec. TP holds a Canada Research Chair (Tier 2) in Human Genomics.

**Corresponding author:**

Daniel Sinnett, PhD

Department of Hematology-Oncology, CHU Sainte-Justine Research Center

3175 Côte Sainte-Catherine, Montréal, Québec, Canada, H3T 1C5,

Email: daniel.sinnett@umontreal.ca; phone: +1 514 345-4931 ext. 3282; fax: +1 514 345-4731

**Conflict of interest disclosures:**

The authors disclose no potential conflicts of interest.

**Manuscript features:**

- Abstract word count: 224
- Text word count: 4994
- Number of Figures: 4
- Number of Tables: 3
- Number of references: 51
ABSTRACT

B-cell precursor acute lymphoblastic leukemia (pre-B ALL) is the most common pediatric cancer. Although the genetic determinants underlying disease onset remain unclear, epigenetic modifications including DNA methylation are suggested to contribute significantly to leukemogenesis. Using the Illumina 450k array we assessed DNA methylation in matched tumor-normal samples of 46 childhood pre-B ALL patients, extending single CpG-site resolution analysis of the pre-B ALL methylome beyond CpG-islands (CGIs). Unsupervised hierarchical clustering of CpG-site neighborhood, gene, or miRNA gene associated methylation levels separated the tumor cohort according to major pre-B ALL subtypes, and methylation in CGIs, CGI shores, and in regions around the transcription start site was found to significantly correlate with transcript expression. Focusing on samples carrying the t(12;21) *ETV6-RUNXI* fusion we identified 119 subtype-specific high-confidence marker CpG-loci. Pathway analyses linked the CpG-loci associated genes with hematopoiesis and cancer. Further integration with whole transcriptome data showed the effects of methylation on expression of 17 potential drivers of leukemogenesis. Independent validation of array methylation and sequencing-derived transcript expression with Sequenom Epityper technology and quantitative real-time RT-PCR, respectively, indicates >80% empirical accuracy of our genome-wide findings. In summary, genome-wide DNA methylation profiling enabled us to separate pre-B ALL according to major subtypes, to map epigenetic biomarkers specific for the t(12;21) subtype, and through a combined methylome and transcriptome approach to identify downstream effects on candidate drivers of leukemogenesis.
INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most frequent cancer in children, accounting for approximately 25% of all pediatric cancers.(1) B-progenitor cell (pre-B), the most prevalent form of ALL in children (~80% of cases), arises from an immature lymphoid progenitor cell that acquires genomic lesions leading to malignant transformation.(1, 2) Cytogenetic analyses detect primary gross chromosomal abnormalities in 75% of pre-B ALL cases, allowing further classification into subtypes based on recurrent translocations or aneuploidies.(3) The t(12;21) resulting into the ETV6-RUNX1 fusion, and high hyperdiploidy (> 50 chromosomes) each account for approximately 25% of cases. Less frequent translocations include t(9;22), generating the BCR-ABL fusion (2-4% frequency), t(1;19) TCF3-PBX1 (2-6%), and t(4;11) MLL-AF4 (1-2%).(3) Current treatment regimens for childhood ALL rely on risk classification systems incorporating molecular genetic subtype information, and risk-adapted therapies have improved overall survival rates. However, due to dismal cure rates in high-risk groups, relapse, and drug resistance resulting in treatment failure, pre-B ALL remains the leading cause of cancer-related death in children.(1, 4) In addition, over two-thirds of survivors develop chronic or late-occurring health problems linked to their disease and its treatment.(5, 6)

DNA methylation is a frequent epigenetic modification that occurs almost exclusively in the context of cytosine methylation, and is crucially involved in the regulation of gene expression.(7) Promoter hypermethylation of tumor suppressor genes, hypomethylation of oncogenes, and alterations of micro-RNA (miRNA) gene methylation have been recognized to favor cancer development.(7, 8) Although primary genetic lesions are considered important initiating events in leukemogenesis, it is now evident that cooperating epigenetic alterations
including DNA methylation modifications contribute significantly to the phenotype of leukemia cells and disease progression.\(^{(8)}\)

DNA methylation in childhood ALL has been intensively studied and methylation profiles have been used to distinguish ALL from healthy samples, other leukemias, or ALL subtypes. \(^{(9-23)}\) Single CpG-site resolution studies focused on analysis of CpG-islands (CGIs), promoter regions, and regions around the transcription start site (TSS), and recent large-scale approaches screening methylation levels of thousands of protein-coding genes associated with these regions found inverse correlation with gene expression in ALL cell lines and primary cells.\(^{(9,14-18,20,22-24)}\) However, recent studies in colon cancer and induced pluripotent stem cells identified substantial alterations of DNA methylation to occur outside of CGIs.\(^{(25,26)}\)

The ALL methylome assessed earlier might overlook important ALL subtype classifiers and prognostic markers. The aim of our study was to provide a detailed map of childhood pre-B ALL methylation, and to identify novel pre-B ALL subtype-specific methylation markers within and outside the classic CpG regions. We further investigated the downstream effects of DNA methylation alterations to find clues of pathways perturbed in leukemia, and intersected methylome and transcriptome data to identify methylation events affecting putative genes driving leukemogenesis.
MATERIALS AND METHODS

Patient samples

Bone marrow or peripheral blood samples isolated at diagnosis (leukemia material prior to treatment) and remission (matched normal material following treatment, derived 1 month to 2 years after diagnosis) from 46 children diagnosed with pre-B ALL were analyzed in this study (Table 1). All patients were of French-Canadian origin in order to decrease genetic heterogeneity. The tumor material contained high levels of leukemic blasts (Table 1), whereas remission samples were blast-free. Patients were from the established Quebec childhood Acute Lymphoblastic Leukemia (QcALL) cohort, and were diagnosed between 1999 and 2010 at the Sainte-Justine University Health Center (UHC), Montreal. The Sainte-Justine UHC Institutional Review Board approved the research protocol and informed consent was obtained from all participating individuals and/or their parents.

DNA methylation analysis

Genomic DNA (gDNA) was isolated from frozen cell pellets using a genomic DNA purification kit (Gentra Systems) and 500 ng of it was used for bisulfite conversion employing the EZ DNA Methylation Kit (Zymo Research) according to manufacturer’s protocols. The modified gDNA was processed as described in the Infinium Assay Methylation Protocol Guide Rev. C (November 2010), and analyzed on Infinium HumanMethylation450 BeadChips (Illumina), measuring DNA methylation at single CpG-site resolution based on genotyping of C/U polymorphisms. As measure of methylation we chose the $\beta$-value, ranging from 0 (no methylation at any allele) to 1.0 (complete methylation of both alleles). We classified the methylation status of probes with $\beta < 0.3$ as unmethylated, with $0.3 \leq \beta \leq 0.7$ as hemi-
methylated, and with $\beta > 0.7$ as methylated. (27) For genomic annotation of BeadChip probes, RefSeq gene and CGI annotations were obtained from the UCSC hg18 reference genome (NCBI Reference Sequence Database Release 36), whereas miRNA annotations were obtained from miRBase (http://www.mirbase.org) version 17. We excluded probes located on the Y chromosome, binding multiple genomic regions, or binding targets with known sequence variants (Table S1). The methylation data have been deposited in NCBI's Gene Expression Omnibus: GSE38235.

**Array methylation validation with the Sequenom Epityper**

Array methylation was validated by quantitative analyses of cleavage products of bisulfite-converted DNA by MALDI-TOF MS (28) using gDNA from twelve patients whose material was also analyzed in the array experiment (Table 1). For PCR, target-specific primer pairs were designed using the EpiDesigner tool (http://www.epidesigner.com) and 50 ng bisulfite converted gDNA was used as input. Duplicates and unreliable detected CpG units having high or low mass were removed from the analysis.

**RNA expression analysis**

Total RNA of 17 pre-B ALL tumor samples (Table 1) was extracted from cell pellets using the mirVana Kit (Life Technologies). Six $\mu$g RNA (RNA integrity number > 6) spiked with ERCC RNA Spike-In Controls (Life Technologies) were rRNA-depleted using the RiboMinus Eukaryote kit (Life Technologies), and cDNA libraries were prepared using the SOLiD Total RNA-seq Kit (Life Technologies). Whole-transcriptome paired-end sequencing (50 bp x 35 bp) was performed using the SOLiD 4 System (Life Technologies). RNA expression
levels were assessed as normalized FPKM (fragments per kilobase per million sequenced reads) by running Cufflinks in quantification-only mode, using the UCSC RefSeq database (downloaded November 1, 2010) with additional coordinates for the ERCC spike-in control as reference (Table S2).

RNA expression validation by quantitative real-time RT-PCR (qPCR)

Tumor RNA from the same twelve patients validated for array methylation was used for expression validation by qPCR (primer sequences given in the supplementary material). First-strand cDNA synthesis was performed using the Verso cDNA Synthesis Kit (Thermo Scientific) with anchored oligo-dT primers according to manufacturer’s protocol. qPCR assays were run on the Rotor-Gene 6000 real-time rotary analyzer (Corbett Life Sciences) using the Platinum SYBR Green qPCR SuperMix-UDG (Life Technologies) according to manufacturer's protocol with the following cycling conditions: 4 min at 95°C, 40 cycles × (20 sec at 95°C, 30 sec at 55°C, and 30 sec at 72°C) followed by the dissociation protocol at 72°C. Transcript expression was analyzed in triplicates. 18S and GAPDH were used for normalization.

Statistical data analysis

Data analyses were performed in R version 2.12.2 (29) applying standard functions and the “lattice” and “RColorBrewer” packages. Heat maps were generated with the “heatmap” function, wherein “hclust” was employed to compute the unsupervised clustering analyses with average linkage as method of clustering and one minus the correlation coefficient (Spearman’s rho) as dissimilarity structure. WGCNA was performed in R version 2.14.2 by employing the WGCNA R package for weighted correlation network analysis.(30) Default settings were applied
and unsigned networks were constructed; the soft-thresholding power was adjusted to 7, the minimum module size set to 30.

**Ingenuity pathway analysis (IPA)**

IPA (Ingenuity Systems) content version 12402621 was used. We performed core analyses using approach-specific reference datasets, considered direct and indirect relationships, included endogenous chemicals, and set thresholds to a maximum of 35 molecules per network and 25 networks per analysis. We included all available data sources, selected for the human species, tissue and primary cells, experimentally observed molecules and/or relationships only, and selected stringent filters when applicable.
RESULTS

Global methylation landscapes

Methylation profiles were measured in 46 pre-B ALL French-Canadian patients (Table 1) on Illumina 450k BeadChips. For each patient, tumor material was isolated at diagnosis and the corresponding normal material at remission. Taking into account the 392,904 autosomal probes on the chip that mapped uniquely to the genome and showed no genetic variation, we observed a bimodal median methylation level distribution (Figure S1A-I) with probes being either unmethylated (40.1% in normal/35.7% in tumor) or methylated (47.5% in normal/50.9% in tumor), with a median methylation level of 0.63 in normal and 0.70 in tumor samples (Figure S1J). We annotated autosomal probes with respect to their location (Figure 1A). Each probe was assigned to its CpG-site neighborhood being located in a CGI, CGI shore, CGI shelf, or open sea, and towards the closest TSS of RefSeq and miRBase annotated genes located within 1.5 kb or 200 bp upstream of the TSS (TSS1500 or TSS200, respectively), or intragenic. For RefSeq genes, intragenic regions were further subcategorized in 5’UTR, exon 1, gene body, and 3’UTR. Irrespective of disease status we detected location-specific CpG-site methylation trends: CGIs were predominantly unmethylated, shores showed dichotomous methylation, and shelves and open sea were essentially methylated. In terms of RefSeq gene context, sites within the TSS1500, TSS200, 5’UTR, and exon 1 were predominantly unmethylated, whereas the remaining intragenic sites were mostly methylated. Sites located in miRNA gene promoter and intragenic regions tended to be methylated (Figure S1K-L). The X-chromosomal median methylation level distribution allowed a clear gender-based sample separation with hemi-methylated patterns characteristic for X-chromosome inactivation detected in females (Figure S1M-O), and median methylation levels of 0.54 in remission samples derived from females (0.64
in tumor) and 0.25 in males (0.41 in tumor). Overall, the increase in median tumor methylation levels is driven by (i) a decrease in unmethylated and concordant increase in methylated sites, and (ii) a quantitative increase towards higher methylation levels in predominantly methylated samples (Figure S1A-L).

Across samples, individual CpG-site methylation variation was low for the majority of probes: in every CpG-site context more than 56.5% (up to 98.9%) of probes displayed a standard deviation (SD) smaller than 0.1 (Table S3). However, we observed significant differences between tumor and normal material (Table S3). A SD < 0.1 was observed for more than 95.3% of CpG-sites in every CpG-site context in normal samples, compared to only 73.6% in tumor samples. In contrast, high CpG-site methylation variation (SD > 0.2) was observed for at least 7.4% of the probes when summed in tumor material, and was rarely observed in normal samples.

Despite these differences we note that DNA methylation patterns are tissue specific. The analyzed tumor material consisted mainly of leukemic blasts which resemble lymphoid progenitor cells within the early developmental stages of B-lymphocytes. In contrast, the corresponding normal material was a mixture of untransformed lymphocytes and monocytes. Therefore, some of the detected methylation alterations might arise from differences in tissue origin rather than be true disease associated methylation changes, and the lower cross-sample CpG-site methylation variation in normal samples might be a result of blunted variability in heterogeneous tissues. Also, patients undergo chemotherapy and remission samples might exhibit treatment-biased methylation patterns which outlast the disease. To circumvent these biases we decided to conduct comparisons between tumor samples to identify tumor and subtype specific methylation alterations.
Methylation profile-based pre-B ALL subtype stratification

Sample-dependent CpG-site methylation level variation should be most revealing for methylation profile-based sample differentiation into specific subgroups. Autosomal probes demonstrating highest variance among pre-B ALL tumor samples and passing filters were selected (Table S4). Unsupervised hierarchical clustering of the detected methylation levels clearly separated samples carrying the t(12;21) translocation from other pre-B ALL subtypes (Figure 1B). A second broad cluster was formed by ten high hyperdiploid (HHD) samples aligned next to four samples categorized as displaying no abnormal finding (NAF). One sample categorized as t(12;21) aligned next to the t(12;21) sample with additional trisomy 21, with both of these samples exhibiting a methylation profile distinct from other t(12;21) tumors. We performed the same analysis using methylation data from the normal samples: these clustered independently of their tumoral counterpart’s cytogenetic characteristics (Figure S3A), indicating that ALL subtypes have distinct global CpG-methylation signatures irrespective of the patients’ germline genetic background.

We applied the same clustering approach to methylation levels detected in individual CpG-site neighborhoods and gene regions (Table S4, Figure S2A-P). In all CpG-site contexts we obtained subtype separations similar to those described above: t(12;21) samples formed a separate cluster, and an additional cluster was formed by the HHD samples. Clustering of methylation levels detected on the X-chromosome followed the same separation patterns (Figure S2Q-R); however, interpretation for males was difficult as only two individuals carried the t(12;21) translocation. Normal samples showed no subtype-dependent clustering in different CpG-contexts (Figure S3B-S).
Subtype stratification based on miRNA gene methylation profiles

Deregulation of miRNA expression in acute leukemia has been associated with altered miRNA gene promoter methylation levels.(8) The BeadChip covers 683 autosomal miRNA gene loci with a total of 3126 probes (Table S1). We tested whether the detected methylation levels displaying highest cross-sample variance allow hierarchical clustering-based subtype separation (Table S4). Twelve of the thirteen samples carrying the t(12;21) translocation formed a distinct cluster, which aligned next to a second cluster consisting of nine HHD samples interspersed with two NAF samples (Figure 1C). Clustering of region-specific methylation levels did not improve separation efficiency (Figure S2S-T). Normal samples clustered irrespective of their tumoral counterpart’s subtype (Figure S3T-V).

Correlation of DNA methylation and gene expression levels

DNA methylation is involved in the regulation of gene expression.(7) We observed a decrease in median methylation levels for CpG-sites located within 1kb of the TSS (Figure S4A). When we compared methylation levels and gene expression in 17 patient samples whose transcriptomes had been determined by RNA-sequencing (Table 1), we found inverse correlation between methylation at CpG-sites located within the TSS1500 and gene expression. The unmethylated / methylated ratio of the highest median expression quartile of transcripts is close to nine (Figure S4B), whereas for the lowest expression quartile this ratio dips to two (Figure S4C). We quantified the methylation / expression correlation in different CpG-site regions determining Spearman’s rank correlation rho for each patient sample (Figure 1D). We observed the highest inverse correlations for CpG-sites located within exon 1, and slight but significant positive correlations for the 3’TUTR. Further inverse correlations were observed in the TSS1500,
TSS200, 5’UTR, CGIs, CGI shores, and gene bodies. The detected methylation and expression levels in CGI shelves were not correlated (Table S5 and Table S6).

**t(12;21)-specific DNA methylation patterns and gene expression**

Unsupervised clustering clearly separated samples carrying the t(12;21) translocation from other molecular subtypes (Figure 1B, Figure S2). Weighted gene correlation network analysis (WGCNA) is based on unsupervised clustering and can be used to link clustering-derived modules of highly interconnected genes to external sample traits.\(^{(30)}\) It has recently been applied to investigate correlation patterns among CpG-sites.\(^{(31)}\) We used WGCNA to identify modules of interconnected CpG-sites related to the t(12;21) subtype. As input data we used probes demonstrating highest cross-sample methylation level variance (as for Figure 1B) and excluded the t(12;21) sample with the additional trisomy 21. Figure 2A displays the cluster dendrogram generated by WGCNA with assigned modules; the brown module displayed highest correlation with the t(12;21)-subtype (rho = 0.98, p = 7e-169, Figure 2B). This module contains 240 probes, 124 of which displaying lower and 116 displaying higher median methylation levels in samples carrying the t(12;21) translocation when compared to the other samples (Table S7A). These probes were associated with 133 genes; their annotations are given in Table S8A.

In addition to the clustering-based WGCNA approach we carried out a Mann-Whitney rank sum test to identify significantly altered DNA methylation patterns specific for the t(12;21) subtype. In this analysis we included all probes that passed our filters and compared probe methylation between samples with and without the t(12;21) translocation. We selected significantly differing probes (false discovery rate (FDR) = 5%) with median β-values < 0.3 in one and > 0.7 in the other group. This led to the identification of 169 hypo- and 140
hypermethylated probes in the t(12;21) subtype (Table S7B-C), associated with 85 and 75 genes, respectively. Annotations of probe-associated genes are given in Table S8B-C. A common set of 119 probes and 62 genes were identified by both approaches and are highlighted in Table S7A and S8A, respectively. Less than 50% of the differentially methylated CpG-sites are located in CGIs or around the TSS (Figure 2C-D).

No probes in the brown module were associated with miRNA genes. In the Mann-Whitney approach, the only identified probe associated with a miRNA, cg19463199, indicated hypomethylation in the TSS1500 of the miRNA-922 gene in addition to its association with KIAA0226. However, miRNA gene methylation-based clustering separated t(12;21) from other subtypes (Figure 1C). To look for evidence of differential methylation at miRNA loci we lowered stringency of our probe selection criteria (FDR 10%, Δβ > 0.25) and identified two additional CpG-sites that confirmed hypomethylation of miRNA-922 and five other differentially methylated miRNA loci within the t(12;21) subtype (Table 2).

To identify methylation alterations that translate into gene expression differences in the t(12;21) subtype, we selected probes located in regions with observed highest absolute median correlation coefficients (Table S5; |median ρ| > 0.1) and compared the methylation trend of associated genes with RNA-sequencing derived transcript expression estimates of the available six t(12;21) and eleven other subtype patient samples. Of the 133 probe-associated genes identified by the WGCNA approach, 49 and 43 were classified as uniformly hypo- and hypermethylated, respectively (Table S9A). Of these, seven hypomethylated genes (IGF2BP1, DSC3, FAM19A1, TMED6, BEST3, SIDT1, OPN3) and one hypermethylated gene (SPSB1) displayed significantly altered expression levels in t(12;21) (Table 3). In every case, significant expression alteration inversely correlated with methylation trend.
Mann-Whitney analysis identified 85 genes as hypo- and 75 genes as hypermethylated specifically in the t(12;21)-subtype, of which 60 hypo- and 52 hypermethylated genes were associated with probes in regions with high absolute correlation (Table S9B-C). 16 hypomethylated genes (IGF2BP1, DSC3, FAM19A1, TMED6, TCFL5, CHL1, EPOR, SOX11, BEST3, FUCA1, SIDT1, HLA-DPB1, MYH10, CPNE2, NRN1, GRK7) and one hypermethylated gene (FBXL15) showed significantly altered transcript expression levels (Table 3). In all instances, hypomethylation was associated with increased transcript expression. However, for hypermethylated gene FBXL15 transcript expression was slightly elevated in t(12;21) samples (Table 3). Of the 17 identified hypomethylated genes, six were jointly identified by both approaches (IGF2BP1, DSC3, FAM19A1, TMED6, BEST3, SIDT1).

Overall, the median expression levels of genes hypermethylated in t(12;21) (β > 0.7) and unmethylated in the other subtypes (β < 0.3) did not differ significantly (p=0.87 for WGCNA, p=0.65 for Mann-Whitney; Figure 2E). In contrast, we observed significantly increased median transcript expression levels only for hypomethylated genes in t(12;21). This suggests that the unmethylated genes in the other subtypes are expressed at low levels similar to methylated genes.

Pathway analysis of differentially methylated genes in t(12;21)

Ingenuity Pathway Analysis (IPA) was carried out to determine biological and disease-associated pathways affected by differential methylation in t(12;21). Genes differentially methylated in regions displaying high absolute correlation between methylation and expression levels were considered (see above). The WGCNA modules contain both hypo- and hypermethylated genes. We carried out a core analysis including all differentially methylated
genes (Table S9A) and weighted gene hypo- or hypermethylation as transcriptional up- or downregulation, respectively, with all probe-associated genes considered in the WGCNA approach used as reference dataset. “Cell Cycle, Cancer, Gastrointestinal Disease” was identified as the top associated network, followed by “Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking” (Figure S5A-B, Table S10A).

In the Mann-Whitney approach we generated two gene lists separating hypo- and hypermethylated genes. We carried out IPA for genes of each list separately (Table S9B-C), weighted differential methylation as previously, and used probe-associated genes interrogated in the Mann-Whitney approach as reference dataset. “Cell Cycle, Cell Death, Cancer” was identified as top network associated with hypomethylation (Figure S5C and Table S10B), and “Cellular Development, Cellular Growth and Proliferation, Hematological System Development and Function” with hypermethylation (Figure S5D, Table S10C).

Finally, we investigated networks generated by genes jointly identified in WCGNA and Mann-Whitney approaches (Table S9B-C); the top associated network was “Cell Death, Cell-To-Cell Signaling and Interaction, Hematological System Development and Function” (Figure S5E, Table S10D).

Validation of t(12;21)-specific CpG-site methylation and gene expression

We employed the Epityper technology and qPCR to validate array-derived CpG-site methylation and RNA-seq-derived gene expression levels, respectively, in twelve previously analyzed patient samples (Table 1). Methylation validation was targeted to CpG-sites associated with genes specifically methylated and expressed in t(12;21) (IGF2BP1, DSC3, FAM19A1, BEST3, TCFL5, SIDT1, TMED6, SPSB1). For the selected array-interrogated TMED6-associated
CpG-site the Epityper experiment failed. In all other instances, array methylation significantly correlated (median rho = 0.882) with Epityper methylation (Figure 3A). Moreover, previously observed significant methylation differences between samples carrying the t(12;21) translocation and other subtypes were confirmed for all interrogated CpG-sites, except for one (of two) FAM19A1-associated site at chr3:68,136,174 and the SPSB1-associated site at chr1:9,300,572 (Figure 3B). Failed validation of t(12;21)-specific methylation at these loci might be due to false positive discovery, or, in case of SPSB1, blurred by additional CpG-sites interrogated in the Epityper CpG unit. Since the Epityper technology quantifies CpG-site methylation on up to 600 bp long stretches of template, we investigated methylation of CpG-sites in close proximity to the array interrogated sites. This identified t(12;21)-specific methylation of SPSB1, and multiple novel loci displaying t(12;21)-specific methylation were discovered (Figure 3B), further supporting characterized subtype-specific gene methylation patterns.

Due to limited RNA availability transcript expression validation was targeted to a subset of the methylation validated genes (IGF2BP1, DSC3, FAM19A1, TCFL5, SIDT1, TMED6). For all genes, we observe high inverse correlations between RNA-seq-derived and qPCR-detected transcript expression (Figure 4A), and previously observed subtype-specific expression was validated for all except TMED6 (Figure 4B).

**DISCUSSION**

We studied 46 childhood pre-B ALL sample pairs derived at diagnosis (tumor) and after disease remission (normal) from the QcALL cohort. Importantly, we studied primary material as culturing of cells might introduce methylation artifacts. For each sample we assessed the methylation status of 402,842 CpG-sites across the whole genome, including CpG-site contexts.
not previously assessed in ALL. We note that the Illumina 450K BeadChip interrogates 485,577 CpG-sites, but we chose a stringent approach and discarded probes targeting multiple genomic regions or containing known SNPs to minimize non-epigenetic variation. Whole transcriptome sequencing was carried out for 17 of the tumor samples. Array methylation and sequencing-derived transcript expression were validated with Sequenom Epityper technology and qPCR, respectively. These independent tests indicate >80% empirical accuracy of our genome-wide findings (10/12 methylation and 5/6 expression observations validated).

Unsupervised hierarchical clustering of methylation levels displaying high cross-sample variance separated the tumor cohort according to major pre-B ALL subtypes. Overall, our clustering results were similar to those obtained in a previous study conducted by Milani et al. investigating 1536 CpG-sites.(15) The CpG-sites interrogated in the Milani study were located in CGIs and around the TSS; their results validate our findings in the corresponding regions, and bring more weight to the clustering results we obtained for previously uninvestigated regions. We report for the first time pre-B ALL subtype separation based on the clustering of methylation levels detected in CGI shores or CGI shelves, associated with individual gene regions, or miRNA genes. Unexpectedly, subtype separation efficiencies achieved in these novel regions were similar to the ones obtained previously, indicating that cross-sample methylation level variance specific to pre-B ALL spans a variety of genomic elements.

DNA methylation in CGIs, CGI shores and at gene promoters is associated with gene silencing;(7) we indeed detect strong inverse correlations between methylation and transcript expression levels in these regions. In contrast, intragenic methylation is assumed to positively correlate with gene expression.(7) However, we observe a positive correlation only for
methylation levels detected in 3’UTR, whereas levels in 5’UTR and exon 1 exhibit high, and in the gene body slight, negative correlations.

As samples carrying the t(12;21) translocation segregated most clearly from the other subtypes we focused further analyses on this subtype with the aim to identify subtype-specific epigenomic biomarkers and potential cooperating epigenetic lesions driving leukemogenesis. The t(12;21) translocation creates the in-frame ETV6-RUNX1 fusion oncogene (syn. TEL-AML1), and the chimeric protein is assumed to act as regulation resistant transcriptional repressor of RUNX1 target genes.(33) However, recent studies investigating fusion protein-dependent genome-wide expression changes report gene repression and activation in similar amounts for protein-coding (34) and miRNA (35) genes.

We employed two methodically distinct approaches, WGCNA and Mann-Whitney, to identify differential probe methylation in the t(12;21)-subtype. We identified a total of 430 unique CpG-sites specifically related to the t(12;21)-subtype, of which 119 were jointly identified by both approaches. None of these CpG-sites have previously been identified to be differentially methylated in pre-B ALL and all these sites might function as novel epigenetic biomarkers specific for the t(12;21) subtype. Their ultimate suitability, however, will have to be confirmed in larger study cohorts. Furthermore, over 50% of these sites are not located in CGIs or around the TSS, indicating that CpG-site regions previously not investigated at single-site resolution might harbor valuable biomarkers for pre-B ALL.

In the Mann-Whitney approach we also identified t(12;21)-specific CpG-sites associated with miRNA genes. The most statistically significant hit in the hypermethylated cohort was miRNA-320a, which was identified in a recent study (35) to be downregulated together with miRNA-494 in t(12;21)-leukemia. Mechanistically, both miRNAs are coupled to the repression
of survivin, their inhibition promoting anti-apoptotic signaling. A second significantly hypermethylated miRNA gene in t(12;21), miRNA-200c, was shown to exhibit strong tumor suppressor properties partially by interacting with the epigenetic regulator and polycomb group protein BMI1,(36) which fulfills essential roles in both normal and leukemic stem cells (37).

CpG-sites identified as t(12;21)-specific were associated with 231 genes (62 were jointly identified by both approaches). Of these, the methylation signatures of BMP4, CELSR1, DSC3, and PON2 were previously identified as subtype classifiers for t(12,21). (15) To our knowledge the remaining 227 genes are novel differentially methylated gene classifiers of the t(12;21)-subtype.

Correlation analysis irrespective of molecular subtype identified differential methylation of BMP4, CELSR1, DSC3, and PON2 to significantly inversely correlate with transcript expression in ALL.(15) We investigated methylation alterations translating into expression differences specific for the t(12;21) subtype, and identified one hyper- and 17 hypomethylated genes with inversely correlating transcript expression. Previous studies carrying out expression profiling of pre-B ALL identified eight of the hypomethylated genes (DSC3, EPOR, FUCA1, HLA-DPB1, TCFL5, NRN1, IGF2BP1, BEST3) as cohort-specific subclassifiers upregulated in the t(12;21)-subtype.(38-42) Transcriptional upregulation of SOX11 was previously identified in t(12;21)- or TCF3 rearrangement-carrying ALL, and for t(12;21) upregulation was accompanied by a demethylated promoter.(43) To our knowledge FAM19A1, CHL1, SIDT1, CPNE2, GRK7, MYH10, and OPN3 are novel biomarkers identified as specifically upregulated in the t(12;21)-subtype, and for all genes except SOX11, demethylation as mechanism underlying augmented expression was previously unknown. The RNA-seq-determined t(12;21)-specific upregulation of
TMED6 expression could not be validated by qPCR, presumably due to a limited validation cohort size.

Several genes hypomethylated and upregulated in t(12;21) have been identified to have a functional impact on tumorigenesis. In pre-B leukemic cells, ETV6-RUNX1 binding to the EPOR promoter was shown to promote cell survival. IGF2BP1 is suggested to enhance tumor cell proliferation by sustaining elevated MYC expression, TCFL5 downregulation is associated with reduced cell viability of human colorectal cancer cell lines, and CHL1 was shown to be differentially expressed during the development of several human cancers.

We identified only one gene, SPSB1, both hypermethylated and downregulated in samples carrying the t(12;21) translocation (Epityper validation of the array-interrogated probe failed; however, neighboring CpG-sites display t(12;21)- specificity; Figure 4B). No link between SPSB1 and leukemia has been previously reported, making it a novel biomarker gene in t(12;21).

Overall, we observed significantly increased median transcript expression levels only for hypomethylated genes in t(12;21), whereas hypermethylation did not significantly impact transcription. An explanation for this might be provided by studies investigating non-hematological cancers, which indicate that CGIs become de novo methylated at an early stage of tumorigenesis and that most newly targeted genes are already repressed before transformation. Mechanistically, abnormally expanding progenitor cells with flexible Polycomb-group protein-mediated transcriptional repression of CGI-associated genes as observed in embryonic stem cells, might shift towards stabler silencing via DNA methylation. This shift from dynamic to static transcriptional repression might prevent subsequent transcriptional activation needed for signal transduction and differentiation.
As the ETV6-RUNX1 fusion gene was identified to directly bind the *EPOR* promoter (44) we hypothesized that this interaction might be generally coupled to promoter demethylation. We tested for it but did not detect significant enrichment of hypomethylated genes identified in our approach when comparing to published RUNX1 chromatin-immunoprecipitation (ChIP)-seq data generated in K562 lymphoblasts.(49) However, K562 cells are derived from chronic myelogenous leukemia, and to ultimately clarify this hypothesis fusion protein ChIP-seq in pre-B ALL needs to be carried out.

Both ETV6 and RUNX1 are known regulators of hematopoiesis (33, 50). IPA analysis identified the functional network “Hematological System Development and Function” to be significantly associated with differential gene methylation in samples carrying the t(12;21) translocation. Other significant associations include “Cancer” and cellular functions deregulated during carcinogenesis, e.g., “Cell Cycle”, “Cell Death”, and “Cellular Growth and Proliferation”. *TNF*, a central hub in several of these generated gene networks, is known to be critically involved in these functions.(51) The most significant networks associated with Mann-Whitney-derived hypo- and hypermethylated genes identified *TP53* as central hub. *TP53* itself is seldom found to be targeted by genetic alterations in ALL, however, components of the *TP53* pathway have been shown to be frequently mutated in ALL,(2) and a recent study reported frequent epigenetic inactivation of *TP53* pathway genes in ALL (21).

Overall, this study illustrates the power of methylation profiling to classify leukemic subtypes, and to identify subtype-specific methylation biomarkers. However, most methylation changes do not alter gene expression detectably, so differential methylation analysis seems suitable for epigenetic biomarker discovery, but as tool to understand leukemogenesis it needs to be coupled with other genomic data (e.g. RNA-sequencing). We demonstrate that integrative analysis of
global methylation and transcriptome alterations in pre-B ALL subtypes can be applied to study
downstream effects of individual genomic rearrangements in tumors. Additional work on the
group of hypomethylated and upregulated genes in samples carrying the t(12;21) translocation
we identified in this study could reveal novel causally important processes in leukemogenesis.
ACKNOWLEDGEMENTS

The authors are indebted to all the patients and their parents who consented to participate in this study. The methylation assays were performed at the McGill University and Génome Québec Innovation Centre, Montréal, Canada. Whole transcriptome sequencing was performed at the Child Health Genomics Platform of the Sainte-Justine UHC Research Center.

AUTHORSHIP CONTRIBUTIONS

DS, TP, and SB contributed to the conception and design of the study. JFS and CR performed sample and library preparation and VS, RV, AD, and SB performed data analysis for the transcriptome sequencing. SHC and SB carried out DNA bisulfite conversions. SB and BG performed the remaining data and statistical analyses. SB wrote the paper. DS and TP contributed to the interpretation of the data. JH provided administrative, technical or material support and provided supervision. JH, DS, and TP were involved in critical revision of the manuscript for important intellectual content. All authors approved the final version.
REFERENCES


FIGURE LEGENDS

Figure 1: Cluster analysis of pre-B ALL sample-derived methylation values for subtype stratification, and correlation of methylation and gene expression levels. (A) Probe annotation regions as outlined in the text. Each probe was assigned to its CpG-site neighborhood and towards the closest TSS of RefSeq and miRBase annotated genes. (B-C) Heat maps illustrate an unsupervised hierarchical clustering analysis of (A) autosomal (B) miRNA gene associated methylation levels exhibiting highest cross-sample variance including all study samples. The dissimilarity structure was determined as one minus Spearman’s rank correlation coefficient and average linkage was used as method of clustering. In the dendrogram on the left each horizontal line represents a sample color-coded according to clinically determined molecular subtype. The color key is given in the legend below Figure 1B. In the top dendrogram each vertical line represents a CpG-site, with detected methylation levels colored according to the scale bar below Figure 1B. Methylation levels detected at (B) 3929 probes (top 1%) and (C) 63 probes (top 2%) were used for heat map generation. (D) Spearman’s rank correlation rho was determined for methylation and transcript expression levels in each indicated CpG-site context. Black diamonds represent individual patient sample-derived Spearman’s rho values, red diamonds the median rho within the specified region.

Figure 2: DNA methylation and expression patterns specific for samples carrying the t(12;21)-translocation. (A) Weighted Gene Correlation Network Analysis (WGCNA)-derived methylation level clustering dendrogram obtained by average linkage hierarchical clustering, with dissimilarity based on topological overlap. Each vertical line in the dendrogram represents one CpG-site, with highly interconnected CpG-sites grouped together in branches. Individual
branches are assigned to colored modules determined by the “Dynamic Tree Cut” function, and identified modules with very similar methylation profiles were merged (Merged dynamic). As input data methylation values detected at the 3929 autosomal probes displaying highest cross-sample variance were employed (same as used to generate Figure 1B). (B) The brown module was identified to be most significantly associated with the t(12;21) subtype. Shown is a scatter plot of methylation probe significance versus brown module membership. Absolute Pearson’s rho (cor) with p-value (p) is indicated above the plot. (C-D) Location of probes identified by the WGCNA or Mann-Whitney approach as t(12;21)-specifically hypo- and hypermethylated annotated to (C) CpG-site neighborhood or (D) regional gene context. (E) Significance of median transcript expression level alterations between hypo- and hypermethylated genes in t(12;21) and other samples. Mann-Whitney rank sum tests were carried out to determine statistical significance between median expression levels of hypo- and hypermethylated genes identified in Table S9. P-values derived for genes identified in the WGCNA and Mann-Whitney approaches are highlighted in yellow and blue boxes, respectively; significant differences (p < 0.05) are highlighted in bold.

**Figure 3: Comparison of array- and Sequenom Epityper-derived methylation.** (A) Epityper CpG unit quantitative methylation values are plotted against array-derived beta values. Genes, interrogated CpG-site location(s), and array-probe ID(s) are displayed. When the Epityper interrogated CpG unit comprised more than one CpG-site: *the array-derived beta values of both CpG-sites were averaged and compared to the Epityper quantified CpG unit comprising these same two CpG-sites; **the Epityper quantified CpG unit comprised two neighboring CpG-sites, of which only the one highlighted in black was interrogated by the array. Spearman’s rank
correlation rho (ρ) with significance level (p) is displayed. (B) Sequenom Epityper-derived methylation of genes identified to be specifically methylated and expressed in t(12;21). For each Epityper-interrogated CpG unit the methylation status of individuals carrying (orange circles) and individuals not carrying (blue circles) the t(12;21)-translocation is displayed (group-specific median methylation indicated by an X). CpG unit comprising CpG-site locations are indicated on the left, CpG-sites also interrogated by the array are highlighted in green. Mann-Whitney p-values comparing methylation of t(12;21)-carriers to non-carriers are indicated on the right, insignificant values (p>0.05) are shaded in grey.

Figure 4: Comparison of RNA-seq- and qPCR-derived transcript expression. (A) For the indicated genes, RNA-seq-derived transcript expression levels (FPKM) are plotted against qPCR ΔCt values (Ct target gene - Ct housekeeping gene (18S or GAPDH)). Samples carrying the t(12;21)-translocation are highlighted in orange, samples of other subtypes in blue; “n.d.” on the ΔCt axis indicates that target gene expression was not detected by qPCR. Spearman’s rank correlation rho (ρ) with significance level (p) derived from correlating FPKM to ΔCt (for each housekeeping gene separately) is displayed. (B) qPCR-derived median ΔCts detected for the indicated genes across samples carrying (orange) or not carrying (blue) the t(12;21)-translocation when normalized to 18S (left) or GAPDH (right). Stars indicate statistical significance (Mann-Whitney p < 0.05).
Table 1: Histopathological and clinical characteristics of pre-B ALL patients.

<table>
<thead>
<tr>
<th>Subtype</th>
<th>ID</th>
<th>Sex</th>
<th>Blast rate [%]</th>
<th>Age [month]</th>
<th>Height [cm]</th>
<th>Weight [kg]</th>
<th>WBC [10E9/L]</th>
<th>Risk</th>
<th>Events</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperdiploidy (HHD)</td>
<td>356</td>
<td>M</td>
<td>63</td>
<td>124</td>
<td>137</td>
<td>35.8</td>
<td>9.8</td>
<td>High</td>
<td>R, D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>372</td>
<td>M</td>
<td>87</td>
<td>99</td>
<td>133</td>
<td>30.3</td>
<td>2.1</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>390</td>
<td>F</td>
<td>72</td>
<td>90</td>
<td>120</td>
<td>21.1</td>
<td>10.0</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>618</td>
<td>F</td>
<td>97</td>
<td>36</td>
<td>94</td>
<td>12.8</td>
<td>2.9</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>663</td>
<td>F</td>
<td>86</td>
<td>28</td>
<td>96</td>
<td>15.2</td>
<td>25.1</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>670</td>
<td>M</td>
<td>93</td>
<td>33</td>
<td>98</td>
<td>14.7</td>
<td>5.5</td>
<td>Standard</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>761</td>
<td>M</td>
<td>88</td>
<td>30</td>
<td>96</td>
<td>15</td>
<td>3.2</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>764</td>
<td>M</td>
<td>75</td>
<td>55</td>
<td>112</td>
<td>21.6</td>
<td>59.2</td>
<td>High</td>
<td>R, R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>768</td>
<td>M</td>
<td>95</td>
<td>53</td>
<td>100</td>
<td>16</td>
<td>4.9</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>775</td>
<td>F</td>
<td>98</td>
<td>41</td>
<td>102</td>
<td>15.7</td>
<td>50.7</td>
<td>High</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>777</td>
<td>F</td>
<td>73</td>
<td>33</td>
<td>93</td>
<td>13.6</td>
<td>61.7</td>
<td>High</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>801</td>
<td>M</td>
<td>93</td>
<td>48</td>
<td>108</td>
<td>19.3</td>
<td>3.9</td>
<td>Standard</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>815</td>
<td>M</td>
<td>80</td>
<td>63</td>
<td>106</td>
<td>19.2</td>
<td>6.1</td>
<td>Standard</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>819</td>
<td>M</td>
<td>97</td>
<td>48</td>
<td>98</td>
<td>16.8</td>
<td>98.2</td>
<td>High</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td>Trisomy 5</td>
<td>595</td>
<td>M</td>
<td>73</td>
<td>155</td>
<td>161.5</td>
<td>45</td>
<td>48.8</td>
<td>High</td>
<td>R, D</td>
<td></td>
</tr>
<tr>
<td>Hypodiploidy</td>
<td>605</td>
<td>M</td>
<td>88</td>
<td>141</td>
<td>148</td>
<td>48</td>
<td>30.2</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(12;21)</td>
<td>373</td>
<td>F</td>
<td>75</td>
<td>55</td>
<td>96.5</td>
<td>13</td>
<td>208.0</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>375</td>
<td>F</td>
<td>86</td>
<td>32</td>
<td>88</td>
<td>13.3</td>
<td>141.4</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>379</td>
<td>F</td>
<td>93</td>
<td>28</td>
<td>90</td>
<td>12.2</td>
<td>4.5</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>392</td>
<td>M</td>
<td>55</td>
<td>84</td>
<td>123</td>
<td>23.6</td>
<td>11.6</td>
<td>Standard</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>411</td>
<td>F</td>
<td>87</td>
<td>57</td>
<td>110.5</td>
<td>18.1</td>
<td>3.8</td>
<td>Standard</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>420</td>
<td>M</td>
<td>94</td>
<td>57</td>
<td>107.5</td>
<td>19.4</td>
<td>127.0</td>
<td>High</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>443</td>
<td>F</td>
<td>90</td>
<td>32</td>
<td>94.5</td>
<td>14.2</td>
<td>103.9</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>580</td>
<td>F</td>
<td>94</td>
<td>52</td>
<td>105</td>
<td>16.6</td>
<td>35.3</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>614</td>
<td>F</td>
<td>75</td>
<td>48</td>
<td>105</td>
<td>17.1</td>
<td>33.1</td>
<td>High</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>657</td>
<td>F</td>
<td>62</td>
<td>31</td>
<td>90</td>
<td>11.2</td>
<td>17.1</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>685</td>
<td>F</td>
<td>86</td>
<td>63</td>
<td>107</td>
<td>19.7</td>
<td>83.5</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>753</td>
<td>F</td>
<td>95</td>
<td>61</td>
<td>111</td>
<td>18.2</td>
<td>2.9</td>
<td>Standard</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td></td>
<td>817</td>
<td>F</td>
<td>93</td>
<td>67</td>
<td>110</td>
<td>16.9</td>
<td>3.3</td>
<td>Standard</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>t(12;21) &amp; trisomy 21</td>
<td>718</td>
<td>F</td>
<td>66</td>
<td>142</td>
<td>130</td>
<td>44.2</td>
<td>30.1</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(9;22)</td>
<td>790</td>
<td>M</td>
<td>75</td>
<td>171</td>
<td>170</td>
<td>60.1</td>
<td>14.3</td>
<td>High</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>825</td>
<td>F</td>
<td>91</td>
<td>138</td>
<td>157</td>
<td>43.3</td>
<td>65.5</td>
<td>High</td>
<td>R</td>
<td></td>
</tr>
<tr>
<td>t(1;19)</td>
<td>424</td>
<td>M</td>
<td>94</td>
<td>46</td>
<td>104</td>
<td>19</td>
<td>15.2</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(4;11)</td>
<td>757</td>
<td>F</td>
<td>97</td>
<td>127</td>
<td>149</td>
<td>35.3</td>
<td>7.7</td>
<td>High</td>
<td>R, V</td>
<td></td>
</tr>
<tr>
<td>t(9;12) &amp; monosomy 7</td>
<td>752</td>
<td>M</td>
<td>30</td>
<td>31</td>
<td>86</td>
<td>13.2</td>
<td>37.8</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No abnormal finding (NAF)</td>
<td>381</td>
<td>F</td>
<td>67</td>
<td>32</td>
<td>99</td>
<td>16.9</td>
<td>27.7</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>384</td>
<td>M</td>
<td>94</td>
<td>102</td>
<td>130</td>
<td>30</td>
<td>3.1</td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>431</td>
<td>M</td>
<td>98</td>
<td>144</td>
<td>156.5</td>
<td>44.4</td>
<td>7.1</td>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>579</td>
<td>M</td>
<td>87</td>
<td>75</td>
<td>127</td>
<td>25</td>
<td>183.1</td>
<td>High</td>
<td>R, D</td>
<td></td>
</tr>
</tbody>
</table>

Downloaded from cancerres.aacrjournals.org on April 15, 2017. © 2013 American Association for Cancer Research.
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>608</td>
<td>M</td>
<td>72</td>
<td>47</td>
<td>107</td>
<td>20</td>
<td>81.5</td>
</tr>
<tr>
<td>617</td>
<td>F</td>
<td>77</td>
<td>46</td>
<td>111</td>
<td>20</td>
<td>23.6</td>
</tr>
<tr>
<td>762</td>
<td>F</td>
<td>97</td>
<td>133</td>
<td>156</td>
<td>34.8</td>
<td>274.6</td>
</tr>
<tr>
<td>767</td>
<td>F</td>
<td>79</td>
<td>81</td>
<td>128</td>
<td>23.6</td>
<td>3.6</td>
</tr>
<tr>
<td>794</td>
<td>M</td>
<td>89</td>
<td>168</td>
<td>156</td>
<td>37.2</td>
<td>280.6</td>
</tr>
<tr>
<td>802</td>
<td>M</td>
<td>92</td>
<td>58</td>
<td>107</td>
<td>19.8</td>
<td>40.5</td>
</tr>
<tr>
<td>820</td>
<td>M</td>
<td>31</td>
<td>101</td>
<td>96</td>
<td>15.6</td>
<td>19.3</td>
</tr>
</tbody>
</table>

1. Details on the cytogenetic analyses are given in the supplementary material.
2. Data were obtained at diagnosis.
3. WBC, white blood cell count.
4. The risk groups were determined and the patients were treated according to the Dana-Farber Cancer Institute Childhood ALL consortium protocols effective at the time of diagnosis.
5. R, relapse; D, death. For two relapse cases death did not occur within the follow up period.
6. Additional datasets available: R, RNA-sequencing; V, Epityper and qPCR.
Table 2: miRNA genes differentially methylated in samples carrying the t(12;21) translocation.

<table>
<thead>
<tr>
<th>miRNA1</th>
<th>probe ID2</th>
<th>Chr3</th>
<th>Location3</th>
<th>miRNA context4</th>
<th>CpG-site neighbrhd5</th>
<th>t(12;21) med. β6</th>
<th>others med. β7</th>
<th>p-value8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>hypomethylated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-mir-922</td>
<td>cg19463199</td>
<td>3</td>
<td>198886256</td>
<td>TSS1500</td>
<td>open sea</td>
<td>0.289</td>
<td>0.868</td>
<td>8.62E-07</td>
</tr>
<tr>
<td></td>
<td>cg10579986</td>
<td>3</td>
<td>198885848</td>
<td>TSS200</td>
<td>open sea</td>
<td>0.503</td>
<td>0.922</td>
<td>8.77E-06</td>
</tr>
<tr>
<td></td>
<td>cg15435900</td>
<td>3</td>
<td>198885797</td>
<td>gene body</td>
<td>open sea</td>
<td>0.558</td>
<td>0.899</td>
<td>2.07E-06</td>
</tr>
<tr>
<td>hsa-mir-320b-1</td>
<td>cg04390865</td>
<td>1</td>
<td>117015533</td>
<td>TSS1500</td>
<td>s.shelf</td>
<td>0.679</td>
<td>0.927</td>
<td>4.71E-05</td>
</tr>
<tr>
<td>hsa-mir-4747</td>
<td>cg11970254</td>
<td>19</td>
<td>4883663</td>
<td>TSS1500</td>
<td>CGI</td>
<td>0.455</td>
<td>0.922</td>
<td>5.23E-05</td>
</tr>
<tr>
<td><strong>hypermethylated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hsa-mir-320a</td>
<td>cg08264885</td>
<td>8</td>
<td>22159482</td>
<td>TSS1500</td>
<td>s.shore</td>
<td>0.402</td>
<td>0.133</td>
<td>5.83E-07</td>
</tr>
<tr>
<td>hsa-mir-1976</td>
<td>cg23955417</td>
<td>1</td>
<td>26753516</td>
<td>TSS1500</td>
<td>open sea</td>
<td>0.841</td>
<td>0.304</td>
<td>4.71E-05</td>
</tr>
<tr>
<td>hsa-mir-200c</td>
<td>cg16642299</td>
<td>12</td>
<td>6941644</td>
<td>TSS1500</td>
<td>open sea</td>
<td>0.703</td>
<td>0.282</td>
<td>5.53E-05</td>
</tr>
</tbody>
</table>

1 miRNA name (miRBase version 17).
2 Illumina ID of CpG-site interrogating probe associated with the miRNA gene.
3 Chromosome number (Chr) and genomic coordinate (Location) of the cytosine interrogated by the probe.
4 Location of the probe with respect to the miRNA gene.
5 Location of the probe in the CpG-site neighborhood.
6 Median methylation level across samples carrying the t(12;21) translocation.
7 Median methylation level across all other samples.
8 Mann-Whitney p-value for difference in methylation between t(12;21) and other samples.
Table 3: Genes displaying significant hypo- or hypermethylation and differential expression in samples carrying the t(12;21) translocation.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession</th>
<th>Approach</th>
<th># of CpG-sites</th>
<th>med. fpkm</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEST3</td>
<td>NM_032735</td>
<td>W+M</td>
<td>1</td>
<td>43442</td>
</tr>
<tr>
<td>CHL1</td>
<td>NM_006614</td>
<td>M</td>
<td>n/a</td>
<td>6477</td>
</tr>
<tr>
<td>CPNE2</td>
<td>NM_152727</td>
<td>M</td>
<td>n/a</td>
<td>64912</td>
</tr>
<tr>
<td>DSC3</td>
<td>NM_024423</td>
<td>W+M</td>
<td>6</td>
<td>15275</td>
</tr>
<tr>
<td>EPOR</td>
<td>NM_000121</td>
<td>M</td>
<td>n/a</td>
<td>10508</td>
</tr>
<tr>
<td>FAM19A1</td>
<td>NM_213609</td>
<td>W+M</td>
<td>2</td>
<td>43213</td>
</tr>
<tr>
<td>FBXL15</td>
<td>NM_024326</td>
<td>M</td>
<td>n/a</td>
<td>6991</td>
</tr>
<tr>
<td>FUCA1</td>
<td>NM_000147</td>
<td>M</td>
<td>n/a</td>
<td>90807</td>
</tr>
<tr>
<td>GRK7</td>
<td>NM_139209</td>
<td>M</td>
<td>n/a</td>
<td>4198</td>
</tr>
<tr>
<td>HLA-DPB1</td>
<td>NM_002121</td>
<td>M</td>
<td>n/a</td>
<td>38738</td>
</tr>
<tr>
<td>IGF2BP1</td>
<td>NM_006546</td>
<td>W+M</td>
<td>6</td>
<td>35148</td>
</tr>
<tr>
<td>MYH10</td>
<td>NM_005964</td>
<td>M</td>
<td>n/a</td>
<td>19942</td>
</tr>
<tr>
<td>NRN1</td>
<td>NM_016588</td>
<td>M</td>
<td>n/a</td>
<td>38910</td>
</tr>
<tr>
<td>OPN3</td>
<td>NM_014322</td>
<td>W</td>
<td>n/a</td>
<td>30895</td>
</tr>
<tr>
<td>SIDT1</td>
<td>NM_017699</td>
<td>W+M</td>
<td>1</td>
<td>17969</td>
</tr>
<tr>
<td>SOX11</td>
<td>NM_003108</td>
<td>M</td>
<td>n/a</td>
<td>28661</td>
</tr>
<tr>
<td>SPSB1</td>
<td>NM_025106</td>
<td>W</td>
<td>n/a</td>
<td>3708</td>
</tr>
<tr>
<td>TCFL5</td>
<td>NM_006602</td>
<td>M</td>
<td>n/a</td>
<td>126736</td>
</tr>
<tr>
<td>TMED6</td>
<td>NM_144676</td>
<td>W+M</td>
<td>1</td>
<td>55113</td>
</tr>
</tbody>
</table>

1 Gene name according to HUGO Gene Nomenclature Committee.
2 RefSeq transcript accession number.
3 Probe-identifying approach: W, WGCNA; M, Mann-Whitney; W+M, both approaches.
4 Number of gene associated CpG-sites identified in the WGCNA approach (Table S7A).
5 Number of gene associated CpG-sites identified in the Mann-Whitney approach (Table S7B-C).
6 Gene methylation trend in t(12;21).
7 Median transcript expression across samples carrying the t(12;21) translocation.
8 Median transcript expression across all other samples.
9 Mann-Whitney p-value for difference in transcript expression between t(12;21) and other samples.
10 Hypermethylation in t(12;21) is associated with slightly increased expression.
t(12;21)-specific expression was not validated by qPCR.
**Figure 4**

**A**

**IGF2BP1 (NM_006546)**

- \( p_{18S} = -0.848 \) / \( p_{18S} = 3.9 \times 10^{-3} \)
- \( p_{GAPDH} = -0.932 \) / \( p_{GAPDH} = 2.5 \times 10^{-4} \)

**DSC3 (NM_024423)**

- \( p_{18S} = -0.422 \) / \( p_{18S} = 1.7 \times 10^{-1} \)
- \( p_{GAPDH} = -0.257 \) / \( p_{GAPDH} = 4.2 \times 10^{-1} \)

**TMED6 (NM_144676)**

- \( p_{18S} = -0.874 \) / \( p_{18S} = 3.1 \times 10^{-4} \)
- \( p_{GAPDH} = -0.839 \) / \( p_{GAPDH} = 1.2 \times 10^{-3} \)

**SIDT1 (NM_017699)**

- \( p_{18S} = -0.629 \) / \( p_{18S} = 3.2 \times 10^{-2} \)
- \( p_{GAPDH} = -0.622 \) / \( p_{GAPDH} = 3.5 \times 10^{-2} \)

**TCFL5 (NM_006602)**

- \( p_{18S} = -0.902 \) / \( p_{18S} < 2.2 \times 10^{-16} \)
- \( p_{GAPDH} = -0.839 \) / \( p_{GAPDH} = 1.2 \times 10^{-3} \)

**FAM19A1 (NM_213609)**

- \( p_{18S} = -0.619 \) / \( p_{18S} = 1.2 \times 10^{-1} \)
- \( p_{GAPDH} = -0.881 \) / \( p_{GAPDH} = 7.2 \times 10^{-3} \)

**B**

- **normalized to 18S**
  - \( t(12;21) \) to 18S
  - others to 18S

- **normalized to GAPDH**
  - \( t(12;21) \) to GAPDH
  - others to GAPDH
Integration of high-resolution methylome and transcriptome analyses to dissect epigenomic changes in childhood acute lymphoblastic leukemia

Stephan Busche, Bing Ge, Ramon Vidal, et al.

Cancer Res  Published OnlineFirst May 30, 2013.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-4367

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2013/05/30/0008-5472.CAN-12-4367.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.