Abstract: Molecular epidemiology is the study of genetic and environmental risk for disease, with much effort centered on cancer. Childhood leukemia occurs in nearly a third of all patients newly diagnosed with pediatric cancer. Only a small percentage of these new cases of childhood leukemia are associated with high penetrant hereditary cancer syndromes. Childhood leukemia, especially acute lymphoblastic leukemia, has been associated with a dysregulated immune system due to delayed infectious exposure at a young age. Identical twins with childhood leukemia suggest that acute lymphoblastic leukemia begins in utero and that the concordant presentation is due to a shared preleukemia subclone via placental transfer. Investigation of single nucleotide polymorphisms within candidate genes find that leukemia risk may be attributed to population-based polymorphisms affecting folate metabolism, xenobiotic metabolism, DNA repair, immunity, and B-cell development. More recently, genome-wide association studies for leukemia risk have led investigators to genes associated with B-cell development. When describing leukemia predisposition due to hereditary cancer syndromes, the following 6 categories become apparent on the basis of biology and clinical presentation: (1) genetic instability/DNA repair syndromes, (2) cell cycle/differentiation syndromes, (3) bone marrow failure syndromes, (4) telomere maintenance syndromes, (5) immunodeficiency syndromes, and (6) transcription factor syndromes and pure familial leukemia. Understanding the molecular epidemiology of childhood leukemia can affect the treatment and tumor surveillance strategies for these high risk patients and their family members.

Key Words: childhood leukemia, acute lymphoblastic leukemia, cancer predisposition, inherited cancer syndrome, genetic instability, DNA repair, inherited leukemia, familial leukemia, molecular epidemiology, genome wide association study

(J Investig Med 2015;63: 00-00)

Molecular epidemiology is the study of the genetic and environmental causes of disease and both their interactions together to understand clinical risk, outcome, and prevention of disease. The field of molecular epidemiology has advanced rapidly with the introduction of the genomic era, especially in the field of cancer.1–3 Through many different types of genomic investigations, we have learned a tremendous amount about the molecular contribution to disease distribution. The field of molecular epidemiology continues to grow and adapt at a rapid pace while new sequencing technologies are introduced into studies. This has been especially true in childhood leukemia, where new advances in technology have allowed for the relatively recent identification of genetic risk factors for disease in both general population and high risk, inherited populations.

From the Department of Pediatrics, Department of Oncological Sciences, University of Utah, Salt Lake City, UT. Received March 19, 2015, and in revised form April 6, 2015. Accepted for publication April 6, 2015. Reprints: Joshua D. Schiffman, MD, Huntsman Cancer Institute, 2000 Circle of Hope, Salt Lake City, UT 84112. E-mail: Joshua.Schiffman@hci.utah.edu. Supported by the Edward B. Clark, MD, Chair in Pediatric Research and the Primary Children’s Hospital Pediatric Cancer Program, funded by the Intermountain Healthcare Foundation and the Primary Children’s Hospital Foundation. Copyright © 2015 by The American Federation for Medical Research. ISSN: 1081-5589. DOI: 10.1097/JIM.0000000000000204

Applying Molecular Epidemiology in Pediatric Leukemia

Joshua D. Schiffman, MD

Childhood leukemia comprises nearly a third of all new cancer diagnoses in children and adolescents, making it one of the most common forms of pediatric cancer.4–6 Despite occurring so commonly, only a very small percentage of children diagnosed with leukemia are thought to be due to familial or hereditary cancer syndromes; nevertheless, this small fraction of cases has proven very informative to our understanding of childhood cancer and even has impacted clinical management. In contrast to the single-gene, high penetrant familial disorders, many new studies being published describe the presence of recurring, low penetrant risk alleles or single nucleotide polymorphisms (SNPs) that may contribute to leukemia risk in children.8–10 The vast majority of childhood leukemia is acute lymphoblastic leukemia (ALL), which can be classified by immune cell phenotype as B-cell ALL (the most common) and T-cell ALL (less common and typically more aggressive). The other type of childhood leukemia includes acute myeloid leukemia (AML). As described later, each type of childhood leukemia can be associated with a different hereditary cancer syndrome, and sometimes a single syndrome can cause multiple leukemia subtypes. In addition to helping understand individual disease risk, along with possible clinical implications, the application of molecular epidemiology in leukemia helps shed light on the underlying biology of one of the most common childhood cancers.

ALL Epidemiology

One of 2000 children will develop ALL, which translates to more than 3250 new cases of childhood acute leukemia diagnosed annually in the United States.11,12 Acute lymphoblastic leukemia is slightly more frequent in boys versus girls11 and in Hispanic and non-Hispanic whites versus African Americans,11 and most commonly presents in children between the ages of 2 to 6 years, the so called “common ALL.”11 Many theories exist to the causes of childhood leukemia, but no single unifying theory has yet been able to explain all cases. The general consensus now seems to point to a dysregulated immune response to infection as contributing to leukemia risk, due to either lack of infectious exposure, genetic impairment of response, or a combination of both of these factors. The population mixing hypothesis by Kinlen11,12 bases leukemia risk on exposure to demographic mixing and subsequent introduction of a viral infection to previously unexposed individuals in the perinatal period. The delayed infection hypothesis by Greaves15,16 describes leukemia risk due to delayed exposure to a common infection and an evolutionary mismatch between immune system exposure and modern, hygienic lifestyle. Fascinatingly, Greaves et al.17 have demonstrated preleukemic clones in dried newborn blood spots at birth, suggesting that the first preleukemic “hit” occurs before delayed infectious exposure. Greaves et al.18 also have described concordant ALL in identical (monozygotic) twins with shared identical translocation breakpoints, suggesting again that the first hit of ALL occurs in utero and, in the case of twins, the preleukemia clone has crossed across the placenta. Considering this molecular epidemiological evidence, an accepted model of ALL risk follows a combination of chance, exposure, and inherited genetic variation leading to in utero initiation followed by postnatal promotion and finally
leukemia. As covered below, the process of leukemia risk can be accelerated in the setting of hereditary cancer syndromes.

Candidate Gene Approaches

Initially, while investigation began into the molecular epidemiology of childhood leukemia, one of the approaches was to explore specific “candidate genes” that could contribute to leukemia risk. Using technological approaches available at the time, study investigators most often looked at SNPs within single genes that were thought to be involved in the process of leukemia development and progression. When reviewing the hundreds of published studies available, the majority of these candidate genes associated with the biology of ALL can be divided into the following 5 main categories: (1) folate metabolism/transport, (2) xenobiotic metabolism/transport, (3) immune function, (4) DNA repair, and (5) cell cycle. Many of these studies have mixed and even conflicting results, demonstrating the difficulty in identifying risk genes for cancer. Nevertheless, several candidate genes seem to suggest an association with ALL risk and include MTHFR C677T (folate metabolism), 1–29 CYP1A1 TP235C (xenobiotic metabolism), 28,29 GSTM1 deletion (xenobiotic metabolism), 29–34 NAT2*5 (xenobiotic metabolism), 30–38 XRCC1 G28152A (DNA repair), 39,40 and HLA-DRB4 (encoding HLA-DR53 immune antigen). 41–43

Genome-Wide Association Studies

The introduction of SNP microarrays offered the possibility of studying hundreds of thousands, sometimes millions, of SNPs and their cancer risk in a simultaneously agnostic approach in what has become known as the genome-wide association study (GWAS). Several GWAS have been performed in the past few years on DNA from thousands of children diagnosed with leukemia. These GWAS have discovered SNPs within the following genes associated with growth regulation, hematopoiesis, and lymphocyte development: IKZF1 (7p12.2), CDKN2A (9p21.3), ARID5B (10q21.2), and CEBPE (14q11.2) genes. 44–49 These findings, seen in children from European descent, lead to almost a 3-fold risk for leukemia and are among the strongest cancer susceptibility variants identified through GWAS. 50 In multietnic populations (including African Americans and especially Hispanic Americans), other risk alleles have been identified, such as ARID5B, CEBPE, BM11-PIP4KA, and hyperdiploid subtype. 51–55 Most recently, GATA3 was identified through GWAS to be a risky allele for ALL diagnosed in adolescents and young adults. 56 Although these GWAS findings probably account for less than 10% of genetic variation in ALL risk, they still suggest that genetic factors play a strong role in the development of childhood ALL. 55

Inherited Predisposition Syndromes

When discussing the leukemia-associated inherited cancer syndromes, it is helpful to divide these syndromes into the following 6 main categories based on biological functions and affected pathways: (1) genetic instability/DNA repair syndromes, (2) cell cycle/differentiation syndromes, (3) bone marrow failure syndromes, (4) telomere maintenance syndromes, (5) immunodeficiency syndromes, and (6) transcription factor syndromes including pure familial leukemia. See Table 1 for the list of associated syndromes and genetic mutations for each category. For the purposes of this report, we will focus on the first category of genetic instability and DNA repair syndromes.

Genetic Instability/DNA Repair Syndromes

Although leukemia is typically not the primary malignancy often seen in these syndromes, it still plays an important role in cancer risk. Also, once diagnosed, individuals with these types of syndromes need to have treatment tailored to avoid excessive toxicity from their chemotherapy and radiation treatment. These syndromes offer an excellent example of how the application of molecular epidemiology in pediatric leukemia can impact patient care.

Li-Fraumeni syndrome (LFS) is due to TP53 mutations and is associated with multiple cancer types including sarcomas, breast, and bone cancer. 56,57 This is a highly penetrant cancer syndrome and occurs in the population at an estimated prevalence of 1/5000 to 1/20,000. 58,59 Acute lymphoblastic leukemia, AML, and bone marrow myelodysplastic syndrome (MDS) have been reported, with hematological malignancies occurring about 1% to 3% of the time. 59–62 Nearly half of patients with hypodiploid ALL might have germline TP53 mutations, 63,64 making it prudent to check for LFS in any patients diagnosed with hypodiploid ALL regardless of family history. Although leukemia surveillance is still being studied in LFS, some have recommended annual complete blood counts as part of a biochemical screening program for early cancer detection. 65

Biallelic mismatch repair syndrome is caused when two of the following mismatch DNA repair alleles are inherited: MLH1, MSH2, MSH6, and PMS2. Normally just associated with Lynch syndrome (hereditary nonpolyposis colon cancer), 66 patients with biallelic (homoyzogous) alterations in the mismatch repair genes have multiple café-au-lait spots, pediatric brain tumors, and an especially high rate of pediatric hematological malignancies including both ALL and AML. 67–71 The population prevalence of Lynch syndrome is high at 1/440; therefore, the biallelic mismatch repair syndrome would be estimated to be as low as 1/775,000. Any child presenting with leukemia who also has several café-au-lait spots and/or a family history of colorectal cancer should be considered for testing for biallelic mismatch repair syndrome.

Fanconi anemia is often an autosomal recessive (AR) disorder of chromosomal breakage due to germline mutations in one of the 15 complementation groups, including FANCA-C, FANCD1-2, FANCE-G, FANCI-J, and FANCL-P. 72–75 Typically, Fanconi anemia is diagnosed through a combination of bone marrow failure and its distinct physical characteristics including short stature, microcephaly, microphthalmia, epicanthal folds, dangling thumbs, ureteral defects, congenital dislocated hips, and rocker bottom feet. 72 Fanconi anemia diagnosis can be complicated as up to 25% of affected individuals who do not display significant dysmorphology, and the disorder has high genetic heterogeneity. 72 Bone marrow failure often occurs between ages 5 to 15 years, 72 and just under 10% may develop AML or MDS. 73,76 The cumulative probability of leukemia is nearly 40% by age 30 years, and MDS has a 50% cumulative incidence by age 50 years. 73,75 Patients with Fanconi anemia remain very sensitive to DNA-damaging agents, especially radiation, and treatment often will be modified from standard care due to the high rate of secondary malignancy. Recommendation for surveillance for hematological malignancies in Fanconi anemia including measuring complete blood counts along annual bone marrow evaluation for changes in morphology, cellularity, and cytogenetics. 77

Ataxia telangiectasia is caused by ATM mutations, and patients experience progressive ataxia with central nervous system degeneration, growth deficiency, ocular and facial telangiectasia, immunodeficiency, and a very high risk for hematological malignancies. 78,79 Patients with ATM mutations experience a 70-fold increase in leukemia risk, mostly ALL. 79 Due to an extreme sensitivity to
<table>
<thead>
<tr>
<th>Leukemia Predisposition Category</th>
<th>Syndrome</th>
<th>Gene(s)</th>
<th>Inheritance Pattern</th>
<th>Leukemia Type</th>
<th>Leukemia Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA repair/genetic instability</td>
<td>LFS</td>
<td>TP53</td>
<td>AD</td>
<td>ALL, MDS, AML</td>
<td>1%–3%</td>
</tr>
<tr>
<td></td>
<td>Biallelic mismatch repair syndrome</td>
<td>MLH1, MSH2, MSH6, PMS2</td>
<td>AR</td>
<td>ALL, AML</td>
<td>Unknown, but high</td>
</tr>
<tr>
<td></td>
<td>Werner syndrome</td>
<td>WRN</td>
<td>AR</td>
<td>AML, MDS</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td>Rothmund-Thomson</td>
<td>RECQL4</td>
<td>AR</td>
<td>MDS</td>
<td>Unclear</td>
</tr>
<tr>
<td></td>
<td>Bloom syndrome</td>
<td>BLM</td>
<td>AR</td>
<td>AML, AML, MDS</td>
<td>15%</td>
</tr>
<tr>
<td>Fanconi anemia</td>
<td>FANCA-C, FANCD1-2, FANC-E, FANCI-J, FANCL-P</td>
<td>AR except for FANC-B, which is XL</td>
<td>MDS/AML</td>
<td>7% MDS, 9% AML 500-fold AML</td>
<td></td>
</tr>
<tr>
<td>Ataxia telangiectasia</td>
<td>ATM</td>
<td>AR</td>
<td>ALL</td>
<td>ALL, T-cell lymphoblastic lymphoma/AML</td>
<td>70-Fold leukemia</td>
</tr>
<tr>
<td>Nijmegen breakage syndrome</td>
<td>NBS1</td>
<td>AR</td>
<td>ALL</td>
<td>ALL, T-cell lymphoblastic lymphoma/AML</td>
<td>Unclear</td>
</tr>
<tr>
<td>Noonan syndrome</td>
<td>PTPN11, SOS1, KRAS, NRAS, RAF1, BRAF, SHOC2, MEPK1</td>
<td>AD</td>
<td>TMD, JMMIL, CMMIL, ALL</td>
<td>Unknown, but high</td>
<td></td>
</tr>
<tr>
<td>CBL syndrome</td>
<td>CBL</td>
<td>AD</td>
<td>JMMIL</td>
<td>JMMIL, AML</td>
<td>Unknown</td>
</tr>
<tr>
<td>Neurofibromatosis type 1</td>
<td>NF1</td>
<td>AD</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>11% MDS 200-fold to 500-fold JMMIL</td>
</tr>
<tr>
<td>Diamond Blackfan anemia</td>
<td>RPS19, RPS24, RPS17, RPL35A, RPL5, RPL11, RPS7, RPS26, RPS10, GATA1</td>
<td>De novo and AD</td>
<td>MDS/AML, AML</td>
<td>MDS/AML, AML</td>
<td>5%</td>
</tr>
<tr>
<td>Shwachman-Diamond Anemagakaryocytic thrombocytopenia</td>
<td>SBDS</td>
<td>AR</td>
<td>MDS/AML, ALL</td>
<td>MDS/AML</td>
<td>5%–24%</td>
</tr>
<tr>
<td>Thrombocytopenia and absent radii</td>
<td>RBM8A Del 1q21.1</td>
<td>AR</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>Unknown, rare reports</td>
</tr>
<tr>
<td>Severe congenital neutropenia/Kostmann syndrome</td>
<td>ELANE, G6PC3, GFI1, HAX1, CSF3R</td>
<td>AD, AR</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>8%–25%</td>
</tr>
<tr>
<td>Dyskeratosis congenital</td>
<td>CTC1, DKC1, TERC, TERT, TINF2, NOP10, NHP2, WRAP53</td>
<td>XL, AD, AR</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>3%–33%</td>
</tr>
<tr>
<td>Immunodeficiency</td>
<td>WAS</td>
<td>XL</td>
<td>ALL</td>
<td>ALL</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>BTK</td>
<td>XL</td>
<td>ALL</td>
<td>ALL</td>
<td>Unknown, rare</td>
</tr>
<tr>
<td>Transcription factor</td>
<td>CEBPA</td>
<td>AD</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>Unknown, younger onset</td>
</tr>
<tr>
<td>Familial AML due to CEBPA mutations</td>
<td>RUNX1</td>
<td>AD</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>35% AML, young onset</td>
</tr>
<tr>
<td>Familial platelet disorder</td>
<td>GATA2</td>
<td>AD</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>50%</td>
</tr>
<tr>
<td>Familial PAX5 syndrome</td>
<td>PAX5</td>
<td>AD</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>Unknown, but high</td>
</tr>
<tr>
<td>Familial SH2B3 syndrome</td>
<td>SH2B3</td>
<td>AR</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>Very high, early onset</td>
</tr>
<tr>
<td>Familial mosaic monosomy 7</td>
<td>Unknown</td>
<td>Unknown</td>
<td>MDS/AML</td>
<td>MDS/AML</td>
<td>10% TMD, 1%–2% ALL-AML</td>
</tr>
<tr>
<td>Down syndrome</td>
<td>Trisomy 21</td>
<td>De novo</td>
<td>TMD, AML, ALL</td>
<td>TMD, AML, ALL</td>
<td>10% TMD, 1%–2% ALL-AML</td>
</tr>
</tbody>
</table>

CMML indicates chronic myelomonocytic leukemia; JMMIL, juvenile myelomonocytic leukemia; TLBL, T-cell lymphoblastic lymphoma; TMD, transient myeloproliferative disorder; XL, X-linked.
ionizing radiation and risk for secondary malignancies, treating oncologists need to modify therapy accordingly. Despite the high risk for leukemia, no consensus on hematologic surveillance has been recommended for patients with Ataxia telangiectasia although patients should be aware of signs of malignancy including weight loss, bruising, and localized pain or swelling. Nimegen breakage syndrome is caused by germline mutations in the NBS1 gene, responsible for DNA double-strand break repair in the same pathway as ATM with some of the same chromosomal breakage patterns seen in cells from patients with Ataxia telangiectasia. 73 Patients with Nimegen breakage syndrome have distinctive dysmorphology, growth deficiency, immunodeficiency, cognitive impairment, and increased cancer risks up to 40% to 50%. 74,75,76 Most frequently, patients with Nimegen breakage syndrome will develop lymphoma although a smaller percentage of individuals will develop ALL. 77 Also similar to Ataxia telangiectasia, patients with Nimegen breakage syndrome need specifically tailored cancer treatment due to their extreme sensitivity to radiation and chemotherapy. Anyone found to have Nimegen breakage syndrome should be monitored for general signs of malignancy. 82

Bloom syndrome is an AR disease due to BLM germline mutations, which is a helicase gene integral for double stranded DNA break repair. 84 Bloom syndrome lymphocytes reveal a high frequency of characteristic sister chromatid exchanges and quadrangular configurations. 85,86 Ashkenazi Jews carry the c.2207_2212delinsTAGATTC in BLM with an estimated carrier frequency of 1/100 due to a founder effect. 77,85 With less than 300 cases reported, our knowledge about the natural history of this syndrome comes from the Bloom Syndrome Registry. 87 Cancer is the most common cause of death, and patients have 25% cancer risk with multiple different tumor types developing at an early age of onset around 25 years. 77 Twenty-one cases of acute leukemia were documented. Among the 168 registered patients with Bloom syndrome, 21 patients were documented acute leukemias (ALL, 6; AML, 6; biphenotypic, 2; and unspecified/other, 7). 87 Similar to the other DNA repair syndromes described, patients with Bloom syndrome are sensitive to radiation and chemotherapy and therefore require specifically tailored treatment. 85

In addition to the above syndromes related to genetic instability and DNA repair dysfunction, several other hereditary cancer syndromes have also been described with leukemia predisposition as either a major or minor component. As reported previously, many of these can be grouped into one of the following remaining categories: cell cycle/differentiation syndromes, bone marrow failure syndromes, telomere maintenance syndromes, immunodeficiency syndromes, and transcription factor syndromes with pure familial leukemia. A comprehensive description of these other syndromes with associated leukemia risk is beyond the scope of the current report, but the reader is directed to the 2 following reviews by Seif (2011) and Stiegitz and Loh (2013) for an excellent summary of leukemia predisposition. Understanding the genetic risk for childhood leukemia is very important to identify children, and their family members, who may be at risk for hereditary cancer predisposition. As discussed above, this has clear implications for treatment strategy and avoidance of therapy that may lead to secondary malignancies. In addition, if other family members are found to harbor the same genetic mutations, then they can be appropriately managed through early clinical screening and surveillance. In summary, molecular epidemiology has identified several important genetic causes of leukemia risk for both the general population and those with inherited cancer syndromes. Understanding this connection has increased our knowledge about the biology and development of leukemia, as well as provided important insight into the appropriate clinical management of these patients. While our genomic technologies continue to improve, even more information will be learned about leukemia predisposition and how to use this knowledge to impact the care of children and adults with acute leukemia.

REFERENCES


