Application Details

Funding Opportunity:
Project Scheme: 2016 1st Live Pilot (2016-03-01)

Applicant:

Surname	Given Names	Participant Type
Goldenberg	Anna	Independent Researcher - New/Early Career Investigator

Institution	Faculty	Department
Hospital for Sick Children (Toronto)	Faculty of Genetics and Genome Biology	

Telephone	Fax	E-mail
6478082692		anna.goldenberg@sickkids.ca

Title:
Identifying integrative biomarkers of risk and resilience to childhood psychopathology

Primary location where research to be conducted: Hospital for Sick Children (Toronto)

Faculty:

Department:

Institution which will administer project funds (Institution Paid):
Hospital for Sick Children (Toronto)

Budget:
Total Requested Amount: $770000

Certification Requirements

☑ Human subjects ☐ Human stem cells ☐ Animals ☐ Biohazards

☐ Environmental Impact ☐ Containment Level _____

☐ Clinical Trial

☐ Contains a randomized trial

☐ In order to carry out the proposed research in this application, an exemption from Health Canada under Section 56 of the Controlled Drugs and Substances Act is required. Should my application be approved, I understand that I will need to seek an exemption from Health Canada and provide this exemption to CIHR before any funding will be released for this application.

Partnered/Integrated Knowledge Translation (iKT) projects: Special consideration

Knowledge translation or commercialization project requires a partner AND/OR a knowledge user? ☐ Yes ☑ No

☐ Partner and knowledge user ☐ Partner only ☐ Knowledge user only

Other Project Information

☐ For statistical purposes, does this application propose research involving Aboriginal people?

☑ Are sex (biological) considerations taken into account in this study?

☐ Are gender (socio-cultural) considerations taken into account in this study?

Please describe how sex and/or gender considerations will be considered in your research proposal:
Our study collects psychiatric and genetic and epigenetic measurements on both male and female children. There is a great variability among trajectories of mental disorders and in some cases, sex of a child makes a difference to their trajectory. Sex of each child in the study will be used as one of the covariates in every model we develop. The results will be analyzed and the conclusions regarding how much influence there is on the trajectory will be assessed and established. Ultimately, this will be one of the important findings of our study.

Other Applicants

Surname		Given Names
Kobor		Michael
Role		**Participant Type**
Co-Applicant		Independent Researcher - Senior Investigator
Institution	**Department**	**Faculty**
University of British Columbia	Medical Genetics	Faculty of Medicine

Surname		Given Names
Meaney		Michael
Role		**Participant Type**
Co-Applicant		Independent Researcher - Senior Investigator
Institution	**Department**	**Faculty**
McGill University/Université McGill		

Surname		Given Names
Mostafavi		Sara
Role		**Participant Type**
Co-Applicant		Independent Researcher - New/Early Career Investigator
Institution	**Department**	**Faculty**
University of British Columbia	Statistics	Faculty of Arts and Science

363979
Goldenberg, Anna
Hospital for Sick Children (Toronto)

Descriptors *

DNA methylation, computational medicine, data integration, disease trajectory, machine learning, mental disorders

Areas of Research *	**Classification Codes ***
Primary GENOMICS, PROTEOMICS, AND BIOINFORMATICS	Primary MULTIDISCIPLINARY
Secondary MENTAL HEALTH	Secondary MENTAL AND BEHAVIOURAL DISEASES

Themes *	**Suggested Institutes ***
1st Social/Cultural/Environmental/Population Health	1st Neurosciences, Mental Health and Addiction
2nd Biomedical	2nd Human Development, Child and Youth Health
3rd	3rd
4th	4th

*** For Administrative Purposes Only**

Suggested External Referee(s)*		
Name Eric, Nestler	E-mail	eric.nestler@mssm.edu
Area of Expertise neuroscience, epigenetics, psychiatry		
Name Greiner, Russel	E-mail	russ.greiner@gmail.com
Area of Expertise		
Name Gunnar, Megan	E-mail	gunnar@umn.edu
Area of Expertise social-emotional and brain development, neurobiology		
Name Hanna, Greg	E-mail	ghanna@umich.edu
Area of Expertise child and adolescence, genetics		
Name Kennedy, James	E-mail	james.kennedy@camh.ca
Area of Expertise psychiatry, molecular genetics		
Name	E-mail	
Area of Expertise		
Name	E-mail	
Area of Expertise		
Name	E-mail	
Area of Expertise		
Name	E-mail	
Area of Expertise		
Name	E-mail	
Area of Expertise		
Name	E-mail	
Area of Expertise		
Name	E-mail	
Area of Expertise		

*** For Administrative Purposes Only**

Lay Title and Lay Abstract

Lay Title:
Identifying early indicators of children at risk for mental disorders

Lay Abstract:
Mental disorders constitute the largest contributor to the global burden of disease as measured using the disability-adjusted life years index. The most common mental disorders, including major depression and conduct disorders show a peak age of onset in later childhood and adolescence thus derailing the quality of life and productivity of individuals over entire lifetimes. However, by identifying at risk children at an early age we have an opportunity to intervene and prevent many of such mental disorders. The challenge is in effectively identifying truly vulnerable children to be able to intervene in a timely manner. Current programs for identifying at risk children are constructed on evidence linking early life adversity, such as poverty or birth outcomes, and the risk for mental illness. These factors predict mental illness at the level of the population, but are inefficient at the level of the individual due to the considerable variability in outcomes: many children born early, small, or into poverty are healthy and productive. The goal of this project is to identify robust and predictive biological and psychological measurements that help to predict which children are at risk early on. In particular, by assaying molecular markers that capture the biological mechanisms associated with specific psychiatric outcomes, we will identify factors that differentiate those who do and do not succumb to childhood mental illness given the same set of known early life risk factors. This study will thus inform targeted, and personalized interventions aimed at both reducing the severity of onset of these disorders and the negative outcomes that accompany them, such as suicide and substance abuse.

Complete Summary

Mental disorders are the largest contributor to the global burden of disease as measured by the disability-adjusted life years index. The most common mental disorders, such as depression and drug addiction, show peak age of onset in later childhood and adolescence, derailing one's quality of life and productivity over an entire lifetime. Mental disorders occur in response to both heritable genetic influences and environmental conditions. The latter are reflected in epigenetic signals (such as DNA methylation) which might serve as the basis for identifying vulnerability at the individual level. The goal of this project is to find robust biomarkers for identifying at risk children early by combining early life socio-emotional data and genomic measurements acquired in a non-invasive manner.

We have collected PsychChip data and longitudinal behavioral and DNA methylation data for two cohorts totalling over a thousand children. This rich data allows us to explore two sets of novel biomarkers: socio-behavioral trajectories, defining each child as a sequence of her socio-behavioral measurements, and DNA methylation markers. Jointly these indicators constitute an unprecedented level of fine-grained data that will underlie our models for personalized predictors for risk of mental illness.
The main challenge in identifying robust predictors of diagnosis is computational: integration of longitudinal clinical data with high-dimensional genomic measurements requires novel methods that can address heterogeneity in the data and alleviate the multiple testing burden to increase statistical power. In Aim 1 of this project we will develop novel trajectories of socio-behavioral assessments from birth to 72 months old, by combining longitudinal psychometric indices across multiple tests and time points. In Aim 2 we will identify variable DNA methylation regions that are predictive of diagnosis within each socio-behavioral trajectory while using prior knowledge about the genome structure to reduce the burden of multiple testing. In Aim 3 we will assess the power of the novel markers compared to the standard cross-sectional psychiatric indicators and genetic biomarkers to establish the most powerful non-invasive markers.

We have assembled a diverse team of experts to ensure the success of this project. Dr Goldenberg is a machine learning expert, a leader in developing data integration methods and models of heterogeneity in complex human diseases. Dr Mostafavi is a statistician, expert in modeling psychiatric diseases using genomic data. Dr Meaney is an expert in biological psychiatry, neurology, and neurosurgery whose studies had impact on public policy and economic health. Dr Kobor is a leading molecular biologist working to decipher the relationship between environmental exposures, gene regulation and disease across the lifecourse.

This grant will develop innovative methodology that will take advantage of uniquely rich longitudinal epigenetic, environmental, and phenotypic data to identify integrative biomarkers that will improve on the current state-of-the-art early detection of mental disorders. We will ascertain factors that differentiate those who do and do not succumb to childhood mental illness given the same set of known early life risk factors. This study will thus inform targeted, and personalized interventions aimed at both reducing the severity of onset of these disorders and their negative outcomes, such as suicide and substance abuse.

Complete Application - Quality of the Idea

Mental disorders develop in response to genetic and environmental conditions. Epigenomic biomarkers, such as methylation, are inherited, respond to the environment and have also been found to be useful indicators of mental health risk. Developmental factors such as childhood behavior and social and emotional well being also point to a child's risk of mental illness. While these sources of information have been extensively studied independently, no studies have taken advantage of the true richness of these data sources jointly to enhance mental disorder risk prediction of risk on an individual basis.

Our goal is to identify integrative biomarkers that are predictive of patterns in social-emotional development across early childhood through late adolescence associated with increased lifetime risk of psychiatric disorders. To this extent we will develop a novel computational framework as follows:

Aim 1. Generate longitudinal early childhood social-emotional trajectories for each child and determine trajectory subtypes (i.e. prominent patterns of social-emotional development in kids, scientifically novel and clinically relevant);

Aim 2. Develop predictive model of social-emotional trajectory subtypes determined in Aim 1 by integrating DNA methylation and genotype data from each individual; (novel computational advance with translational potential)

Aim 3. Identify DNA methylation probes that persistently predict social-emotional behavior trajectories over time and are thus likely to be most useful as clinical tests (novel biological finding with translational potential).

The tools and the analysis pipeline developed in this project will help clinicians predict children who are likely to develop a mental illness later in life.

Complete Application - Importance of the Idea

Context. Mental illness costs the Canadian economy $49 billion per year [1]. More than two thirds of mental health issues have their onset during childhood or adolescence [2]. Identifying children at risk for mental illness later in life and predicting the type of illness is not easy. There are no blood or genetic 'tests' for mental illness. Instead, psychiatrists and health professionals use individual's and his/her family's clinical history to take note of persistent symptoms and make a diagnosis [3]. This process can place a timely and accurate diagnosis out of reach, leading many with mental illness to go untreated for >10 years [4].

Knowledge gap. There exist interventions that can improve the health outcomes of people with mental illness, especially when provided early on. For the targeted treatments to reach the individuals that need them early, we need better ways of effectively identifying children at risk for mental illness, and accurately predicting changes in their symptoms over time.

Project contributions. This project will identify the biological measures and biomarkers we need to 'test' for and predict a child's social-emotional development in early childhood and adolescence in order to assess their risk of developing mental illness. Our results will allow clinicians to provide appropriate intervention(s) early, thereby improving the health outcomes of at-risk individuals. Specific advances made by this project will be:

More accurate definitions of typical and atypical early childhood social-emotional trajectories and corresponding subtypes. Predicting an illness requires that its characteristics are well defined. This project will develop a platform to define subtypes of childhood social-emotional development using longitudinal, multifactorial trait-based measures. These subtypes will be informative of natural divisions in social-emotional development. Using psychiatric diagnoses identified across childhood and adolescence, we will be able to evaluate the way in which certain disorders may show distinct patterns of development. Subtypes discovered through this study will advance psychiatric knowledge of the early behavioral signs of mental illness. This information can be applied in clinical psychiatric assessments of children to improve diagnostic precision.

Identification of genetic biomarkers that predict early childhood trajectories. These biomarkers will be valuable in the clinic and in the health research community: 1) A clinician can measure and use these to assess and inform an individual about the expected progression of their mental health - improving health outcomes. 2) The location of the identified biomarkers and their proximity to specific genes and regulatory regions will advance knowledge of the biological origins of psychiatric differences among individuals. 3) Our findings will implicate both short and long term predictors and biological contributors of psychopathology helping to unravel the complex relationship between genomics and social-emotional development. These findings will also generate candidate biological targets for novel treatments.

Impact on the health research community. The easy-to-use computational tools developed to ascertain longitudinal subtypes and their associated biomarkers will be made publicly available to the larger research community. Providing this software will advance the work of the larger Canadian health research community.

Complete Application - Approach

Identifying the environmental, biological and behavioral precursors that differentiate individuals with psychiatric disorders is a predominant goal in psychiatry[5–7]. Finding consistent biological associations with psychiatric disorders is challenging because they are heterogeneous disorders of polygenic and multifactorial origin[8],[9],[10]. An approach proposed by the National Institute of Mental Health (NIMH) is to use continuous traits that cut across diagnostic boundaries [11]. In our work [12], we found that integrating genetic and epigenetic data identifies more clinically relevant subtypes. DNA methylation (DNAm) is known to differentiate patients with and without psychosis [13]. DNAm biomarkers are useful because they capture inherited patterns and changes due to environmental exposures. They also provide clues to regulatory changes in gene expression and biological pathways that open the door to targeted psychiatric therapies[14, 15].

We hypothesize that, by combining trajectories of social-emotional measures and genome-wide genetic and DNA methylation data, we can identify clinically relevant subtypes of childhood and adolescent development and the corresponding biomarkers that predict traits underlying psychiatric disorders.

To link subtypes to genetics, we will: **Aim 1**. Generate longitudinal childhood-adolescent social-emotional development (SED) subtypes; **Aim 2**. Integrate multidimensional genetic and DNAm data to build a model that predicts SED; **Aim 3**. Identify clinically relevant biomarkers that robustly predict SED subtypes. The combined expertise of our team and richness of our longitudinal birth cohort data are a unique opportunity to uncover these patterns and impact the lives of Canadians now and in the future.

Data. This project uses 5 longitudinal birth cohorts: MAVAN (Maternal Adversity, Vulnerability and Neurodevelopment; Meaney PI, Kobor co-PI)[16], QNTS (Quebec Newborn Twin Study), WSFW (Wisconsin Study of Families and Work; Kobor co-PI)[17, 18], GUSTO (Growing Up in Singapore Toward Healthy Outcomes; Meaney co-PI)[19], and ALSPAC (Avon Longitudinal Study of Parents and Children)[20]. We have genetic (Illumina PsychArray BeadChip) and methylation (Illumina Infinium HumanMethylation450 BeadChip in whole blood ALSPAC, buccal in other datasets) data across all individuals. Since it is not possible to collect brain tissues for large-scale psychiatric genomic studies; accessible tissues must be used. Buccal DNAm profiles were shown to closely resemble brain DNAm profiles[21]. The psychiatric evaluation data contains information from the Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS;QNTS), the electronic Preschool Age Psychiatric Assessment (ePAPA;MAVAN), the Structured Clinical Interview for DSM Disorders (SCID;GUSTO), and the Developmental Well Being Assessment and the Clinical Interview Schedule-Revised (DAWBA and CIS-R;ALSPAC). Social-emotional data includes Infant Behavioral Questionnaire, Early Childhood Behavioral Questionnaire, Infant Toddler Social Emotional Assessment (ITSEA), Child Behavior Checklist (CBCL), Social Behavior Questionnaire, Adaptive Social Behavior Inventory, Children's Behavior Questionnaire, Strengths and Difficulties Questionnaire, Preschool Behavior Questionnaire and the Health and Behavior Questionnaire. Home environment and physical health measures are from parental questionnaires. We will also conduct a follow-up acquisition between the ages of 13-15 years by 2019 in MAVAN. Fig 1 details acquisition timepoints for each cohort and data type.

Relevant published methodology. We developed a novel algorithm to combine various patient (or individual) based data, the Similarity Network Fusion (SNF) method. SNF successfully identifies clinically relevant subtypes of 5 different cancers [12]. The key innovation of SNF is to represent each type of information available for the individuals (e.g. internalizing and externalizing behaviors) as a network where each individual is a node and each edge represents similarity of these individuals using each type of available information, respectively (Fig 2). We will use this approach to integrate trajectory similarities stemming from each individual trait measurement (Aim 1).

Preliminary Data

Generating univariate trajectories of SED measures. Composite internalizing and externalizing scores from ITSEA and CBCL were transformed to percentiles using the relative rank of 95 children from MAVAN (a proof-of-concept sample). Each of their internalizing and externalizing trajectories were clustered using a combined distance and shape dissimilarity metric[22]. We considered from 2 to log2(n)=6 clusters and identified the number of clusters that produced the largest silhouette statistic[23]. 5 and 3 clusters were optimal for externalizing and internalizing behaviors, respectively. Distinct trajectory patterns are evident in each cluster (Fig 3), indicating the existence of true subtypes in the data.

Predicting future points in trajectories. To confirm the clinical potential of our approach, we built models each corresponding to a social-behavioral trajectory subtype (see Fig 3). We used Gaussian Processes (GPs)[24], powerful non-parametric nonlinear generative Bayesian models, to predict subtype-specific patient-level trajectories. We fit 8 GPs corresponding to each subtype. Our approach accurately captures trajectories and future observations (Fig 4), using subtype specific variation among the individuals (Fig 3). We expect this result to be strengthened by incorporating biomarkers into the model.

Predicting subtypes based on DNAm data. Using DNAm probes as input and internalizing and externalizing subtypes (Fig 3) as output, we randomly split the data into 70% train and 30% test sets. On the training set, we did probe selection by selecting the top 0.5% of methylation probes most associated with the subtype assignments. Seven classifiers were trained on the training set: Decision tree, Random forest, Lasso, Ridge, and Elastic net linear regression, SVM (linear kernel), SVM (radial basis function kernel). Performance of the trained classifiers was evaluated on the test set. We repeated this procedure 100 times to obtain confidence intervals of subtype prediction performance, and obtain accurate subtype classification (Fig 5). We will further improve subtype classification/prediction by Integrating SNPs and DNAm data with novel, biologically driven classifier(s) (see Aim 2).

Research Plan

The overall goal of this research is to develop a risk model for childhood psychiatric disorders based on longitudinal trait-generated subtypes and their genetic and epigenetic signatures.

Aim 1. Generate longitudinal social-emotional development subtypes (led by Dr Goldenberg, PI)
Rationale: Groups of individuals with or at risk for a particular psychiatric disorder such as depressive disorders and anxiety disorders display persistently different SED patterns from their peers across time[26–29]. For this reason, we believe that robust, biologically and clinically relevant subtypes can be generated by integrating longitudinal SED scores.

Method: We will identify univariate trajectories for each individual in WSFW and QNTS cohorts where traits have been measured from infancy 0yo and 2yo to adolescence 17.5yo and 14yo, respectively (Fig 1), for internalizing, externalizing, inattention, hyperactivity, impulsivity, and prosocial behavior scores. We will use a bootstrapping approach to compute similarity for all possible pairs of individuals in our data over 1000 random draws using a combined distance and shape similarity metric[22]. We will then use SNF[12] to combine all similarity matrices each based on univariate SED obtaining multivariate subtypes. This will generate "core clusters" within the QNTS and WSFW cohorts providing us with robust reference subtypes that represent strong SED patterns. For each subtype, we will develop a GP model that can predict the full SED for each individual within that subtype (see prelim. data and Fig 4).

Evaluation: The global subtypes established with WSWF and QNTS cohorts will be used to identify the subtype of each individual in the MAVAN (PI Meaney, coPI Kobor), GUSTO (coPI Meaney) and ALSPAC cohorts. We will use all but the the last observation for each of the individuals to identify their subtype(s), and then use the GP corresponding to their subtype to predict the last observation for each and evaluate

prediction error.

Following obtaining subtypes, psychiatric disorder diagnoses in MAVAN, GUSTO, and ALSPAC will be mapped to predicted trajectories using psychiatric evaluations in childhood (age 5.5 yrs) and adolescence (age 14-15.5 yrs). This will allow us to characterize how psychiatric disorders distribute and occur over time within each subtype.

Challenges and mitigating strategies: We will need to carefully account for all the confounders when integrating across multiple cohorts. We will test all obtained subtypes for confounding by cohort (batch effect), home environment, ethnicity and measured demographics. If any subtypes are confounded by these measures, we will first regress the effect out and perform trajectory clustering on the residuals of the trajectory scores. Additionally, if our GP does not significantly outperform the baseline predictor of the last observation, we will examine more adaptive GP kernels[30].

Aim 2. Integrate multidimensional genetic and DNAm data to build a model that predicts SED (led by Dr Mostafavi, co-PI)
Rationale: Identifying traits relevant to disease subtypes increases the power of biomarker detection[31]. Clinically relevant biomarkers indicate an individual's symptoms as well as their diagnostic class[32–35]. In this Aim, we will develop a method to identify genetic and epigenetic biomarkers that robustly predict individual SED for each subtype defined in Aim 1.

Method: We have implemented the framework for testing standard machine learning classifiers using DNAm as input and subtypes as output (Fig 5). We will incorporate genetic information and implement a more biologically driven algorithm to more powerfully identify biomarkers. We have developed an approach based on linear mixed-effects models (LMM) and kernel machines[36] designed to analyze high dimensional (epi-)genomic data [37]. To handle ~500,000 methylation probes, we use a "kernel" approach within the LMM framework to represent a group of correlated and spatially co-located CpGs as one kernel. Our approach accounts for correlation among nearby DNAm probes and reduces a large multiple testing burden.

To integrate genotype data, we will compute a kernel for a set of SNPs that jointly predict a given gene's gene expression (as expression quantitative trait loci (eQTLs) [38]). We will use PrediXcan [39] to predict gene expression for multiple brain tissues. We will thus have two kernels per gene, one based on DNAm probes and another based on eQTLs, and will get a ranking over predictive ability of the kernels (DNAm probes and eQTLs for a given gene), which will enable us to identify the most promising biomarkers.

Evaluation: We will test our approach on 1000 individuals from ALSPAC to establish the robustness and predictive power of our approach. The DNAm data in ALSPAC (derived from whole blood samples) will be considered separately from the other 4 cohorts. it is an ideal test case due to the large sample size. We will test our method in a 5-fold cross validation setting and make all the kernel parameter choices using this data. Additionally, we will be able to evaluate the robustness of the blood associated biomarkers in a cross validation setting.

Challenges and Mitigating Strategies: Cell type heterogeneity is a major contributor to inter-individual differences in DNAm data. We will address this challenge using reference-free computational approaches [40].

Aim 3. Identify clinically relevant biomarkers that robustly predict SED subtypes
Rationale: To ensure the clinical utility of our findings, we will confirm that identified biomarkers predict an individual's future trajectory irrespective of when the biomarker is measured.

Method: We will collect DNAm data from 178 individuals between the ages of 13-15 years for whom we already have the early DNAm data[41]. With this additional data for the MAVAN cohort, we will have 1 cohort with repeated DNAm measures in early childhood (GUSTO), 1 cohort with repeated DNAm measures before and after puberty (MAVAN), and 2 cohorts with repeated measures after puberty (WSFW, QNTS) (Fig 1).

We will perform gene selection for each available time point across our data collections with the method developed in Aim 2. We will obtain a set of predictive genes and corresponding probes per subtype per time point across cohorts, and identify those that are stably predictive across time. We will rank our probes from most predictive over time to being predictive only for a particular time point, placing the most clinically relevant probes at the top of the list.

Evaluation: Candidate biomarkers that are predictive across multiple cohorts and multiple timepoints before and after age 13 will be considered the most stable and clinically useful measures. Candidate biomarkers that are predictive across multiple cohorts within childhood or adolescence only will be considered stable and replicated for that developmental period. Finally, biomarkers that are predictive in only 1 sample and over one or more time point will remain candidates in need of further investigation.

Challenges and mitigation strategies: Nearly half of our DNAm data is from GUSTO (57% Chinese, 25% Malay, 18% Indian). We will correct for ancestry and/or other confounding variables related to socio-economic status (SES), by using genetic information to derive ethnic similarity between all individuals in all cohorts and the SES information collected in the home environment questionnaire data as random effects in our model.

Deliverables and Impact

We will generate subtypes of early childhood social-emotional development using longitudinal, multifactorial trait-based measures [years 1-3, led by Dr Goldenberg and implemented by a Goldenberg lab postdoc (PD1)]. These subtypes will parallel natural SED divisions that persist across ethnic and diagnostic boundaries. We will develop and make publicly available an easy-to-use novel computational method that integrates DNAm and genetic data to predict SED subtypes (years 1-4, led by Dr Mostafavi and implemented by a Mostafavi lab postdoc). We will find biomarkers that robustly predict the identified SED subtype trajectories and their future behavioral scores (years 3-4, led by Dr Goldenberg, co-supervised by PD1, and implemented by Ms Erdman, Goldenberg lab graduate student). The location of the identified biomarkers and their proximity to specific genes and regulatory regions will provide insight into the biological etiology of SED and candidate biological targets for novel medications.

Finally, we will assess the predictive accuracy of our candidate DNAm biomarkers when measured across childhood and adolescence, the period during which most psychiatric disorders have their onset. Our findings will refine the transient and sustained mental health outcomes within our SED subtype trajectories. We will also implicate short and long term predictors and biological contributors of psychopathology to unravel the complex relationship between genomics, SED and psychiatric outcomes.

Complete Application - Expertise, Experience and Resources

Participants:

Name	Role	Institution
Anna, Goldenberg	Nominated Principal Applicant	Hospital for Sick Children (Toronto)
Department	Faculty	Contribution
N/A	Faculty of Genetics and Genome Biology	8

Name	Role	Institution
Michael, Kobor	Co-Applicant	University of British Columbia
Department	Faculty	Contribution
Medical Genetics	Faculty of Medicine	2

Name	Role	Institution
Michael, Meaney	Co-Applicant	McGill University/Université McGill
Department	Faculty	Contribution
N/A	N/A	2

Name	Role	Institution
Sara, Mostafavi	Co-Applicant	University of British Columbia
Department	Faculty	Contribution
Statistics	Faculty of Arts and Science	5

Complete Application - Expertise, Experience and Resources

Expertise. Dr Goldenberg is a computer scientist, an expert in machine learning and is internationally recognized for integrating omic and clinical data. Dr Mostafavi (co-PI) is a statistician and a Canada Research Chair (II) in Computational Biology. She is an expert in modeling psychiatric diseases using genomic data. Dr Meaney (Co-Applicant) is a scientific director at the Ludmer Centre for Neuroinformatics and Mental Health at the Douglas Mental Health University Institute, James McGill Professor and Director of the Program for the Study of Behaviour, Genes and Environment at McGill. He is an expert in biological psychiatry, neurology, and neuroscience. Dr Meaney will play an integral role in the project by providing data (PI MAVAN, co-PI GUSTO) and critical expertise on developmental traits and psychiatric disorders. Dr Kobor is an expert on the relationships of environmental exposures and disease. Dr Kobor (Co-Applicant) has collected DNA methylation on some of the largest social-behavioral cohorts across Canada and will provide data (co-PI on MAVAN, GUSTO, WSFW), his expertise in epigenetic profiling, processing and relating DNA methylation to social-behavioural traits.

Experience. Dr Goldenberg will lead the interdisciplinary project team. She leads multiple interdisciplinary projects linking genetic information to clinical data, and produces open-source software widely used by the research community[12, 31, 42, 43]. Dr Mostafavi led and participated in a variety of collaborative projects that produced methods for robust and accurate genomic data analyses (e.g., DGN[44], GeneMANIA[45, 46]). Dr Meaney leads several studies that capture neurocognitive and socio-behavioral trajectories of childhood development and provides psychiatric assessment for these children. Drs Kobor and Dr Meaney are co-applicants on multiple grants studying human epigenetics in the context of developmental human trajectories. Dr Goldenberg collaborates with Dr Meaney and has already received data from his lab.

Team. Dr Goldenberg will co-ordinate this interdisciplinary team. A postdoc (P1) and graduate student Lauren Erdman in Dr Goldenberg's lab will be responsible for Aims 1 and 3, collaborating with Dr Meaney's team developing trajectories of social-behavioral traits. Dr Mostafavi will supervise a postdoc (P2) to develop the methodology and perform the analysis in Aim 2 that will collaborate with Dr Kobor's lab on the analysis of the DNA methylation data.

Meetings. There will be bi-weekly meetings to discuss progress on the grant between the labs of Drs Goldenberg and Mostafavi. Bi-monthly Skype meetings with Drs Meaney and Kobor labs to present tutorials on the computational tools being developed. Finally, there will be yearly get together meetings for all the trainees involved to get hands on experience having both the computational and psychiatric experts in the room to facilitate knowledge exchange.

Resources. Dr Goldenberg has access to high-performance computing infrastructure at SickKids Centre for Computational Medicine (10200 threads and 3 PB of storage). This facility is the 2nd largest high speed compute cluster in Canada. Each member of her lab has a new computer, presently 8 computers total. Dr Mostafavi's lab is equipped with high performance mac desktop computers for algorithm development and data visualization, her lab also has access to the WestGrid computer facility for scalable computation.

Complete Application - References

1. Deraspe, R: Current Issues in Mental Health in Canada: The Economic Impact of Mental Illness, http://www.parl.gc.ca/Content/LOP/ResearchPublications/2013-87-e.pdf, (2013).

2. The human face of mental health and mental illness in Canada, (2006).

3. Martin, E.A...Lewandowski, K.E: Social functioning and age across affective and nonaffective psychoses. J. Nerv. Ment. Dis. 203, 37–42 (2015).

4. Wang, P.S...Ustün, T.B: Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry. 6, 177–185 (2007).

5. Cuthbert, B.N., Insel, T.R: Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 11, 1–8 (2013).

6. Alkadhi, K: Brain Physiology and Pathophysiology in Mental Stress. International Scholarly Research Notices. 2013, (2013).

7. Blair, R.J.R: Neurobiological basis of psychopathy. Br. J. Psychiatry. 182, 5–7 (2003).

8. Sullivan, P.F...Neale, M.C: Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch. Gen. Psychiatry. 60, 1187–1192 (2003).

9. Gottesman, I.I., Gould, T.D: The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatry. 160, 636–645 (2003).

10. Inoue, K., Lupski, J.R: Genetics and genomics of behavioral and psychiatric disorders. Curr. Opin. Genet. Dev. 13, 303–309 (2003).

11. Insel, T...Wang, P: Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry. 167, 748–751 (2010).

12. Wang, B...Goldenberg, A: Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods. 11, 333–337 (2014).

13. Dempster, E.L...Mill, J: Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum. Mol. Genet. 20, 4786–4796 (2011).

14. Heyn, H., Esteller, M: DNA methylation profiling in the clinic: applications and challenges. Nat. Rev. Genet. 13, 679–692 (2012).

15. Levenson, V.V: DNA methylation as a universal biomarker. Expert Rev. Mol. Diagn. 10, 481–488 (2010).

16. O'Donnell, K.A...Meaney, M.J, MAVAN Research Team: The maternal adversity, vulnerability and neurodevelopment project: theory and methodology. Can. J. Psychiatry. 59, 497–508 (2014).

17. Hyde, J.S….Clark, R: MATERNITY LEAVE AND WOMEN'S MENTAL HEALTH. Psychol. Women Q. 19, 257–285 (1995).

18. Essex, M.J...Kobor, M.S: Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. Child Dev. 84, 58–75 (2013).

19. Soh, S....Saw, S.M., GUSTO Study Group: Cohort profile: Growing Up in Singapore Towards healthy

Outcomes (GUSTO) birth cohort study. Int. J. Epidemiol. 43, 1401–1409 (2014).

20. Golding, J., ALSPAC Study Team: The Avon Longitudinal Study of Parents and Children (ALSPAC)--study design and collaborative opportunities. Eur. J. Endocrinol. 151 Suppl 3, U119–23 (2004).

21. Smith, A.K...Binder, E.B.: DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. Am. J. Med. Genet. B Neuropsychiatr. Genet. 168B, 36–44 (2015).

22. Chouakria, A.D., Nagabhushan, P.N: Adaptive dissimilarity index for measuring time series proximity. ADAC. 1, 5–21 (2007).

23. Rousseeuw, P.J: Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Comput. Appl. Math. 20, 53-65. J. Comput. Appl. Math. 20, (1987).

24. Rasmussen, C.E., Williams, C.K.I: Gaussian Processes for Machine Learning. the MIT Press (2006).

25. Roberts, S...Aigrain, S: Gaussian processes for time-series modelling. Philos. Trans. R. Soc. Lond. A. 371, 20110550 (2013).

26. Rende, R.D: Longitudinal relations between temperament traits and behavioral syndromes in middle childhood. J. Am. Acad. Child Adolesc. Psychiatry. 32, 287–290 (1993).

27. Warner, M.B ...Grilo, C.M: The longitudinal relationship of personality traits and disorders. J. Abnorm. Psychol. 113, 217–227 (2004).

28. Blanchard, J.J., ..., Brown, S.A: Diagnostic differences in social anhedonia: a longitudinal study of schizophrenia and major depressive disorder. J. Abnorm. Psychol. 110, 363–371 (2001).

29. Schachar, R., Logan, G.D: Impulsivity and inhibitory control in normal development and childhood psychopathology. Dev. Psychol. 26, 710 (1990).

30. Herlands, W...Xing, E.: Scalable Gaussian Processes for Characterizing Multidimensional Change Surfaces, http://arxiv.org/abs/1511.04408, (2015).

31. Colak, R...Goldenberg, A: JBASE: Joint Bayesian Analysis of Subphenotypes and Epistasis. Bioinformatics. 32, 203–210 (2016).

32. Kapur, S...Insel, T.R: Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol. Psychiatry. 17, 1174–1179 (2012).

33. Rotzinger, S; Kennedy, S on on behalf of the CAN-BIND Team: The Canadian Biomarker Integration Network for Depression (CAN-BIND): Looking Deeper into Major Depressive Disorder. Issues. 2, (2013).

34. Niciu, M.J...Zarate, C.A., Jr: Biomarkers in mood disorders research: developing new and improved therapeutics. Rev Psiquiatr Clin. 41, 131–134 (2014).

35. Scarr, E...Dean, B: Biomarkers for Psychiatry: The Journey from Fantasy to Fact, a Report of the 2013 CINP Think Tank. Int. J. Neuropsychopharmacol. 18, yv042 (2015).

36. Liu, D...Ghosh, D: Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. Biometrics. 63, 1079–1088 (2007).

Complete Application - References

37. Ng, B...Mostafavi, S: Hidden Statistical Concerns with Interaction Kernel Machines.

38. Lonsdale, J...Moore, H.F: The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).

39. Gamazon, E.R...Im, H.K, GTEx Consortium: PrediXcan: Trait Mapping Using Human Transcriptome Regulation. (2015).

40. Zhang, B., Baosen, Z., Trott, M.D: Reference-free audio matching for rendezvous. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (2010).

41. Illumina: Infinium MethylationEPIC BeadChip. (2015).

42. Mezlini, A.M...Goldenberg, A: Identifying cancer specific functionally relevant miRNAs from gene expression and miRNA-to-gene networks using regularized regression. PLoS One. 8, e73168 (2013).

43. Goldenberg, A...Morris, Q: Unsupervised detection of genes of influence in lung cancer using biological networks. Bioinformatics. 27, 3166–3172 (2011).

44. Mostafavi, S., Battle, A...Levinson, D.F: Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing. Mol. Psychiatry. 19, 1267–1274 (2014).

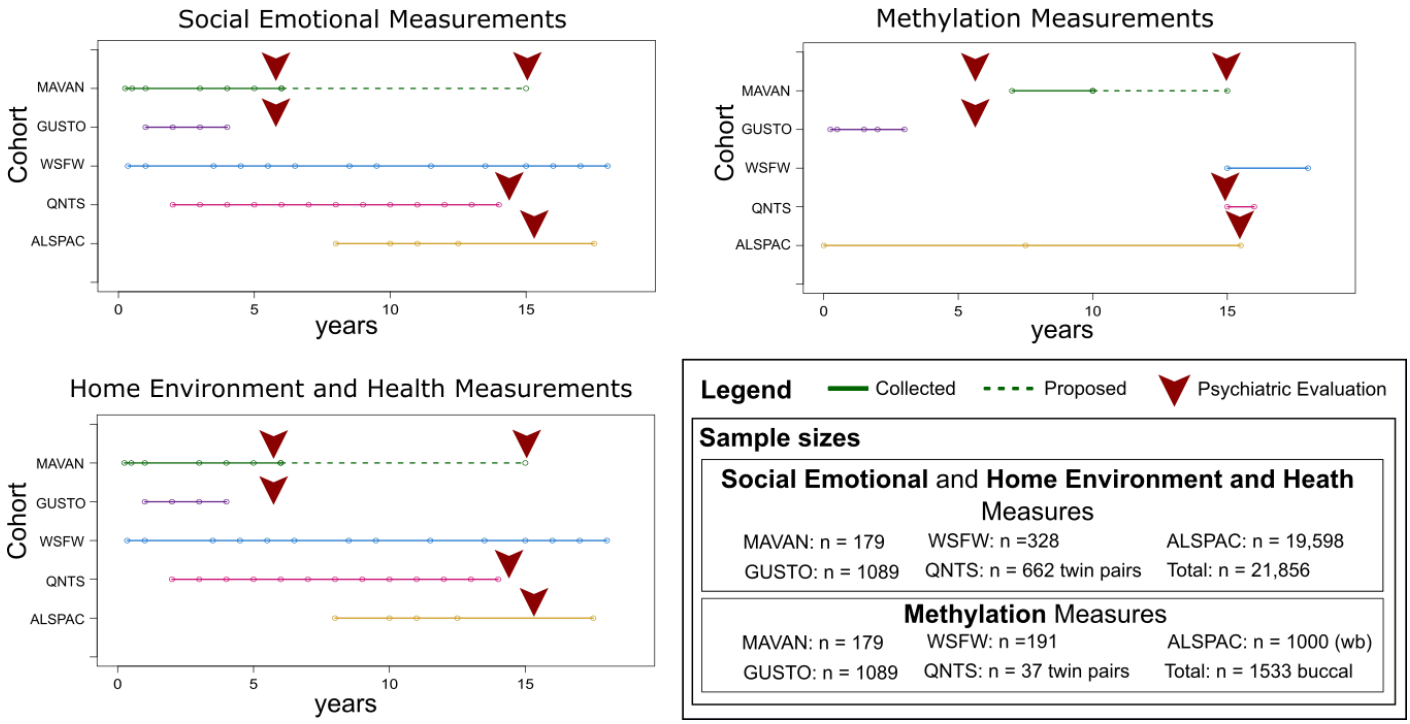

Figure 1. Overview of the data we will use in this project. This study uses five longitudinal birth cohorts across four different countries, Canada (MAVAN, QNTS), Singapore (GUSTO), The United States (WSFW), and Britain (ALSPAC). Social-emotional, home environment, and health measurements are based on questionnaire data. Methylation measurements are from buccal samples in all but ALSPAC which uses whole blood (wb) methylation.

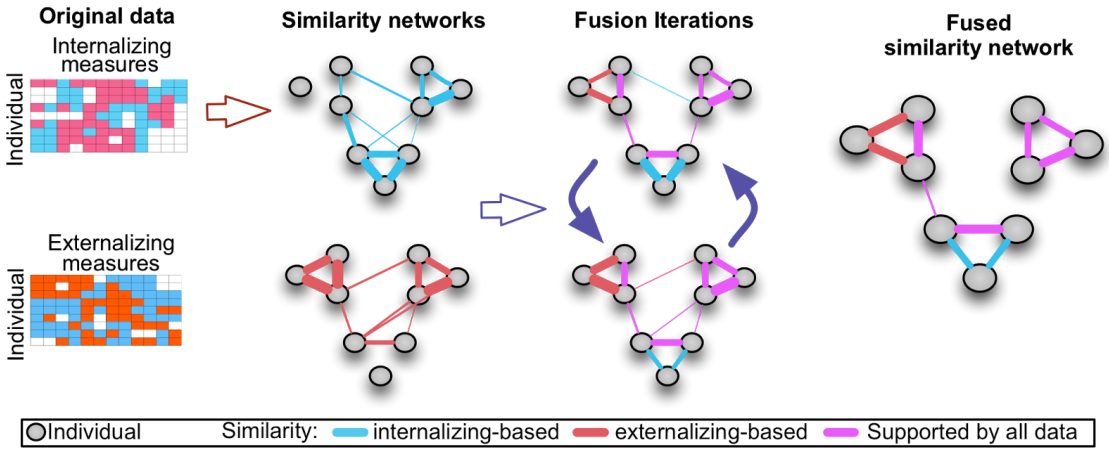

Figure 2. Similarity Network Fusion (SNF) - a toy example. SNF is a data integration method developed by Dr Goldenberg. It will be used to integrate different individual-based trait measures in Aim 1. This method integrates data by iteratively amplifying shared cluster boundaries between the different data types and dampening non-shared cluster boundaries. The resulting fused network represents clusters that are supported by strong divisions in one or more data type. We can construct a similarity network based on internalizing measurements over time, each edge representing similarity in the internalizing trajectories of the individuals over time, and another one based on externalizing measurements over time.

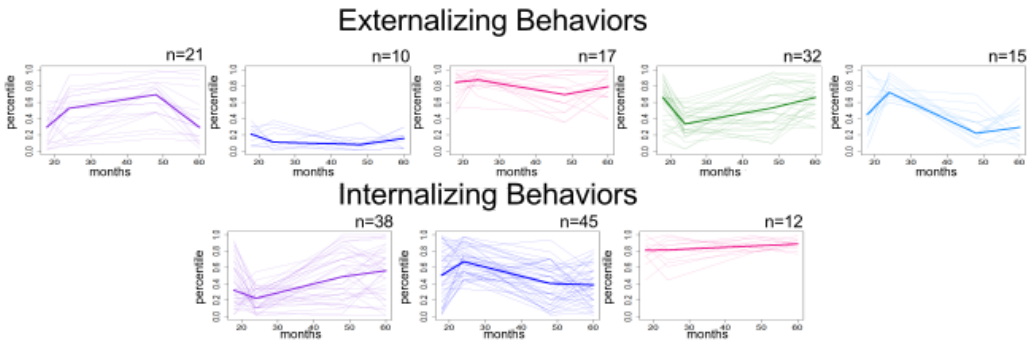

Figure 3. Univariate trajectories of social-emotional trait-based measures. Trajectories of internalizing and externalizing behavior (using ITSEA and CBCL) across early childhood for 95 individuals in the MAVAN cohort for whom preliminary methylation data was available. Even in a small sample, distinct patterns over time are distinguishable.

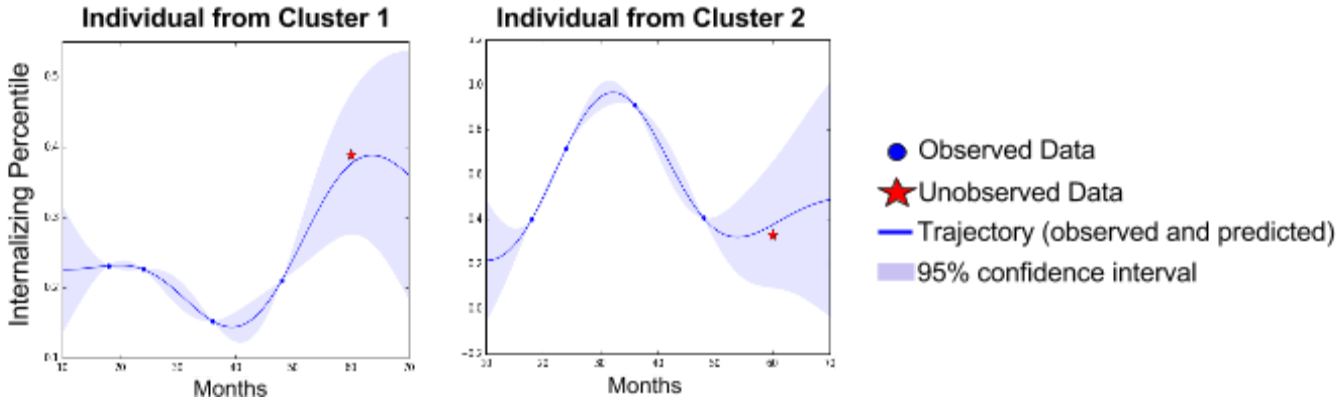

Figure 4. Applying subtype specific Gaussian Processes to predict future observations. The trajectories of 2 individuals from Internalizing clusters 1 and 2 (first and second from the left). We fit one GP with 35 of 38 individuals in the internalizing-behavior subtype 1 (1st from left), another with 40 of 45 individuals in the internalizing-behavior subtype 2 (2nd from left). We then used these GPs to predict future scores (shown as a star) for the 60th month for one of the children whose data was not used to fit GP parameters.

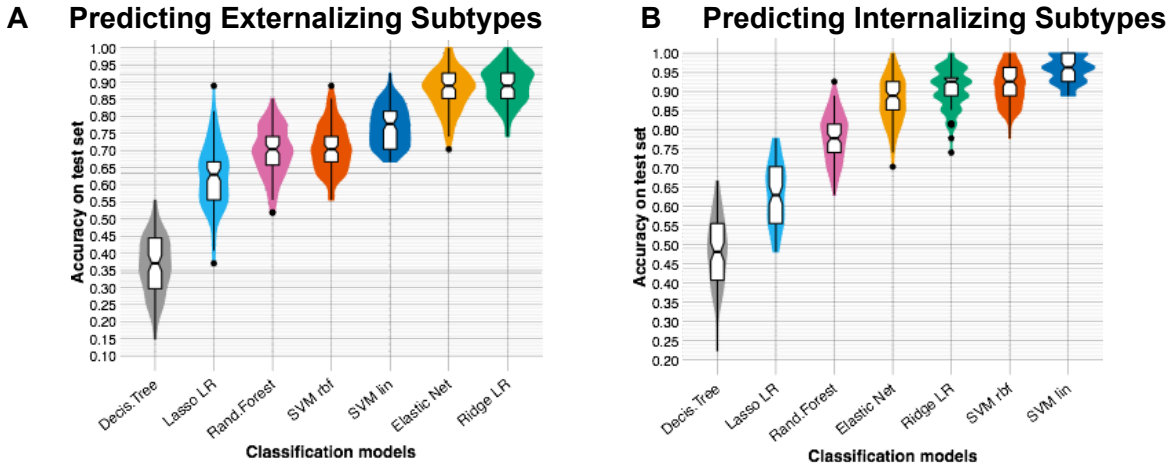

Figure 5. Predicting socio-behavioral subtypes. We trained 7 classifiers to predict externalizing (A) and internalizing (B) subtypes as identified in Fig 2 based on non-time dependent methylation probes. Classifiers are ranked based on accuracy with 92% and 96% being the maximum accuracy achieved for externalizing and internalizing subtypes respectively.