Genetic Risk Factors for Childhood Leukemia

GETA Symposium
3 March, 2009

Anand Chokkalingam, Ph.D.
School of Public Health,
University of California, Berkeley
anandc@berkeley.edu
Outline

- Background
- Review from literature
- Northern California Childhood Leukemia Study (NCCLS) experience
- Challenges
- Future directions
Genetic susceptibility

Q: What is genetic susceptibility?

A: Heritable factors that increase risk of a given disease

Usually one or more genes, or gene variations

May work in concert with

Other genetic factors, AND/OR

Environmental and lifestyle factors

Degree of involvement of other factors depends on penetrance
Penetrance

- **Penetrance**
  - **High**: rare, but high risk (e.g., BRCA1, RR~5)
    - Major part of short causal pathway to disease
  - **Low**: common, but low risk (RRs of ~1.3-1.8)
    - Minor part of long causal pathway to disease
  - Low-penetrance genetic factors likely to comprise the bulk of inherited cancer risk

3 March, 2009
Rationale for genetic susceptibility to CL

- Early age of onset
- Risk in twins
  - Mostly intraplacental metastasis, not highly penetrant risk allele
  - Suggests low penetrance susceptibility alleles
Review
Candidate pathways in published reports

- Folate metabolism
- Xenobiotic (exogenous chemical) transport and metabolism
- Immune function
- DNA repair
Folate metabolism & ALL

- **Folate**
  - Essential micronutrient, modulates balance between accuracy of DNA synthesis and DNA methylation
  - Deficiency can induce chromosomal damage and fragile chromosomal sites → carcinogenesis
  - Maternal supplementation during pregnancy may reduce risk

- **MTHFR, 2 loss-of-function variants: 14 reports**
  - 677C>T: null effect or modest risk reduction (OR~0.9)
  - 1298A>C: null effect

3 March, 2009
Xenobiotic metabolism & ALL

To do harm, exogenous chemicals must
- enter cells
  - Membrane transporters (e.g., MDR1)
- be metabolized into harmful species
  - Phase 1, bioactivation enzymes (e.g., CYPs)
  - Phase 2, detoxification enzymes (e.g., GSTs, NQO1)

Transporters
- MDR1: 4 reports
  - 3435C>T: null risk
- Phase 1, bioactivation
  - CYP1A1: 7 reports
    - 6235T>C: inconsistent
Xenobiotic metabolism & ALL

- Phase 2, detoxification
  - GSTM1 (detoxifies PAHs): 13 reports
    - deletion: null to modestly increased risk
  - GSTT1 (detoxifies epoxides and halomethanes): 13 reports
    - deletion: null risk
  - GSTP1: 7 reports
    - I105V: null risk
  - NQO1 (anti-oxidant, detoxifies quinones): 7 reports
    - 609C>T: null risk
    - 465C>T: null risk (2 reports)
Primary reports of gene main effects: AML

MTHFR
DHFR
Folic acid metabolism
Cys metabolism
Cys metabolism
DNA repair
Immune function
Other

3 March, 2009
Evaluating the evidence

- HuGENet – Human Genome Epidemiology Network
- HuGENet Encyclopedia: synopsis of evidence for genetic associations of complex disease
- CL as one of several prototype encyclopedia entries
- Venice meeting (2006): to develop criteria for rapid evaluation of evidence to facilitate encyclopedia effort

3 criteria:
- Amount of evidence
- Replication
- Protection from bias

Letter grades (A, B, C) for each – AAA is ideal
Venice criteria evaluation for ALL

Pilot Results

Result of preliminary review
- None have reached A status in any criterion
- Only MTHFR and GSTM1 rank BBB
- All else have a C in at least one criterion

Evaluation of criteria in progress

Next steps:
- Refine criteria
- Develop systems for
  - Consistent assignment and adjudication of grading
  - Updating as new evidence is published
Summary

- Few genes have been studied to date for ALL, even fewer for AML
- Many high-probability candidates remain unexamined or unreplicated
- Entire candidate pathways with very strong biological plausibility remain poorly studied (e.g., immune function, DNA repair)
- Reports to date do not ensure good coverage of variation within a gene
HapMap Project
Publicly available SNP data

Data on linkage between SNPs in different populations

3 March, 2009
Haplotype-based analysis

- Uses HapMap and similar data
- Permits:
  - Maximal coverage of variation within genes
  - Minimum number of SNPs
Haplotype-based analysis

- Uses HapMap and similar data
- Permits:
  - Maximal coverage of genes
  - Minimum number of SNPs
NCCLS experience
Northern California Childhood Leukemia Study

1995-present

- Population-based case-control study
- Incident cases ascertained from 9 pediatric hospitals in N. & C. California
- Controls individually matched (DOB, sex, Hispanic status, and maternal race)
- 42% Hispanic
NCCLS Biospecimen collection

**Background**

**Review**

**NCCLS**

**Challenges**

**Future Directions**

- Buccal Cells (98%)
- Buccal Cells (98%)
- Bone Marrow (cases only – 86%)
- Peripheral Blood (cases only – 86%)
- Archived Newborn Blood (85%)
- Peripheral Blood
- Urine
- Mothers of children <8 at diagnosis/reference
- Peripheral Blood

3 March, 2009
NCCLS Biospecimen collection

- Buccal Cells (98%)
- Urine
- Mothers of children <8 at diagnosis/reference
- Peripheral Blood
- Archived Newborn Blood (85%)
- Bone Marrow (cases only – 86%)
- Peripheral Blood (cases only – 86%)

3 March, 2009
Large-scale Genotyping

Objective: to comprehensively examine ~200 candidate genes in subset of available children (464 cases, 464 controls)

Custom Illumina 1536-plex

- 183 genes
  - Haplotype tagging SNPs
  - Literature SNPs

Ancestry Informative Markers

- Adjust for genetic ancestry
Candidate Pathways

- Immune function: 36%
- DNA repair: 17%
- Anti-oxidant/dietary: 17%
- Ancestry: 1%
- Oxidative stress: 4%
- Xenobiotic transport: 25%

Percent of 183 included genes
NCCLS genotyping

Current status – Preliminary analyses

- Adaptive immunity and ALL (208 SNPs in 29 genes)
  - 19 SNPs with nominal p<0.05
  - Single *IL12A* SNP (p<0.00001)
  - Haplotype sliding window analyses: regions in *CD28*, *FCGR2A*, *GATA3*, *IL2RA*, *IL12A*, *STAT4*, and *STAT6*

- Xenobiotic transport and metabolism and ALL (251 SNPs in 42 genes)
  - Individual SNP associations for *ABCB1*, *ABCC1*, and *CYP2C8* genes.
  - Haplotype sliding window analyses: regions in *ABCC1*, *ABCB1*, *CYP1A1*, *CYP1A2*, *CYP2B6*, *CYP2C8*, *IDH1*, *UGT1A1*, *UGT1A7*, and *UGT1A9*.
  - Potential interaction: *CYP2C8* risk haplotype and self-reported exposure to paints and solvents
Challenges to genetic susceptibility research in childhood leukemia

Statistical power and sample size issues
Maternal genes
Replication
Publication Bias
Statistical power

Required sample size to detect G main effects
(Power=0.80, \( \alpha=0.05 \), log-additive inheritance)

- OR(G)=1.2
- OR(G)=1.3
- OR(G)=1.4
- OR(G)=1.5
- OR(G)=1.7
- OR(G)=2.0

Risk allele frequency vs. II cases (or controls)

3 March, 2009
Sample size

- Sample sizes limited by
  - Low incidence
    - 4.5/100,000 person-years
  - Disease heterogeneity
    - Subtype groups
    - “Lumping vs. splitting”
  - Availability of high-quality DNA
- Published genetic studies typically 100-300 cases
  - Larger studies expected
Interactions

- Small effect sizes (OR~1.2-1.5) point to effects of GxE interactions
- Critical to understanding susceptibility
- Requires
  - High-quality exposure data
  - Even larger sample sizes for sufficient power
Maternal genes

**Background**

- Maternal genes

**Challenges**

- Maternal exposures
  + Maternal genes

  **in utero** environment

  **Child’s genes**

  **Childhood Leukemia**

**Future Directions**

- 3 March, 2009
Other major challenges

- Replication of results
  - Multiple studies
  - Multiple populations
- Publication bias
  - When null results are not presented or published
CLIC

- Childhood Leukemia International Consortium
  - www.clic.berkeley.edu
  - To date: 14 case-control studies in 10 countries
  - Over 9,000 cases

- Purpose
  - Replication of findings
  - Coordinated publication (address publication bias)
  - Data pooling (improve statistical power)
  - Collaborative research
GWAS

- **Genome-Wide Association Studies**
  - Allows exploration of genome beyond candidate genes
  - 300K-1M variants across the genome

- **Issues**
  - Sample size
  - Replication strategy
  - Technical requirements
    - DNA quality (unamplified, genomic DNA)
    - DNA quantity (~1 microgram)
  - Cost

3 March, 2009
GWAS:
Tiered replication approach

300K-1M SNPs

Stage 1
Discovery
(n~800-1000 cases)

15K-50K SNPs

Stage 2
Replication 1
(n~1200-1500 cases)

<3K SNPs

Stage 3
Replication 2
(n~1200-1500 cases)

20-100 SNPs

Sequencing/
fine mapping

Genes associated
with ALL
Acknowledgments

UCB
- Patricia Buffler
- Catherine Metayer
- Kevin Urayama
- Ghislaine Scelo
- Jeffrey Chang
- Melinda Aldrich
- Lisa Barcellos
- Martyn Smith

UCSF
- Joe Wiemels
- John Wiencke
- Helen Hansen

CDC
- Muin Khoury

NCI
- Stephen Chanock

UKCCS
- Malcolm Taylor