

CCSS Study Analysis Concept Proposal Genetics Working Group

1. Study Title: Epigenomic Profiling of Metabolic Outcomes in Childhood Leukemia Survivors

2. Working Group and Investigators: Genetics Working Group and Chronic Diseases Working Group

Proposed Investigators include:

Name	email	Institution
Philip Lupo	Philip.lupo@bcm.edu	Baylor College of Medicine
John Belmont	jbeltmont@bcm.edu	Baylor College of Medicine
Kala Kamdar	kykamdar@txch.org	Baylor College of Medicine
Smita Bhatia	sbhatia@coh.org	City of Hope National Medical Center
Kevin Oeffinger	oeffingk@mskcc.org	Memorial Sloan Kettering Cancer Center
Greg Armstrong	Greg.Armstrong@stjude.org	St. Jude Children's Research Hospital
Leslie Robison	Les.Robison@stjude.org	St. Jude Children's Research Hospital
Yutaka Yasui	yyasuiua2@gmail.com	University of Alberta
Wendy Leisenring	wleisenr@fhcrc.org	Fred Hutchinson Cancer Research Center

3. Background and rationale:

Acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy, accounting for 25% of all childhood cancers. Combination chemotherapy for two to three years is the primary treatment for all patients, and in a subset, cranial radiotherapy (CRT) is given. Although children with ALL rarely survived in the 1950s, remarkable advances in therapy have led to the present 80% cure rate. However, childhood cancer survivors face unique long-term health problems caused by the curative therapies they received. A particular problem among childhood ALL survivors is treatment-related obesity. For instance, in an assessment from the Childhood Cancer Survivor Study (CCSS), ALL survivors were at increased risk for obesity, compared to population-based, age-specific, and gender-specific norms (females: odds ratio [OR]=1.5; 95% confidence interval [CI], 1.2-1.8; males: OR=1.2; 95% CI, 1.0-1.5).¹ Previous studies have identified only a few risk factors associated with treatment-related obesity in long-term ALL survivors (e.g., CRT). However, these factors do not account for the inter-individual variation in treatment-related obesity. Because of obesity-related consequences in childhood ALL survivors, it is critical to identify at-risk subpopulations and develop targeted interventions.

Epigenetics: Epigenetics refers to events that alter the phenotype of a cell or organism that do not involve changes in DNA sequence or copy number. DNA methylation, histone modifications, and microRNA (miRNA) exert epigenetic influences on gene expression patterns and can affect other processes as well, including DNA replication, repair, or recombination. DNA methylation is the best studied and most easily measured of the epigenetic modifications. Hypomethylation and hypermethylation of CpG islands, frequently located in the control region of genes, may alter gene activity. Furthermore, there is evidence in animal models that DNA methylation regulates body composition. For instance, when the *agouti* gene in mice is completely unmethylated, the mouse has yellow fur and is obese, prone to diabetes and cancer. When the *agouti* gene is methylated, the mouse has brown fur, is normal weight, and has a lower risk of disease. *These mice are genetically identical and have identical diets, however there are differences in DNA methylation (and therefore gene expression) due to exposures during critical periods.*² Therefore, it is suspected that environmental stimuli during critical periods (e.g., cancer therapy during childhood) can lead to differences in DNA methylation, and thus, gene expression. There is also evidence in humans that DNA methylation is associated with obesity. Specifically, DNA methylation differences in key genes are associated with obesity in the general population. In a genome-wide methylation analysis on seven obese cases and seven lean controls, obese cases had a CpG site in the *UBASH3A* gene that was hypermethylated compared to lean controls ($p=5\times 10^{-6}$) and a CpG site in the *TRIM3* gene that was hypomethylated compared to lean controls ($p=2\times 10^{-5}$).³ *Based on this evidence, we hypothesize that ALL therapy may lead to epigenetic modifications and ultimately obesity in survivors.*

Genetics and Epigenetics: Patterns of DNA methylation appear to be under genetic control and allele-specific methylation is a widespread phenomenon.⁴⁻⁷ For instance, there is a strong effect of common single nucleotide polymorphisms (SNPs) on the methylation status of the *FTO*⁸ and *MCHR1*⁹ gene promoters. Additionally, both *FTO* and *MCHR1* have been associated with body mass index (BMI) as a quantitative trait and obesity in large genome-wide association studies (GWAS). Gene variants that affect methylation often arise from CpG polymorphisms that are the direct targets of methylation. *Therefore, an important step in making causal inferences related to DNA methylation and obesity is the incorporation of genetic data, specifically, determining SNPs that influence methylation status.*

Significance: Dynamic models are needed to explain the complexities of treatment-related obesity. Our study will be one of the first to integrate genetic and epigenetic data to explore the biology underlying the increased risk of obesity in ALL survivors. Ultimately, the identification of those at the greatest risk of developing treatment-related obesity through the discovery of novel biomarkers will aid in the implementation of targeted interventions, alterations in therapeutic strategies, and the potential use of epigenetic therapy to lessen these impact of these late effects.

4. Specific Aims/Objectives/research hypotheses:

Our goal is to use genome-wide DNA methylation (epigenomic) profiles as a novel biomarker of obesity in childhood ALL survivors. We hope this information will allow for the identification of at-risk subpopulations and the development of targeted interventions. **Our overall hypothesis** is that childhood ALL therapy induces consistent and recognizable epigenetic modifications at loci critical for normal body weight homeostasis in a subset of patients, ultimately leading to excessive weight gain in these survivors. As these epigenetic modifications may be under genetic control, our **secondary hypothesis** is that underlying individual genetic variation influences the risk of acquiring these adverse alterations in epigenetic regulation at a subset of loci. Assessing SNP-methylation pairs will allow us to distinguish methylation changes that act upstream of obesity onset rather than occur as a consequence of obesity.¹⁰

We have a timely opportunity to explore these hypotheses in a well-characterized population of long-term ALL survivors within the CCSS. This CCSS Career Development Award (CDA) will serve as a pilot study to launch a junior investigator's path in survivorship research and will be used as preliminary data for a larger grant application. We have secured institutional funds to conduct the DNA methylation profiling. Furthermore, we will leverage the genome-wide SNP array data as part of the CCSS initiative (PI: Kamdar) to conduct an integrative epigenomic/genomic analysis to identify loci that play a direct causal role in the obesity phenotype.¹⁰ Our specific aims are:

Aim 1: Determine if gene-specific DNA methylation status is associated with obesity in ALL survivors by conducting genome-wide DNA methylation profiling. *Hypothesis: DNA methylation changes are correlated with treatment-related obesity in ALL survivors. We will characterize DNA methylation profiles using a genome-wide approach among 48 obese ALL survivors and 48 normal-weight ALL survivors.*

Aim 2: Identify obesity susceptibility genes in ALL survivors through an integrated epigenomic/genomic analysis. *Hypothesis: An integrative approach exploring epigenotype-genotype associations will allow causal inferences about pathways contributing to obesity among ALL survivors. We will perform methylation quantitative trait loci (mQTL) analyses to explore epigenotype-genotype associations in the context of treatment-related obesity.*

5. Analysis Framework:

- Outcome of interest: The primary outcome of interest is body mass index (BMI), computed from self-reported height and weight in the CCSS questionnaire at the third timepoint (2007). Normal weight will be defined as 18.5-24.9 kg/m² and obesity will be defined as ≥ 30 kg/m².
- Exposure of interest: The primary exposure of interest is DNA methylation status. We will request 1 micrograms of DNA from CCSS. Each purified DNA sample from saliva will be bisulfite converted and then subjected to genome-wide DNA methylation profiling of 485,577 methylation sites using the Illumina Infinium 450K Methylation BeadChip platform (Illumina, San Diego, CA, USA) at Baylor College of Medicine. The coverage of this assay is extensive and includes: 96% of human RefSeq genes, including promoter, 5' and 3' regions; 95% of CpG islands and flanking regions (>26,000 at 19 sites/islands, >52,000 "shores" i.e. $\geq 2,000$ bp upstream and

downstream from islands, >52,000 “shelves” $\geq 4,000$ bp upstream and downstream from islands); CpG sites outside islands/outside coding regions; non-CpG methylated sites in human stem cells; differentially methylated sites identified in tumor vs. normal; and miRNA promoter regions. Fluorescence intensities at the methylated and unmethylated bead sites will be assessed using the Illumina iScan system, yielding β values, which estimate the methylation level of each locus. The β value is a continuous variable that is obtained by dividing the intensity of the methylated signal by the combined intensity, and β values may range from 0 (completely unmethylated) to 1 (completely methylated). External funding will be used to cover DNA methylation costs.

- Subject population: Eligibility criteria for our study will consist of: 1) prior diagnosis of childhood ALL, 2) availability of saliva-extracted DNA, 3) no history of bone marrow transplantation and 4) genotype data included as part of the CCSS GWAS. As this is a pilot study, we will sample two populations ($n=48$ in each group) from the CCSS: 1) survivors of ALL who are obese (i.e., ≥ 30 kg/m²), i.e., “cases” and 2) survivors of ALL with a normal BMI based on the NHLBI guidelines (i.e., 18.5-24.9 kg/m²), i.e., “controls”. More specifically, we will obtain four groups (total $n=96$):
 - Obese ALL survivors who had CRT doses of ≥ 18 Gy ($n=24$)
 - Normal-weight ALL survivors who had CRT doses of ≥ 18 Gy ($n=24$)
 - Obese ALL survivors who did not have CRT ($n=24$)
 - Normal-weight ALL survivors who did not have CRT ($n=24$)

We will match on age at diagnosis, gender, race/ethnicity, and length of follow up in order to reduce heterogeneity between the groups. This selection will be done based on BMI at the third timepoint, in 2007. Previous studies using “extreme” outcomes have been effective in identifying biological predictors of outcome in spite of limited sample sizes (i.e., comparable to our own).^{3,11}

- Exploratory variables:

The following information will be requested from CCSS questionnaire data:

- Height at Baseline, Follow-up 2003, and Follow-up 2007
- Weight at Baseline, Follow-up 2003, and Follow-up 2007
- Age at diagnosis
- Age at Baseline, Follow-up 2003, and Follow-up 2007
- Gender
- Race/ethnicity
- Treatment history, including:
 - Dexamethasone exposure
 - Cranial radiation therapy exposure and dose

- Statistical analysis:

Aim 1 (DNA Methylation Status and Obesity): For Aim 1, we will evaluate the influence of DNA methylation status on the outcome of interest (i.e., obesity). Specifically, the main exposure of interest will be the β of CpG DNA methylation at a given loci (i.e., the independent variable) and the outcome of interest will be obesity as described above. Covariates will be explored in preliminary analysis but are likely to include: gender, age at diagnosis, time since diagnosis, cancer diagnosis, treatment history, and hybridization array/batch. Logistic regression analyses will be conducted to identify CpG sites that are associated with obesity. These analyses will be carried out using R.

Type I error due to multiple comparisons will be controlled using the false discovery rate (FDR) method of Benjamini and Hochberg.¹² To better understand the functional mechanisms related to obesity, we will evaluate the gene content of differentially methylated loci. We will perform separate analyses using three methodologies: association of Gene Ontology (GO) terms with each pattern of expression, the DAVID

bioinformatics database, and Ingenuity Pathway Analysis software. To make comparisons between sets, we will take differences between the standardized scores determined for each gene set. Those loci that demonstrate differential methylation will identify genes deserving of further investigation as potential determinants of treatment-related obesity in survivors.

Aim 2 (integrated epigenomic/genomic analyses): For Aim 2, we will perform a methylation quantitative trait loci (mQTL) analysis. Specifically, we will evaluate the association between SNP loci within 1 Mb on either side of each methylation loci (i.e., *cis* association analyses) identified in Aim 1, after adjusting for covariates evaluated in Aim 1. Typical genes will encompass 10-30 SNP markers and 2 Mb intervals have >100. Genotyping of SNPs is currently being conducted on ALL survivors using the Illumina Omni5 BeadChip and will be available for this study. Using the software package PLINK, the main analysis will rely on linear regression of the methylation loci where genotypes (allele count of 0, 1, or 2 for the less common allele) are the primary independent variables of interest. We will also permute the data to adjust the p-values for each trait for the number of SNPs tested.

An adjusted p-value will be calculated as the fraction of simulations that produce an association for any SNP at least as significant as the most significant *cis* association observed in the original data. A permutation-based method, using label swapping of the traits, is an appropriate method of test correction for these analyses. To correct for the number of traits that are assessed, an FDR threshold will be explored based on the empirical p-values using *multtest* package in R. We will use a multidimensional scaling (MDS) method to correct for population structure. MDS will be applied to the full genotype data to infer continuous axes of genetic variation, producing dimension reduction that maximizes the explained variability (first few eigenvectors of the between-sample covariance matrix). Genetic effects arising from ethnic differences will be accounted for using: (1) stratification of analysis restricted to single MDS-defined groups and (2) local ancestry inference within each genomic analysis interval and stratification on continental origin of the specific haplotypes.

By assessing the effect of methylation status on single SNP associations, causal inferences can be better evaluated. A subset of SNP-methylation pairs will be selected based on significant association of the SNP with methylation status, correlation of quantitative methylation with obesity, and marginal association of the SNP with obesity. For SNPs that tag a true causal association with obesity, the marginal association will be reduced or completely lost after inclusion of methylation status (i.e., conditional independence).

- Power analysis:

For Aim 1, we used R to determine the different combinations of the effect size (mean difference between obese and normal-weight individuals divided by the standard deviation) and number of true prognostic genes under which at least 20 biomarkers can be obtained with 5% FDR from multiple tests of 485,577 sites with 48 obese and 48 normal-weight samples (**Table 1**).

For Aim 2, power calculations were made using Quanto software, version 1.2.4. Assuming an additive model of inheritance, $\alpha=10^{-5}$, and a minor allele frequency of 0.1, we have 80% power to detect a locus that explains at least 25% of the variance (**Figure 1**). These effect sizes correspond to the range observed by Wang et al.³ for DNA methylation and obesity and by Zhang et al.⁶ and Gibbs et al.¹³ evaluating DNA methylation as a heritable trait. *While our sample size is relatively small, this is a timely and unique opportunity to study these effects in childhood cancer survivors. Additionally, our integrative approach is more powerful than traditional methods in discovering obesity susceptibility loci.*

Effect size	No. true prognostic sites
0.25	1292
0.50	116
0.75	33
1.00	22
1.25	21

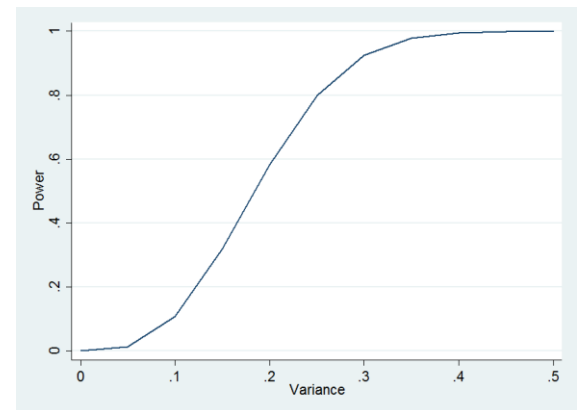


Figure 1. 80% Power to detect ~25% of variance in 96 subjects when $\alpha=10^{-5}$

6. References

1. Meacham LR, Gurney JG, Mertens AC, et al. Body mass index in long-term adult survivors of childhood cancer: a report of the Childhood Cancer Survivor Study. *Cancer* 2005;103:1730-9.
2. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 1999;23:314-8.
3. Wang X, Zhu H, Snieder H, et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med* 2010;8:87.
4. Bell JT, Pai AA, Pickrell JK, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome biology* 2011;12:R10.
5. Schalkwyk LC, Meaburn EL, Smith R, et al. Allelic skewing of DNA methylation is widespread across the genome. *American journal of human genetics* 2010;86:196-212.
6. Zhang D, Cheng L, Badner JA, et al. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* 2010;86:411-9.
7. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome research* 2010;20:883-9.
8. Bell CG, Finan S, Lindgren CM, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS one* 2010;5:e14040.
9. Stepanow S, Reichwald K, Huse K, et al. Allele-specific, age-dependent and BMI-associated DNA methylation of human MCHR1. *PLoS one* 2011;6:e17711.
10. Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet* 2012;44:841-7.
11. Harismendy O, Bansal V, Bhatia G, et al. Population sequencing of two endocannabinoid metabolic genes identifies rare and common regulatory variants associated with extreme obesity and metabolite level. *Genome Biol* 2010;11:R118.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 1995;57:289-300.
13. Gibbs JR, van der Brug MP, Hernandez DG, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010;6:e1000952.