Memorandum of Understanding

Topic: Participation in the Psychiatric Genomics Consortium (PGC4)

Version: 10 October 2022

Web site: http://pgc.unc.edu

Executive Summary

- Describes the next set of aims for the PGC.
- Updates prior PGC MOUs from 2008, 2011, 2015, and 2018.
- We request that all prior PGC researchers and new participants indicate agreement with this MOU.
- We will continue to collect cases (and controls) with GWAS genotypes for all disorders. We will use the GWAS and CNV pipelines for integrated quality control and analysis as the basis for analyses for each disorder and then across disorders.
- A new major intention of the PGC is facilitation of additional analyses. New mega-analyses for the basic case-control definitions occur every 2-4 years for each group – these are effortful and procedurally complex. An increasingly large portion of PGC papers are based on follow-on analyses. The PGC supports and encourages investigator-initiated/driven papers.
- The PGC is committed to diversifying sample collections and advancing methods for trans-ancestry analysis.

Introduction

We created the Psychiatric Genomics Consortium (PGC, http://www.med.unc.edu/pgc) in 2007 to conduct mega-analyses for attention-deficit hyperactivity disorder (ADHD), autism (AUT), bipolar disorder (BIP), major depressive disorder (MDD), and schizophrenia (SCZ). At later dates, we added new groups for eating disorders (ED), obsessive-compulsive disorder/Tourette syndrome (OCD/TS), posttraumatic stress disorder (PTSD), anxiety disorders (ANX), Alzheimer's disease (AD), substance use disorders (SUD), and suicide (SUI).

The PGC is one of the most innovative experiments in the history of psychiatry. We have unified much of the field to enable rapid progress in elucidating the genetic basis of psychiatric disorders. We have 800+ investigators and >400K participants. The PGC has attracted a cadre of outstanding scientists whose careers center on our work.

Now in its 13th year, the Psychiatric Genomics Consortium is one of the largest, most innovative, and productive experiments in the history of psychiatry. The central idea of the PGC is to use a global cooperative network to advance genetic discovery in order to identify biologically, clinically, and therapeutically meaningful insights. The PGC continues to unify the field and attract outstanding scientists to its central mission (800+ investigators from 150+ institutions in 40+ countries). PGC work has led to ***320 papers***, many in high-profile journals (*Nature* 3, *Cell* 5, *Science* 2, *Nat Genet* 27, *Nat Neurosci* 9, *Mol Psych* 37, *Biol Psych* 25, *JAMA Psych* 12). The full results from all PGC papers are freely available, and our findings have fueled analyses by non-PGC investigators and fostered the careers of many junior scientists.

This is the fifth version of this MOU (PGC4), and describes the next set of scientific aims. We wish to continue this highly productive collaboration. Large amounts of new data will be available to the PGC in the next five years. We have developed a rigorous set of approaches that are yielding discoveries. We thus have the unique opportunity to rapidly and efficiently increase our knowledge of common and rare variation in order to understand the causes and comorbidities of major psychiatric disorders.

More information

Updated information (including FAQs and introductions to the PGC for new members) can be found on the PGC web site, http://www.med.unc.edu/pgc .

PGC4 Specific Aims

To advance discovery and impact, we propose a new phase for the PGC. Large amounts of new data are coming in the next five years (from diverse sources). We can thus rapidly and efficiently increase our knowledge of the fundamental bases of major psychiatric disorders via five *Specific Aims*:

Aim 1: *Advancing genetic discovery for severe psychiatric disorders*

a. Increase discovery power via GWAS meta-analyses for severe psychiatric disorders
b. Efficiently and accurately incorporate genetic data/results from biobank studies into PGC meta-analyses
c. Increase the ancestral diversity of PGC meta-analyses through outreach and novel methods

Aim 2: *Advancing discovery across genetic architecture: integrating common and rare genetic variation*

a. Extend the rigorous work of the PGC CNV group to all 11 primary disorders
b. Expand PGC work on whole genome sequencing (WGS) in densely-affected pedigrees to find rare variants
c. Most studies consider GWAS (Aim 1), rare CNV (Aim 2a), and rare exome sequencing in isolation. We will pinpoint specific genes for psychiatric disorders via integration of common and rare variation
d. In CNV carriers, investigate the effects of common variation on the psychiatric diagnosis and clinical symptoms

Aim 3: *Advancing discovery beyond standard diagnostic definitions*

a. Develop and deploy phenotyping instruments to enable well-powered, global trans-diagnostic studies
b. Use genetics to address unanswered questions about psychiatric disorders (Aims 1a, 1c, 3a, 3c). For example, do few vs many genetic factors underlie clinical presentations? How do genetic effects impact common clinical outcomes? How do genetic effects vary through life?
c. Using the PGC's global network, promote new studies of highly severe or treatment-resistant clinical cases to enrich for causative alleles and increase translational potential

Aim 4: *Increase impact of genomic discovery for novel therapeutic and preventative opportunities*

a. Illuminate biology by fine-mapping and integrating functional genomic data with PGC results (Aims 1-2)
b. Identify potentially modifiable causal risk factors for psychiatric disorders using Mendelian Randomization
c. Combine epidemiological and polygenic modeling to predict important outcomes and enable clinically meaningful patient stratification

Aim 5: *Increase impact by extending and formalizing outreach to different communities*

a. Accelerate therapeutic development via PGC-led engagement with industry/academic medicine discovery groups
b. Use digital media to communicate PGC findings in real time
c. Develop, distribute, and update educational resources for patients, families, and medical professionals
d. Establish a scientific advisory board (SAB) to maximize PGC alignment with multiple end-users

Supporting new projects

The PGC supports all types of research! The PGC is open to new ideas and project proposals beyond the PGC4 Specific Aims. We have a structure for approval of new projects within each working group, as well as across groups. We encourage new ideas and have a dynamic organization which can adapt to new methodology and new discoveries.

We fully support investigator-driven ideas for new analyses. Via the PGC data access committee (DAC) and the Data Receiving Committee (DRC), we have streamlined the process for creating secondary analysis proposals, gaining approval from disorder groups, and accessing PGC study data.

An increasingly large portion of PGC papers are based on follow-on analyses. The PGC supports and encourages these investigator-initiated and driven papers.

We routinely provide letters of support for investigator-driven grants (including training awards) for funding agencies worldwide.

Principles of Collaboration

The PGC is centered around an altruistic overarching purpose: leveraging large-scale cooperation for mega-analyses of genomic data to yield definitive knowledge of the etiology of critically important psychiatric disorders.

There are several basic ideas behind the PGC, and all participants need to understand and agree with these tenets.

- Above all else, the basic premise of the PGC is that more can be accomplished together than separately.
- We have always stressed that individual groups should participate in the PGC at the time that is right for them. For many, this is after their initial manuscript has been accepted for publication.
- From the beginning, the PGC has tried to be as inclusive, democratic, transparent, and rapid as possible. No single individual or group dominates.
- The PGC is a global organization and aims to include individuals from around the world and support researchers from all countries who wish to join.
- Individual investigators can become involved to the extent they desire in the workings of the PGC. This usually means the disease group to which they contributed data, the statistical analysis group, the CNV group, the outreach committee, and the cross-disorder group.
- We have evolved an efficient approach to solve disputes. First, we discuss the issue thoroughly. Our experience is that we almost always come to a reasonable consensus (e.g., to test whether one approach is better than another). Second, if no consensus is reached after sufficient discussion, we vote (majority rule, one vote per PI). This has been required only a few times in since 2007. There have been no instances of deadlock or “institutional paralysis”.
- Please remember the cardinal rule of consortia: “no surprises”. If you are doing work that potentially competes with work being pursued by the PGC or by PGC colleagues, please disclose all potential conflicts so that solutions can be found.

Organizational Structure

The PGC organization structure is simple. The PGC consists of a coordinating committee, a data access committee, 12 disorder working groups (one each for ADHD, AUT, ANX, ALZ, BIP, MDD, SCZ, ED, PTSD, SUD, OCD/TS, SUI), the statistical analysis group, the CNV group, the DRC, the Cross Population Group (CPG) and the cross-disorder group, and regional representatives (e.g., Latin America, India, Africa, Asia).

All important decisions are made at the working group level. The role of the coordinating committee is only to adjudicate issues of relevance to the whole consortium and to plan new directions in a coordinated manner. The main tasks of the coordinating committee are to integrate the efforts of the working groups and to secure the resources needed for the project.

<u>Coordinating committee</u>

An admittedly US-centric metaphor for the relationship between the coordinating committee and all groups is the recurrent issue through US history, the tension between Federalism and “states’ rights”. We believe that the best science will emerge from the PGC if balance is strongly shifted towards “states’ rights.” The “federal” coordinating committee has a non-intrusive and facilitating role, and all other decisions are delegated to the scientists who understand the issues best.

Disorder workgroups

The eleven disorder working groups are listed below. These groups will regularly produce papers.

Group	Chairs
ADHD	Ben Neale, Barbara Franke
ALZ	Danielle Posthuma, Ole Andreassen
ANX	Jack Hettema, Jürgen Deckert, Thalia Eley
AUT	Elise Robinson, Anders Børglum
BIP	Ole Andreassen
CPG	Hailiang Huang, Laramie Duncan, Roseann Peterson
ED	Cynthia Bulik, Gerome Breen, Laura Huckins
MDD	Andrew McIntosh, Cathryn Lewis
OCD/TS	Manuel Mattheisen, Jeremiah Scharf
PTSD	Karestan Koenen, Kerry Ressler, Caroline Nievergelt, Murray Stein
SCZ	Mick O'Donovan, James Walters
SUD	Arpana Agrawal, Howard Edenberg, Joel Gelernter
SUI	Anna Docherty, Niamh Mullins, Douglas Ruderfer

It may be helpful for additional disorder groups to be granted "provisional" membership in the PGC (e.g., the Alzheimer's group). Any such groups will only directly participate in PGC activities after until GWAS data are obtained or in progress. We will assist groups in obtaining funding: if a PI provides the coordinating committee with a letter/email of assurance that the research group agrees to this MOU, and providing an outline of a planned project which falls into the domain of PGC's interests, the coordinating committee will provide the PI with a letter of support for a grant application. The letter will affirm the group's agreement to participate in the PGC and will briefly describe PGC's framework and principles.

CNV group

The CNV group co-chairs are Drs. Jonathan Sebat and Steve Scherer. This group is tasked with developing a best-practices QC and analysis pipeline for CNV intensity data.

Cross-disorder workgroup

The cross-disorder group is chaired by Drs Jordan Smoller and Ken Kendler and conducts analyses that cross-cut most of the PGC disorders. Many of these analyses intend to identify genomic regions whose associations undercut traditional DSM- or ICD-based classification systems. This group has representatives from each of the working groups. In addition to the important activities of this group, many cross-disorder analyses are conducted via collaborations between working groups or as secondary analyses.

Statistical analysis/computing workgroup

This group is chaired by Dr. Ripke. Its role is to apply the "ricopili" GWAS QC, analysis, and imputation pipeline. This pipeline has explicit written protocols for uploading data to the cluster farm in the Netherlands that we use for quality control, imputation, analysis, and bioinformatics.

The Network and Pathway analysis sub-group (led by Drs. Breen and Holmans in coordination with the central analysis team) conducts analyses of GWAS results to identify common biological processes or functionally connected gene sets implicated from the GWAS data within and across PGC disease studies.

Computation and data warehousing for the PGC are non-trivial. We have been fortunate to secure the use of a cluster computer in the Netherlands for the PGC. The Genetic Cluster Computer PI is Dr. Danielle Posthuma (a member of the analysis group). The Linux-based GCC

(http://www.geneticcluster.org) has 630 nodes (each with 2 or 8 processors), and has sufficient CPU capacity and storage space for all work needed for the PGC meta-analyses (including imputation and permutation).

The Cross Population Special Interest Group is chaired by Dr Laramie Duncan, Hailiang Huang and Rosann Petersen.

Data access committee

The DAC is chaired by Dr Lea Davis. They will produce a streamlined process by which to obtain analyst access to genotype data for approved secondary analyses.

Data receiving committee

The DRC is chaired by Dr. Lea Davis. The committee produced a cohesive workflow to eliminate duplication and problem areas.

Including Additional Samples

It is highly desirable to include all comparable genomic data relevant to the Specific Aims in a mega-analysis. The organizing committee and/or its working groups will develop initial guidelines for determining which studies are sufficiently comparable to the other studies to be included (in terms of ancestry, phenotypes, diagnoses, and genotyping). Each working group has created their own individual policies, please contact the working group to learn more. As the membership expands, the participants may choose to modify these criteria. Some studies may share genome-wide summary statistics where data sharing restrictions preclude sharing of individual level data.

Participants will be required to provide phenotype and genotype data to meet the criteria that have been established, and to demonstrate that ethical approvals and informed consent documents are consistent with the project. An inclusive strategy is explicit—we will try to include all relevant genomic data for these disorders.

Publication Policy & Dissemination of Results

The PGC working groups determine their own publication and authorship policies.

However, a typical approach circa 2022 is: (a) list all authors with the last author being “xxx Working Group of the Psychiatric Genomics Consortium”; (b) at the time of submission, posting the manuscript on medRxiv,; and (c) posting the GWAS summary statistics for all SNPs that was used for the manuscript on the PGC downloads web page when the manuscript is accepted for publication.

Please note that PGC participants retain the capacity to publish on their own samples and to engage in any sort of additional research and collaborations they choose.

Authorship

The PGC groups each have explicit authorship policies. We initially used byline authorships, but many groups have shifted to individual authorship policies. Please refer to your workgroup’s authorship policy for details.

Data Sharing Within the PGC

Being part of the PGC requires contributing genomic and phenotypic data on individuals with psychiatric disorders to PGC analyses. The coordinating committee’s policy is that it would be ideal to carry out meta-analyses with individual-level data from each study. However, some investigators may be prohibited from sharing individual level data due to consent issues, ethical review committee ban, or by their national law.

It is anticipated that the default position for groups seeking full PGC membership and voting rights will be that individual genotype data will be contributed (unless there is a compelling and unavoidable legal or ethical prohibition). Any other reasons for contributing summary data will be considered on a case-by-case basis, as will the degree of membership offered to the contributing group, and the level of data access available to them.

Analysts from participating groups can get access to genomic data appropriate to the disorder and to the aims in which they participate. The majority of analyses will occur on the Genetic Computing Cluster which has a full suite of analytic software (and additional software can be added). Downloading or transfer of any individual or group data to any other computer or data storage device in any manner is prohibited.

Data Deposit in Controlled-Access Repositories

Individual genotype and phenotype data are shared with the PGC. The PGC does not deposit individual data—this is up to the data owners.

The PGC strongly encourages participating groups to deposit de-identified genotype and phenotype data into controlled-access repositories to maximize progress via sharing with the scientific community with appropriate safeguards to limit access to qualified scientists and for approved uses. However, data deposit in a controlled-access repository is not required for participation in the PGC.

Confidentiality, Obligations, and Intellectual Property

Never forget the cardinal rules of consortia: “no surprises” and “treat others as you wish to be treated”.

All information discussed in PGC meetings and calls is confidential and cannot be used for competing purposes unless discussed and approved by the group. All genotype and phenotype data shared should be kept strictly confidential within each group.

When the analyses for a specific aim are completed, all groups should feel free to discuss the results of the analysis freely and to conduct follow-up experiments. In the interests of maximal progress on these idiopathic disorders, we encourage pre-publication sharing of follow-up experiments.

The intention is to publish results as quickly as possible. Each group commits not to use any other group’s data or the joint analysis results in publications before the initial publication of the joint analysis.

We encourage a responsible approach to management of intellectual property derived from downstream discoveries that is consistent with the recommendations of the NIH's Best Practices for the Licensing of Genomic Inventions[1] and the NIH Research Tools Policy.[2] The management of patent applications in a manner that might restrict use of the joint findings and that could substantially diminish the value and public benefit provided by these resources is discouraged.

There is a shared commitment to protect the confidentiality of data and to protect the joint analysis activity by ensuring that no use of the data can be published in advance of an agreed-upon group publication and/or data release. At the same time, we also recognize that each participating group (either individually or as consortia) are actively pursuing follow-up genetic and functional work. Thus, individual parties performing additional experiments on genes identified or replicated in exchanges of data are not in violation of this agreement. However, we reaffirm that results of downstream experiments of genes identified in the meta-analysis cannot be published in advance of meta-analysis release or publication and we affirm a desire that purely genetic confirmation experiments (i.e., typing

[1] http://www.ott.nih.gov/policy/genomic_invention.html

[2] http://ott.od.nih.gov/policy/research_tool.html

in additional replication samples) be done as a component of a continued collaboration between our groups.

PGC and GDPR

Indication of Agreement

It is impractical to require formal signatures of participants and institutions across many different countries.

For PGC1, PGC2, and PGC3 we required participants to give their word that they would agree to this MOU. It is thus a manner of personal honor and integrity. This simple practice worked well, and will be continued in PGC4.

If you wish to participate in PGC4, please indicate your agreement with this MOU via email to the chair of the group you are joining.