Project Viva Genetic Data Policy

Background/Rationale

In recent years, there have been ongoing advances in technologies and knowledge in the field of genetic determinants of health and common conditions affecting mothers and children, such as obesity, cardiometabolic disorders, allergies and respiratory diseases. Microarrays covering millions of genetic variants across the human genome are now available at a relatively affordable price, and knowledge of the human genome has expanded our ability to impute data to offer broad coverage of known common and rare genetic variants across the whole genome.

The Project Viva genome-wide genotyping dataset was completed using Illumina Infinium Core Exome-24 genetic microarrays for 834 mothers and 798 children who had provided consent for genetic analysis of biospecimen samples. We can use these rich datasets for analyses within the Project Viva cohort and investigate genetic determinants of maternal and child health. In addition, Project Viva contributes to larger analyses by offering the possibility to replicate findings from other cohorts, or participate in meta-analyses involving data from several cohorts. Project Viva benefits from participation in these studies as a result of increased opportunities to: (1) establish and build on collaborations with others in the field; (2) contribute to manuscripts published in high-impact journals; (3) train junior investigators in genetic research and meta-analytic methods; and (4) use the study data and resources more effectively by contributing to pooled studies, which often produce higher-quality results than single-cohort studies limited by inadequate power.

This policy applies to use of **de-identified, genetic** data from Project Viva. Individual-level data will be coded with Project Viva study ID and will be available only to IRB-approved Viva analysts and key project staff. Viva analysts will tabulate individual-level data into summary statistics for analysis by approved investigators and consortia. In genetics, results of associations between genetic variants and phenotypes of interest are aggregated (usually expressed as betas or odds ratios) and presented as a "summary statistic" for each genetic variant analyzed in that population. Any proposals requiring PHI or other identifiers (including dates) will be governed by a separate policy (see the *Project Viva Policies* document for more information on Grant Applications, Analyses, Ancillary Studies and Publications/Authorship).

Sources of Data

Study Population

This Project Viva Genetics study protocol includes data from a subset of the Project Viva cohort. Specifically, this policy applies to data from both mother and child participants who:

a. Provided blood cell samples at one or more in-person visits
b. Consented to use of their biospecimen samples for genetic analysis

All procedures for data collection from Project Viva participants are approved under protocols #235301 and #1830598. Project Viva biospecimen samples are labeled with study ID number, but no other identifiers.

The genetic data includes all genotyping that was conducted in prior targeted genetic analyses (see **E**arly **G**rowth **G**enetics (EGG) IRB protocol #233502 "EGG 95th Percentile BMI Study") and genotype data derived from the Illumina Infinium Core Exome-24 microarrays conducted with University of North Carolina funding under protocol #868145: Choline, Genetic Polymorphisms and Cognition in Children. Individual-level data in these genetic datasets are identified by Study ID only

Phenotypes

Project Viva has collected detailed phenotypes concerning maternal and child health during pregnancy, at delivery, and in post-partum (mother), and during childhood. The list of phenotypes and health outcomes collected are available in protocols #235301 and #1830598.

Genetic Analyses fall into one of two categories:

I. Project Viva specific analyses: Viva investigators leading studies that investigate the genetic determinants of maternal and child health in the Viva population

II. Meta-analyses and Replication Analyses: Project Viva contributing summary data to meta-analyses or for replication analyses led by non-Viva investigators, including in the context of consortium

Process for Proposing a Genetic Study

I. Project Viva specific analyses

Analysis Plans

All Viva investigators proposing a genetic study are required to present an analysis plan at a monthly Viva Co-Investigator meeting for scientific review and approval. To protect our participants and the quality and integrity of Project Viva data, Project Viva Principal Investigators and/or designated co-investigators must approve all requests. Following the presentation, the Decision-Making Group (DMG) will either: 1) approve the analysis plan; 2) recommend that the proposing author revise and resubmit the plan and/or present again at a future meeting; or 3) decline Project Viva participation in the proposed analysis.

- The proposing investigator can receive de-identified, tabulated, preliminary results from Viva's Senior Statistical Analyst to prepare the proposal.
- We recommend communication with a Viva biostatistician before presenting the proposal.
- The proposing investigator provides the title of the analysis plan to Viva's Project Manager at least one week prior to the designated Co-Investigator meeting, and submits an electronic copy of the analysis plan and any handouts by 8:00 a.m. on the morning of the meeting.

Specific guidelines for analysis proposals and a thorough explanation of the proposal review, and approval process – including a list of required documentation from investigators – can be found in the **Policies for Using Project Viva Data** on ProjectViva.org.

Requesting and Receiving Genetic Summary Results

Upon obtaining scientific approval and meeting all the above requirements, the investigator may request summary result from the Viva analyst; the Project Viva Program Manager should be copied on this request.

Upon receiving Project Viva genetic summary results, the proposing investigator agrees to follow all policies as outlined in this document.

II. Consortia and non-Viva investigators conducting meta-analyses

Analysis Plans and Scientific Approval

Investigators not directly affiliated with Project Viva and colleagues from external cohorts and multiple-cohort consortia may also request Project Viva to conduct replication analyses or meta-analyses of genetic associations.

Consortia have a process by which consortium investigators review newly proposed projects. Once a project is approved by the respective consortium, the lead investigator requests tabulated results (also called "summary results"), often in a pre-specified format, from replication and/or all participating cohorts. Approval of a proposal by the external consortium does not confer a commitment from Project Viva to contribute to the analysis. When a consortium proposes a new project, the Project Viva liaison to that consortium will provide a copy of the request and analytic plan to Project Viva PI Emily Oken and/or Co-PI Marie-France Hivert. These investigators will review each proposal for scientific merit and will determine whether Project Viva has the required data elements and analytic resources to contribute to the proposed analysis.

Independent investigators and colleagues from other cohorts who are not part of a consortium follow a similar approval process whereby the proposing investigator communicates their request and provides an analysis plan for approval by Project Viva's PI, Dr. Emily Oken and/or Co-PI, Marie-France Hivert. The Viva PIs review the proposal for scientific merit and determine whether Project Viva has the required data elements and resources necessary to contribute to the analysis.

Required Documentation

Upon approval of a request for Project Viva genetic replication or meta-analysis participation, the Project Viva Project Manager or Viva analyst will contact the proposing investigator to request the following documentation:

1. Acknowledgement of and agreement to abide by this *Project Viva Genetic Data Policy*
2. Documentation of IRB or equivalent ethics approval of the proposed project by the investigator's home institution (if required by that institution), or a determination that approval is not required.
3. Documentation of the investigator's human subjects training certification, if required by his/her home institution.
5. Investigator's country of citizenship
6. Investigator's home institution and country of home institution
7. Documentation that the analysis has been registered in PROSPERO or another public protocol repository

Use of Data

Investigators may view summary results that are provided by the Project Viva analyst.

Project Viva IRB-approved analysts are responsible for creating summary results from genetic data and phenotypic data stored in the Viva Data Repository, in accordance with the analytic plan provided by the lead investigator and approved by the Viva Pis. The data shared with investigators for analysis will be in the form of summary results only; there will be no individual-level data. Analysts may also assist investigators with non-genetic analyses if necessary. Analyst(s) providing summary results may be included as co-author(s) on papers as appropriate.

The investigator shall destroy Project Viva summary files created and distributed by the Viva analyst within one year after resulting manuscripts are accepted for publication. This timeframe allows the investigator to respond to any changes or request for additional results based on the initial manuscript review.

To further safeguard participant privacy and confidentiality, the Viva analyst will:

- Include only necessary results in summary tables.
- Analysts will share tabulated (summary) results via encrypted emails or secure data transfer technology approved by HPHC, such as the "Box" collaboration application.
- Emails containing results will include the following message:

  "The recipient has read Project Viva's Policies and agrees to abide by them. The recipient agrees to use or disclose the data only for the purpose requested, and for no other purpose. The recipient agrees to use appropriate safeguards to prevent any use or disclosure of the data. The recipient will report to Project Viva's Program Manager any violation of this agreement or Viva policies."

Data Analyses for Approved Analysis Plans

Genetic data analyses

Genetic association analyses and genetic replication and meta-analyses will be performed by IRB-approved Viva analysts to protect genetic information of participants. Project Viva will only provide summary results, i.e., statistics resulting from analysis of individual-level data. For example, the Project Viva analyst would perform analyses on the data to generate a mean and standard deviation, or a beta coefficient and p-values from a regression model. The analyst will provide these statistics, rather than an individual-level dataset, to the proposing investigator. All analysts working with Viva genetic data must be approved by the HPHC IRB on this genetics protocol. In most cases, proposals will also require information on covariates. The Project Viva analyst will use phenotypic data stored in Project Viva's general Data Repository (approved

under protocol #228471) in combination with genetic data to generate the requested summary statistics.

Before proceeding with the publication process for an approved analysis, investigators are expected to follow additional steps regarding providing data updates to the Decision-Making Group, and having manuscript drafts reviewed. Details on these steps can be found in the **Policies for Using Project Viva Data document at ProjectViva.org, in sections IV.A.4, IV.B, IV.C, and V.**

Data Storage and Security

Genetic datasets with individual-level data are identified by Study ID only. The Project Viva Data Manager stores these datasets in a secure folder on the Harvard Pilgrim Health Care Institute J:/ drive, with access limited only to authorized users. Genetic data is also stored on the Harvard University O2 server, which is Harvard Security Level 3 data compliant. The O2 server is protected by two-factor authentication. Only approved analysts included on this protocol will have access to these genetic datasets to conduct analyses of individual-level data and to generate summary statistics in response to requests. Analysts will share tabulated (summary) results using secure technology (secure emails, SFTP sites, or other secure data transfer technology available).

All e-mails containing results or data will include the following message:

> *The recipient has read Project Viva's Policies and agrees to abide by them. The recipient agrees to use or disclose the data only for the purpose requested, and for no other purpose. The recipient agrees to use appropriate safeguards to prevent any use or disclosure of the data. The recipient agrees to destroy any dataset provided by the Lead Research Analyst within the time frame outlined in these policies. The recipient will not attempt to re-identify the data or merge the data with non-approved datasets. The recipient will report to Project Viva's Program Manager any violation of this agreement or Viva Policies.*

Investigators receiving Viva genetic summary results will be expected to abide all Project Viva policies as outlined in this document. Any shared data or results may only be used for the purpose originally requested and approved. Project Viva's policies require that all Project Viva data and tabulated results are destroyed by the recipient within one year after acceptance of their final manuscript for publication.

Tracking Approved Requests

The Viva Program Manager and/or the Viva Analyst will track all requests for contributions of Project Viva data. The following information will be maintained:

- Project title
- Lead investigator name, title and contact information
- Lead investigator institution

- Aims and hypotheses
- Type of data (variables or summary statistics) requested
- IRB approval from investigator's home institution (if required by that institution)
- Record of completion of training in human subjects research (if required by the investigator's home institution)
- Acknowledgement of *Project Viva's Genetic Data Policy*
- Date of transfer of data or summary results to the proposing investigator
- Resulting publication

Project Viva will make this information available to the HPHC IRB upon request.

NOTE: This policy provides general guidance. Each proposal will be considered individually by Project Viva's operational leadership; specific requirements may differ from what is listed above.

Project Viva Staff Contact Information

Name	Title	Email	Phone
Sarah Cohan	Program Manager	Sarah_Cohan@hphci.harvard.edu	(617) 867-4968
Sheryl Rifas-Shiman	Lead Research Analyst	Sheryl_Rifas@hphci.harvard.edu.org	(617) 867-4824