**Working Title:**

**Integrated omics analysis in Parkinson’s disease: genome-wide analysis of epigenetics and metabolomics provides links to Parkinson’ disease**

Kimberly C Paul\*, Gary Chen\*, Cynthia Kusters, Janet Sinsheimer, Jeff Bronstein, Beate Ritz

\*Co-First Author

**Objective:**

1. Identify blood-based methylation and metabolomic quantitative trait loci (meQTLs and metQTLs) and integrate them with PD-related genetic variants, epigenetic variants, and metabolites discovered in GWAS, EWAS, and MWAS.
2. Assess enrichment of QTLs observed in PD patients: (a) did PD patients have more QTLs than expected; (b) are QTLs enriched for known biologic pathways (KEGG/GO/etc GSEA)

“Identifying meQTL variants and linking them to disease-associated genetic variants from GWAS would pinpoint molecular mechanisms underlying genetic susceptibility to human diseases that are due, at least in part, to altered epigenetic regulation. Additionally, it could help explain the molecular consequences of non-protein-coding, disease-associated genetic variants from GWAS.”

<https://www.nature.com/articles/s41467-019-12228-z>

**Analysis Outline:**

1. Perform genome-wide association testing of genetic variants with whole blood DNA methylation and untargeted serum-based metabolites from 750 participants of the PEG Parkinson’s Study (541 discovery cohort, 209 replication cohort) to identify cis-meQTLs (methylation QTLs), trans-meQTLs, and metQTLs (metabolite QTLs).
   1. QTL hotspots (regions of the genome that have more QTLs than expected/others) *I am not sure how we would do this exactly for cis meQTLs. However, using nature paper above for defining hotspots as regulatory centers for trans assoc, I generated Table 6*
   2. Colocalization of QTLs (meQTLs, metQTLs, and eQTLs from GTeX), what proportion of QTLs influence a single omic layer versus influencing multiple (ie most meQTLs are also eQTLs, metQTLs?) *Table 7 gets at this at least partially (i.e. overlap between meQTL and metQTL). Not sure which subset of the data from GTEx we would specifically want to compare against.*
2. Link the QTLs to GWAS summary statistics for Parkinson’s disease from the meta-GWAS
   1. Assess whether there are more QTLs among GWAS SNPs than expected (ie x% of all SNPs are QTLs, y% of GWAS SNPs (p<0.05) are QTLs, overrepresented?) *Results from hypergeometric hypothesis tests are listed in Table 4*
   2. Do PD genes have more QTLs among patients than expected *Should we run a simple logistic regression model with a predictor variable as number of significant cis QTLs?*
   3. Colocalization? which GWAS SNPs colocalize with a QTL *Should we just select a few candidate probes that were interesting and tease out which of the cis SNPs also explained PD risk using the Approximate Bayes Factors of the coloc program? This would not be practical genome wide*
   4. Mendelian randomization (MR) to identify putatively causal CpGs for PD? (<https://www.nature.com/articles/s41467-019-12228-z>) *Sounded like a low priority based on last meeting but I may have misunderstood.*
3. Enrichment of QTLs for pathways, etc (KEGG/GO/etc): meQTLs only, metQTLs only, and combined? *(Figure 1 has a DAG from GREAT for GO enrichment among our significant meQTLs in PEG1) Should we present differently. MetQTLs did not turn up anything interesting using GREAT.*

PPMI as replication cohort now? Or leave out as they don’t have the metabolomics?

Results so far showing a lot of immune pathways, should we add a small section of how QTLs are related to immunoplex panel (measured levels of 38 immune factors: cytokines, chemokines, growth factors, etc.)?

**Results:**

*(age, sex, AIMs ancestry (though we are limiting to Euro, still can show fractional ancestry), etc)*

*(Sup?)*

Table 1: Study Information N

*Include indicator for cis or trans (large table of all SNP~CpG results, FDR<0.05 or maybe even p<0.05 for supplement)PEG1 discovery, PEG2 test*

Table 2: cis meQTL output

Table 3: meQTL annotations, including PD GWAS SNP/p-value if so, gene meQTL located in, and gene region (TSS, UTR, etc)

Table 4: meQTL enrichment, including PD GWAS SNPs,

*known meQTL:known eQTL (Gtex) do we have any unique meQTLs?*

Table 5: known meQTL: Link to BIOS meQTLs

Table 6: trans meQTL output

*meQTL enrichment for gene / genome region, PD pathways, ?*

Table 7: GO enrichment among cis meQTLS found in PEG1 and PEG2

Figure 1: Manhattan plot of cis SNP-CpG results, FDR<.05

Figure 2: Manhattan plot of trans SNP-CpG results, FDR<.05

**Table 1 Study Information**

|  |  |  |
| --- | --- | --- |
|  | **PEG 1** | **PEG 2** |
| Sample size (female/male) | 231/310 | 64/145 |
| Age (mean, range) | 69 (35-92) | 71 (46-92) |

**Supplementary Table 1** Fractional Ancestry Information among four clusters1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **PEG 1 clusters** | | | | **PEG 2 clusters** | | | |
| Self-reported ethnicity | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Latino | 0.038 | 0.490 | 0.470 | 0.002 | 0.034 | 0.492 | 0.472 | 0.001 |
| White | 0.011 | 0.007 | 0.976 | 0.006 | 0.009 | 0.017 | 0.970 | 0.004 |
| Asian | 0.280 | 0.717 | 0.002 | 0.001 | 0.279 | 0.718 | 0.002 | 0.001 |
| Native American | 0.024 | 0.008 | 0.965 | 0.003 | 0.005 | 0.003 | 0.985 | 0.008 |
| Other |  |  |  |  | 0.003 | 0.002 | 0.991 | 0.004 |

1 ethnicity clusters defined in STRUCTURE software program

**Table 2: List of ten cis meQTLs ranked by statistical evidence, discovered in PEG1 and replicated in PEG2**

(Supplementary table lists all results at FDR<0.05)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | **PEG1** | | | | **PEG2** | | | |  |  |
| **SNP** | **gene probe** | **allele** | **p-value** | **FDR** | **beta** | **SE** | **p-value** | **FDR** | **beta** | **SE** | **Meta Z** | **Meta p-value** |
| GSA-rs1040961 | cg17707870 | G | 2.65E-242 | 2.50E-234 | 0.46 | 0.01 | 1.25E-83 | 1.23E-76 | 0.48 | 0.01 | -37.21 | 5.44E-303 |
| rs10010994 | cg17858192 | C | 9.19E-232 | 4.35E-224 | 0.38 | 0.01 | 9.10E-75 | 2.70E-68 | 0.39 | 0.01 | -35.92 | 1.46E-282 |
| exm2267473 | cg09084244 | G | 1.33E-217 | 4.20E-210 | 0.43 | 0.01 | 2.73E-72 | 5.70E-66 | 0.42 | 0.01 | -34.98 | 4.71E-268 |
| rs10184015 | cg02502145 | A | 6.88E-215 | 1.63E-207 | 0.46 | 0.01 | 9.63E-80 | 4.72E-73 | 0.48 | 0.02 | 35.49 | 6.01E-276 |
| rs2532925 | cg04145681 | G | 9.34E-215 | 1.77E-207 | 0.47 | 0.01 | 2.21E-76 | 7.74E-70 | 0.46 | 0.02 | -35.2 | 2.33E-271 |
| GSA-rs1035142 | cg07227024 | T | 1.79E-205 | 2.83E-198 | 0.43 | 0.01 | 1.97E-75 | 6.23E-69 | 0.44 | 0.01 | 34.62 | 1.10E-262 |
| GSA-rs10750097 | cg12556569 | G | 1.30E-203 | 1.75E-196 | 0.38 | 0.01 | 3.05E-80 | 1.57E-73 | 0.41 | 0.01 | -34.94 | 1.65E-267 |
| rs2883456 | cg11144103 | C | 2.55E-203 | 3.01E-196 | 0.44 | 0.01 | 1.28E-93 | 4.18E-86 | 0.47 | 0.01 | -36.03 | 3.11E-284 |
| rs1939015 | cg10306192 | G | 7.31E-200 | 7.68E-193 | 0.49 | 0.01 | 3.28E-93 | 8.04E-86 | 0.54 | 0.01 | -35.81 | 7.87E-281 |
| rs1043793 | cg02502145 | A | 2.07E-198 | 1.95E-191 | 0.45 | 0.01 | 1.71E-47 | 3.06E-42 | 0.44 | 0.02 | 31.49 | 1.35E-217 |
| rs4777684 | cg02074316 | A | 1.01E-195 | 7.94E-189 | 0.42 | 0.01 | 6.15E-78 | 2.62E-71 | 0.44 | 0.01 | 34.32 | 4.15E-258 |

**Table 3: List of ten meQTLs ranked by statistical evidence, discovered in PEG1 and replicated by Nall GWAS**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **meQTL evidence** | | | **GWAS evidence** | | | |
| **type** | **snp** | **allele** | **probe** | **cohort** | **pvalue** | **FDR** | **beta** | **beta** | **se** | **pvalue** | **gene** |
| cis | rs72660967 | T | cg06961873 | peg1 | 1.79E-158 | 2.92E-152 | 0.408418 | 0.0439 | 0.019 | 0.0206 | missing |
| cis | rs10903129 | A | cg06961873 | peg1 | 7.88E-158 | 1.22E-151 | -0.409488 | -0.0435 | 0.0168 | 0.009549 | missing |
| cis | rs3087801 | G | cg10776061 | peg1 | 3.55E-157 | 5.42E-151 | 0.37696 | -0.0401 | 0.0185 | 0.03017 | missing |
| cis | rs9276490 | A | cg18572898 | peg1 | 1.70E-153 | 1.78E-147 | 0.41004 | -0.0448 | 0.0224 | 0.04595 | missing |
| cis | rs9649865 | A | cg25543264 | peg1 | 4.91E-152 | 4.65E-146 | 0.385467 | 0.0505 | 0.0223 | 0.02349 | missing |
| cis | rs166849 | G | cg13143872 | peg1 | 5.40E-150 | 4.65E-144 | 0.312391 | -0.0391 | 0.0192 | 0.04131 | missing |
| cis | rs10191694 | A | cg15147113 | peg1 | 2.74E-148 | 2.23E-142 | 0.108973 | 0.0717 | 0.0235 | 0.002248 | missing |
| cis | rs7799245 | G | cg11957130 | peg1 | 7.39E-146 | 4.95E-140 | 0.440158 | 0.0685 | 0.0246 | 0.005368 | missing |
| cis | rs9276490 | A | cg07389699 | peg1 | 1.04E-145 | 6.40E-140 | 0.342294 | -0.0448 | 0.0224 | 0.04595 | missing |
| cis | rs9689 | A | cg06961873 | peg1 | 2.71E-135 | 1.34E-129 | -0.392911 | -0.0392 | 0.0175 | 0.02491 | missing |

**Table 4: Enrichment for cis meQTLs among PD SNPs based on hypergeometric test**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cohort** | **Total SNPs** | **Total meQTLs** | **PD risk SNPs** | **MeQTLs among PD SNPs** | **p-value** |
| PEG1 | 263704 | 33430 | 11920 | 1966 | 2.6e-35 |
| PEG2 | 272674 | 19037 | 12120 | 1208 | 5.0e-36 |

**Table 5: Top ten meQTLs ranked by statistical evidence, also found in BIOS repository**

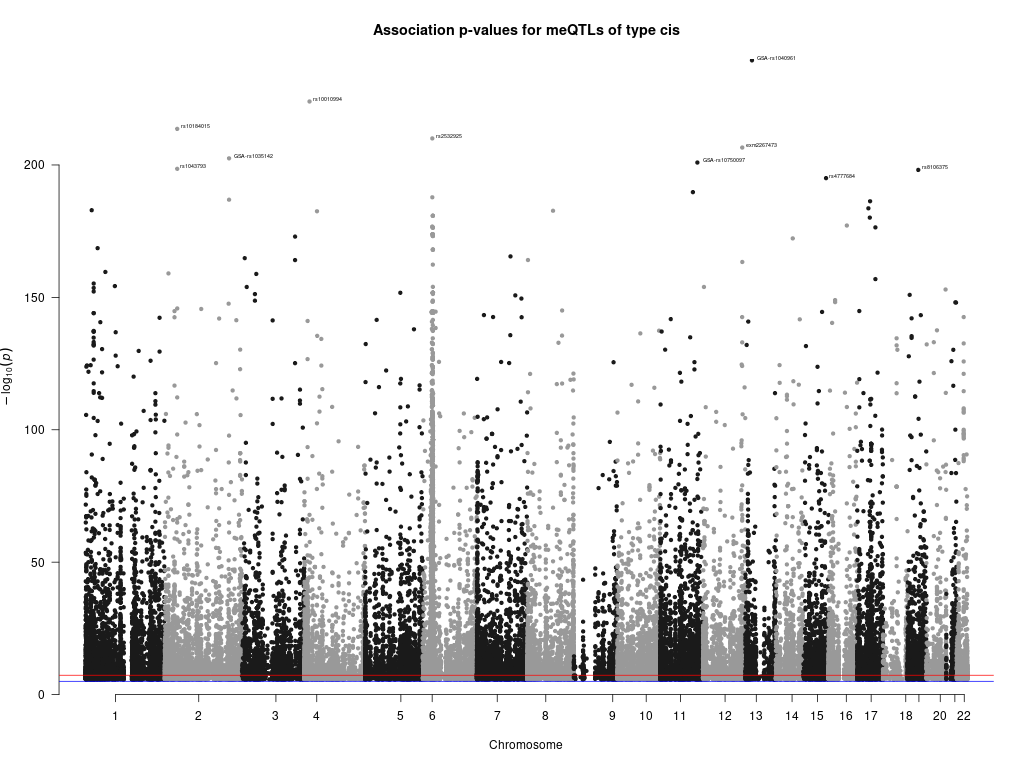
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SNP** | | **Probe** | |  |  |  |  |  |
| **Chr** | **ID** | **Position** | **ID** | **Position** | **Alleles** | **Tested Allele** | **Z-score** | **p-value** | **HGNCName** |
| 13 | rs11616903 | 50424677 | cg04520693 | 50425377 | C/T | T | 72.6668802 | 3.27167E-310 | RNY4P30 |
| 4 | rs4406078 | 57521325 | cg19978674 | 57523802 | T/C | T | 72.1343579 | 3.27167E-310 | HOPX |
| 2 | rs12475055 | 102878891 | cg23719516 | 102870432 | A/C | A | -71.8785276 | 3.27167E-310 | AC007248.6 |
| 11 | rs515449 | 122047290 | cg13160852 | 122047052 | G/A | A | -71.3891567 | 3.27167E-310 | RP11-820L6.1 |
| 8 | rs10086636 | 126604637 | cg08775595 | 126608666 | C/T | T | -71.3193873 | 3.27167E-310 | RP11-136O12.2 |
| 4 | rs11730732 | 119761678 | cg11663691 | 119759924 | A/G | A | 71.2074002 | 3.27167E-310 | SEC24D |
| 8 | rs10090179 | 144339053 | cg19504605 | 144339082 | G/A | A | -71.1133419 | 3.27167E-310 | ZFP41 |
| 8 | rs1454616 | 126611214 | cg08775595 | 126608666 | T/C | C | -71.0101274 | 3.27167E-310 | RP11-136O12.2 |
| 14 | rs10483261 | 22356551 | cg16702660 | 22361306 | T/C | C | -70.9943176 | 3.27167E-310 | TRAV8-4 |
| 17 | rs9894429 | 79596811 | cg21028142 | 79581735 | C/T | T | 70.9090052 | 3.27167E-310 | NPLOC4 |

**Table 6: Top 10 hotspots for trans meQTL associations discovered in PEG1 and replicated in PEG2 (FDR<.05).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **meQTL** | **Total trans** | **Meta Z-score** | **Meta p-value** | **Associated trans probes** |
| rs4324798\_G | 10 | 16.109 | 2.19E-58 | {'cg24838316', 'cg15671450', 'cg23681866', 'cg21438527', 'cg04559908', 'cg26121931', 'cg15411272', 'cg14345882', 'cg03403996', 'cg23892028'} |
| rs889688\_G | 9 | 9.016 | 1.95E-19 | {'cg25100848', 'cg26487966', 'cg21140353', 'cg03969070', 'cg23873523', 'cg01325958', 'cg11143796', 'cg02201215', 'cg16383001'} |
| rs7766356\_T | 9 | -11.118 | 1.02E-28 | {'cg26021304', 'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg12035144', 'cg23892028'} |
| rs35128564\_T | 8 | -13.232 | 5.72E-40 | {'cg26021304', 'cg06458771', 'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272'} |
| rs116591906\_A | 8 | -11.333 | 9.04E-30 | {'cg26021304', 'cg06458771', 'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272'} |
| GSA-rs13193532\_A | 8 | -13.339 | 1.37E-40 | {'cg26021304', 'cg06458771', 'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272'} |
| rs7739915\_G | 7 | 13.388 | 7.11E-41 | {'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg25046571'} |
| rs2765220\_T | 7 | -12.899 | 4.54E-38 | {'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg25046571'} |
| rs237997\_G | 7 | 10.887 | 1.33E-27 | {'cg15383120', 'cg07332563', 'cg11235426', 'cg05064044', 'cg01516881', 'cg03395511', 'cg21548813'} |
| rs17312661\_A | 7 | -10.858 | 1.82E-27 | {'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg23892028'} |
| rs172329\_T | 7 | -12.844 | 9.27E-38 | {'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg25046571'} |
| GSA-rs417919\_A | 7 | -12.844 | 9.27E-38 | {'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg25046571'} |
| GSA-rs209181\_G | 7 | 13.333 | 1.49E-40 | {'cg24838316', 'cg23681866', 'cg15671450', 'cg21438527', 'cg26121931', 'cg15411272', 'cg25046571'} |
| exm2272434\_C | 7 | -9.386 | 6.22E-21 | {'cg15383120', 'cg11235426', 'cg07332563', 'cg05064044', 'cg01516881', 'cg03395511', 'cg21548813'} |

**Table 7 GO Enrichment**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **# GREAT version 4.0.4** | **Species assembly: hg19** | **Association rule: Basal+extension: 5000 bp upstream, 1000 bp downstream, 1000000 bp max extension, curated regulatory domains included** |  |  |  |  |  |  |  |  |  |  |
| **# Term Name** | **Binom Rank** | **Binom Raw P-Value** | **Binom FDR Q-Val** | **Binom Fold Enrichment** | **Binom Observed Region Hits** | **Binom Region Set Coverage** | **Hyper Rank** | **Hyper FDR Q-Val** | **Hyper Fold Enrichment** | **Hyper Observed Gene Hits** | **Hyper Total Genes** | **Hyper Gene Set Coverage** |
| MHC class II protein complex | 1 | 0 | 0 | 141.1948 | 719 | 0.04185099 | 2 | 0.01530776455 | 2.394192 | 14 | 15 | 0.001936108 |
| MHC protein complex | 1 | 0 | 0 | 161.9402 | 1592 | 0.09266589 | 3 | 0.0182384463866667 | 2.180425 | 17 | 20 | 0.002350989 |
| clathrin-coated endocytic vesicle membrane | 1 | 0 | 0 | 6.930553 | 717 | 0.04173458 | 11 | 0.0454348944909091 | 1.656695 | 31 | 48 | 0.004287097 |

**Figure 1: Manhattan plot of all PEG1 meQTL cis associations (FDR<.05)**

**Figure 2: Manhattan plot of all meQTL trans associations in PEG1 (FDR<.05)**

