Memorandum of Understanding

Genomic Aggregation Project in Sweden (GAPS)

9 June 2016

Background

In the last decade, genomic studies have increased from hundreds of subjects at one site to global consortia with hundreds of thousands of subjects. This has dramatically improved power and has yielded large numbers of genetic associations for many human traits (e.g., type 2 diabetes, cancers, cardiovascular disease, immune diseases, schizophrenia, educational attainment). While tremendous advances have been made, genetic heterogeneity across samples continues to hinder these efforts.

Sweden has a relatively homogeneous population and comprehensive national registers. We estimate that there are now GWA data on ~250,000 Swedes, WES on 25,000, and WGS on 3,000. A Swedish genomic "mega-consortium" offers many benefits: a Swedish universal control group; Sweden-specific imputation panels; electronic medical record research; studies of register-based phenotypes; GxE studies of traditional epidemiological exposures with genomic data; revealing the population history of Sweden; and finding people with knockouts for important human genes. *This could easily expand to population scale precision medicine.*

This is an obvious idea. A key step is to create a framework that respects the needs and agendas of PIs who put enormous effort into their studies. We thus propose creation of a modern genomics consortium. PIs will never lose control of their data, and will always know how their data are being used. PIs can choose to leave the consortium at any time with their data erased within 24 hours.

Initial procedural and scientific aims (tentative, subject to review and approval by GAPS):

1	***Consortium building: create organizational and scientific framework—a modern genomics consortium that is transparent, participatory, and democratic.***
	• Recruit studies willing to participate (ethics consistent, N>500 subjects, existing SNP array data) • Finalize a Memorandum of Understanding (MOU, an informal agreement on governing principles)
2	***Infrastructure creation: develop the necessary informatics and computational organization.***
	• Use secure UPPMAX compute cluster (at Uppsala University) • Document ethical approval. Database allowable uses for each study • Create database for study metadata and subject characteristics and phenotypes • Process GWA data using an existing and robust QC and imputation pipeline • Establish working groups on imputation references, *HLA* imputation, chrY, CNVs (etc.)
3	***Initial studies to establish consortium viability and scientific value (finalized in discussion with GAPS PIs)***
	• Tentative ideas. Publish three papers (e.g., large Sweden GWA of educational attainment, conscription data (BMI, blood pressure), and fertility. Papers on these have been in top journals • Population allele frequencies could enable case-only studies. UNICORN (Universal Control Repository Network, PMID: 27087321) can be leveraged for genomic studies without individual control data. • Develop and disseminate Swedish imputation reference

Completion of these aims will accomplish multiple intentions. The scientific output is important of itself and as is outstanding pilot data for grants. We would also show that this is a viable and "real" consortium. Critically, it could form the basis for population-scale precision medicine. There are now

similar large studies in progress in the US, UK, Denmark, Finland, Estonia, and Iceland. With some organization and cooperative spirit, Sweden could improve on the work of all of these studies.

Basic expectations

This MOU describes the basic expectations of participants. Many are now familiar with genomics consortia. The usual rules and key expectations for members are straightforward. There are four basic rules:

1. "treat others as you would like to be treated"
2. "no surprises"
3. "don't use other people's data against them to gain a competitive advantage"
4. "no direct competition on the core purpose of the consortium".

If you participate in GAPS, you are free to participate in other consortia. As with the "no surprises" rule, just let people know.

Governance

The intention is that this will be a transparent, participatory, and democratic organization. We will have regular meetings/teleconferences. Study PIs (one per group) will form a management team to guide activities.

Data control and data security

UPPMAX (Uppsala Multidisciplinary Center for Advanced Computational Science) is Uppsala University's high-performance compute resource for large-scale storage and genomic computing. UPPMAX is part of the Swedish National Infrastructure for Computing. We propose a secure part of UPPMAX for GAPS.

A limited set of GAPS analysts will be able to access the genomic data. Who these people are and what they are doing will be transparent. With funding, we intend that these individuals would be employed directly by GAPS and will do genomic analyses for GAPS investigators.

The basic control of data shared with GAPS will be codified with a Data Transfer Agreement (DTA). A DTA is a legal and binding contract between UPPMAX and the PI's institution. It controls how the data shared with GAPS data can and cannot be used. It will contain a critical provision – if a PI chooses to leave GAPS, all data that were shared will be erased within 24 hours.

Data processing

Upon sharing of uncleaned/raw genomic data, it will be processed using the "ricopili" pipeline developed by the Psychiatric Genomics Consortium. The author of "ricopili" is Dr. Stephan Ripke of Charité Universitätsmedizin Berlin, and he is a collaborator on this project. Following standardized quality control, the data will be imputed to the best current reference. Participating studies will get back their own cleaned and imputed data that can be used for whatever purpose they choose. Data from all participating studies will be aggregated for GAPS analyses.

Analyses

There will be an initial set of analyses that will be agreed upon by GAPS management team. After GAPS gets established, any PI or group member can propose a GAPS analysis. Analysis plans should be approved by the PIs prior to initiating studies involving their samples. Proposals will be circulated to the relevant PIs in advance of conducting the analyses and should be returned with any comments

or concerns within two weeks. No comments within two weeks means approval. If there are substantial issues with a proposal, the proposal is tabled for discussion at the next call. Investigators can ask to become part of the team for an analysis and proposers should include them.

Authorship

We will follow standard and widely accepted criteria for authorship on scientific papers (ICMJE criteria http://www.icmje.org). They require: (a) substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (b) drafting the work or revising it critically for important intellectual content; AND (c) final approval of the version to be published; AND (d) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

The collection PIs will generally meet these conditions and be named authors when their samples were used. Other authors will be included as appropriate given their involvement in that project.

Presentations prior to publication (i.e. conferences and conference calls) should be approved in advance. Abstracts, posters, and presentations to be submitted for conferences should be circulated to the PIs and coauthors at least one week in advance. Feedback should be returned promptly. No response within one week indicates approval.

Dispute resolution

We will have an efficient approach to solve any disputes that might arise. First, we discuss the issue thoroughly. Our experience is that we can almost always find consensus (e.g., to test whether one approach is better than another). Reasonable people behaving reasonably can usually find reasonable solutions. Second, if no consensus is reached after sufficient discussion, we vote (majority rule, one vote per contributing group). Third, if the dispute remains unresolved, we agree to discuss the issue with an impartial mediator in order to broker a resolution to the dispute.

Indication of agreement

This is a “gentleman’s agreement”, relying on the honor of the members for fulfillment. If you would like to participate in GAPS, indicate your agreement with this MOU by sending an email to Sarah Bergen (sarah.bergen@ki.se).