Molecular Analysis of Central Nervous System Disease Spectrum in Childhood Acute Lymphoblastic Leukemia

Chindo Hicks[1,2], Jitsuda Sitthi-Amorn[3], Jessica Douglas[3], Ritika Ramani[4], Lucio Miele[1], Vani Vijayakumar[5], Cynthia Karlson[3], James Chipeta[6] and Gail Megason[3]

[1]Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA, USA. [2]Department of Public Health Sciences, University of Lusaka, Lusaka, Zambia. [3]Children's Cancer Center, University of Mississippi Medical Center, Jackson, MS, USA. [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. [5]Department of Radiology, University of Mississippi Medical Center, Jackson, MS, USA. [6]Department of Pediatrics and Child Health, University of Zambia, Lusaka, Zambia.

ABSTRACT: Treatment of the central nervous system (CNS) is an essential therapeutic component in childhood acute lymphoblastic leukemia (ALL). The goal of this study was to identify molecular signatures distinguishing patients with CNS disease from those without the disease in pediatric patients with ALL. We analyzed gene expression data from 207 pediatric patients with ALL. Patients without CNS were classified as CNS1, while those with mild and advanced CNS disease were classified as CNS2 and CNS3, respectively. We compared gene expression levels among the three disease classes. We identified gene signatures distinguishing the three disease classes. Pathway analysis revealed molecular networks and biological pathways dysregulated in response to CNS disease involvement. The identified pathways included the *ILK, WNT,* B-cell receptor, *AMPK, ERK5*, and *JAK* signaling pathways. The results demonstrate that transcription profiling could be used to stratify patients to guide therapeutic decision-making in pediatric ALL.

KEYWORDS: acute lymphoblastic leukemia, central nervous disease spectrum, gene expression

CITATION: Hicks et al. Molecular Analysis of Central Nervous System Disease Spectrum in Childhood Acute Lymphoblastic Leukemia. *Clinical Medicine Insights: Oncology* 2016:10 5–15 doi: 10.4137/CMO.S18180.

TYPE: Original Research

RECEIVED: June 25, 2015. **RESUBMITTED:** January 04, 2016. **ACCEPTED FOR PUBLICATION:** January 07, 2016.

ACADEMIC EDITOR: William C. S. Cho, Editor in Chief

PEER REVIEW: Two peer reviewers contributed to the peer review report. Reviewers' reports totaled 437 words, excluding any confidential comments to the academic editor.

FUNDING: Louisiana State University Health Sciences Center's School of Medicine, Hyundai Hope on Wheels and the CHEW Foundation, the University of Mississippi Medical Center's Cancer Institute, and the Children's Cancer Center funded this project. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: chick3@lsuhsc.edu

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert single-blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

Introduction

Childhood acute lymphoblastic leukemia (ALL) is the most common malignancy in children.[1] It accounts for 80% of all newly diagnosed leukemia cases and affects ~3000 children in the United States.[2] With five-year event-free survival of >80% now, current efforts are focusing on optimal risk-directed treatment.[3–5] It is estimated that contemporary use of risk-directed effective systemic chemotherapy and central nervous system (CNS)-directed treatment protocols may further increase the five-year event-free survival rate that is currently approaching 90%.[3–5] However, this figure is deceptive for those children with unfavorable features as defined by biology or poor response to treatment and for whom the probability of cure is much lower, which is well below 50% in some groups.[3] This limited progress must be balanced against the recognition that patients with CNS disease account for 30%–40% of initial relapses in some clinical trials.[3] Although the use of CNS-directed treatment has led to 5%–10% reduction in CNS relapse in childhood ALL, relapse remains an important cause of morbidity and mortality, occurring in up to 6% of the patient population.[1] Moreover, the obvious therapeutic benefit of CNS-directed treatment using either cranial radiation, intrathecal methotrexate, or a combination thereof has been tempered by the recognition of long-term neurotoxicity, which has the potential to impair the quality of life in some patients.[3] Therefore, there is an urgent need for the discovery of molecular markers to stratify patients to guide treatment decisions. Precise assessment of CNS disease involvement is essential for current treatment planning to avoid overtreatment or undertreatment of pediatric patients with ALL.[5]

A major risk factor associated with an increased risk of CNS relapse is the presence of leukemic blast cells in the cerebrospinal fluid (CSF).[5] Leukemia patients have been considered to be at risk of CNS disease if blast cells are detected in their CSF.[5–7] Currently, patients are classified as CNS1, CNS2, and CNS3.[5–7] Under this classification system, CNS1 indicates no detectable blast cells in CSF, CNS2 indicates the presence of <5 leukocytes per μL with detectable blast cells in cytocentrifuged preparation of CSF, and CNS3 indicates the presence of overt CNS leukemia, in which is the sample contains >5 white blood cell (WBC) counts per μL with identifiable blast cells in CSF.[5–7] CNS classification determines the intensity of CNS-directed therapy, which may vary depending on the patient's

risk status. Although this classification scheme is routinely used in a clinical setting to stratify patients, the molecular mechanisms underpinning this classification scheme have not been well characterized. Understanding the molecular mechanisms underlying the CNS disease spectrum is essential for improving the survival rate of children who relapse and to avoid overtreatment and undertreatment of ALL patients.

Advances in transcription profiling using microarray technology have made possible molecular classification of ALL subtypes.[8–10] Our group recently used transcription profiling to stratify pediatric ALL patients according to marginal residue disease (MRD).[11] However, currently, there is a knowledge gap about the molecular mechanisms underpinning the CNS disease spectrum in pediatric ALL. The objective of this study was to identify molecular signatures distinguishing patients with CNS disease from those without CNS disease and the molecular networks and biological pathways dysregulated in response to CNS disease involvement. Our working hypothesis was that molecular perturbation significantly differs between patients diagnosed with CNS disease and those without CNS disease. We further hypothesized that differences in gene expression levels between patients with and without CNS disease affect molecular networks and biological pathways associated with ALL.

Materials and methods

Patient population and sources of gene expression data. We used publically available gene expression data generated from a total of 207 pediatric ALL patients diagnosed with and without CNS disease.[10] Patient samples for this study were obtained from the Children's Oncology Group (COG) Clinical Trial P9906. COG P9906 enrolled 272 eligible high-risk B-precursor ALL patients between March 15, 2000, and April 25, 2003. All patients were treated uniformly with a modified augmented Berlin-Frankfurt-Münster Study Group regimen.[10] This trial targeted a subset of newly diagnosed high-risk ALL patients who had experienced a poor outcome in prior studies.[10] Patients with CNS disease or testicular leukemia were eligible for the trial, regardless of age or WBC count at diagnosis. Patients with very high-risk features (*BCR-ABL1* or hypodiploidy) were excluded, whereas those with low-risk features (trisomies of chromosome 4 or 10; t(12;21) *ETV6-RUNX1*) were excluded unless they had CNS disease or testicular leukemia. We also used previously cryopreserved residual pretreatment leukemia specimens that were available for a representative cohort of 207 of the 272 (76%) registered patients. With the exception of differences in presenting WBC count, these 207 patients were highly similar in all other clinical and outcome parameters to all the 272 patients accrued to this trial.[10] Treatment protocols were approved by the National Cancer Institute and COG participating institutions through their institutional review boards.[10] Informed consent for clinical trial registration, sample submission, and participation in these research studies was obtained from all patients or their guardians in accordance with the Declaration of Helsinki.

For gene expression profiling, RNA was extracted and purified from 207 pretreatment diagnostic samples with >80% blasts (131 bone marrow and 76 peripheral blood) as previously described.[10] RNA was extracted using Trizol reagent, and RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent). cDNA was synthesized using a T-7 linked oligo-dT primer, and cRNA was then synthesized with biotinylated UTP and CTP. The labeled RNA was fragmented and hybridized to HG_U133A_Plus2.0 oligonucleotide microarrays (Affymetrix) after RNA quantification, cDNA preparation, and labeling according to Affymetrix protocols. Signals were scanned using the Affymetrix GeneChip Scanner. The data were processed and normalized using Affymetrix's Microarray Analysis Suite (MAS 5.0). All the data were normalized to log 2.

Data analysis. Consistent with the current treatment protocols,[5–7] we partitioned the original data set into three data sets representing the three disease classes: CNS1 ($N = 160$), CNS2 ($N = 16$), and CNS3 ($N = 20$). We analyzed the partitioned data by comparing the gene expression levels among the patient groups: CNS1 versus CNS2, CNS1 versus CNS3, and CNS2 versus CNS3. The significant differences in gene expression levels between patient groups were tested using a permutation *t*-test. We applied the false discovery rate (FDR) procedure to correct for multiple hypothesis testing.[12]

One of the goals of this study was to identify genes that are predictive of CNS disease. We used four inherent steps to this process: identification of genes distinguishing disease classes as outlined above, model selection, prediction assessment, and validation. Due to the small sample size in patients diagnosed with CNS disease, we did not partition the data into test and validation sets, as such an approach would lead to bias resulting from sampling errors. Therefore, instead, we used the leave-one-out cross-validation as our prediction and validation model procedure to identify genes with predictive power.[13] This approach has been successfully used in gene expression data analysis to mitigate and eliminate this bias.[13,14] For each analysis, the genes were ranked based on the *P*-values and the FDR.

To ensure that the results were comparable across the three analysis conducted, we used the threshold of $P < 0.01$ and an FDR of <1% to select the most significantly differentially expressed genes in each comparison. Differential expression analyses were performed using GenePattern[15] and Pomelo II software packages.[16] We performed network and pathway analysis and visualization using the Ingenuity Pathway Analysis (IPA) system.[17] Gene identifiers/symbols approved by the Human Genome Organization's Nomenclature Committee were mapped onto networks and canonical pathways using the network and pathway design and analysis modules as implemented in the IPA. The probability scores and the log *P*-values were calculated to assess the likelihood and reliability

of correctly assigning the genes to the correct networks and pathways, respectively. Functional relationships, biological processes, and cellular components in which the genes are involved were assessed using the Gene Ontology (GO) information as implemented in the IPA and Affymetrix databases.

Results

Differences in gene expression levels among patient groups. We compared gene expression levels among patients with and without CNS disease involvement. First, we compared gene expression levels between patients with CNS1 and CNS2. We hypothesized that gene expression levels significantly differ between the two patient groups. After correcting for multiple hypothesis testing, this analysis produced a signature of 279 highly significant ($P < 0.01$) differentially expressed genes. A signature of 30 most highly significant ($P < 0.0009$) differentially expressed genes distinguishing the two patient groups is listed in Table 1. The signature included the genes *ATXN1* and *GTF3C3* involved in DNA binding; *ZNF343, RNASE11, SMC5*, and *PDPK1* involved in nucleotide binding; *TCERG1L, VANGL1, FBXO31, RIMS3, TGS1, WHSC1L1, PRRC2A, AP3M1, GOLGA2, RSF1, CD14, MYH10*, and *LCAT* involved in protein binding; and *ADARB1, ADARB2, RSL1D1*, and *CUX1* involved in RNA binding. In addition, the signature included the genes *CMKLR1* involved in signal transduction, *MAPKBP1* involved in translation initiation factor activity, and *ALDH1L2* involved in catalytic activity. The signature also included the genes *C19ORF52, LOC645513*, and *SEL1L3* with unknown functions. A complete list of all highly significant differentially expressed genes between the two patient groups is presented in Supplementary Table 1.

Table 1. List of the top most highly significant differentially expressed genes distinguishing patients with CNS1 from patients with CNS2.

GENE SYMBOL	CHROM POSITION	GENE TITLE	*P*-VALUE
SMC5	9q21.11	Structural maintenance of chromosomes 5	4.50×10^{-5}
TCERG1L	10q26.3	Transcription elongation regulator 1-like	4.50×10^{-5}
RSL1D1	16p13.13	Ribosomal L1 domain containing 1	9.00×10^{-5}
ATXN1	6p23	Ataxin 1	1.20×10^{-4}
LCAT	16q22.1	Lecithin-cholesterol acyltransferase	1.30×10^{-4}
ADARB1	21q22.3	Adenosine deaminase, RNA-specific, B1	1.90×10^{-4}
MYH10	17p13	Myosin, heavy chain 10, non-muscle	2.20×10^{-4}
GTF3C3	2q33.1	General transcription factor IIIC, polypeptide 3, 102kDa	2.60×10^{-4}
AP3M1	10q22.1	Adaptor-related protein complex 3, mu 1 subunit	2.70×10^{-4}
ZNF343	20p13	Zinc finger protein 343	3.20×10^{-4}
LOC645513	4q26	Uncharacterized LOC645513	3.50×10^{-4}
WHSC1L1	8p11.2	Wolf-Hirschhorn syndrome candidate 1-like 1	3.60×10^{-4}
CD14	5q31.3	CD14 molecule	3.70×10^{-4}
CMKLR1	12q23.3	Chemokine-like receptor 1	3.80×10^{-4}
GOLGA2	9q34.13	Golgin A2	3.80×10^{-4}
RIMS3	1p34.2	Regulating synaptic membrane exocytosis 3	4.20×10^{-4}
C19orf52	19p13.2	Chromosome 19 open reading frame 52	4.30×10^{-4}
PHF17	4q26-q27	PHD finger protein 17	4.30×10^{-4}
FBXO31	16q24	F-box protein 31	4.70×10^{-4}
CUX1	7q22.1	Cut-like homeobox 1	4.70×10^{-4}
ADARB2	10p15.3	Adenosine deaminase, RNA-specific, B2 (non-functional)	5.20×10^{-4}
RSF1	11q14.1	Remodeling and spacing factor 1	5.20×10^{-4}
RNASE11	14q11.1	Ribonuclease, RNase A family, 11 (non-active)	5.20×10^{-4}
PRRC2A	6p21.3	Proline-rich coiled-coil 2A	5.40×10^{-4}
ALDH1L2	12q23.3	Aldehyde dehydrogenase 1 family, member L2	6.50×10^{-4}
SEL1L3	4p15.2	Sel-1 suppressor of lin-12-like 3 (C. elegans)	7.20×10^{-4}
MAPKBP1	15q15.1	Mitogen-activated protein kinase binding protein 1	7.70×10^{-4}
PDPK1	16p13.3	3-phosphoinositide dependent protein kinase-1	8.70×10^{-4}
VANGL1	1p13.1	VANGL planar cell polarity protein 1	9.10×10^{-4}
TGS1	8q11	Trimethylguanosine synthase 1	9.90×10^{-4}

Patients with CNS3 require intensified CNS-directed therapy. Therefore, the second step in our analysis focused on comparing gene expression levels between patients with CNS1 and patients with CNS3. This analysis revealed a signature of 389 highly significant ($P < 0.01$) differentially expressed genes after correcting for multiple hypothesis testing using the FDR procedure. A signature containing 33 most highly significant ($P < 0.0009$) differentially expressed genes distinguishing the two patient groups is listed in Table 2. The signature included the genes *ZNF549*, *ZNF260*, *GTF2H2*, *PFKP*, *PPIP5K1*, *ACTG1P4*, *LONP2*, *DHX33*, and *CAPRIN2* involved in nucleotide-binding activity. Many other genes involved in various molecular functions were also identified, including: *WFDC10B* involved in peptidase inhibitory activity; *ZDHHC8* with receptor activity; *NXN* involved in serine-type endopeptidase activity; *GRB10* SH3/SH2 involved in adaptor activity; *MCHR1* and *SH2B1* with signal transduction activity; *CLDN19* and *OCLN* with structural molecular and protein binding activity; *RORA* involved in transcription corepressor binding; *LTBP4* having transforming growth factor beta-activated receptor activity; *PURA* involved in translational repressor activity; and *MAGEL2* associated with ubiquitin-protein transferase activity. Other genes in the signature included *ARPC4* involved in actin binding; *ST3GAL3* associated with carbohydrate metabolism, *HSD17B6* and *MST1* with catalytic activity;

Table 2. List of the top most highly significant differentially expressed genes distinguishing patients with CNS1 from patients with CNS3.

GENE SYMBOL	CHROM POSITION	GENE TITLE	*P*-VALUES
LONP2	16q12.1	lon peptidase 2, peroxisomal	2.00×10^{-5}
ZNF549	19q13.43	Zinc finger protein 549	4.00×10^{-5}
PPIP5K1	15q15.3	Diphosphoinositol pentakisphosphate kinase 1	9.00×10^{-5}
ACTG1P4	1p21	Actin, gamma 1 pseudogene 4	1.50×10^{-4}
MYBBP1A	17p13.3	MYB binding protein (P160) 1a	2.10×10^{-4}
PHF2P1	13q11	PHD finger protein 2 pseudogene 1	2.20×10^{-4}
SH2B1	16p11.2	SH2B adaptor protein 1	2.20×10^{-4}
ST3GAL3	1p34.1	ST3 beta-galactoside alpha-2, 3-sialyltransferase 3	3.10×10^{-4}
TCEB3	1p36.1	Transcription elongation factor B (SIII), polypeptide 3	3.10×10^{-4}
PURA	5q31	Purine-rich element binding protein A	3.40×10^{-4}
TM7SF3	12q11-q12	Transmembrane 7 superfamily member 3	3.70×10^{-4}
CAPRIN2	12p11	Caprin family member 2	4.10×10^{-4}
OCLN	5q13.1	Occludin	4.20×10^{-4}
GRB10	7p12.2	Growth factor receptor-bound protein 10	4.30×10^{-4}
ZNF260	19q13.12	Zinc finger protein 260	4.60×10^{-4}
ATXN7L1	7q22.1	Ataxin 7-like 1	5.30×10^{-4}
PNPLA7	9q34.3	Patatin-like phospholipase domain containing 7	5.90×10^{-4}
DHX33	17p13	DEAH (Asp-Glu-Ala-His) box polypeptide 33	6.10×10^{-4}
LTBP4	19q13.1-q13.2	Latent transforming growth factor beta binding protein 4	6.50×10^{-4}
HSD17B6	12q13.3	Hydroxysteroid (17-beta) dehydrogenase 6	6.50×10^{-4}
MCHR1	22q13.3	Melanin-concentrating hormone receptor 1	6.70×10^{-4}
MST1	3p21	Macrophage stimulating 1	7.00×10^{-4}
RORA	15q21-q22	RAR-related orphan receptor A	7.20×10^{-4}
MAGEL2	15q11-q12	MAGE-like 2	7.70×10^{-4}
WFDC10B	20q13.12	WAP four-disulfide core domain 10B	7.80×10^{-4}
CLDN19	1p34.2	Claudin 19	7.80×10^{-4}
VWA8-AS1	13q14.11	VWA8 antisense RNA 1 (head to head)	8.40×10^{-4}
SNHG3	1p35.3	Small nucleolar RNA host gene 3	8.50×10^{-4}
PFKP	10p15.3-p15.2	Phosphofructokinase, platelet	9.00×10^{-4}
TTLL3	3p25.3	Tubulin tyrosine ligase-like family, member 3	9.10×10^{-4}
GTF2H2	5q13.2	General transcription factor IIH, polypeptide 2, 44kDa	9.20×10^{-4}
NXN	17p13	Nucleoredoxin	9.60×10^{-4}
ZDHHC8	22q11.21	Zinc finger, DHHC-type containing 8	9.60×10^{-4}

SNHG3, MYBBP1A, and *TCEB3* involved in DNA binding; *PNPLA7* associated with hydrolase activity; and *TTLL3* with ligase activity. The signature included four genes *ATXN7L1, PHF2P1, TM7SF3,* and *VWA8-AS1* with unknown functions. A complete list of the highly significant ($P < 0.01$) differentially expressed genes between the two patient groups is presented in Supplementary Table 2. Interestingly, there was little overlap between the genes distinguishing patients with CNS1 from patients with CNS2 and genes distinguishing patients with CNS1 from patients with CNS3 disease.

One of the challenges faced by clinicians is stratifying patients with CNS disease into a group that may require standard treatment and a group that may require intensified CNS-directed therapy. This is complicated by the fact that CNS2 follows a variable clinical course with some patients behaving like those with CNS1, while others behave like those with CNS3. To identify a signature that may be predictive of CNS3, we compared gene expression levels between patients with CNS2 and patients with CNS3. We hypothesized that gene expression levels in patients with CNS2 significantly differ from patients with CNS3. After correcting for multiple hypothesis testing, this analysis produced a signature of 418 highly significant ($P < 0.01$) differentially expressed genes. A signature containing the 40 most highly significant ($P < 0.0009$) differentially expressed genes distinguishing the two patient groups is listed in Table 3. The signatures included the genes *EZH1, MYBBP1A, LYL1, AGFG1, HUWE1*, and *ZNF549 ERCC1* involved in DNA binding; *R3HCC1, ACTG1P4, LONP2*, and *PIM2* involved in nucleotide binding; and *CNOT1, CMIP, PRRC2A, OCLN*, and *RSF1* involved in protein binding. The signature also included other genes with a wide range of molecular functions, including *LILRA6* involved in receptor activity, *MCHR1* and *SH2B1* involved in signal transduction activity, *LMAN1L* with neurotransmitter transport activity, *EFEMP2* involved in transmembrane signaling receptor activity, *MAGEL2* associated with ubiquitin-protein transferase activity, *COL24A1* involved in extracellular matrix structural constituent, *FOXO3* associated with core promoter binding, and *EYA3* involved in chromatin binding. The signature also contained a number of genes, including *BAALC, C11orf21, C19orf52, CCDC85C, KIAA1211L, LOC100130370, LOC100288310, LOC100506114, SEL1L3, SOCS2-AS1*, and *TAMM41*, whose molecular functions are unknown. A complete list of all the highly significant ($P < 0.01$) differentially expressed genes between the two patient groups is presented in Supplementary Table 3. Intriguingly, there was little overlap in the genes among the three signatures identified. This demonstrates that molecular classification could be used to stratify pediatric ALL patients diagnosed with CNS disease.

To identify genes uniquely distinguishing patient groups and those which overlap, we used a Venn diagram. The results showing genes uniquely differentially expressed and genes that overlap among the patient groups are shown in Figure 1. Comparison of CNS1 versus CNS2 revealed 229 uniquely differentially expressed genes. Comparison of CNS1 versus CNS3 revealed 254 genes, whereas comparison of CNS2 versus CNS3 revealed 251 genes (Fig. 1). A total of eight genes were significant in both CNS1 versus CNS2 and CNS1 versus CNS3. Another 40 genes were significant in both CNS1 versus CNS2 and CNS2 versus CNS3. The largest overlap involving a signature of 125 genes was found in the comparisons of CNS1 versus CNS3 and CNS2 versus CNS3. This was expected because CNS2 follows a variable clinical course in which some patients tend to behave like those having CNS1 and others like those having CNS3. Overall, only two genes were found to be significantly differentially expressed among all the three comparisons, demonstrating that transcription profiling could be used to stratify patients to guide treatment decisions.

Functional relationships among identified genes. To determine whether the identified genes are functionally related, we performed GO analysis for each set of significantly differentially expressed genes discovered in each of the three comparisons. GO analysis allows characterization of genes according to the GO nomenclature. The GO Consortium has developed three separate ontologies, namely, biological process, molecular function, and cellular component, to describe the attributes of the gene products. Biological process describes the contribution of a gene product to a biological objective, molecular function defines what a gene product does at the biochemical level without specifying where or when the event actually occurs, and cellular component refers to where in the cell a gene product functions. Here, we characterized the genes according to all three GO categories. Because many genes are involved in multiple biological processes, have multiple overlapping molcular functions, and are involved in many cellular components, we have listed these results in Supplementary Tables 4–6 to accommodate all the three pieces of information. The results showing the biological processes, molecular functions, and cellular components in which the genes distinguishing patients with CNS1 from CNS2 are involved are presented in Supplementary Table 4. Results for the genes distinguishing patients with CNS1 from CNS3 and CNS2 from CNS3 are presented in Supplementary Tables 5 and 6, respectively. In each case, we identified sets of functionally related genes with multiple overlapping molecular functions, suggesting that the genes act in concert rather than individually to drive the CNS phenotype.

Network and pathway analysis. To gain insights about the broader biological context in which the genes identified operate, we performed network and pathway analysis separately for each comparison. The rationale for separate analysis is that different treatments are used in treating patients with CNS disease involvement. For example, while standard treatment is used for patients with CNS2, intensified CNS-directed treatment is used for patients with CNS3, suggesting that different pathways may be targeted by each treatment regimen. Thus, we sought to identify molecular networks and biological pathways dysregulated in response to CNS involvement in different

Table 3. List of the top most highly significant differentially expressed genes distinguishing patients with CNS2 from patients with CNS3.

GENE SYMBOL	CHR POSITION	GENE TITLE	*P*-VALUE
SEL1L3	4p15.2	Sel-1 suppressor of lin-12-like 3 (*C. elegans*)	1.00×10^{-5}
LONP2	16q12.1	Lon peptidase 2, peroxisomal	1.00×10^{-5}
C19orf52	19p13.2	Chromosome 19 open reading frame 52	4.50×10^{-5}
ZNF549	19q13.43	Zinc finger protein 549	1.00×10^{-4}
MCHR1	22q13.3	Melanin-concentrating hormone receptor 1	1.15×10^{-4}
LOC100130370	17q25.3	Uncharacterized LOC100130370	1.20×10^{-4}
ERCC1	19q13.32	Excision repair cross-complementing rodent repair deficiency, complementation group 1	1.45×10^{-4}
PIM2	Xp11.23	Pim-2 oncogene	1.60×10^{-4}
C11orf21	11p15.5	Chromosome 11 open reading frame 21	1.80×10^{-4}
PRRC2A	6p21.3	Proline-rich coiled-coil 2A	1.90×10^{-4}
LOC100288310	8q13.3	Uncharacterized LOC100288310	2.15×10^{-4}
FLJ10489	8q22.3	Uncharacterized protein FLJ10489	2.45×10^{-4}
MYBBP1A	17p13.3	MYB binding protein (P160) 1a	2.75×10^{-4}
ACTG1P4	1p21	Actin, gamma 1 pseudogene 4	3.20×10^{-4}
COL24A1	1p22.3-p22.2	Collagen, type XXIV, alpha 1	3.20×10^{-4}
EYA3	1p36	Eyes absent homolog 3 (Drosophila)	3.35×10^{-4}
KIAA1211L	2q11.2	KIAA1211-like	4.25×10^{-4}
EZH1	17q21.1-q21.3	Enhancer of zeste homolog 1 (Drosophila)	4.55×10^{-4}
RAI2	Xp22	Retinoic acid induced 2	5.05×10^{-4}
HUWE1	Xp11.22	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	5.34×10^{-4}
LILRA6	19q13.4	Leukocyte immunoglobulin-like receptor, subfamily A, member 6	5.55×10^{-4}
BAALC	8q22.3	Brain and acute leukemia, cytoplasmic	5.56×10^{-4}
FOXO3	6q21	Forkhead box O3	5.75×10^{-4}
EFEMP2	11q13	EGF containing fibulin-like extracellular matrix protein 2	5.89×10^{-4}
CMIP	16q23	c-Maf-inducing protein	5.89×10^{-4}
LMAN1L	15q24.1	Lectin, mannose-binding, 1 like	6.25×10^{-4}
RSF1	11q14.1	Remodeling and spacing factor 1	6.35×10^{-4}
MAGEL2	15q11-q12	MAGE-like 2	6.75×10^{-4}
SOCS2-AS1	12q22	SOCS2 antisense RNA 1	6.85×10^{-4}
OCLN	5q13.1	Occludin	7.29×10^{-4}
LOC158863	Xq13.1	Uncharacterized LOC158863	7.79×10^{-4}
LYL1	19p13.2	Lymphoblastic leukemia derived sequence 1	7.99×10^{-4}
LOC100506114	19q13.33	Uncharacterized LOC100506114	8.05×10^{-4}
CNOT1	16q21	CCR4-NOT transcription complex, subunit 1	8.15×10^{-4}
R3HCC1	8p21.3	R3H domain and coiled-coil containing 1	8.25×10^{-4}
SH2B1	16p11.2	SH2B adaptor protein 1	8.45×10^{-4}
AGFG1	2q36	ArfGAP with FG repeats 1	8.60×10^{-4}
SLC25A24	1p13.2	Solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 24	9.50×10^{-4}
TAMM41	3p25.2	TAM41, mitochondrial translocator assembly and maintenance protein, homolog (*S. cerevisiae*)	9.55×10^{-4}
CCDC85C	14q32.31	Coiled-coil domain containing 85C	9.60×10^{-4}

patient groups as potential therapeutic targets. The results of network and pathway analysis based on significantly differently expressed genes between patients with CNS1 and those with CNS2 are shown in Figure 2.

Network analysis revealed genes with multiple overlapping functions. Among the genes responsive to CNS involvement found in the network included the genes involved in gene expression, cancer, cellular development, cell morphology,

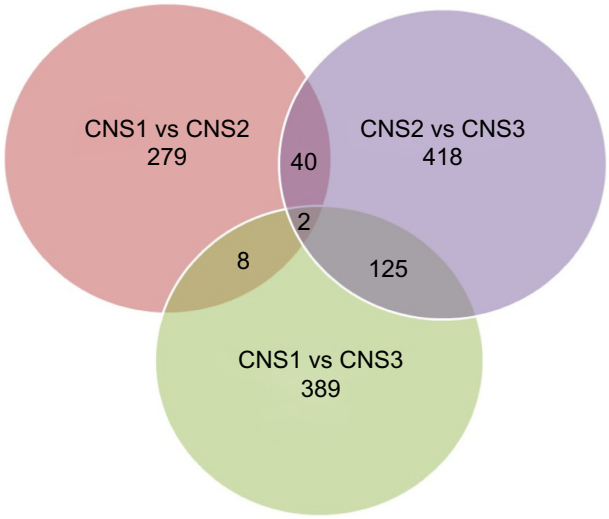

Figure 1. Distribution of significantly differentially expressed genes based on comparison of gene expression levels between the three patient groups. Each quadrant of the Venn diagram represents comparison between the two disease classes and the total number of highly significant ($P < 0.01$) differentially expressed genes resulting from that comparison after controlling for the FDR of <1%. The numbers in the intersections represent the number of genes that were significantly expressed in either two or all three comparisons.

cell-to-cell signaling and interaction, nervous system development and function, and cell cycle. Also identified were genes involved in hematological system development and function, hematopoiesis, cell morphology, cardiovascular system development and function, organ morphology, carbohydrate metabolism, lipid metabolism, and posttranslational modification.

The analysis revealed biological pathways dysregulated by CNS involvement. The most highly significant biological pathways included the ILK ($P = 3.1 \times 10^{-5}$), Wnt/beta-catenin ($P = 7.6 \times 10^{-5}$), ERBB2-ERBB3 ($P = 8.9 \times 10^{-5}$), DHA ($P = 1.3 \times 10^{-4}$), and the B-cell receptor ($P = 2.2 \times 10^{-4}$) signaling pathways. Additionally, the analysis revealed highly significant ($P < 1.0 \times 10^{-4}$) upstream regulators, which included *HTT* ($P = 2.5 \times 10^{-5}$), *TP53* ($P = 1.0 \times 10^{-4}$), *XBP1* ($P = 1.2 \times 10^{-4}$), and *KLC1* ($P = 1.3 \times 10^{-4}$), suggesting that the identified genes may be regulated by other genes.

The results of network analysis using the set of genes identified from a comparison of expression levels between patients with CNS1 and CNS3 are shown in Figure 3. The analysis revealed genes with multiple overlapping functions, which included genes involved in DNA replication, recombination and repair, molecular transport, cellular assembly, and organization. In addition, we identified genes involved in endocrine system development and function, carbohydrate and lipid metabolism, gene expression, cell-to-cell signaling and interactions, amino acid metabolism, cell cycle, and small molecule biochemistry. The analysis revealed biological pathways dysregulated in response to CNS3 disease involvement, including the glycosaminoglycan-protein linkage region biosynthesis pathway ($P = 2.3 \times 10^{-4}$), assembly of RNA polymerase III complex ($P = 1.1 \times 10^{-3}$), heme biosynthesis from uroporphyrinogen-III I pathway ($P = 1.6 \times 10^{-3}$), AMPK signaling pathway ($P = 1.8 \times 10^{-3}$), and heme biosynthesis II pathway ($P = 1.1 \times 10^{-2}$). The analysis also revealed upstream regulators, including *BRD4* ($P = 2.7 \times 10^{-5}$), *FOLR1* ($P = 8.1 \times 10^{-4}$), *TFDP2* ($P = 2.5 \times 10^{-3}$), *SLC2A1* ($P = 3.7 \times 10^{-3}$), and *TFAM* ($P = 3.8 \times 10^{-3}$).

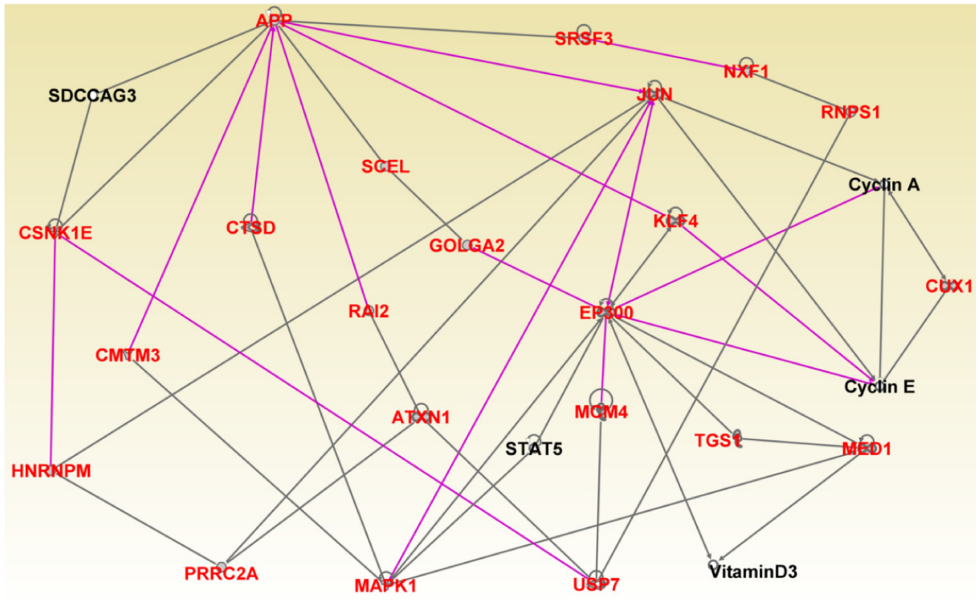

Figure 2. Molecular networks of highly significant differentially expressed genes between patients classified as CNS1 and CNS2 and other functionally related genes.
Note: Differentially expressed genes are shown in red color font and other genes in black font.

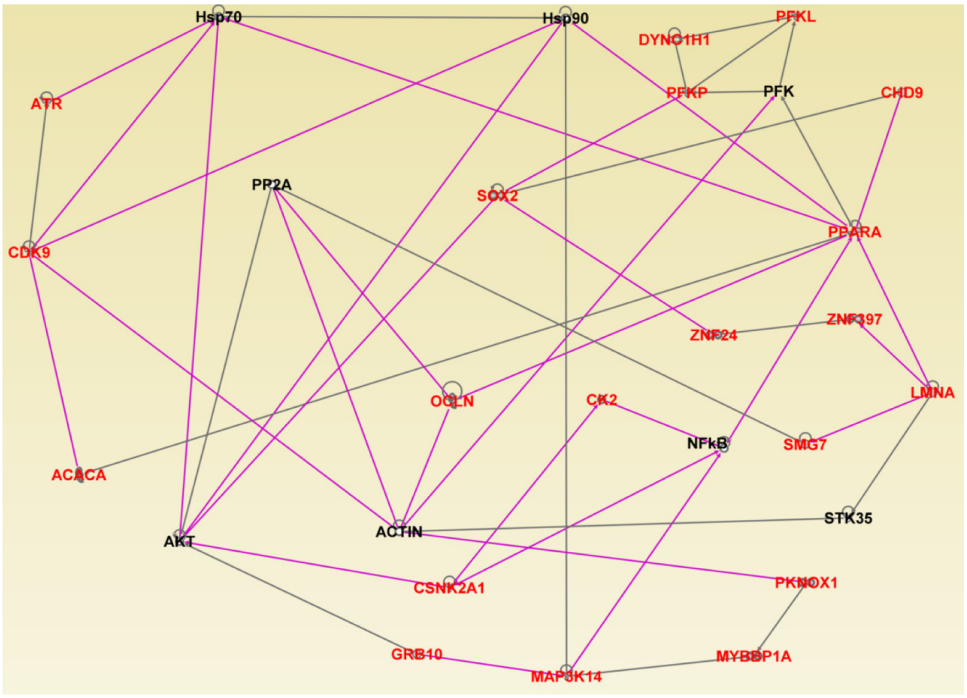

Figure 3. Molecular networks of highly significant differentially expressed genes between patients classified as CNS1 and CNS3 and other functionally related genes.
Note: Differentially expressed genes are shown in red color font and other genes in black font.

The results of network analysis using the set of genes distinguishing CNS2 from CNS3 are shown in Figure 4. This analysis revealed molecular networks containing genes involved in gene expression, cell cycle, embryonic development, hereditary disorder, cardiovascular disease, cell death and survival, cellular development, hematological system development and function, cell morphology, inflammatory response, humoral immune response, protein synthesis, and cancer.

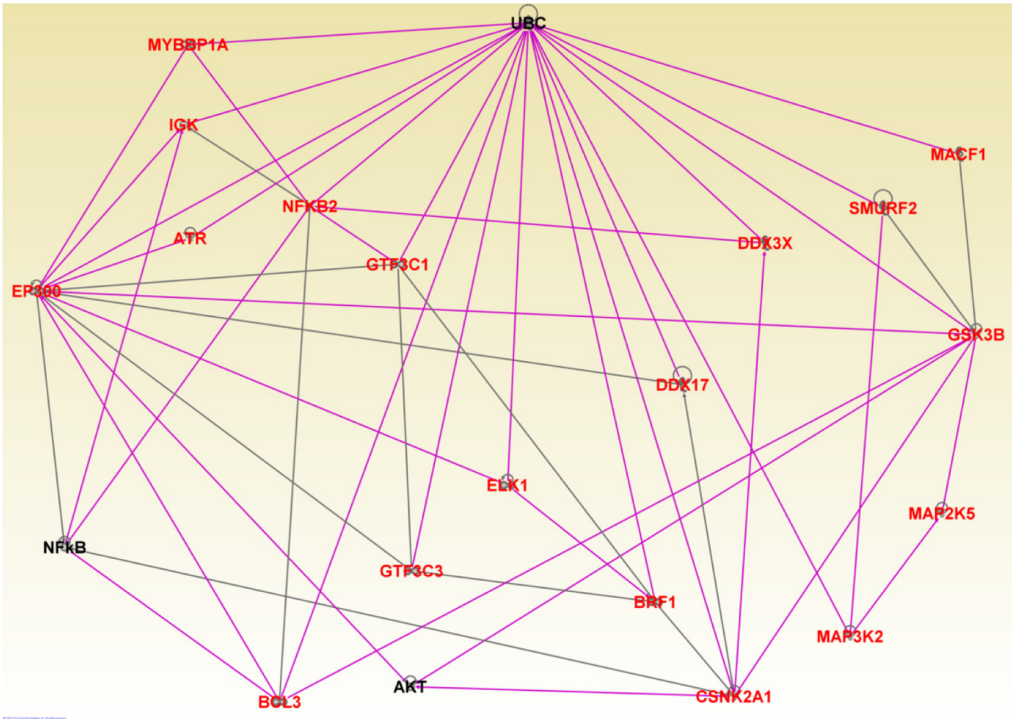

Figure 4. Molecular networks of highly significant differentially expressed genes between patients classified as CNS2 and CNS3 and other functionally related genes.
Note: Differentially expressed genes are shown in red color font and other genes in black font.

Many biological pathways of clinical significance were also identified, including the assembly of RNA polymerase III complex ($P = 1.5 \times 10^{-3}$), dendritic cell maturation (1.7×10^{-3}), ERK5 ($P = 6.4 \times 10^{-3}$), B-cell receptor ($P = 9.4 \times 10^{-3}$), and the role of JAK1, JAK2, and TYK2 ($P = 9.4 \times 10^{-3}$) in interferon signaling pathways. Among the identified pathways were the AKT, NF-κB, and UBC pathways, which have been previously implicated in pediatric ALL. The analysis also revealed upstream regulators, including *NLRP12* ($P = 1.9 \times 10^{-4}$), *DARC* ($P = 2.6 \times 10^{-4}$), *RAET1A* ($P = 2.8 \times 10^{-4}$), *RAET1D* ($P = 2.8 \times 10^{-4}$), and *PSMB9* ($P = 5.3 \times 10^{-4}$) genes. In all, differential expression and network and pathway analysis revealed that genetic alteration in transcript levels leads to measurable changes distinguishing patients with CNS disease from those without the disease. Furthermore, the results revealed that genomic alterations affect the entire molecular networks and biological pathways. This is a significant finding in that such pathways could serve as targets for the development of novel CNS-directed therapeutics as discussed in the next section.

Discussion and clinical implications

This study was undertaken to identify the molecular signatures distinguishing patients without CNS from those with CNS disease and to discover molecular networks and biological pathways dysregulated in response to CNS involvement. In this section, we discuss the potential clinical implication of the results.

The differences in gene expression levels between patients with and without CNS disease demonstrates that transcription profiling could be used to stratify ALL patients to guide treatment decisions. Similarly, the differences in gene expression levels between patients with CNS2 and CNS3 show that patients with CNS involvement could be substratified. This is a significant finding in that patients with CNS disease could be substratified into those requiring standard treatment and those requiring intensified CNS-directed treatment to avoid overtreatment and undertreatment of ALL patients. In practice, patients with CNS3 require intensified CNS-directed therapy.[18–20] This has significant clinical implications since a significant proportion of ALL patients are diagnosed with CNS2.[18,19]

Under the current treatment protocols, predicting whether a newly diagnosed patient with CNS2 will have a clinical course that resembles patients with CNS1 or patients with CNS3 remains challenging. Transcription profiling as demonstrated in this study has the promise to identify subsets of patients with good and bad prognosis and the molecular trajectories that are predictive of clinical course. The importance of discovering potential clinically actionable biomarkers for the CNS disease spectrum using transcription profiling as demonstrated in this study is further amplified by accumulating evidence that cranial radiation currently used for treating patients with CNS disease is a source of long-term mortality and morbidity, as well as long-term endocrine and neurological sequelae.[21]

The discovery of biological pathways dysregulated in response to CNS disease involvement is of particular interest. In this study, the WNT, WNT/beta-catenin, and B-cell receptor signaling pathways were found to be dysregulated in response to CNS2 disease involvement. Activation of WNT signaling in ALL cells is involved in marrow stem cell-mediated drug resistance and improved chemosensitivity in pediatric ALL.[22–24] In addition, activation of the WNT/beta-catenin pathway has been shown to mediate cell growth and survival in B-cell progenitor ALL.[25] Therefore, targeting either the WNT signaling pathway or the WNT/beta-catenin signaling pathway may be an innovative approach to the treatment of pediatric ALL. The discovery of the B-cell receptor signaling pathway in this study is consistent with other studies.[26] Given that the majority of B-cell malignancies are dependent on B-cell receptor signaling,[27] and since all ALL typically originates from pro- and pre-B-cells during B-cell development,[28] targeting the B-cell receptor signaling pathway may have therapeutic benefits. Additionally, since pre-B-cell receptor signaling has been shown to jam the WNT/beta-catenin pathway and induce cell death in B-cell ALL cell lines,[29] targeting both pathways may be an effective approach.

The discovery of the AKT, JAK, and NF-κB signaling pathways in response to CNS3 involvement is of great clinical importance. Since patients with CNS3 require intensified CNS-directed therapy, these pathways could be targeted. For example, the phosphatidylinositol 3-kinase (PI3K), AKT, mammalian target of rapamycin (mTOR) signaling pathway (PI3K/AKT/mTOR signaling pathway) is abnormally activated in childhood ALL.[30] This abnormal activation occurs as a consequence of constitutive activation of AKT,[30] providing a compelling rationale for targeting this pathway. The AKT regulates apoptosis by inhibition of Fas ligand, BCL2-associated death promoter, BCL2-interacting mediator of cell death, BCL2-associated X-protein (BAX), or degradation of P53.[30] This reinforces the potential use of this pathway as a therapeutic target. The NF-κB discovered in this study has clinical significance, because apoptosis resistance to ionization radiation in pediatric B-precursor ALL frequently involves increased NF-κB.[31] Due to lack of clinical information, we did not correlate clinical variables with the discovered pathways. However, it has been shown in the literature that high activation of NF-κB is associated with elevated WBC count and may be responsible for treatment failure in childhood ALL.[32] Taken together, these results suggest that the identified pathways could serve as potential therapeutic targets.

There is very little information in the published literature on the use of transcription profiling to stratify patients with CNS disease spectrum in high-risk pediatric patients with ALL. Cario et al.[33] reported the results of gene expression characterizing ALL with CNS involvement focusing

on overexpression of IL–15. Their study evaluated leukemic gene expression profiles from the bone marrow of 17 CNS-positive and 26 CNS-negative patients who were matched for risk factors associated with CNS disease involvement. The main differences between our study and theirs are that our study examined the entire CNS disease spectrum and discovered pathways dysregulated in response to CNS disease involvement. These issues were not fully addressed in their study. Moreover, our study used a much larger sample size compared to theirs.

This study provides insights into the molecular mechanisms underlying the CNS disease spectrum. However, limitations of the study must be acknowledged. As noted earlier, the sample sizes for CNS2 and CNS3 were small compared to CNS1. We undertook precautionary measures by applying the appropriate methods in data analysis to reduce biases that could arise from sampling errors.[13,14] We did not consider racial and subtype-specific differences. We have previously shown in pediatric ALL that gene expression levels vary and differ between ethnic populations.[34] We have also previously shown that gene expression levels and patterns in patients with MRD can be subtype specific.[11] In this study, we did not examine these variables for two reasons. First, information on these variables was not available, and second, samples sizes for patients with CNS2 and CNS3 were too small to permit substratification of the data. Given these limitations that we readily acknowledge, the findings in this study cannot be generalized to different racial or ethnic groups or to specific subtypes of ALL. In addition, we did not investigate the relationship between CNS disease spectrum and other prognostic factors such as MRD and testicular or WBC count because these variables were not available in the data set used. Due to these limitations, we view this study as exploratory. Further studies that correlate CNS disease with other clinical prognostic variables are recommended.

Conclusions

This study demonstrates that transcription profiling could be used to stratify pediatric patients with and without CNS disease to guide treatment decisions, identify potential clinically actionable biomarkers, and target biological pathways for the development of novel therapeutics.

Author Contributions

Conceived and designed the experiments: CH, GM, JS-A, JD, CK, LM, JC, VV. Analyzed the data: CH, RR, JS-A, JD. Wrote the first draft of the manuscript: AUTHORINITIALS. Contributed to the writing of the manuscript: CH, GM, JC, CK, LM, VV, JS-A. Agree with manuscript results and conclusions: AUTHORINITIALS. Jointly developed the structure and arguments for the paper: AUTHORINITIALS. Made critical revisions and approved final version: AUTHORINITIALS. All authors reviewed and approved of the final manuscript.

Supplementary Materials

Supplementary Table 1. List of significantly differentially expressed genes distinguishing CNS1 from CNS2, estimated *P*-values and FDR.

Supplementary Table 2. List of significantly differentially expressed genes distinguishing CNS1 from CNS3, estimated *P*-values and FDR.

Supplementary Table 3. List of significantly differentially expressed genes distinguishing CNS2 from CNS3, estimated *P*-values and FDR.

Supplementary Table 4. List of significantly differentially expressed genes distinguishing CNS1 from CNS2 and information on the biological processes, molecular functions, and cellular components in which the identified genes are involved.

Supplementary Table 5. List of significantly differentially expressed genes distinguishing CNS1 from CNS3 and information on the biological processes, molecular functions, and cellular components in which the identified genes are involved.

Supplementary Table 6. List of significantly differentially expressed genes distinguishing CNS2 from CNS3 and information on the biological processes, molecular functions, and cellular components in which the identified genes are involved.

REFERENCES

1. Mullighan CG. Molecular genetics of B-precursor acute lymphoblastic leukemia. *J Clin Invest*. 2012;122(10):3407–15.
2. Howlader N, Noone AM, Krapcho M, et al, eds. Childhood cancer. In: Howlader N, Noone AM, Krapcho M, et al, eds. *SEER Cancer Statistics Review,* 1975–2010. Bethesda, MD: National Cancer Institute; 2013:28. [based on November 2012 SEER data submission, posted to the SEER web site].
3. Rabin KR. Attacking remaining challenges in acute leukemia. *N Engl J Med*. 2012;366:1445–6.
4. Pui CH, Campana D, Pei D, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med*. 2009;360:2730–41.
5. Pui CH, Howard SC. Current management and challenges of malignant disease in the CNS in pediatric leukemia. *Lancet Oncol*. 2008;9:257–68.
6. Ranta S, Nilsson F, Harila-Saari A, et al. Detection of central nervous system involvement in childhood acute lymphoblastic leukemia by cytomorphology and flow cytometry of the cerebrospinal fluid. *Pediatr Blood Cancer*. 2015;62(6):951–6.
7. Smith M, Arthur D, Camitta B, et al. Uniform approach to risk classification and treatment assignment for children with acute lymphoblastic leukemia. *J Clin Oncol*. 1996;14:18–24.
8. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002;1:133–43.
9. Ross ME, Zhou X, Song G, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood*. 2003;102:2951–9.
10. Harvey RC, Mullighan CG, Wang X, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010;116:4874–84.
11. Sitthi-Amorn J, Herrington B, Megason G, et al. Transcriptome analysis of minimal residual disease in subtypes of pediatric B cell acute lymphoblastic leukemia. *Clin Med Insights Oncol*. 2015;9:51–60.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57:289–300.
13. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*. 2005;21(15):3301–7.
14. Parker BJ, Günter S, Bedo J. Stratification bias in low signal microarray studies. *BMC Bioinformatics*. 2007;8:326.

15. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38:500–1.
16. Morrissey ER, Diaz-Uriarte R. Pomelo II: finding differentially expressed genes. *Nucleic Acids Res*. 2009;37(suppl 2):W581–6.
17. Ingenuity Pathway Analysis (IPA). Qiagen Inc. http://www.ingenuity.com/products/ipa
18. Laningham FH, Kun LE, Reddick WE, Ogg RJ, Morris EB, Pui CH. Childhood central nervous system leukemia: historical perspectives, current therapy, and acute neurological sequelae. *Neuroradiology*. 2007;49(11):873–88.
19. Mahmoud HH, Rivera GK, Hancock ML, et al. Low leukocyte counts with blast cells in cerebrospinal fluid of children with newly diagnosed acute lymphoblastic leukemia. *N Engl J Med*. 1993;329(5):314–9.
20. Bürger B, Zimmermann M, Mann G, et al. Diagnostic cerebrospinal fluid examination in children with acute lymphoblastic leukemia: significance of low leukocyte counts with blasts or traumatic lumbar puncture. *J Clin Oncol*. 2003;21(2):184–8.
21. Abbott BL, Rubnitz JE, Tong X, et al. Clinical significance of central nervous system involvement at diagnosis of pediatric acute myeloid leukemia: a single institution's experience. *Leukemia*. 2003;17(11):2090–6.
22. Yang Y, Mallampati S, Sun B, et al. Wnt pathway contributes to the protection by bone marrow stromal cells of acute lymphoblastic leukemia cells and is a potential therapeutic target. *Cancer Lett*. 2013;333(1):9–17.
23. Dandekar S, Romanos-Sirakis E, Pais F, et al. Wnt inhibition leads to improved chemosensitivity in paediatric acute lymphoblastic leukaemia. *Br J Haematol*. 2014;167(1):87–99.
24. Wickström M, Dyberg C, Milosevic J, et al. Wnt/β-catenin pathway regulates MGMT gene expression in cancer and inhibition of Wnt signalling prevents chemoresistance. *Nat Commun*. 2015;6:8904.
25. Khan NI, Bradstock KF, Bendall LJ. Activation of Wnt/beta-catenin pathway mediates growth and survival in B-cell progenitor acute lymphoblastic leukaemia. *Br J Haematol*. 2007;138(3):338–48.
26. Chaiboonchoe A, Samarasinghe S, Kulasiri D, Salehi-Ashtiani K. Integrated analysis of gene network in childhood leukemia from microarray and pathway database. *Biomed Res Int*. 2014;2014:278748.
27. Buchner M, Müschen M. Targeting the B-cell receptor signaling pathway in B lymphoid malignancies. *Curr Opin Hematol*. 2014;21(4):341–9.
28. Müschen M. Rationale for targeting the pre-B-cell receptor signaling pathway in acute lymphoblastic leukemia. *Blood*. 2015;125(24):3688–93.
29. Saba NS, Angelova M, Lobelle-Rich PA, Levy LS . Disruption of pre-B-cell receptor signaling jams the WNT/β-catenin pathway and induces cell death in B-cell acute lymphoblastic leukemia cell lines. *Leuk Res*. 2015; 39(11):1220–8.
30. Barrett D, Brown V, Grupp SA, Teachey DT. Targeting the PI3K/AKT/mTOR signaling axis in children with hematologic malignancies. *Paediatr Drugs*. 2012;14(5):299–316.
31. Weston VJ, Austen B, Wei W, et al. Apoptotic resistance to ionizing radiation in pediatric B-precursor acute lymphoblastic leukemia frequently involves increased NF-kappaB survival pathway signaling. *Blood*. 2004;104(5):1465–73.
32. Kamieńska E, Ociepa T, Wysocki M, Kurylak A. Activation of NF-κB in leukemic cells in response to initial prednisone therapy in children with acute lymphoblastic leukaemia: relation to other prognostic factors. *Pol J Pathol*. 2011;62(1):5–11.
33. Cario G, Izraeli S, Teichert A, et al. High interleukin-15 expression characterizes childhood acute lymphoblastic leukemia with involvement of the CNS. *J Clin Oncol*. 2007;25:4813–20.
34. Hicks C, Miele L, Koganti T, et al. Analysis of patterns of gene expression variation within and between ethnic populations in pediatric B-ALL. *Cancer Inform*. 2013;12:155–73.