Genome-wide Studies in Childhood Acute Lymphoblastic Leukemia

Deepa Bhojwani, MD1 and Ching-Hon Pui, MD2

1. Assistant Professor; 2. Professor of Pediatrics, University of Tennessee Health Sciences Center

Abstract

The completion of the human genome project and additional ongoing efforts such as the international haplotype map (HapMap) project have promoted the exponential development of tools to investigate global genomic profiles at the DNA and transcript levels. Rapidly evolving technology now allows the interrogation of 1.8 million genetic markers (including almost one million single-nucleotide polymorphisms [SNPs]) on a single-array platform to study genome-wide variations in copy number and loss of heterozygosity. Likewise, more than 50,000 gene transcripts can be probed on a single microarray chip. This article highlights the use of these tools in childhood acute lymphoblastic leukemia (ALL) in order to better understand the mechanisms of leukemogenesis, develop new prognostic markers, and identify potential therapeutic targets. The integration of genome-wide studies of structural changes in DNA, expression levels of transcripts, microRNA profiles, epigenetics (including DNA methylation and histone acetylation), and re-sequencing efforts can provide further insight to enhance the management of this common childhood malignancy.

Keywords

Acute lymphoblastic leukemia (ALL), genomics

Prognostication

Risk-adapted therapy for ALL is the norm in developed countries, where therapy is tailored according to the predicted risk for relapse.1 Between one and 10 years of age, a leukocyte count <50,000/µl at presentation and genetic features of leukemic cells such as hyperdiploidy (>50 chromosomes and t(12;21) [ETV6-RUNX1] are favorable prognostic indicators, whereas hypodiploidy, t(9;22) [BCR-ABL] and MLL gene re-arrangements are unfavorable features.1 Also, the rapidity of blast clearance during remission-induction treatment as measured by minimal residual disease (MRD) with flow cytometry or polymerase chain reaction is a sensitive prognostic indicator. However, many relapses still occur in patients without known adverse characteristics at presentation, and efforts are focused on identifying additional attributes in both the host and the tumor that can improve the accuracy of current risk-stratification schemes.

After modern high-density genomic tools became accessible, investigators embarked on a search for a gene-expression signature of diagnostic leukemic blasts that could predict outcome. The Affymetrix U133Plus 2.0 platform was used to identify a 48-gene expression signature predicting outcome in a Children’s Oncology Group (COG) study of patients treated on a high-risk protocol (COG1961).2 This signature was validated in three independent cohorts treated by different study groups, but was unable to provide additional benefit for prognostication when age, leukocyte count, and cytogenetics were included in a multivariate analysis. ALL is a heterogeneous disease comprising various genetic subtypes with prognostic implications, thus making it difficult to identify a universal signature for response. For example, Yeoh et al. identified a
prognostic signature in two subtypes of ALL (T-cell and hyperdiploid), although these findings require further validation. The detailed study of genetic lesions and pathways that render chemoresistance within individual known subtypes and the identification of additional subtypes of ALL with prognostic relevance may be more beneficial than identifying a universal signature.

**Identification of Novel Biological Subtypes Associated with Outcome**

It is well-established that gene-expression patterns can be used to accurately classify various subtypes of ALL, as these patterns are influenced by pivotal genetic lesions. However, approximately 25% of cases of childhood ALL have no known genetic characteristics. Unique biological subgroups that are characterized by similarities in gene-expression profiles have been identified in such cases.

**BCR-ABL1-like Acute Lymphoblastic Leukemia**

In a recent study by den Boer et al. of 190 children with ALL treated following German Cooperative ALL protocols, the expression profiles of 30 of 44 children (68%) with precursor B-ALL without defined genetic features (19% of all children with precursor B-ALL) were similar to those of children with the BCR-ABL1 fusion in a class discovery algorithm. The five-year disease-free survival (DFS) for these patients with the so-called BCR-ABL1-like ALL was worse (59.5%) than that for other patients with precursor-B-ALL (84.4%). Fifteen percent of patients from an independent cohort who were enrolled on Dutch Childhood Oncology Group trials showed a similar signature and inferior outcomes. Many patients in this relatively large subtype were traditionally assigned to lower-risk treatment groups, but should actually have been treated with more intensified therapy. *In vitro* drug-resistance assays demonstrated further that BCR-ABL1-like ALL cells were more resistant to L-asparaginase and daunorubicin than other precursor B-ALL cells. Using array comparative genomic hybridization (CGH) methods, they described genetic lesions involving the B-cell developmental pathway in 82% of patients with BCR-ABL1-like ALL. Deletions of IKZF1 were most frequent (17 of 44 patients). As deletions in the B-cell differentiation pathway are common in cases with the BCR-ABL1 fusion, these patients may exhibit similarities in expression patterns.

**Acute Lymphoblastic Leukemia Characterized by Lesions in Ikaros**

Ikaros proteins are transcription factors restricted to the lymphoid lineage and are considered master regulators of lymphocyte development and differentiation. By studying genome-wide copy-number alterations in childhood ALL, Mullighan et al. have systematically shown recurring copy-number alterations in genes involved in the B-cell differenation pathway in approximately 40% of all childhood ALL. There is a very high incidence of IKZF1 lesions (83.7%) in BCR-ABL1-positive ALL, a subtype with poor outcome. Recently, Mullighan et al. showed that the presence of IKZF1 lesions in patients with BCR-ABL-negative ALL also resulted in poor outcomes. Two cohorts were studied: 221 children with high-risk ALL (excluding patients with BCR-ABL1 fusion, hypodiploidy, infants, and induction failures) treated on the COG P9906 study and 258 children (all risk groups) treated at the St Jude Children’s Research Hospital (Memphis, TN). Risk scores based on the copy number status of a subset of genes were used in a supervised principal components analysis and were found to be strongly associated with treatment outcome in both cohorts. Only IKZF1 was significantly associated with poor prognosis in both cohorts. IKZF1 lesions (mutation or deletions) were strong predictors of MRD burden, event-free survival, and overall outcome, even after adjusting for age, leukocyte count at presentation, and cytogenetic subtype. In the COG cohort, the five-year incidence of relapse was 73.4±8% in patients with an IKZF1 lesion versus 25.2±4% for those without a lesion, and the 10-year cumulative risk for relapse in the St Jude cohort for patients with an IKZF1 lesion was 48.4±8% versus 25.6±4% for those without a lesion. Thus, the presence of IKZF1 alterations could potentially be incorporated in future risk-stratification schemes.

It is possible that there is considerable overlap between the biological features of patients in the Mullighan et al. study and those with the BCR-ABL1-like genotype in the den Boer et al. study. Although therapeutic protocols differ, pooling of data generated by all study groups may strengthen the evidence of the existence of high-risk subsets.

**Early T-Cell Precursor Acute Lymphoblastic Leukemia**

Prognostic features for T-Cell ALL have not been as well characterized as for B-lineage disease, but with modern intensive therapy the outcome for T-cell ALL has approached 80%. By using a complementary selection of sensitive and high-resolution tools for molecular studies, Coustan-Smith et al. have recently identified a distinct subset of T-cell ALL termed early T-cell precursor (ETP) ALL that responds poorly to conventional chemotherapy. ETPs are a subset of thymocytes with the potential to differentiate into multiple lineages, including lymphoid and myeloid. Coustan-Smith et al. used a gene-expression signature of normal thymic ETPs to identify a similar profile in malignant cells from 17 of 200 children (12%) with newly diagnosed childhood T-cell ALL. Multiparameter flow cytometry showed that these cases were defined by a characteristic immunophenotype (CD1a-, CD8-, CD5 weak, and stem cell or myeloid markers), and high-resolution SNP array analysis revealed that they were more genomically unstable than cases of typical T-cell ALL. These patients had a markedly inferior outcome: a 72% cumulative risk for treatment failure at 10 years versus 10% for those with typical T-cell ALL. Patients with the ETP phenotype treated on the Italian Pediatric Hematology Oncology Association (AEIO) clinical trial also had a dismal outcome, validating the prognostic implication of this unique biological subtype. On the current frontline ALL trial at St Jude, these patients are candidates for stem-cell transplantation following first remission.

**Prediction of Early Response to Therapy**

The ability to monitor MRD has been one of the most significant contributions to the management of childhood ALL this decade and is the most powerful predictor of overall outcome in contemporary ALL protocols. This sensitive measurement of early treatment response is prognostic for overall DFS after initial diagnosis, post-relapse, and before stem cell transplantation. Subsets of patients at high risk for treatment failure can be identified relatively early, and therapy can be intensified for patients without satisfactory reduction in disease burden at pre-defined time-points, usually within the first few weeks of initiating therapy. Although MRD is an excellent surrogate marker for treatment response, the precise factors contributing to early disease regression are not completely understood. Genome-wide studies have attempted to
Hematological Malignancies

dissect and better comprehend both host and tumor features that play a role. These studies have also attempted to identify at initial diagnosis patients who would be slow responders, so that therapy can be intensified even earlier for patients predicted to have a high MRD.

Attributes of the Tumor
Cario et al.24 used a complementary DNA (cDNA) platform and identified 54 genes that were differentially expressed in patients with standard-risk MRD (not detectable at day 33 of remission induction and week 12 of continuation therapy) and those with high-risk MRD (≥10−3 MRD at week 12). By using the Affymetrix U133A oligonucleotide platform, Flotho et al.25 identified 17 genes associated with MRD at day 19 and 46 of remission-induction therapy, and suggested that the expression of CASP8AP2, a mediator of apoptosis, could be used as a marker of early disease response. Among genes associated with MRD at the early time-point (day 19 of remission induction), they identified 14 genes strongly related to long-term outcome independent of known prognostic variables, such as age, leukocyte count at presentation, and presence of the BCR-ABL1 fusion. After further validation in independent cohorts, the expression of a relatively small set of prognostically relevant genes could be incorporated in early risk-stratification schemes.

Features of the Host
Most genome-wide studies have investigated characteristics of leukemic cells that influence treatment response and outcome. However, an individual’s genetic make-up, especially inherited variations in drug-metabolizing enzymes, transporters, and drug targets, can significantly affect drug disposition and, subsequently, efficacy and toxicity.26 In an elegant study of more than 400,000 host germline polymorphisms in 487 children with ALL, 102 SNPs (representing 71 unique genomic loci) were identified that were significantly associated with MRD at the end of remission-induction therapy.26 Of the 102 SNP genotypes associated with high MRD, 21 were related to decreased exposure to the chemotherapeutic agents methotrexate and etoposide because of either increased clearance of these agents or decreased intracellular accumulation of methotrexate polyglutamates. One of the strongest signals from this genome-wide scan came from five SNPs located in the gene locus for interleukin-15 (IL15), a proliferation-stimulating cytokine. High IL15 gene expression was associated with a high MRD burden. IL15 can protect hematopoietic tumors from glucocorticoid-induced apoptosis in vitro,27 and its high expression in leukemic cells has been previously shown to be related to risk for CNS involvement in childhood ALL.27

Prediction of Drug Sensitivity
Intrinsic resistance to single or multiple chemotherapeutic agents is a significant hindrance to cure, and the molecular determinants of drug resistance have been explored extensively. In vitro drug sensitivity of primary leukemic cells to commonly used antileukemic agents (prednisolone, vincristine, asparaginase, and daunorubicin) has been correlated to their global gene-expression profiles.28 There was minimal overlap between the genes included in the resistance profiles of the four individual drugs, but a combined gene-expression score was significantly associated with outcome. Subsequently, a multidrug resistance phenotype was employed to compute a cross-resistance score that identified patients at high risk for relapse.29 The five-year DFS for cross-resistant patients was 53±10% versus 91±5% for cross-sensitive patients. In addition to identifying patients at high risk for treatment failure, the genes and pathways implicated in these studies may provide rational approaches to bypassing drug-specific and multidrug resistance and eventually to tailoring therapy on the basis of resistance patterns to various agents.

Prediction of Adverse Events
The best characterized and clinically valuable example in the pharmacogenetics of ALL is the influence of polymorphisms in the thiopurine S-methyltransferase (TPMT) gene on hematopoietic toxicity secondary to thiopurines commonly used in ALL.30 Additional candidate-gene approaches have shown that polymorphisms in vitamin D receptor (VDR) fokI site TYMS27 and the PAI 1st are associated with the risk for osteonecrosis and polymorphisms in eight genes associated with hypertension secondary to steroid treatment in ALL.31 With the current availability of genome-wide technologies, adverse-event phenotypes are being correlated to individual genotypes for further individualization of therapy.

Secondary leukemia, particularly secondary acute myeloid leukemia (AML), is an uncommon but devastating adverse event after patients are exposed to chemotherapeutic agents, such as topoisomerase II inhibitors and alkylating agents. These agents must be avoided as much as possible in patients at risk for developing secondary leukemia. To identify at-risk patients, Hartford et al. used genomic tools to study copy number and gene-expression changes in samples from children with ALL exposed to etoposide. They found that 309 germline SNPs differed significantly between patients with ALL who developed secondary leukemia (cases) and those treated on similar protocols but who did not develop secondary leukemia (controls).32 In addition to these inherited variations, 11q23 translocations and multiple cryptic regions of loss of heterozygosity were acquired in the blasts of the secondary leukemia cells. These data were complemented by profiles of genes differentially expressed in diagnostic blasts of cases and controls; pathways related to focal adhesion and Wnt signaling were among those that were affected and were suggested to play a role in the pathogenesis of secondary leukemia.

Biological Insights
The recognition of specific molecular lesions and critical oncogenic pathways paves the way for novel targeted approaches to therapy, such as the development of tyrosine kinase inhibitors for BCR-ABL1-positive hematological malignancies. The study of the global genomic portrait of ALL can provide further insights into the pathways that play a role in leukemogenesis and possibly chemoresistance.

Co-operative Lesions
Fundamental genetic lesions, such as chromosomal translocations, can be easily identified in a large proportion of childhood ALL. However, in most cases leukemogenesis is a multistep process and additional hits are often required, for example deletion of the non-translocated ETV6 (also known as TEL) allele in the development of ETV6–RUNX1-positive leukemia. In a seminal study of global copy number and loss of heterozygosity in 242 cases of childhood ALL, Mullighan et al.4 showed...
that lesions in genes encoding transcriptional regulators of B-lymphoid differentiation occur in more than 40% of cases with precursor B-ALL. The most common target was PAX5, a transcription factor playing a major role in B-lineage commitment. Detailed functional validation experiments and sequencing suggested that PAX5 is a haplo-insufficient tumor suppressor in ALL. Strategies aimed at restoring the function or enhancing the activity of the residual PAX5 allele may be beneficial in ALL with various PAX5 lesions. Mullighan et al. also demonstrated that gross genomic instability was not a feature of most childhood ALL: a mean of 6.4 genomic lesions were present per case, with wide variability within known subtypes. Samples of MLL-re-arranged ALL had less than one additional alteration, consistent with earlier observations that an MLL re-arrangement is strongly leukemogenic and this subtype, which occurs mainly in infants, does not require additional additional abnormalities. In contrast, multiple lesions (mean 11, mainly amplifications) were seen in hyperdiploid cases. Several lesions that have been identified by the systematic search may be passenger mutations, but many may play roles in initiation and maintenance of the leukemic phenotype, and further studies are being conducted to elucidate their function.

**Relapsed Acute Lymphoblastic Leukemia**

Approximately 10–20% of children with ALL ultimately relapse, and their outcome is poor despite extremely intensive therapy, including stem cell transplantation. Timing and site of relapse are prognostic of outcome; early medullary relapse portends the worst outcome. Matched samples from the same patient at diagnosis, remission, and relapse offer the opportunity to study the pathways leading to relapse and to identify novel therapeutic targets. A global gene-expression study of 35 patients at diagnosis and relapse suggested that pathways leading to early and late relapse may differ. Samples from early relapse demonstrated gene-expression patterns that were more similar to their diagnostic counterparts than samples at late relapse, indicating the emergence of a related resistant clone. Early relapse was also characterized by a proliferative signature, with upregulation of genes involved in cell-cycle progression, DNA repair, and anti-apoptosis/cell survival.

Two recent studies of high-resolution copy-number profiling of matched samples have demonstrated clonal evolution from diagnosis to relapse. In these studies, multiple novel copy number losses or gains were observed at the time of relapse. The frequency of deletions in CDKN2A, IKZF1, and EBF1 was higher in the relapsed cohort than in previously published cohorts at initial diagnosis, suggesting a role in resistance. Two patients showed relapse-specific focal deletions at the locus for MSH6. MSH6 is a component of the mismatch repair machinery and has been implicated in resistance to thiopurines, drugs commonly used for prolonged periods in ALL regimens.

Backtracking studies have revealed that in most cases the clone responsible for relapse is present as a minor subclone at initial diagnosis. The relapse clone either evolves from an ancestral pre-leukemic clone or is clonally related to the diagnostic clone. Interestingly, there is a paucity of common shared lesions specific to relapse, but cell-cycle regulation and B-cell development pathways have been frequently targeted at relapse.

**Identification of Therapeutic Targets**

As opposed to a candidate-gene approach, global genomic studies hold the promise of identifying multiple potential targets for therapeutic intervention. Although many recent studies have provided significant insights into the biology of leukemogenesis, it is argued that the progress made in validating and moving therapeutic drugs to the clinic has been relatively slow. However, in addition to identifying specific drug targets as described below, clues obtained from genome-wide studies have provided the basis for developing innovative approaches to circumvent chemoresistance. For example, as several genes involved in glucose metabolism are differentially expressed in steroid-resistant and -sensitive primary ALL cells, a detailed study of the glycolysis pathway was undertaken. Glucocorticoid resistance was associated with increased glucose consumption by cells. Inhibition of glycolysis by synthetic compounds restored sensitivity to glucocorticoids in prednisone-resistant cell lines and primary patient samples.

MLL-re-arranged infant ALL is relatively difficult to treat with conventional chemotherapy, and novel, less toxic targeted agents would be of immense benefit. An important observation was the overexpression of FLT3 in MLL-re-arranged ALL in a global gene-expression study. Activating mutations of the FLT3 tyrosine kinase causing auto-activation are detectable in approximately 30% of cases with AMLs, but global gene expression profiling of MLL-re-arranged infant ALL revealed high levels of FLT3 despite activating mutations being detected in only 3% of cases. Upon phosphorylation, FLT3 triggers downstream pathways important in hematopoietic cell proliferation and survival. Currently, FLT3 inhibitors are being tested in clinical trials for MLL-re-arranged infant ALL. Microarray studies have also revealed low expression of the tumor suppressor FHIT in MLL-re-arranged ALL. FHIT is silenced by methylation and its function can be restored by using demethylating agent, such as decitabine.

Gene-expression profiles of paired leukemic samples from patients at initial diagnosis and at relapse have identified multiple genes upregulated at the time of relapse that may contribute to the poor response to retrieval therapy. One such target is BIRC5, also known as survivin, an inhibitor of apoptosis that is highly expressed on relapsed ALL and other aggressive tumors, such as neuroblastoma. Additional pre-clinical experiments and phase I studies of small-molecule inhibitors and antisense oligonucleotides are ongoing. The selective expression of survivin in tumor tissue makes it an attractive therapeutic target.

**Conclusions**

Although 80–90% of children with ALL can be cured with current multiagent therapy, it may be possible to reduce the intensity of therapy and the subsequent short- and long-term side effects in a subset of patients. On the other hand, the prognosis for children who relapse is extremely poor and innovative strategies to prevent and treat relapse are urgently needed.

High-throughput genomic technologies promise to uncover multiple dimensions of the cancer phenotype. In childhood ALL, novel genetic subtypes have been identified that not only are valuable for outcome prediction but also provide numerous candidate genes that are being examined further to decipher complex pathways underlying leukemia.
Hematological Malignancies

initiation and progression. Furthermore, even within known genetic subtypes, the extensive characterization of co-operative lesions has provided a wealth of information to study the functional consequences of genomic alterations. In addition, the crucial contribution of host germline variations in influencing treatment response and adverse effects has been investigated at a genome-wide level as opposed to a hypothesis-driven candidate-gene approach.

In view of the immense potential of comprehensive genomic technologies in the field of cancer research, various large-scale initiatives such as The Cancer Genome Atlas (TCGA) effort are ongoing. Specifically for childhood ALL, the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) project involves array-based characterization, gene re-sequencing, and high-throughput small-molecule screens to identify and validate potential therapeutic agents (www.target.cancer.gov/). Thus, personalized therapy and targeted approaches will be used in the foreseeable future to treat childhood ALL.