Utilizing big data sets in imaging genetics

Derrek P. HIBAR, PhD | Senior Scientist | Janssen R&D, LLC
What can we learn from identifying the genetic influences on brain structure?

1. **New brain biology**: Understand the development of the human brain
2. **Characterize known disease variants**: A way to localize the effect or develop mechanisms of known disease associated genetic variants
3. **Discover new disease variants**: If genes have a stronger influence on brain structure than disease, imaging genetics may be better powered to detect effects.
Identifying genetic variation, life is hard.

(McCarthy et al., 2008)
Possible research questions:

We have a super rare variant that we found affects mouse hippocampus, does it also affect human hippocampus?

Need a large data set with structural brain imaging and whole genome sequencing data.
Access to 1,902 patients with longitudinal imaging (including sMRI, DTI, and PET) and blood-based biomarker measurements
  – Around 1,400 patients have genome-wide genotyping data available
  – Around 800 patients have whole genome sequencing data available
  – Apply at ida.loni.usc.edu

Tutorials for accessing, performing QC, and conducting your first genome-wide association study (GWAS):
  – http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/presentations/ohbm2015/imggen

Scripts and additional data files for ADNI processing can be found on GitHub:
Possible research questions:

I have identified common variants influencing IQ, are they also associated with brain structure?

You will need a large data set with genome-wide genotyping data and imaging data or imaging meta-analysis data.
- Population study of 500,000 subjects in the UK
- Extensive testing and interviews
  - Thickness of butter/margarine spread on baguettes (Data-Field 20099)
- 10,656 subjects have available brain sMRI data
- Genome-wide genotyping data on all 500k participants will be released Q3 2017
  - 8TB of raw genetic data

- Apply for access:
  - http://www.ukbiobank.ac.uk/register-apply/

- Cost recovery for access: around £2,000 for access
Other Projects:

- **PING** – Pediatric Imaging, Neurocognition and Genetics
  - http://pingstudy.ucsd.edu/
- **PNC** – Philadelphia Neurodevelopmental Cohort
- **PPMI** – Parkinson’s Progression Markers Initiative
  - http://www.ppmi-info.org/
- **HCP** – Human Connectome Project (twin data)
  - http://www.humanconnectome.org/data/
- **GNC** – German National Cohort
  - http://nako.de/informationen-auf-englisch/
Michigan Imputation Server

This server provides a free genotype imputation service. You can upload GWAS genotypes (VCF or 23andMe format) and receive phased and imputed genomes in return. Our server offers imputation from HapMap, 1000 Genomes (Phase 1 and 3), CAAPA and the updated Haplotype Reference Consortium (HRC version r1.1) panel. Learn more or follow us on Twitter.

Sign up now  Login

The easiest way to impute genotypes

Upload your genotypes to our server located in Michigan. All interactions with the server are secured.

Choose a reference panel. We will take care of pre-phasing and imputation.

Download the results. All results are encrypted with a one-time password. After 7 days, all results are deleted from our server.
Steps

• Register for account (might want to get 2-3 per group)
• QC genotypic data (if not already done)
• Make sure data is on build 37 (GRCh37)
• Convert data to vcf & compress
• Upload
  – Select options
• Download
Possible research questions:

I have identified common variants influencing IQ, are they also associated with brain structure?

You will need a large data set with genome-wide genotyping data and imaging data or imaging meta-analysis data.

What if it is not practical for you to “go it alone”?
Advantages of a collaborative framework

- Solving the growing problem of reproducibility in science

  - Essay
  
  Why Most Published Research Findings Are False
  
  John P.A. Ioannidis

  NIH plans to enhance reproducibility
  
  Francis S. Collins and Lawrence A. Tabak discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

  - Research Article
  
  Estimating the reproducibility of psychological science
  
  Open Science Collaboration

- Collaboration leads to new ideas, internal vetting, and testable hypotheses

- We've established a nucleus of highly interested and qualified people to lead groups who are knowledgeable and grateful for collaboration
Common genetic variation has only a **small effect** on brain structure.

In a GWAS, many statistical tests are conducted.

**Huge sample sizes** (>10,000 individuals) are needed in order to find and replicate association of individual genetic variants.

As both imaging and genotyping are expensive, **a consortium is needed**.

> 185 institutions, 300+ co-authors, world-wide consortium countries resulting in ~30,000+ subjects

*Now expanded to multiple multi-national meta-analytic projects:*
- Genetics of brain structure
- Genetics of brain function
- Meta-analytic effects on disease

http://enigma.ini.usc.edu
ENIGMA2: Five novel genetic variants identified; hippocampal and ICV results replicated

Discovery sample: 13,171 individuals
Increasing power with the CHARGE consortium (E2C)

N ~ 33,000

(Adams, et al., under review *Nat Neurosci*; Hibar, et al., in preparation)
ENIGMA3: GWAS of Cortical Thickness and Surface Area

Cortical Surface Area Heritability

Cortical Thickness Heritability

(Eyler et al., TRHG, 2012)
ENIGMA3: Preliminary Findings in ~15k Healthy Subjects

Sarah Medland

Precentral_SurfaceArea -

Pericalcarine_SurfaceArea -
Frontalpole_SurfaceArea -
Temporalpole_SurfaceArea +
Isthmuscingulate_SurfaceArea +
Parsorbitalis_SurfaceArea +
Insula_SurfaceArea -
Bankssts_SurfaceArea -
Entorhinal_SurfaceArea -
Lateraloccipital_SurfaceArea +
Paracentral_SurfaceArea -
Lateralorbitofrontal_SurfaceArea -
Medialorbitofrontal_SurfaceArea +
Parahippocampal_SurfaceArea +
Inferiorparietal_SurfaceArea -
Inferiortemporal_SurfaceArea -
Caudalanteriorcingulate_SurfaceArea +
Cuneus_SurfaceArea -

Postcentral_SurfaceArea -
Supramarginal_SurfaceArea +
Superiorparietal_SurfaceArea +
Parsopercularis_SurfaceArea +
Parstriangularis_SurfaceArea +
Rostralmiddlefrontal_SurfaceArea +
Precuneus_SurfaceArea +
Transversetemporal_SurfaceArea +
Fusiform_SurfaceArea +
Superiorfrontal_SurfaceArea -
Posteriorcingulate_SurfaceArea -
Rostralantieriorcingulate_SurfaceArea -
Lingual_SurfaceArea -
Caudalmiddlefrontal_SurfaceArea -
Superiortemporal_SurfaceArea -
Meditemporal_SurfaceArea -
ENIGMA3: Preliminary Findings in ~15k Healthy Subjects

Round 1 Meta-analysis is complete. Still possible to join! Email: enigma@ini.usc.edu
New and Ongoing Projects in ENIGMA

ENIGMA is an open and collaborative platform for all researchers. Participate or propose an new project!
Update, ENIGMA-CNV
– from genes to phenotypes to brain structures
Ida Elken Sønderby, NORMENT, Norway

CNV and imaging data on ~14,800
30 cohorts
13 countries
14 types of high density SNP arrays
59 scanner sites
Different types of ascertainment

CNV carriers:
- Disease susceptibility
- Low frequency
  - 1 in 50,000
  - 1 in 400
- Huge variation in penetrance, symptoms
- Brain structure on two CNVs so far:
  - 15q11.2 (subtle effects)
  - 16p11.2 proximal
    (ICV + focal effects)
Cortical Thickness and Surface Area ROIs
466 22q11Del Cases vs. 358 Controls
10 international sites

Cortical Thickness

Surface Area

Cohen’s d

-1.0 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0

Left Lateral
Left Medial
Left Lateral
Left Medial
Right Lateral
Right Medial
Right Lateral
Right Medial

Dr. Daqiang Sun
ENIGMA-Vis Browser

- For people without familiarity in programming or genomic information display
- Easy to use tool that allows for comparison of brain imaging GWAS and to user-uploaded associations
- Bulk download also available (> 150 E2 downloads)

http://enigma.ini.usc.edu/enigma-vis/
http://enigma.ini.usc.edu/enigma-vis/enigmavis-dev/

(Novak et al., TRHG, 2011)
ENIGMA-Vis Features: Enrichment

On the fly genetic correlation

File Upload
Option 1: Drag and Drop
Option 2: Select file to upload:

Uploaded File:

Submission Data
Enrichment Method: LD SCORE
Sample Size:
Study Name:
Evaluating genetic overlap for ICV

- No support for ICV as endophenotype for any psychiatric disease.
- We do have support for genetic overlap with adult cognitive function, implying the genetic factors that determine your overall head size in part influence your late in life cognitive function.

(Adams*, et al., in press Nat Neurosci)

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>N total</th>
<th>N cases</th>
<th>$\rho_{\text{genetic}}$</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anthropometric traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult height</td>
<td>253,280</td>
<td>–</td>
<td>0.049</td>
<td>0.039</td>
<td>0.21</td>
</tr>
<tr>
<td>Child head circumference</td>
<td>10,768</td>
<td>–</td>
<td>0.750</td>
<td>0.126</td>
<td>2.5 x 10^{-9}</td>
</tr>
<tr>
<td>Birth length</td>
<td>28,459</td>
<td>–</td>
<td>0.192</td>
<td>0.088</td>
<td>0.029</td>
</tr>
<tr>
<td>Birth weight</td>
<td>26,836</td>
<td>–</td>
<td>0.160</td>
<td>0.086</td>
<td>0.062</td>
</tr>
<tr>
<td><strong>Neurological traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Childhood cognitive function</td>
<td>12,441</td>
<td>–</td>
<td>0.257</td>
<td>0.090</td>
<td>4.2 x 10^{-3}</td>
</tr>
<tr>
<td>Adult cognitive function</td>
<td>53,949</td>
<td>–</td>
<td>0.198</td>
<td>0.059</td>
<td>6.9 x 10^{-4}</td>
</tr>
<tr>
<td>Alzheimer's Disease</td>
<td>54,162</td>
<td>17,008</td>
<td>-0.043</td>
<td>0.098</td>
<td>0.66</td>
</tr>
<tr>
<td>Parkinson's Disease</td>
<td>108,990</td>
<td>13,708</td>
<td>0.335</td>
<td>0.072</td>
<td>3.0 x 10^{-6}</td>
</tr>
<tr>
<td>White matter lesions</td>
<td>17,936</td>
<td>–</td>
<td>0.096</td>
<td>0.079</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Psychiatric traits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>10,263</td>
<td>4,949</td>
<td>0.026</td>
<td>0.071</td>
<td>0.72</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>11,810</td>
<td>6,990</td>
<td>-0.004</td>
<td>0.076</td>
<td>0.95</td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>16,610</td>
<td>9,227</td>
<td>0.005</td>
<td>0.096</td>
<td>0.96</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>17,115</td>
<td>9,379</td>
<td>-0.009</td>
<td>0.058</td>
<td>0.87</td>
</tr>
<tr>
<td>Extraversion</td>
<td>63,030</td>
<td>–</td>
<td>-0.097</td>
<td>0.092</td>
<td>0.29</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>63,661</td>
<td>–</td>
<td>0.070</td>
<td>0.111</td>
<td>0.53</td>
</tr>
</tbody>
</table>
E-Vis: Adding annotations and analysis

- Integrate with functional definitions of the genome (expression quantitative trait loci and epigenetics)
- On the fly genetic overlap methods (genetic correlation, RRHO, Conjunction analysis)

(Nord et al., Neuron, 2015)
Welcome to the GitHub for ENIGMA

The ENIGMA Consortium is an international effort by leaders worldwide. The Network brings together researchers in imaging genomics, neurology and psychiatry, to understand brain structure and function, based on MRI, DTI, fMRI, genetic data and many patient populations.

The best return on our research investments will come from combining our data to achieve the large samples necessary to detect the modest gene effect sizes that we now know are the rule rather the exception for complex traits.

The ENIGMA Network has several goals:

- to create a network of like-minded individuals, interested in pushing forward the field of imaging genetics
- to ensure promising findings are replicated via member collaborations, in order to satisfy the mandates of most journals
- to share ideas, algorithms, data, and information on promising findings or methods
- to facilitate training, including workshops and conferences on key methods and emerging directions in imaging genetics.

Read more about Enigma here
Acknowledgments

NIH Big Data to Knowledge Program (BD2K)

CHARGE Consortium

700+ ENIGMA co-authors and PIs

340 institutions around the world

Individual cohort studies would not be possible without international and national support:

NIH, NHMRC, DCRC, BMBF, Swedish Research Council, German Ministry of Cultural Affairs, Social Ministry of the Federal State of Mecklenburg–West Pomerania, NGFN, Siemens, ISCIII, SENY Fundació, NOW, BBMRI-NL, CBF, Hersenstichting Nederland, Alzheimer’s Australia Dementia Research Foundation, Autism Speaks, NIBIB, NIMH, NIDA, NINDS, NCI, NIA, …