Biobank Project: Progress & Plans

Carlos Cruchaga, PhD

Joel S Perlmutter, MD
Goals for the Biobank

• To extend the current DNA repository by targeting BSP, LD and limb dystonia subjects as well as multiplex families.
  • Sporadic and familial presentations
  • Longitudinal (each visit)

• To develop a centralized repository of other blood-based materials
  • DNA
  • RNA
  • Plasma

• To identify novel genetic and proteomic factors for dystonia risk

• To identify genetic and proteomic factors that influence spread of dystonia
Current Biobank: DNA Available

Older and New sample visits

- Control 1%
- Multi-focal 7%
- Laryngeal 1%
- Hemi 1%
- Generalized 4%
- Other Dystonia Syndrome 0%
- Segmental 20%
- Focal 68%

New sample only

- Control 1%
- Multi-focal 13%
- Hemi 1%
- Generalized 4%
- Other Dystonia Syndrome 27%
- Segmental 27%
- Normal Control 4%
- Focal 50%

Data by Jen Gentsch
Biobank: 2020-2022

# Samples Received per year

<table>
<thead>
<tr>
<th>Year</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samples</td>
<td>10</td>
<td>186</td>
<td>236</td>
<td>84</td>
</tr>
</tbody>
</table>

Samples include: DNA, RNA and plasma

DNA quality

<table>
<thead>
<tr>
<th></th>
<th>Concentration</th>
<th>260/280</th>
<th>260/230</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN</td>
<td>42.97</td>
<td>1.73</td>
<td>0.7</td>
</tr>
<tr>
<td>MAX</td>
<td>370.90</td>
<td>1.92</td>
<td>2.64</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>181.34</td>
<td>1.84</td>
<td>2.02</td>
</tr>
</tbody>
</table>

RNA quality

<table>
<thead>
<tr>
<th></th>
<th>Concentration</th>
<th>RIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIN</td>
<td>5.48</td>
<td>3.60</td>
</tr>
<tr>
<td>MAX</td>
<td>713.00</td>
<td>9.90</td>
</tr>
<tr>
<td>AVERAGE</td>
<td>69.66</td>
<td>7.74</td>
</tr>
</tbody>
</table>

Data by Pat Kohlfeld
Genetic Characterization of DC participants

**RAW**
- Illumina’s Multi-Ethnic Genotyping Array
  - 919 samples, 1,359,498 SNPs
- Illumina Global Screen Array
  - 1344 samples, 653,142 SNPs

**Genotype data QC**
- Filtered for ≥ 98% call rate, HWE P<10^-6
  - 8 samples failed

**Imputation**
- TOPMed Imputation Server
  - A total of 41,499,784 variants passe QC

**IBD and sex checks**
- 12 samples with IBD issues
- 18 have sex mismatches

**Final QC’ed data**
- A total of 2225 samples
- and 41,499,784 variants

Data by Priyanka Gorijala
Polygenic risk scores for Dystonia

1. Obtain GWAS summary statistics (p-values and β’s) in largest possible **discovery sample**
   1. Alzheimer’s disease: Bellenguez et al 2023
   2. Parkinson disease: Kim et al 2022

2. Obtain independent **target sample (this is the Dystonia samples)** with Genome-wide data
   1. Use **SNPs in common** between the two samples.
      Deal with **association redundancy due to LD**.
   2. Restrict to **SNPs with p < 5x10^{-8}** for the disease GWAS
      1. We calculated PRS for AD, PD and FTD
   3. **Construct PRS = Sum of Risk Alleles weighted by β from regression.**

Data by Priyanka Gorijala
Polygenic Risk Score Analysis

Association results of extreme tertiles of the PRS at genome-wide threshold derived for each of the AD, PD, and FTD categories and compared to Dystonia Vs controls.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>2.50% CI</th>
<th>97.50% CI</th>
<th>P value</th>
<th>N Cases</th>
<th>N Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD with APOE PRS</td>
<td>0.698</td>
<td>0.618</td>
<td>0.788</td>
<td>6.505 × 10⁻⁹</td>
<td>2024</td>
<td>7607</td>
</tr>
<tr>
<td>AD without APOE PRS</td>
<td>0.789</td>
<td>0.698</td>
<td>0.892</td>
<td>1.567 × 10⁻⁴</td>
<td>2024</td>
<td>7607</td>
</tr>
<tr>
<td>PD PRS</td>
<td>0.872</td>
<td>0.773</td>
<td>0.984</td>
<td>2.675 × 10⁻²</td>
<td>2024</td>
<td>7607</td>
</tr>
<tr>
<td>FTD PRS</td>
<td>1.054</td>
<td>0.933</td>
<td>1.190</td>
<td>3.961 × 10⁻¹</td>
<td>2024</td>
<td>7607</td>
</tr>
</tbody>
</table>

N – Number of samples in each category, CI – Confidence interval, PRS – Polygenic risk score, APOE – Apolipoprotein.

1. Dystonia cases are depleted for AD risk variants, independently of APOE
2. Dystonia cases are depleted for PD risk variants
3. No differences for FTD

Data by Priyanka Gorijala
Genome-Wide Association studies

- Obtaining GWAS data for all samples.
  - Previous studies by Sun et al, 2021 already identified one genome-wide significant
    - 919 cases and 1,491 controls
    - Chromosome 3 signal is close upstream of COL8A1
    - Defects in COL8A1 are associated with corneal dystrophy and age-related macular degeneration.

- A larger GWAS in collaboration with Dr. Klein is ongoing with > 5K cases

- PRS analyses suggest very low overlap with AD, but some overlap with PD
Goal: identify causal and druggable targets

• We need to go beyond just GWAS/WGS
• Molecularly Phenotype Clinical cohorts
• Generating multiple layers of omic data
Genetics is just the first layer

Plasma

Blood-RNA

Blood-DNA

Function

Translation

Transcription

Phenotype variability

Metabolites

Proteins

mRNA

DNA

METABOLOMICS

PROTEOMICS

TRANSCRIPTOMICS

EPIGENOMICS

GENOMICS
A Metabolomic Study of Cervical Dystonia

Plasma samples from 100 cases with idiopathic cervical dystonia and 100 controls

- 7,346 metabolic features remained after quality control
- 289 significantly associated with case-control status
Other omic studies

• Proteomics:
  • 50 controls and 150 dystonia cases.
    • See poster by Jigyasha Timsina
    • Emory-WU collaboration
    • Manuscript being drafted

• Transcriptomics
  • RNA available for all new samples

• Methylation
  • Biological vs chronological clocks
  • Differential methylation
Summary

• The goals is to obtain samples that will allow deep molecular characterization the DC cohort
  • Genetic (GWAS, WGS), epigenomics (longitudinal), transcriptomics (longitudinal), proteomic (longitudinal), metabolomic and lipidomics

• Deep molecular phenotyping of well clinically characterized cohorts will lead to the identification of:
  • Novel genes and pathways implicated on the diseases
  • A deeper understanding pathologic events
  • Novel molecular phenotypes
  • Novel therapeutic targets

• The multi-omic data (genetic, epigenetic, transcriptomic, proteomic, metabolomic, between others) will allow to a more personalize prediction of disease risk and treatment
Q & A

Biobank (BB) Project
Washington University in St. Louis
Cor PI: Carlos Cruchaga. PhD cruchagac@wustl.edu
Study Coordinator: Jen Gentsch j.gentsch@wustl.edu
Study Coordinator: Joanne Norton nortonj@wustl.edu