Title: Developing a clinical and genetic risk prediction model for diabetes mellitus among survivors of childhood cancer

Working Group: Genetics
Chronic Disease

Investigators:

Melissa Richard, PhD	Melissa.Richard@bcm.edu
Philip Lupo, PhD, MPH	Philip.Lupo@bcm.edu
Smita Bhatia, MD, MPH	sbhatia@peds.uab.edu
Austin Brown, PhD	Austin.Brown@bcm.edu
Les Robison, PhD	Les.Robison@stjude.org
Wendy Leisenring, ScD	wleisenr@fredhutch.org
Kevin Oeffinger, MD	kevin.oeffinger@duke.edu
Yutaka Yasui, PhD	Yutaka.Yasui@stjude.org
Greg Armstrong, MD, PhD	Greg.Armstrong@stjude.org
Sogol Mostoufi-Moab, MD, MSCE	moab@email.chop.edu
Lillian Meacham, MD	lillian.meacham@emory.edu
Danielle Friedman, MD, MS	FriedmaD@mskcc.org
Eric Chow, MD, MPH	ericchow@u.washington.edu
Todd Gibson, PhD	Todd.Gibson@STJUDE.ORG

Background & Rationale

Diabetes mellitus (DM) is a complex heritable metabolic disorder characterized by insulin resistance (type 2) or insulin deficiency (type 1). In the general population, DM is associated with significant morbidity including retinal disease, chronic renal insufficiency, neuropathy, cardiovascular disease, and premature death.[1] Confirmed by several recent studies, survivors of childhood cancer demonstrate an elevated risk of DM.[2-10] Despite being a serious late effect of cancer therapy, susceptibility to DM is highly variable, which limits risk prediction and prevention efforts. Therefore, the objective of this study is to develop a risk prediction model for DM among childhood cancer survivors enrolled in the Childhood Cancer Survivor Study (CCSS) by leveraging both clinical and genetic factors. **We aim to build a simple risk prediction model that ultimately could be used to identify survivors of childhood cancer at highest risk for diabetes upon completion of cancer treatment.**

Diabetes Mellitus in Childhood Cancer Survivors: A recent report by Mostoufi-Moab et al. using data from the CCSS indicated that the risk of DM was nearly two times higher in survivors compared with siblings (relative risk=1.9, 95% confidence interval [CI]: 1.6-2.4).[11] Notably, the risk of DM in survivors of childhood cancer appears to be greatest among those treated with abdominal irradiation and/or total body irradiation (TBI). Two studies reported an increased risk of DM in children with Wilms tumor treated with abdominal irradiation,[3,10] and several studies identified an increased prevalence of DM among survivors exposed to TBI as part of the conditioning regimen for allogeneic bone marrow transplantation (BMT).[2,4,5,7-9] Data from the CCSS provide additional evidence that childhood cancer survivors exposed to abdominal irradiation have an increased risk of DM.[6] Compared to unaffected sibling controls, survivors of high-risk neuroblastoma were 7-times more likely to report DM (odds ratio [OR]=6.9, 95% CI: 3.5-13.9), whereas survivors of Wilms tumor and Hodgkin lymphoma (HL) were twice as likely to report DM (OR=2.1, 95% CI: 1.1-4.0 and OR=2.1, 95% CI: 1.2-3.5, respectively). Furthermore, among these same cancer diagnoses, there was no increased risk of DM in survivors who were spared abdominal irradiation. Among those who received TBI, there was also a strong and increased risk of DM compared to siblings (OR=12.6, 95% CI: 6.2-25.3), particularly among survivors of acute myeloid leukemia (AML) exposed to TBI (OR=17.7, 95% CI: 6.4-49.4).[6] However, these factors do not fully explain why some survivors develop DM, while others do not.[11]

Variability in development of DM in survivors, even among those with similar diagnoses and treatment exposures, suggests that genetic susceptibility is an additional modifying factor. Through a number of genome-wide association studies (GWAS), genetic factors related to risk for DM have been identified.[1,12,13] We know that both type I and type II DM are polygenic traits, with unique genetic loci identified for type I DM[14] and type II DM.[12,13] Based on our preliminary studies, we hypothesize that the genetic risk for DM among survivors may

overlap factors related to risk in the general population but may also include unique loci that pose additional risk based on treatment exposures.

An important gap in clinical care for childhood cancer survivors is incorporating clinical and genetic factors into robust risk prediction models for specific long-term complications, such as DM. Therefore, the objective of this proposal is to develop a clinical and genetic risk prediction model for DM among survivors of childhood cancer. We hypothesize that an integrated clinical and genetic risk prediction model will be superior to risk prediction models that rely on clinical factors alone or genetic factors alone.

Primary Aim

1. **Develop an integrated clinical and genetic risk prediction model for DM among childhood cancer survivors.**
    a. Evaluate the ability of traditional diabetes risk factors measured at survivorship baseline and clinical characteristics of cancer diagnosis and treatment to predict DM in childhood cancer survivors.
    b. Evaluate the ability of genetic variants associated with diabetes in the general population and unique to diabetes risk among childhood cancer survivors to predict DM in childhood cancer survivors.
    c. Determine improvement in risk prediction by the addition of genetic factors to a clinical model.
2. **Validate the integrated clinical and genetic risk prediction for DM among childhood cancer survivors in an independent population.**

Analytic Framework

This analysis will use existing data within the CCSS to address our specific aim. The analytic methods described below will be finalized with input from CCSS statisticians and collaborators.

Study Population: The study population will consist of the 5,173 childhood cancer survivors enrolled in the Original CCSS Cohort (diagnosed 1970-1986) with genotype data available through dbGaP. In this eligible population, there are 394 survivors with self-reported DM as of the June 1, 2017 data release. Validation populations are outlined below (see "Validation").

Outcome: Diabetes cases will be defined in a consistent manner as previous analyses of diabetes in CCSS: those who report being told by a health care practitioner that they have diabetes or the use of medication related to diabetes management (CTCAE Grade 2+). Non-diabetics will be considered CCSS participants who never report diabetes diagnosis or medication in follow-up examinations. Individuals reporting diabetes at baseline will be excluded from primary analyses, but may be considered in secondary or sensitivity analyses.

Demographic and Clinical Risk Factors: In our risk prediction model, we will include traditional risk factors that are considered to modify any individual's risk of diabetes, as well as clinical risk factors for DM identified in the CCSS and other studies:

- Cancer diagnosis
- Year of cancer diagnosis
- Age at cancer diagnosis
- Age at Baseline and Follow-up 1 – Follow-up 5 (baseline age to be used for risk prediction; follow-up age may be used for refining risk prediction in secondary analyses)
- Sex
- Genetically determined ancestry (calculated ancestry–specific principal components)
- Height at Baseline and Follow-up 1 – Follow-up 5 (baseline BMI to be calculated for risk prediction; follow-up measures may be used for refining risk prediction in secondary analyses)
- Weight at Baseline and Follow-up 1 – Follow-up 5 (baseline BMI to be calculated for risk prediction; follow-up measures may be used for refining risk prediction in secondary analyses)
- Radiation therapy field (any, brain, abdominal, and total body) and dose
- Chemotherapy (any, alkylating agents, anthracyclines, corticosteroids) and dose (for alkylating agents and anthracyclines)

The optimal representation for each cancer characteristic will be chosen based on the type of coding that individually offers the best improvement in area under the curve (AUC).

Analysis: Our primary analysis will be the development of a diabetes risk prediction model that incorporates clinical and genetic factors. We will build and present results using **receiver-operating characteristic (ROC) curves** using predicted values from logistic regression models. To build a comprehensive risk prediction model we will employ a forward selection procedure and select covariates based on improvement in AUC. We will begin with traditional risk factors and subsequently include cancer treatment characteristics to build a clinical risk prediction model; our approach is shown in Example Table 1, below. We will retain risk factors that improve AUC by ≥0.01.

Example Table 1. Forward-selection of covariates in a clinical risk prediction model of diabetes in the CCSS. Each row of data represents a model that is inclusive of risk factors named in previous rows.

	Logistic regression model		**Risk prediction**	
Predictor	**OR**	**P-value**	**AUC**	**DeLong's pairwise p-value**
Age at last follow-up, years				
Sex, female vs male				
Body mass index, kg/m^2				
Genetic estimates of ancestry				
Cancer diagnosis, category				
Age at cancer diagnosis, years				
Radiation therapy				
Chemotherapy				

Next, we will develop genetic risk scores (GRSs). Specifically, we will construct both unweighted and weighted GRSs for diabetes based on genetic loci identified in 1) population-based genome-wide association studies (**known loci GRS**) and 2) genome-wide association studies among childhood cancer survivors (**novel loci GRS**). Genetic variants associated with diabetes risk generally have similar effect sizes, which excludes cases of inherited metabolic disorders related to diabetes (see Figure).[13] For this reason, we will construct an unweighted GRS as the sum of risk-conferring alleles at a known and/or novel locus. We will also assess weighted GRSs where the weights are derived from population-based GWAS for the known loci and CCSS discovery population for the novel loci. From our GWAS of diabetes in CCSS we have observed that variants tend to have slightly larger effect sizes than estimated in the general population, but that effect estimates also do not widely vary between individual variants. For these reasons, we anticipate the unweighted GRS will be most appropriate for obtaining robust estimate of diabetes risk prediction in childhood cancer survivors.

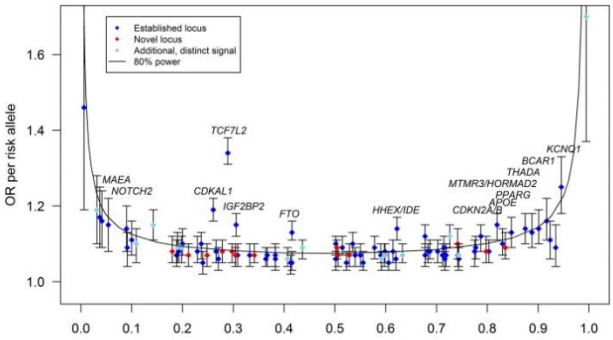

Figure. Genetic effect estimates for diabetes by risk allele frequency in the DIAGRAM consortium.

In the general population, there are 69 previously identified genetic loci that contribute to diabetes risk (see Appendix). **It is well-established that increasing the number of risk-conferring genetic loci, even below statistically significant thresholds, improves phenotype risk prediction**.[15] Therefore, we will explore risk prediction for loci **suggestive** of diabetes association in our CCSS discovery population and have identified 29 genetic loci (P<1e-6) that may explain a cancer survivor's predisposition to develop diabetes. For each unique risk locus and source (population-based vs. childhood cancer survivors), we will count the number of risk-conferring alleles across all loci to create an unweighted and weighted GRS. Any known variants that are identified in both populations will only be counted in the known loci GRS. From the best clinical risk prediction model, we will then calculate the improvement in AUC for three types of GRS: 1) known loci GRS only, 2) novel loci GRS only, and 3) known loci GRS and novel loci GRS. We will also assess the AUC for unadjusted models of the known loci GRS and novel loci GRS individually. Any model with AUC >0.7 will be considered to have good predictive value for diabetes risk among childhood cancer survivors. DeLong's pairwise p-value will

be calculated to statistically test the fit of models with a GRS to the best clinical model. The best clinical model and GRS will be validated in an independent sample of childhood cancer survivors.

Power: We used the easyROC webtool (http://www.biosoft.hacettepe.edu.tr/easyROC/) to calculate power to show non-inferiority of the GRS-extended prediction model compared to the clinical risk prediction model. We have 99% power to detect a 0.05 improvement in AUC given our available sample size in the CCSS. For validation of the clinical and GRS-extended prediction models, we need at least 15 cases and 75 controls to have 80% power to show AUC of 0.7.

Validation: We propose two populations for validation. (1) *CCSS Expansion Cohort* – approximately 3,000 childhood cancer survivors diagnosed between 1987 and 1999 are included in the CCSS Expansion Cohort with whole genome sequencing data now publicly available. Assuming a modest 2% prevalence of DM in a childhood cancer survivor population, we anticipate 60 cancer survivors will have developed DM (16 treated with abdominal irradiation; 44 not treated with abdominal irradiation). Dr. Armstrong (co-investigator) will facilitate the incorporation of data from this assessment. (2) *St. Jude LIFE Study* – an ongoing cohort of 3,006 long-term survivors being following at St. Jude Children's Research Hospital. There are currently 760 participants in the St. Jude LIFE Study who have an abnormal glucose metabolism based on clinical assessment. Dr. Robison (co-investigator on this application) will facilitate the incorporation of data from this assessment. An alternative validation strategy is to divide the participating cohorts into random subsets for discovery and validation. This approach may be employed if temporal trends between the cohorts seems evident.

REFERENCES

1 Flannick, J. & Florez, J. C. Type 2 diabetes: genetic data sharing to advance complex disease research. *Nature reviews. Genetics* **17**, 535-549, doi:10.1038/nrg.2016.56 (2016).

2 Baker, K. S. *et al.* Diabetes, hypertension, and cardiovascular events in survivors of hematopoietic cell transplantation: a report from the bone marrow transplantation survivor study. *Blood* **109**, 1765-1772, doi:10.1182/blood-2006-05-022335 (2007).

3 Cicognani, A. *et al.* Abnormal insulin response to glucose following treatment for Wilms' tumor in childhood. *European journal of pediatrics* **156**, 371-375 (1997).

4 Hoffmeister, P. A., Storer, B. E. & Sanders, J. E. Diabetes mellitus in long-term survivors of pediatric hematopoietic cell transplantation. *Journal of pediatric hematology/oncology* **26**, 81-90 (2004).

5 Lorini, R. *et al.* Hyperinsulinemia in children and adolescents after bone marrow transplantation. *Bone marrow transplantation* **15**, 873-877 (1995).

6 Meacham, L. R. *et al.* Diabetes mellitus in long-term survivors of childhood cancer. Increased risk associated with radiation therapy: a report for the childhood cancer survivor study. *Archives of internal medicine* **169**, 1381-1388, doi:10.1001/archinternmed.2009.209 (2009).

7 Neville, K. A., Cohn, R. J., Steinbeck, K. S., Johnston, K. & Walker, J. L. Hyperinsulinemia, impaired glucose tolerance, and diabetes mellitus in survivors of childhood cancer: prevalence and risk factors. *The Journal of clinical endocrinology and metabolism* **91**, 4401-4407, doi:10.1210/jc.2006-0128 (2006).

8 Shalitin, S. *et al.* Endocrine dysfunction and parameters of the metabolic syndrome after bone marrow transplantation during childhood and adolescence. *Bone marrow transplantation* **37**, 1109-1117, doi:10.1038/sj.bmt.1705374 (2006).

9 Taskinen, M., Saarinen-Pihkala, U. M., Hovi, L. & Lipsanen-Nyman, M. Impaired glucose tolerance and dyslipidaemia as late effects after bone-marrow transplantation in childhood. *Lancet (London, England)* **356**, 993-997, doi:10.1016/s0140-6736(00)02717-3 (2000).

10 Teinturier, C. *et al.* Diabetes mellitus after abdominal radiation therapy. *Lancet (London, England)* **346**, 633-634 (1995).

11 Mostoufi-Moab, S. *et al.* Endocrine Abnormalities in Aging Survivors of Childhood Cancer: A Report From the Childhood Cancer Survivor Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **34**, 3240-3247, doi:10.1200/jco.2016.66.6545 (2016).

12 Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47, doi:10.1038/nature18642 (2016).

13 Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888-2902, doi:10.2337/db16-1253 (2017).

14 Onengut-Gumuscu, S. *et al.* Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nature genetics* **47**, 381-386, doi:10.1038/ng.3245 (2015).

15 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649, doi:10.1093/hmg/ddy271 (2018).

APPENDIX

Appendix Table. Genetic index variants for loci associated with diabetes in population-based genome-wide association studies: UK Biobank, and DIAGRAM 2016[12] and 2017[13] meta-analyses.

SNP	Chr	Pos	PVAL	COHORT
rs41463147	1	120554134	5.13E-09	UKBB
rs340882	1	214145731	1.25E-08	UKBB
rs4846569	1	219771721	8.80E-09	DIAGRAM_2017
rs1260326	2	27730940	2.16E-10	UKBB
rs17387355	2	36465267	1.74E-08	UKBB
rs6757251	2	43734847	1.90E-10	DIAGRAM_2017
rs10193447	2	60552476	1.30E-08	DIAGRAM_2017
rs2972145	2	227101309	6.32E-11	UKBB
rs11712037	3	12344730	8.60E-13	DIAGRAM_2017
rs1496653	3	23454790	8.62E-10	UKBB
rs2014830	3	50172397	2.75E-09	UKBB
rs7428936	3	64710850	1.00E-08	DIAGRAM_2017
rs11708067	3	123065778	8.80E-13	DIAGRAM_2017
rs4402960	3	185511687	2.70E-25	DIAGRAM_2017
rs7651090	3	186996086	2.00E-11	DIAGRAM_2016
rs1046319	4	6304286	1.51E-16	UKBB
rs60780116	4	185708807	7.40E-08	DIAGRAM_2017
rs6885132	5	14768092	1.06E-09	UKBB
rs28650790	5	55861464	7.40E-10	DIAGRAM_2017
rs2307111	5	75003678	6.23E-08	UKBB
rs116454070	5	101837362	8.46E-08	UKBB
rs78408340	5	102338739	2.38E-11	UKBB
rs41302867	6	7240876	5.79E-12	UKBB
rs7451008	6	20673880	3.80E-37	DIAGRAM_2017
rs1264372	6	30769726	7.70E-09	UKBB
rs2260051	6	31591918	1.03E-15	UKBB
rs3134603	6	32126002	2.08E-16	UKBB
rs9273363	6	32626272	9.88E-38	UKBB
rs116647495	6	33614871	1.20E-08	UKBB
rs11759026	6	126792095	5.80E-10	DIAGRAM_2017
rs4719433	7	15065003	1.84E-11	UKBB
rs849142	7	28185891	3.97E-15	UKBB
rs10954284	7	130114298	1.20E-08	DIAGRAM_2016
rs13262861	8	41508577	5.42E-13	UKBB
rs3802177	8	118185025	1.70E-17	DIAGRAM_2017
rs62530366	8	145536056	1.90E-08	DIAGRAM_2017
rs10965247	9	22132729	4.89E-21	UKBB
rs9410573	9	84311800	8.55E-13	UKBB
rs10760280	9	126112812	7.30E-08	DIAGRAM_2017
rs11257659	10	12309269	2.70E-08	DIAGRAM_2017
rs810517	10	80942620	1.30E-12	DIAGRAM_2017
rs10882098	10	94444793	1.40E-26	DIAGRAM_2017

rs535931506	10	97334990	7.11E-08	UKBB
rs34872471	10	114754071	5.28E-125	UKBB
rs2292626	10	124186714	1.80E-12	DIAGRAM_2017
rs2237895	11	2857194	2.90E-17	UKBB
rs5213	11	17408404	4.23E-11	UKBB
rs1061810	11	43877934	5.30E-09	DIAGRAM_2017
rs11602873	11	72460762	4.79E-10	UKBB
rs7933855	11	92323970	1.30E-09	DIAGRAM_2016
rs76895963	12	4384844	1.05E-16	UKBB
rs4931479	12	27948615	8.03E-08	UKBB
rs2612069	12	64501559	7.70E-08	DIAGRAM_2016
rs2258238	12	66221060	1.89E-09	UKBB
rs56348580	12	121432117	2.50E-08	DIAGRAM_2017
rs500443	13	80748024	2.59E-11	UKBB
rs34715063	15	38873115	9.56E-08	UKBB
rs4774420	15	62117975	2.70E-08	DIAGRAM_2017
rs952471	15	77776498	4.00E-10	DIAGRAM_2017
rs9936385	16	52376670	4.70E-11	DIAGRAM_2016
rs1558902	16	53803574	4.70E-25	DIAGRAM_2017
rs2917677	16	69750849	2.92E-09	UKBB
rs8056814	16	75252327	3.70E-11	DIAGRAM_2017
rs2925979	16	81534790	2.70E-08	DIAGRAM_2017
rs78761021	17	9780387	5.50E-08	DIAGRAM_2017
rs4239217	17	36098987	1.51E-12	UKBB
rs3111316	19	13038415	6.09E-08	UKBB
rs429358	19	45411941	1.40E-10	DIAGRAM_2017
rs2023681	22	30599562	3.90E-09	DIAGRAM_2017