**Genomic mapping of neuroimaging endophenotypes of psychiatric disorders**

**Project Description**

1. Specific Aims and Significance

It has been recently shown that a substantial proportion (50%) of autism risk is captured by common genetic variants each of small effect (Gaugler et al 2014). This result suggest that in many cases it will not be one or a few genes/mutations that determine the presence of an autistic phenotype, but a genomic, highly diluted, massively polygenic accumulation of frequent variants. We have recently shown that neuroanatomical variability in the IMAGEN cohort is also captured by frequent variants (Toro et al 2014). We could then try to map the genetic correlation between autism risk and neuroimaging, which would allow us to determine which brain structures are those whose genetic bases are more strongly overlapping with those of autism. Although it may not be possible to obtain biological information about psychiatric disorders through the detection of major genes, genomic mapping of neuroimaging phenotypes could allow us to obtain relevant biological information for psychiatric disorder with massively polygenic architectures.

We have the following three specific aims for our proposed research.

1. We propose to estimate genetic correlations between risk to psychiatric disorders and neuroimaging by combining cohorts of genotyped patients, for example, the Autism Genome Project in the case of autism, and cohorts of MRI scanned, genotyped controls, such as IMAGEN.
2. We plan to combine the ENIGMA cohort with the IMAGEN cohort in our analysis, which would potentially produce a total sample size of about 20,000 subjects to ensure statistical power to reliably detect small to moderate genetic correlations.
3. We propose new statistical methodology for genomic mapping of neuroimaging endophenotypes of psychiatric disorders. We extend the genome-wide complex trait analysis (GCTA) of Yang et al. (2011), which focuses on linear association between phenotype and genome. Our new method permits nonlinear association, and also incorporates nonlinear interactions of genes, by utilizing least squares kernel machine and specially designed kernel for genomic features.

[Significance] In many cases, the risk to psychiatric conditions such as autism or schizophrenia may be not mediated by a few major genes, but by the widely distributed effect of thousands of small effect variants. In those cases, our method could provide a mean to obtain relevant biological information on the nervous systems possibly involved (neuroimaging endophenotypes), despite the impossibility to find major genes.

2. Material and Methods

2.1. IMAGEN and ENIGMA consortium

The data for this project already exist in the IMAGEN (Schumann et al., 2010) and ENIGMA (Thompson et al, 2014) consortia. Both R. Toro and JB Poline are members of these consortia and have access to the data (the project described has been submitted and approved by both IMAGEN and ENIGMA). These consortia give access to either the primary data (IMAGEN) or the subjects genetic relation matrix (ENIGMA). With four thousands subjects or more, these data will provide us with enough power to detect small effects (heritability or genetic correlations) with small confidence interval.

2.2. Genome-wide complex trait analysis

Lexin

2.3. Least squares kernel machine and Wright-Fisher kernel

Lexin

2.4. Other extensions

Lexin

2.5. Previous work and preliminary results

We have recently shown, using the MRI and genetic data collected by IMAGEN, that an important part of the diversity of neuroanatomical phenotypes is captured by thousands of common SNPs, each of small-effect (Toro et al 2014). The strong heritability of phenotypes such as intracranial volume (~50%) or brain volume (~45%) allowed us to obtain statistically significant estimations of the variance captured by SNPs. However, the standard errors were large, ~20%. Power estimations show that a cohort of 4,000 subjects will be required to decrease standard error to ~10%, and 8,000 subjects would be required to further decrease standard error to ~5%. The extra statistical power would not only allow us to have better estimations of the amount of neuroanatomical variance captured by SNPs, but also to detect more subtle effects than those we could detect with 1,765 subjects. We propose here to lead an effort to replicate and extend our previous results. Through the ENIGMA consortium we could potentially have access to a cohort of 20,000 subjects.

Contrary to GWAS analyses, genomic complex trait analysis (GCTA) requires direct access to the genotyping data. The analysis approach is then a mega-analysis, instead of a meta- analysis. Thanks to the expertise that we have acquired with the analysis of the IMAGEN cohort, we have now the knowledge and the infrastructure to perform this mega-analysis. We would also benefit from the work already done in the context of the participation of IMAGEN to the ENIGMA 1 project (ICV, BV, Hipp) and the ENIGMA 2 project (subcortical structures).

Lexin

3. Research Team

* Lexin Li is a new faculty in the department of Biostatistics at Berkeley. He has experience in both imaging and genetic data.
* JB Poline is a researcher at the Brain Imaging Center at Berkeley. He has been the lead of the biostatistics and bioinformatics subproject of the IMAGEN consortium and has collaborated with ENIGMA on several projects.
* Roberto Toro is a tenured researcher at Pasteur in a human genetic laboratory specialized in autism research. R. Toro has a wide experience on imaging and imaging genetics. The three individuals gather the necessary experience and expertise for the success of this project.

4. Work Plan

* First kick off meeting at the Pasteur Institute, Paris. Objectives of the meeting:
  + Set up the collaboration tools (github account, common repositories, etc)
  + Develop a simulation dataset to validate estimation procedure (R. Toro, JB Poline, L. Li)
  + Bring participants up to date: genetic correlation current estimation procedures,
* During month 1-8, we will accomplish the following tasks.
  + Develop a simple simulation tool for validation
  + Obtain first genetic kindship/GRM matrix from IMAGEN and ENIGMA
  + Develop standard and regularized genetic correlation estimation procedures
* During month 9-12:
  + Test the regularized estimation procedures on both simulated and actual kinship matrices. Compare results with standard GCTA approach. Publish the results.

5. References

JL Stein, SE Medland, A Arias Vasquez, DP Hibar, RE Senstad, AM Winkler, R Toro, et al (2012) Common genetic polymorphisms are associated with human hippocampal and intracranial volumes. Nature Genetics.

Hibar, D. P., J. L. Stein, M. E. Renteria, A. Arias-Vasquez, S. Desrivières, N. Jahanshad, R. Toro, et al. 2015. “Common Genetic Variants Influence Human Subcortical Brain Structures.” Nature advance online publication (January). doi:10.1038/nature14101.

Gaugler, Trent, Lambertus Klei, Stephan J. Sanders, et al. 2014. “Most Genetic Risk for Autism Resides with Common Variation.” Nature Genetics 46 (8): 881–85. doi:10.1038/ng.3039.

Toro R, Poline JB, Huguet G, Loth E, et al. (2014) Genomic architecture of human neuroanatomical diversity. Mol Psychiatry, Accepted doi: 10.1038/mp.2014.99

Schumann, G., Loth, E., Banaschewski, T., Barbot, et al. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Molecular Psychiatry *15*, 1128–1139.

Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging and Behavior 1–30.

## Budget

* 2 meetings
  + One in Paris (kick off meeting of one week within the first 2 months of the project Lexin Li + JB Poline) : Airfare : 2\* 1500USD, lodging 5 nights for L. Li (500USD) : Total: 3500USD.
  + One in Berkeley : reviewing progress 8 months after the start of the project. R. Toro + G. Dumas: 3500$
* Publications fees: 2k$
* International conference: imaging genetics Irvine, Human Brain Mapping? 2k$ (one participant from Berkeley, one from Paris)
* Small computing hardware (hard disk, memory): 1k$

------------------------------- Scrap notes + to do list ---------------------------

TODO

1. The filled-out cover sheet available on our website.

2. A description of the project, including a timetable for comple-

tion, not to exceed THREE pages, including references. The de-

scription should indicate how the project contributes to scientific

knowledge and research methods within the applicants’ fields,

the nature of the collaboration proposed, the involvement of

junior scholars (detailing their names and roles in the project),

and the proposed impact of the collaboration, both in scientific

terms and institutionally. This description must be in French in

the French file and in English in the English file.

3. A letter of intent, signed by both the French and American proj-

ect coordinators indicating their commitment to work together

and to observe a proposed calendar of expenses not to exceed

ONE page. This letter can be either in French or in English; a trans-

lation is not necessary, but it must be included in both files.

4. A detailed budget, indicating a breakdown of the budget re-

quest by category (airfares, lodging, etc.) and an account of oth-

er sources of funding (potential and actual)—not to exceed ONE

page. The budget can be either in French or in English, and a

translation is not necessary. The same document can be used in

both files (French and English) but both files must be complete.

5. A curriculum vitae of each project coordinator not to exceed

TWO pages each. The resumes can be either in French or in En-

glish, and a translation is not necessary. The same documents can

be used in both files (French and English) but both files must be

complete.

<http://fbf.berkeley.edu/Application2015.html>

<http://fbf.berkeley.edu/files/FBF-Application-Guidelines-2015-Berkeley.pdf>

Lexin Li: <http://sph.berkeley.edu/lexin-li>

Possible statistical question:

Estimation of h2 / genetic correlation ? Regularization of genetic correlation ? efficient algorithm for estimation ? estimation per pathways

Analysis of the Genetic Relationship Matrix: treelet covariance smoothing?

Decompose the genetic relationship matrix into overlapping genomic/regulomic pathways?

Biological

We propose to estimate ***genetic correlations*** between risk to psychiatric disorders and neuroimaging by combining cohorts of genotyped patients (for example, the Autism Genome Project in the case of autism) and cohorts of MRI scanned, genotyped controls, such as IMAGEN. Our estimations of statistical power show that ~4,000 subjects will be required to detect reliably moderate genetic correlations. We will have to combine the IMAGEN cohort with additional cohorts. This project is a collaboration with ENIGMA consortium which will provide with other cohorts than the IMAGEN cohort.