

# Thesis Figures

Main and supplementary, in order of appearance

Mary B Makarious

# Figures

- [Figure 1: Manolio Plot for Parkinson's Disease](#)
- [Figure 2: Six whole gene \*SNCA\* duplications were identified in the UK Biobank cohort](#)
- [Figure 3: Five full \*SNCA\* deletions and one \(likely\) partial \*SNCA\* deletion \(D\) were identified in the UK Biobank cohort](#)
- [Figure 4: Partial validation of genomic events of the genotyping array data using exome sequencing data](#)
- [Figure 5: Graphical representation of the analytical process for conducting large-scale rare variant burden testing in Parkinson's Disease](#)
- [Figure 6: Principal component plots per dataset included in large-scale rare variant burden meta-analyses](#)
- [Figure 7: Age distribution per dataset included in large-scale rare variant burden meta-analyses](#)
- [Figure 8: Analysis workflow schematic for the GWAS in African and African Admixed individuals](#)
- [Figure 9: Workflow diagram with case/control breakdown per dataset for GWAS in African and African Admixed Individuals](#)
- [Figure 10: African cohort with 1000 Genome populations](#)
- [Figure 11: African Admixed cohort with 1000 Genome populations](#)
- [Figure 12: Age distributions of cohorts involved in studying susceptibility of risk in the African and African admixed populations](#)
- [Figure 13: African Parkinson's disease risk GWAS](#)
- [Figure 14: African Admixed Parkinson's disease risk GWAS](#)
- [Figure 15: African and African Admixed GWAS Meta-analysis assessing Parkinson's disease risk](#)
- [Figure 16: \*GBA1\* - rs3115534 Genotypes versus age at Parkinson's disease onset](#)
- [Figure 17: LocusZoom plot displaying African and African Admixed Parkinson's disease GWAS Meta-analysis](#)
- [Figure 18: GCase activity analyses performed on \*GBA1\* - rs3115534-GG, rs3115534-GT, and rs3115534-TT carriers](#)
- [Figure 19: GCase activity analyses performed on \*GBA1\* - rs3115534-GG, rs3115534-GT, and rs3115534-TT carriers](#)
- [Figure 20: LocusZoom plots of \*GBA1\* in AFR/AAC, EUR, EAS, AMR populations](#)
- [Figure 21: Beta-beta plot comparison of African versus African Admixed estimates for PD known risk loci identified in Europeans](#)
- [Figure 22: Density plots showing polygenic risk score distributions in the African and African Admixed individuals using the 90 Parkinson's disease risk loci](#)
- [Figure 23: Comparative Odds Ratio Analysis Across Different Cohorts for rs3115534-G Variant in Parkinson's Disease Studies](#)
- [Figure 24: Miami Plot comparing European versus African and African admixed GWAS meta-analysis](#)
- [Figure 25: Power calculations for the meta-GWAS in African and African admixed populations](#)
- [Figure 26: Population attributable risk comparison for \*GBA1\* known coding variants in the EUR population versus the novel \*GBA1\* intronic variant in the AFR population](#)
- [Figure 27: GenoML Workflow](#)
- [Figure 28: PPMI and PDBP Age Distribution](#)
- [Figure 29: Workflow and data summary for multi-omics of PD risk prediction using ML study](#)
- [Figure 30: Correlation matrix of top 5% contributing features in ML model](#)
- [Figure 31: Receiver operating characteristic curves and case probability density plots in withheld training samples at default thresholds comparing performance metrics in different data modalities from the PPMI dataset](#)
- [Figure 32: Receiver operating characteristic and case probability density plots in the external dataset \(PDBP\) at validation for the trained and then tuned models at default thresholds](#)
- [Figure 33: Feature importance plots for top 5% of features in XGBoost surrogate of best combined \\*omics model](#)
- [Figure 34: Network plot of nominated genes following best combined multi-model ML prediction model](#)
- [Figure 35: Misclassified case as a healthy control using the best multi-modal model](#)

# Effect Size

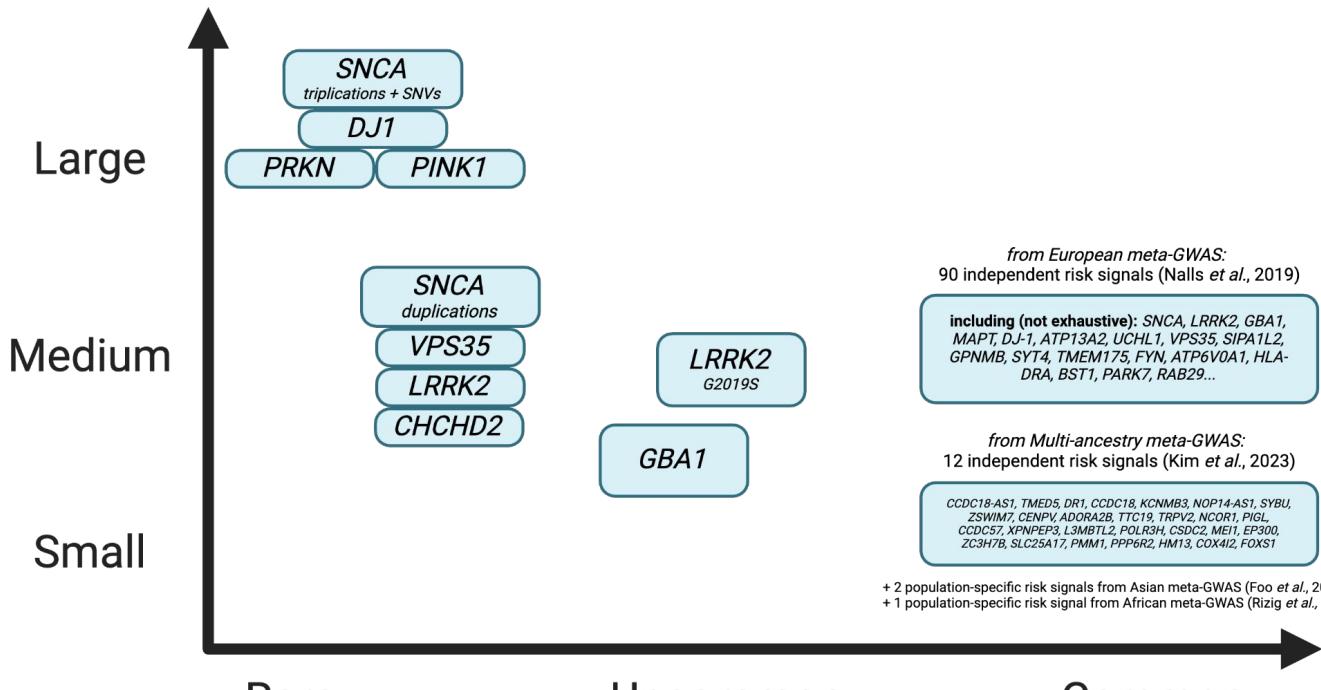
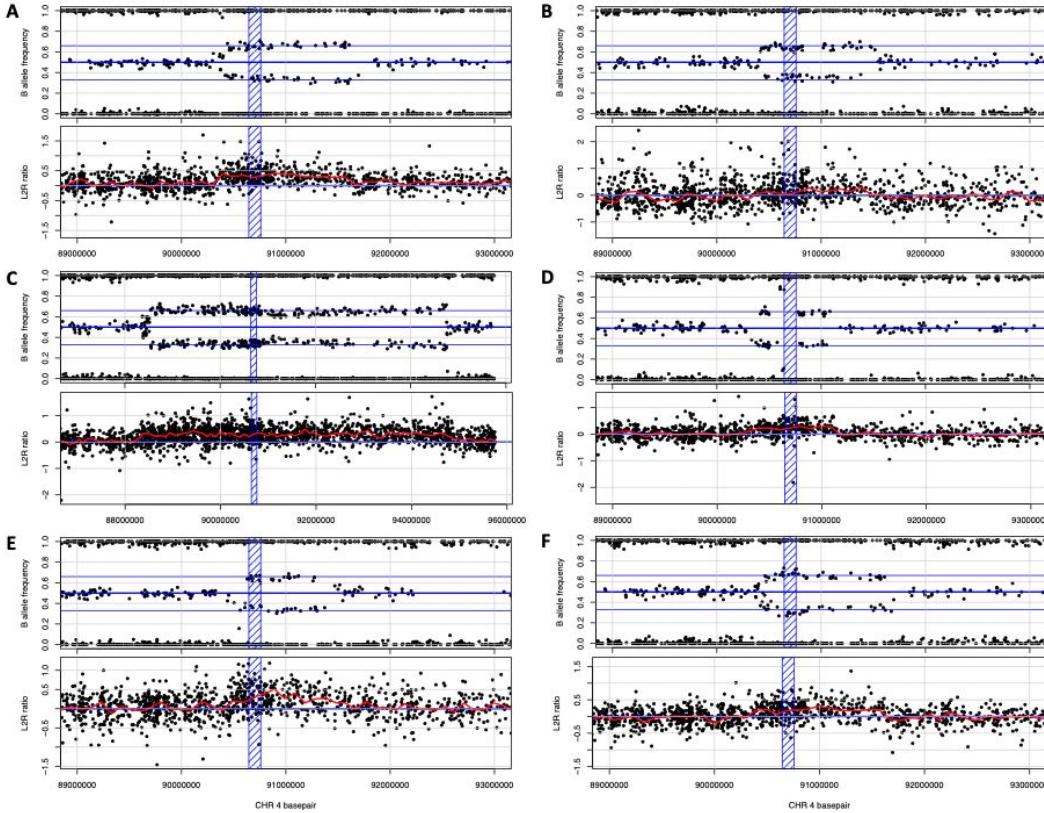
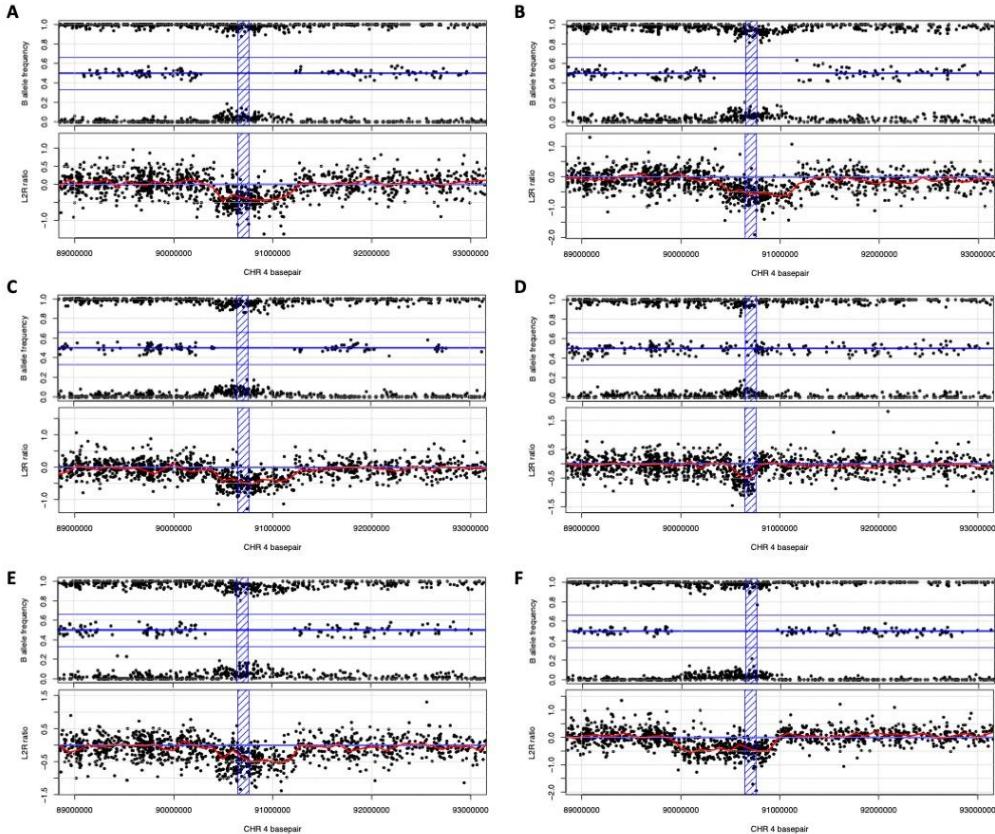


Figure 1: Manolio Plot

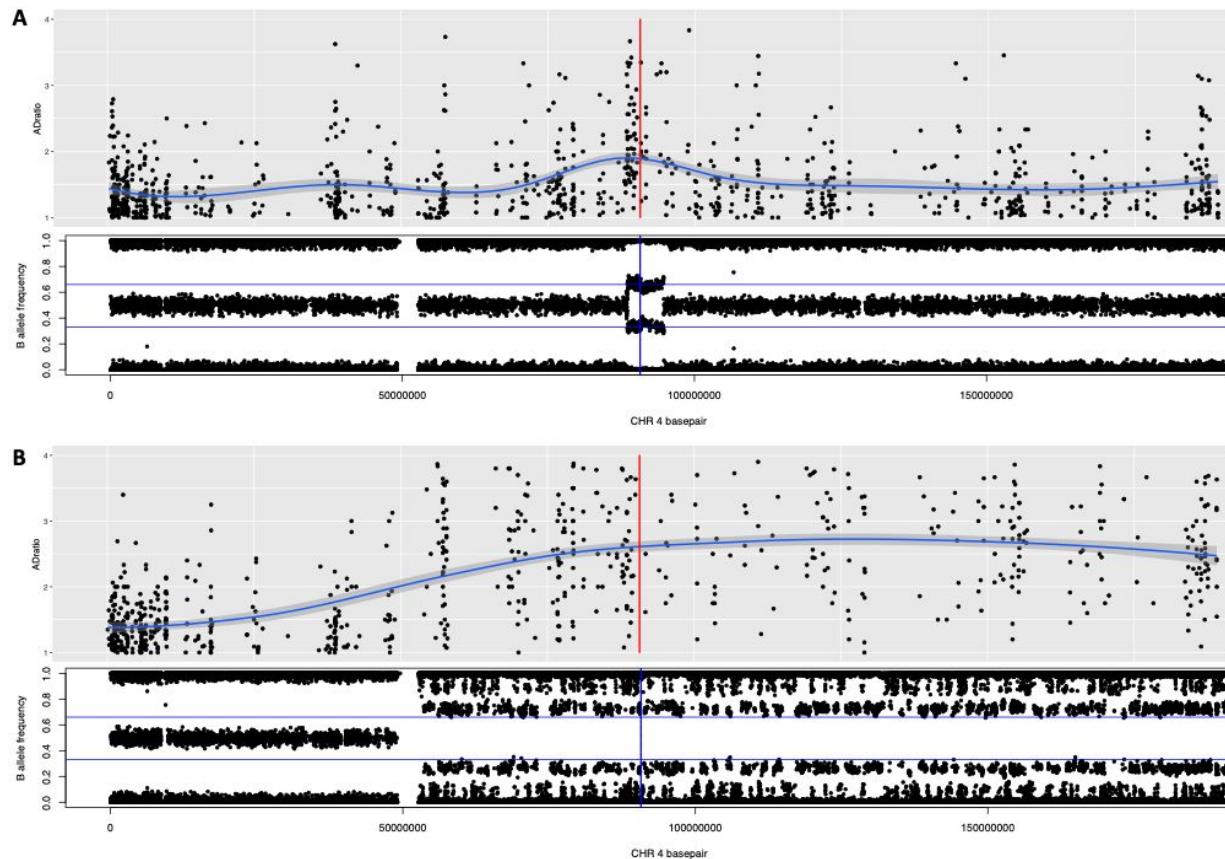
Chapter 1



**Figure 2:** Six whole gene *SNCA* duplications were identified in the UK Biobank cohort  
*Chapter 2*



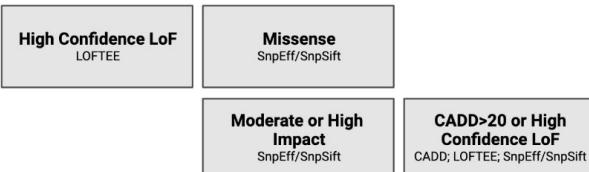
**Figure 3:** Five full *SNCA* deletions and one (likely) partial *SNCA* deletion (D) were identified in the UK Biobank cohort  
Chapter 2



**Figure 4:** Partial validation of genomic events of the genotyping array data using exome sequencing data. A) Duplication in the *SNCA* gene region (subject #dup3), B) Partial duplication of the long arm of chromosome 4 (subject #comp8). Red and blue vertical line represents the *SNCA* gene body.

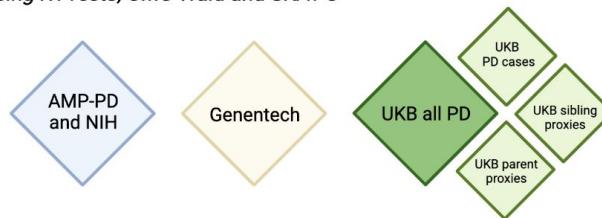
### 1) Annotation and Variant Groups

Using VEP (LOFTEE; CADD), SnpEff, and SnpSift



### 2) PD Risk Burden Analysis per Dataset

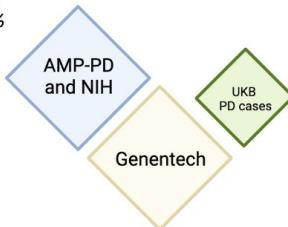
Using RVTests; CMC Wald and SKAT-O



### 3) PD Risk Burden Meta-analysis

Cases and Controls

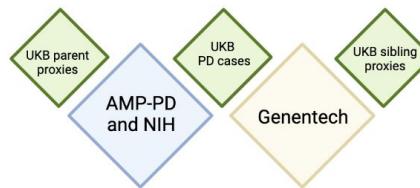
MAF <0.1% and <1%



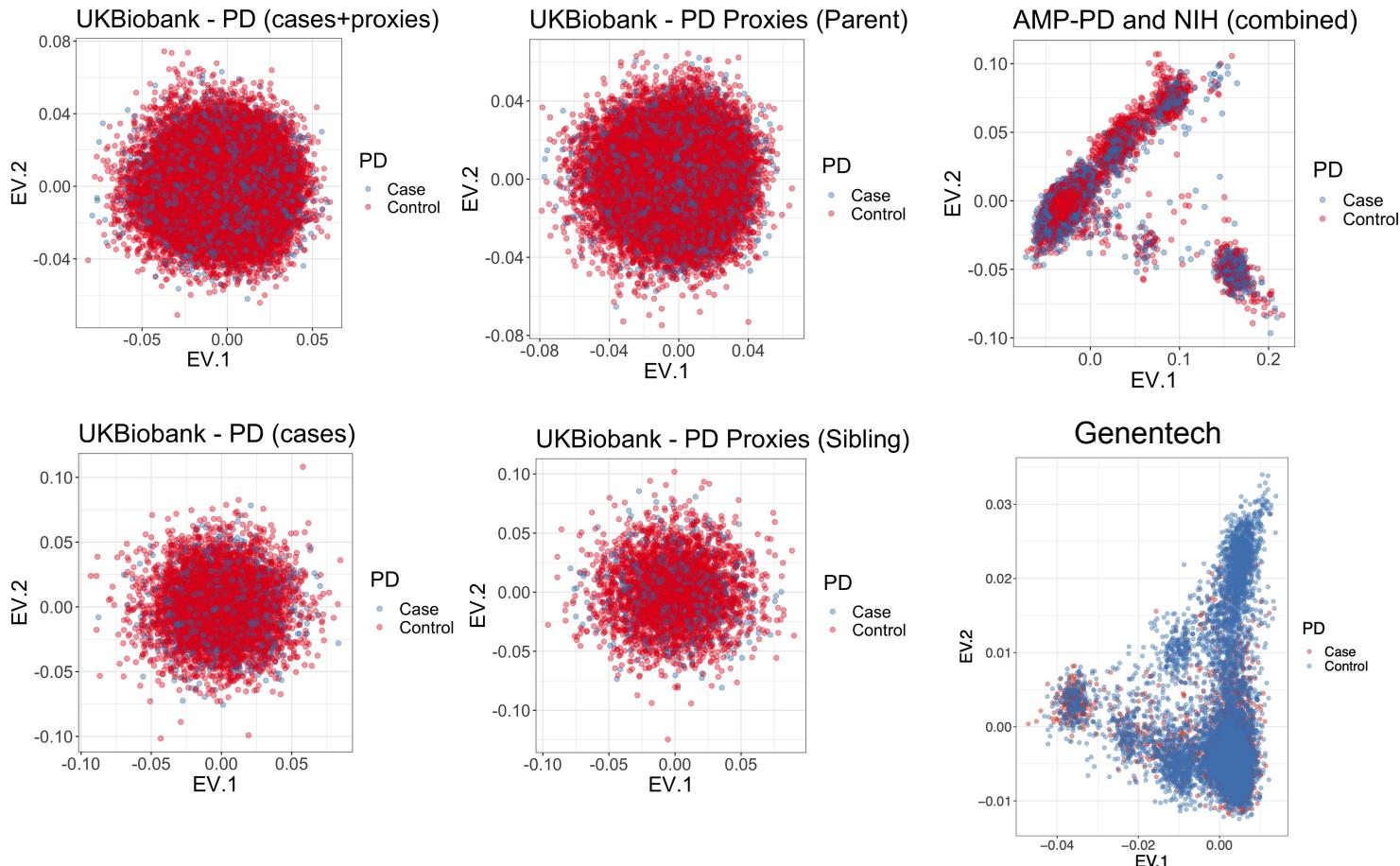
### 4) PD Risk Burden Meta-analysis

Cases, Proxies, and Controls

MAF <0.1% and <1%

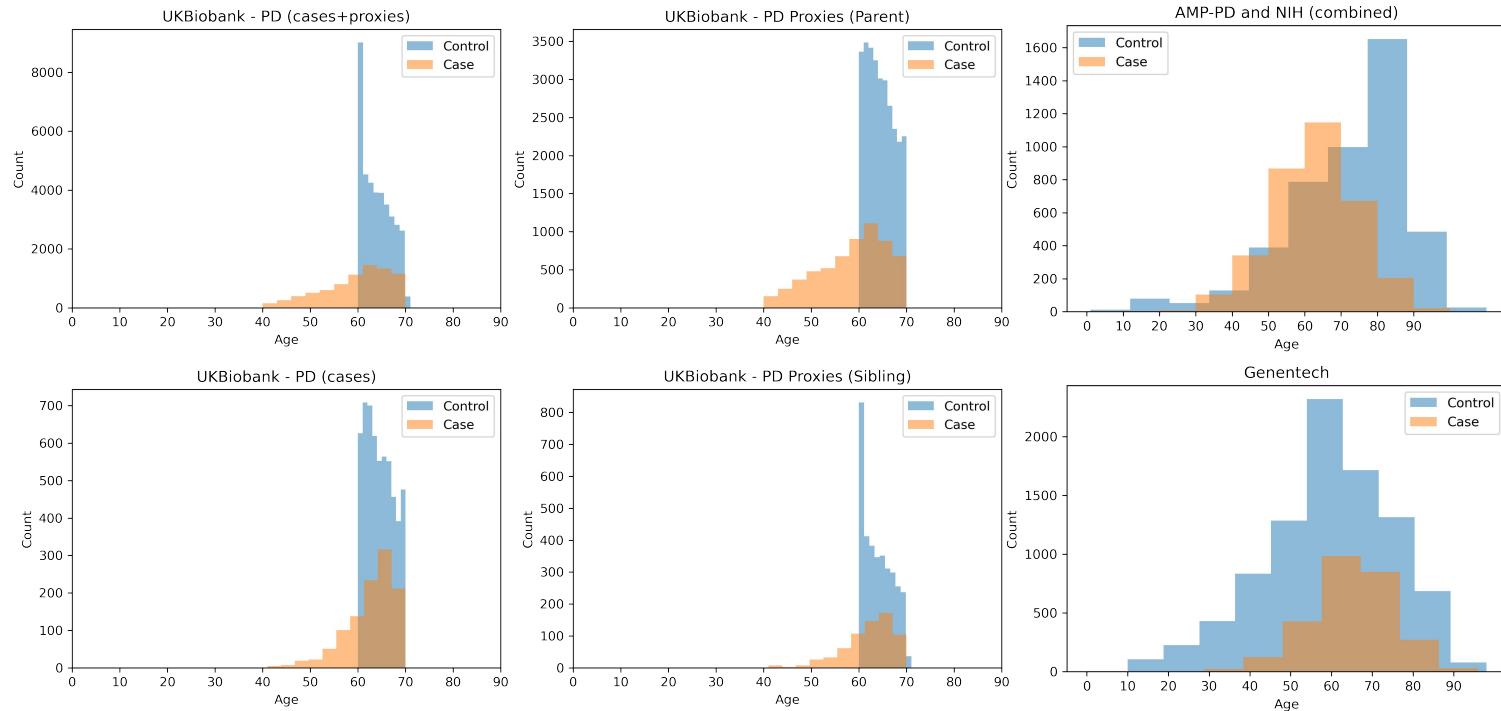


**Figure 5:** Graphical representation of the analytical process for conducting large-scale rare variant burden testing in Parkinson's Disease  
Chapter 3



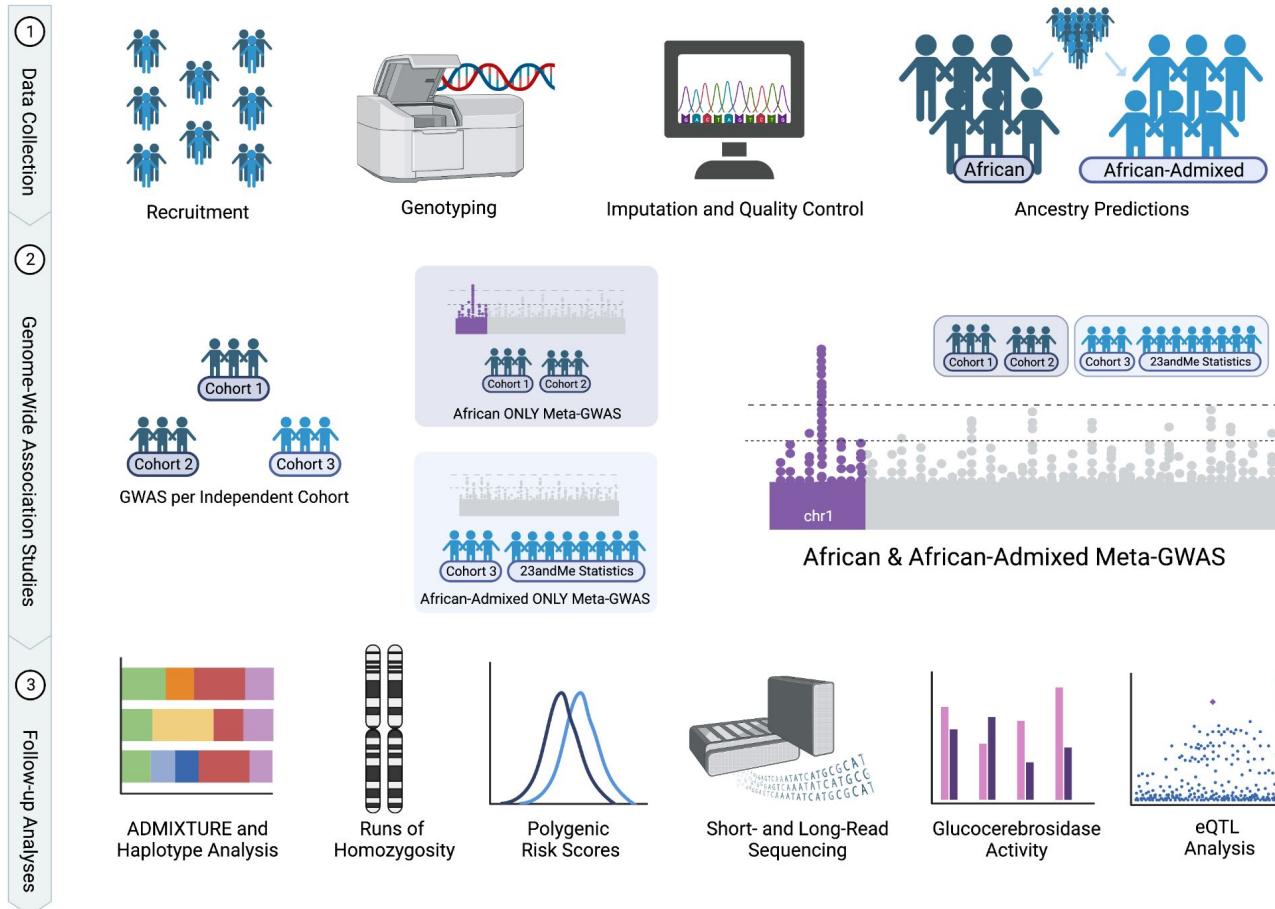
**Figure 6:** Principal component plots per dataset included in large-scale rare variant burden meta-analyses

Chapter 3



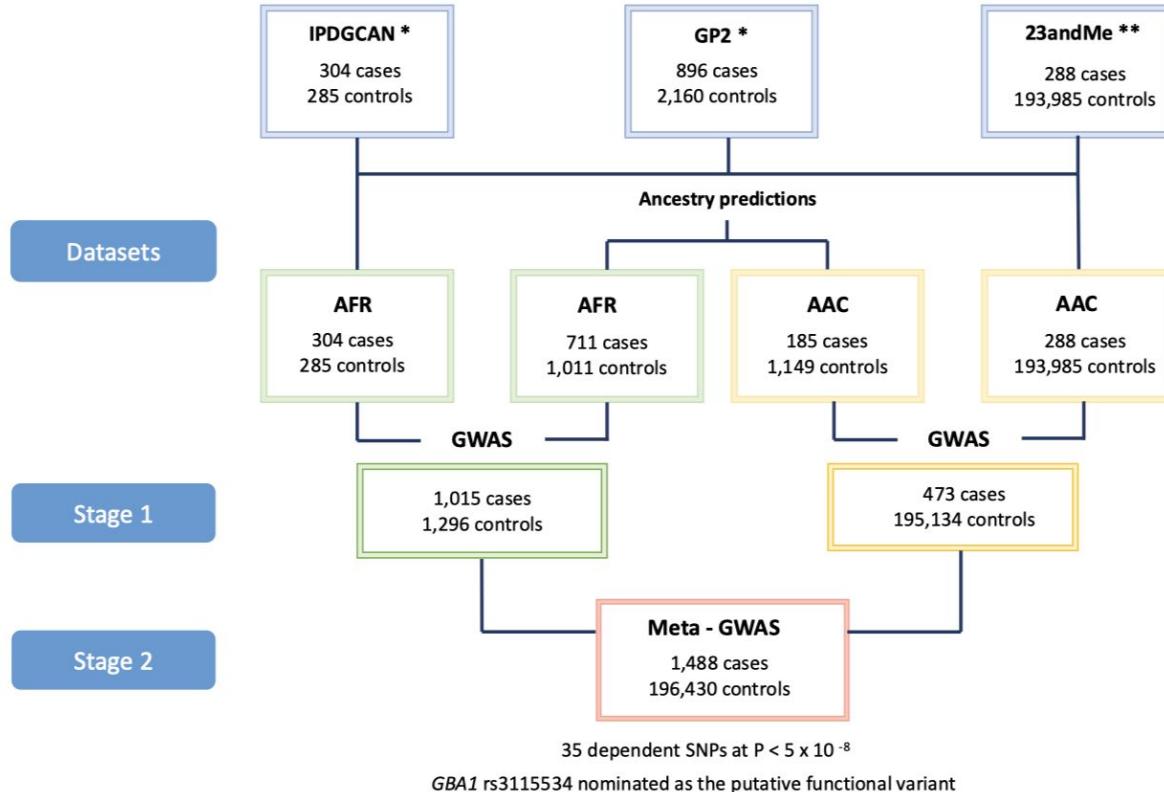
**Figure 7:** Age distribution per dataset included in large-scale rare variant burden meta-analyses

Chapter 3



**Figure 8:** Analysis workflow schematic for the GWAS in African and African Admixed individuals

Chapter 4



**Figure 9:** Workflow diagram with case/control breakdown per dataset for GWAS in African and African Admixed Individuals

\* Genotyped samples underwent quality control procedures before taking part in this study

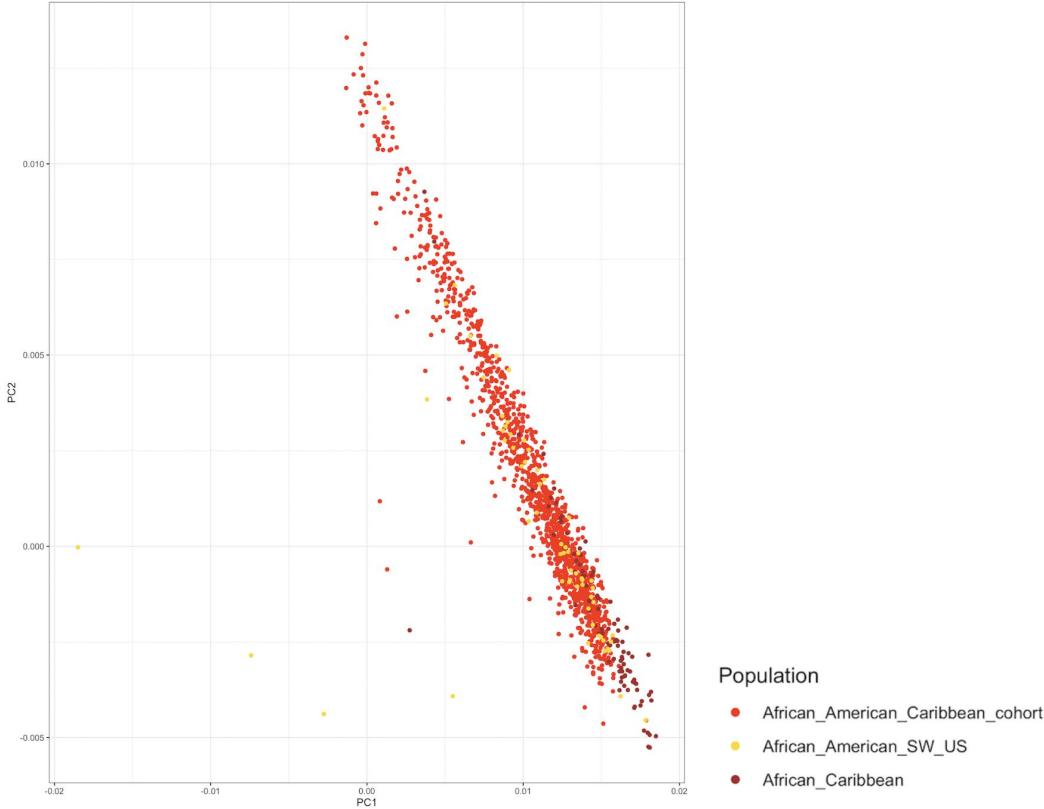
\*\* Summary statistics were obtained via collaboration with 23andMe



**Figure 10:** African cohort with 1000 Genome populations

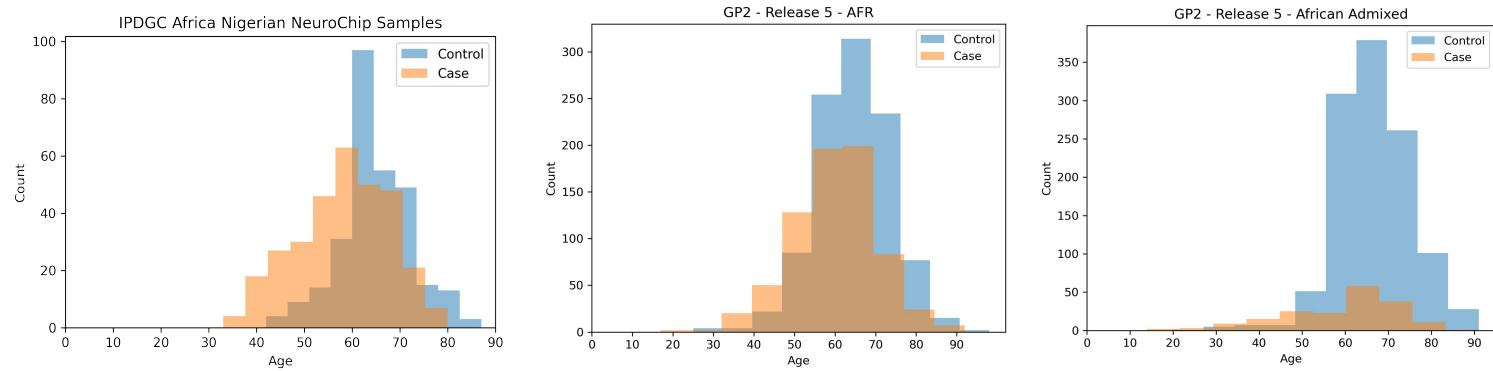
Principal component analysis plots displaying African samples under study (red) with 1000 Genome African sub-populations

Chapter 4

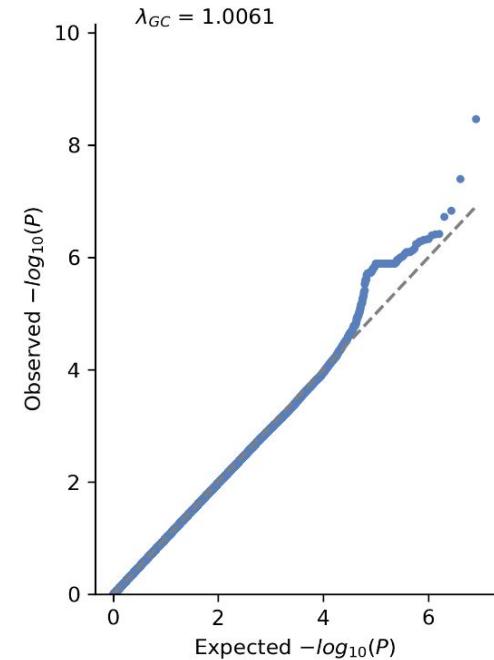
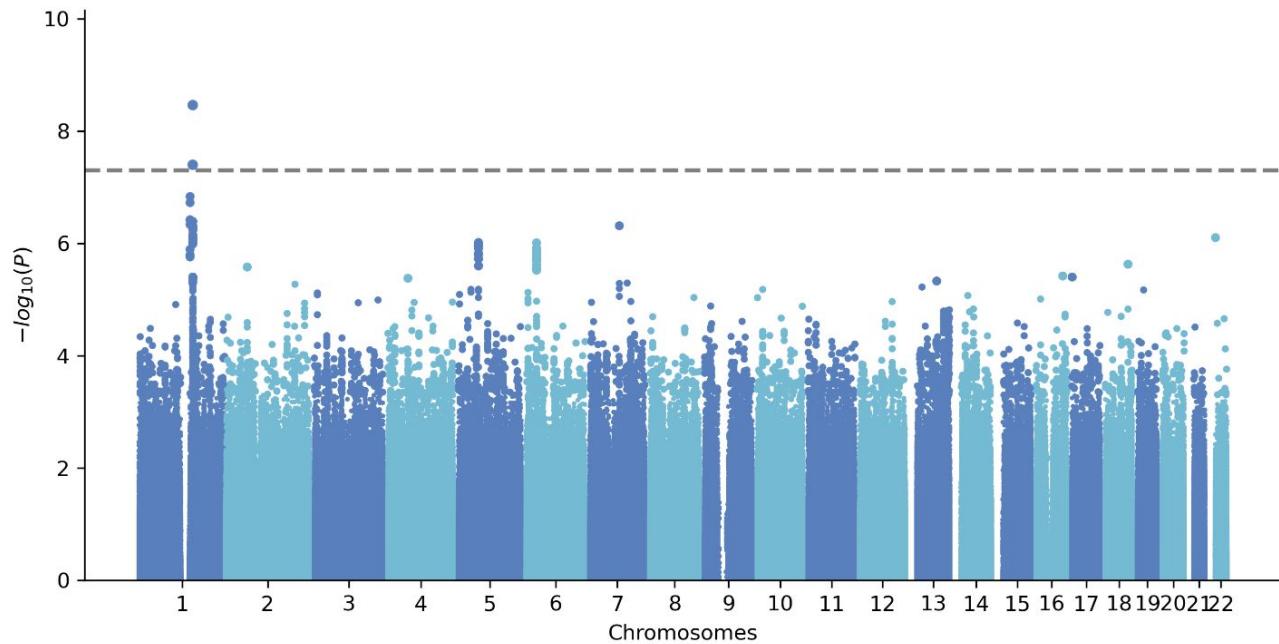


**Figure 11:** African Admixed cohort with 1000 Genome populations

Principal component analysis plots displaying African admixed samples under study (red) with 1000 Genome populations including Africans from the Southwest of the US (yellow) and African Caribbeans (brown)

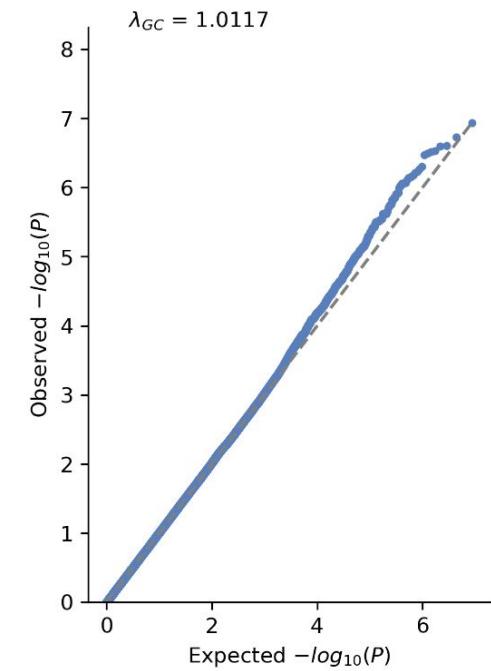
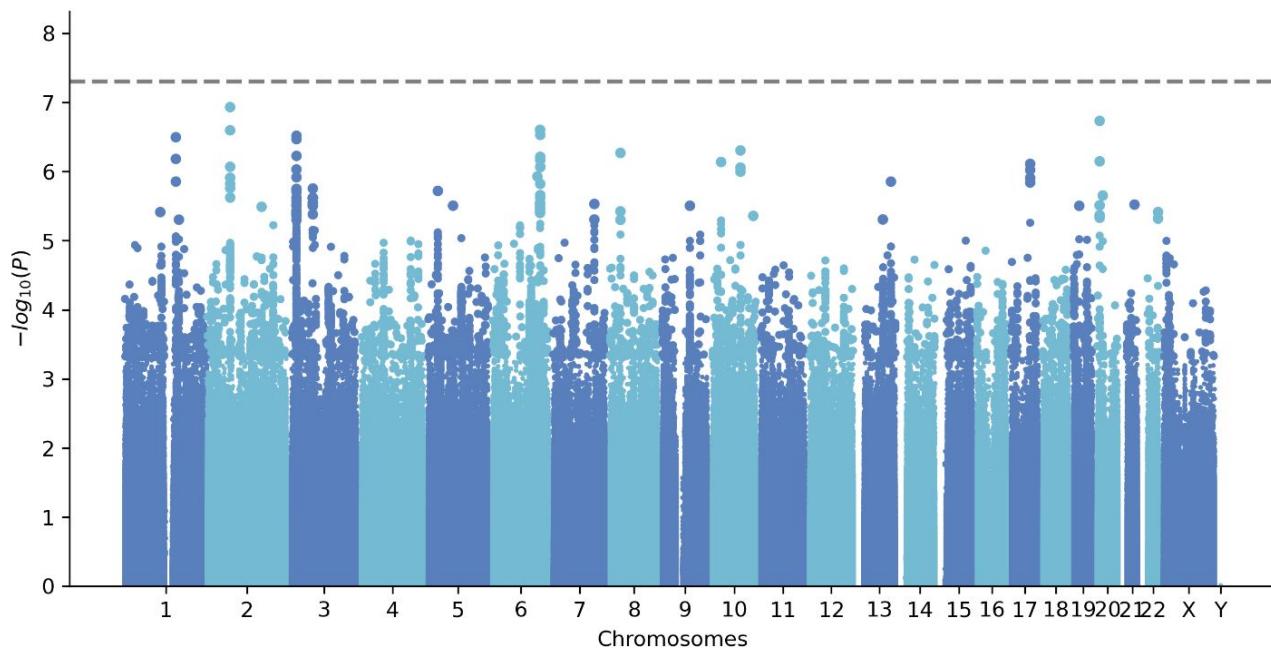


**Figure 12:** Age distributions of cohorts involved in studying susceptibility of risk in the African and African admixed populations  
 IPDGC: International Parkinson's disease Genomics Consortium; GP2: Global Parkinson's Genetics Program; AFR: African  
*Chapter 4*



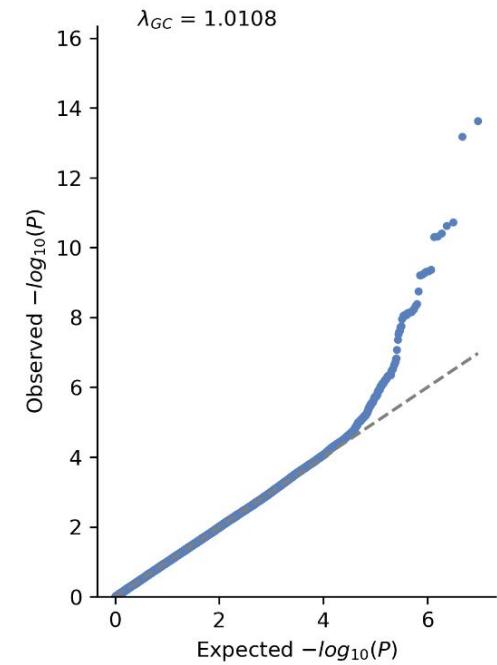
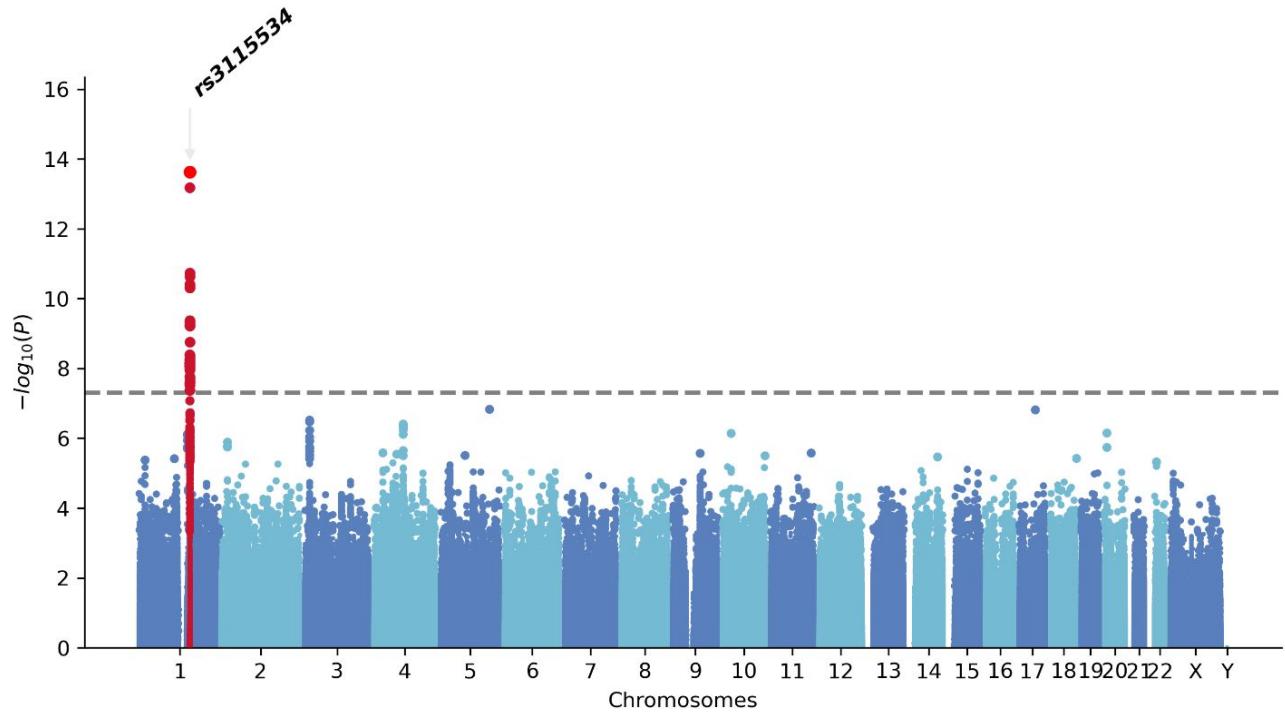
**Figure 13:** African Parkinson's disease risk GWAS

Manhattan plot displaying the association statistical significance as  $-\log_{10}(p\text{-value})$  in the y-axis against chromosomes in the x-axis at a genome scale (Bonferroni correction highlighted at 5E-8). Quantile-quantile plot displaying genetic data distribution;  $\lambda_{GC}$ : Lambda value representing genomic inflation



**Figure 14:** African Admixed Parkinson's disease risk GWAS

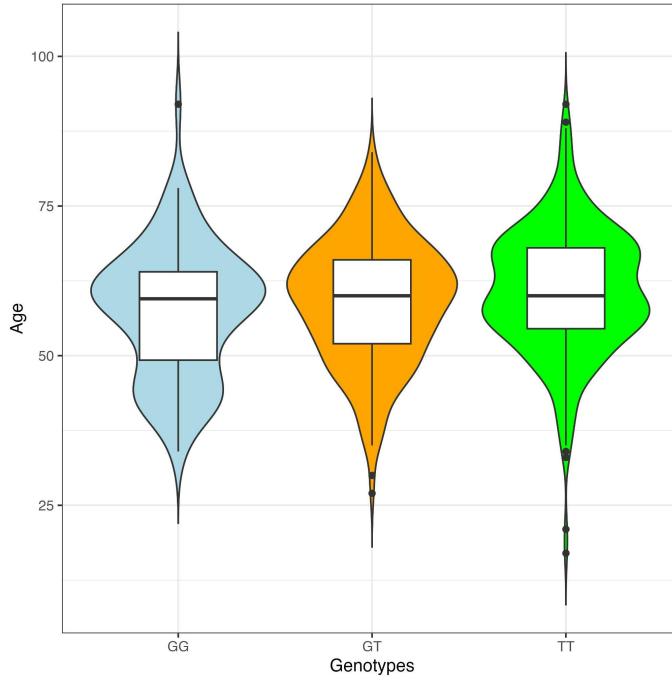
Manhattan plot displaying the association statistical significance as  $-\log_{10}(p\text{-value})$  in the y-axis against chromosomes in the x-axis at a genome scale (Bonferroni correction highlighted at  $5E-8$ ). Quantile-quantile plot displaying genetic data distribution;  $\lambda_{GC}$ : Lambda value representing genomic inflation



**Figure 15:** African and African Admixed GWAS Meta-analysis assessing Parkinson's disease risk

Manhattan plot displaying the association statistical significance as  $-\log_{10}(p\text{-value})$  in the y-axis against chromosomes in the x-axis at a genome scale (Bonferroni correction highlighted at  $5\text{E}-8$ ). Quantile-quantile plot displaying genetic data distribution;  $\lambda_{GC}$ : Lambda value representing genomic inflation

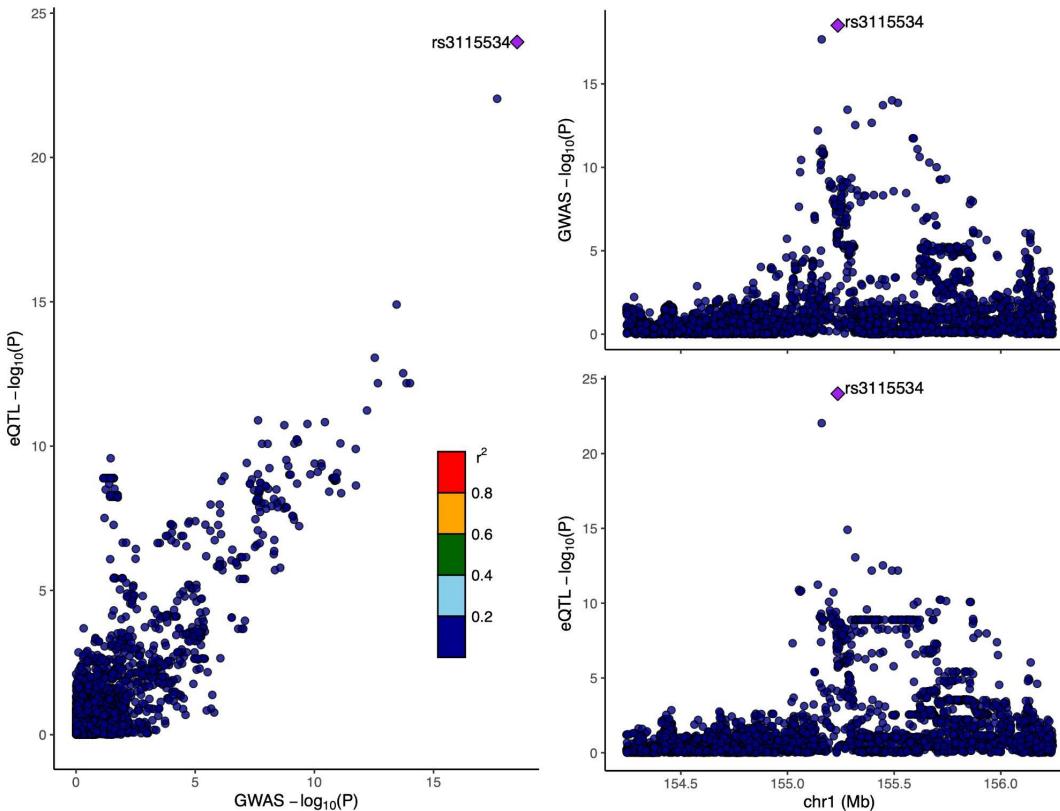
Chapter 4



**Figure 16:** *GBA1* - rs3115534 Genotypes versus age at Parkinson's disease onset

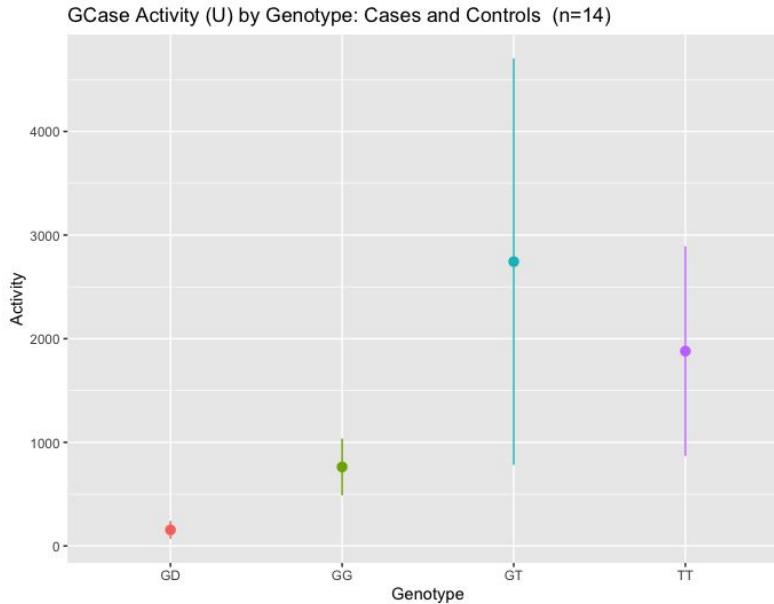
rs3115534-GG versus age at onset; BETA =-1.96, SE = -0.64, P=0.002; rs3115534-GT versus age at onset; BETA =-2.28, SE =0.85 , P=0.007

Chapter 4

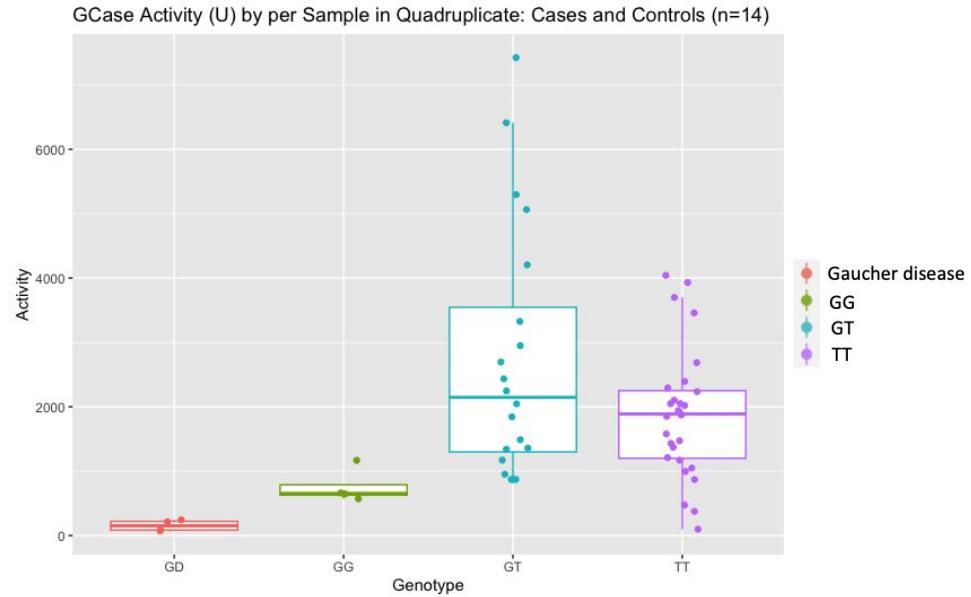


**Figure 17:** LocusZoom plot displaying African and African Admixed Parkinson's disease GWAS Meta-analysis summary statistics versus African expression quantitative trait locus summary statistics from blood (by Kachuri et al., 2023)  
Chapter 4

## GCase Activity (U) by Genotype Average (All)



## GCase Activity (U) per Sample by Genotype (All)



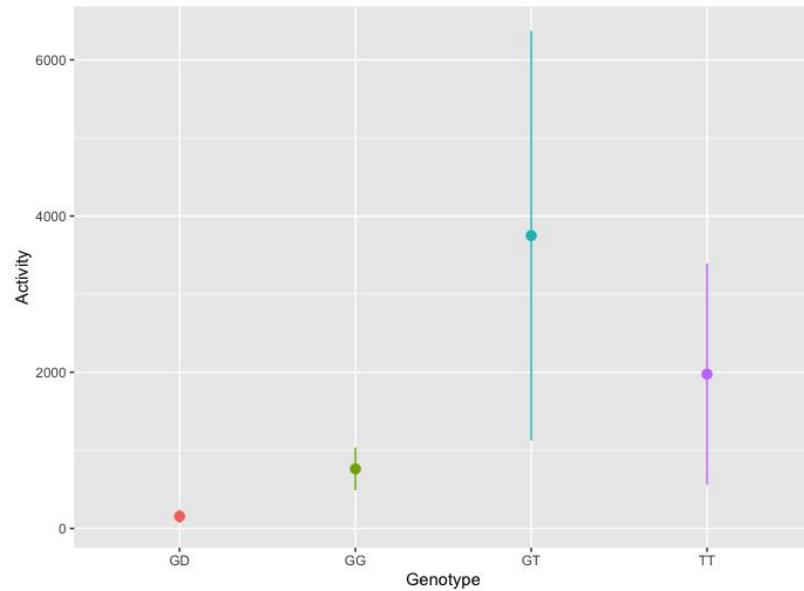
**Figure 18:** GCase activity analyses performed on *GBA1* - rs3115534-GG, rs3115534-GT, and rs3115534-TT carriers

A fluorometric 4-MU assay was used to measure GCase activity in 14 lymphoblastoid cell lines, including a type I Gaucher disease (GD) patient as a positive control. A) Samples were aggregated by rs3115534 genotype and average activity. Values are represented by mean and standard deviation. B) All 14 samples were run in quadruplicate. Samples were screened for known *GBA1* pathogenic mutations that could bias these estimates. A total of two carriers (one heterozygous for *GBA1* p.I320S and one heterozygous for *GBA1* p.T75del) were removed from further analyses.

Chapter 4

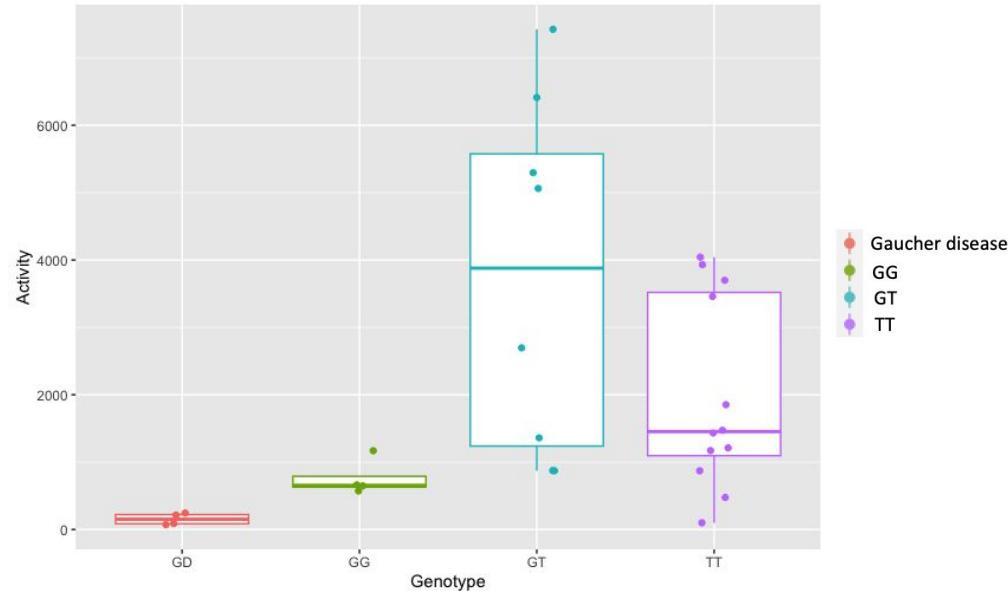
## GCase Activity (U) by Genotype Average (PD Only)

GCase Activity (U) by Genotype: PD Cases Only (n=7)



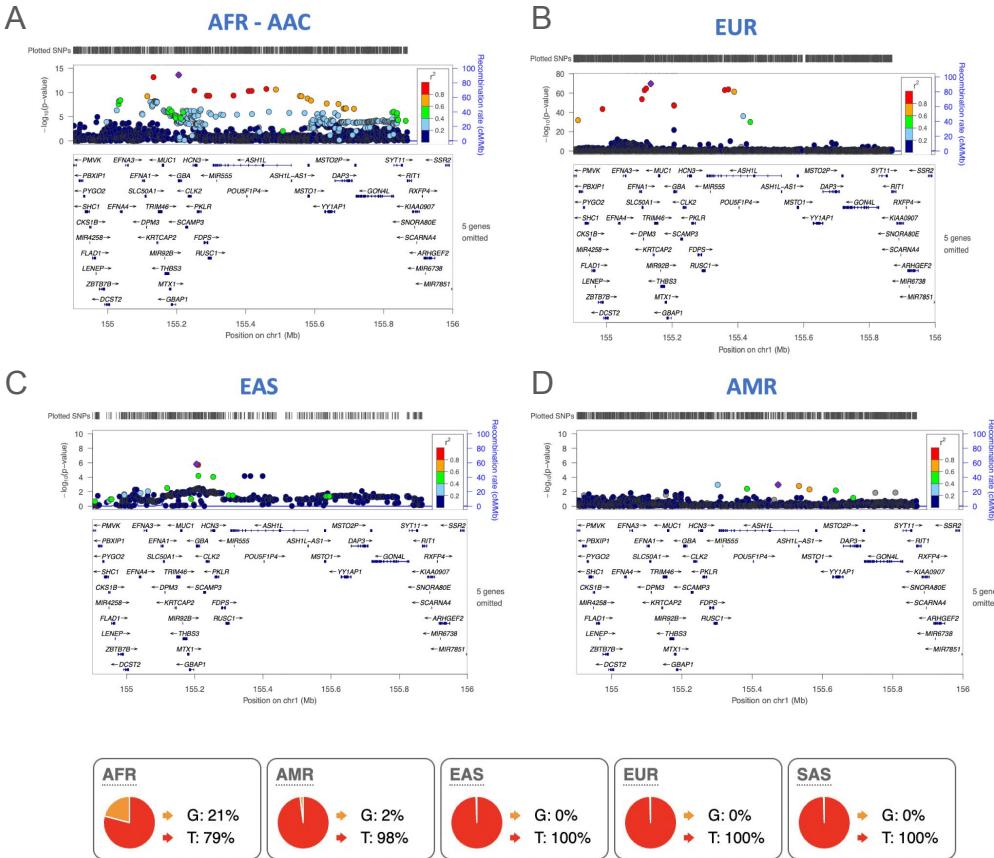
## GCase Activity (U) by Genotype per Sample (PD Only)

GCase Activity (U) by per Sample in Quadruplicate: PD Cases Only (n=7)



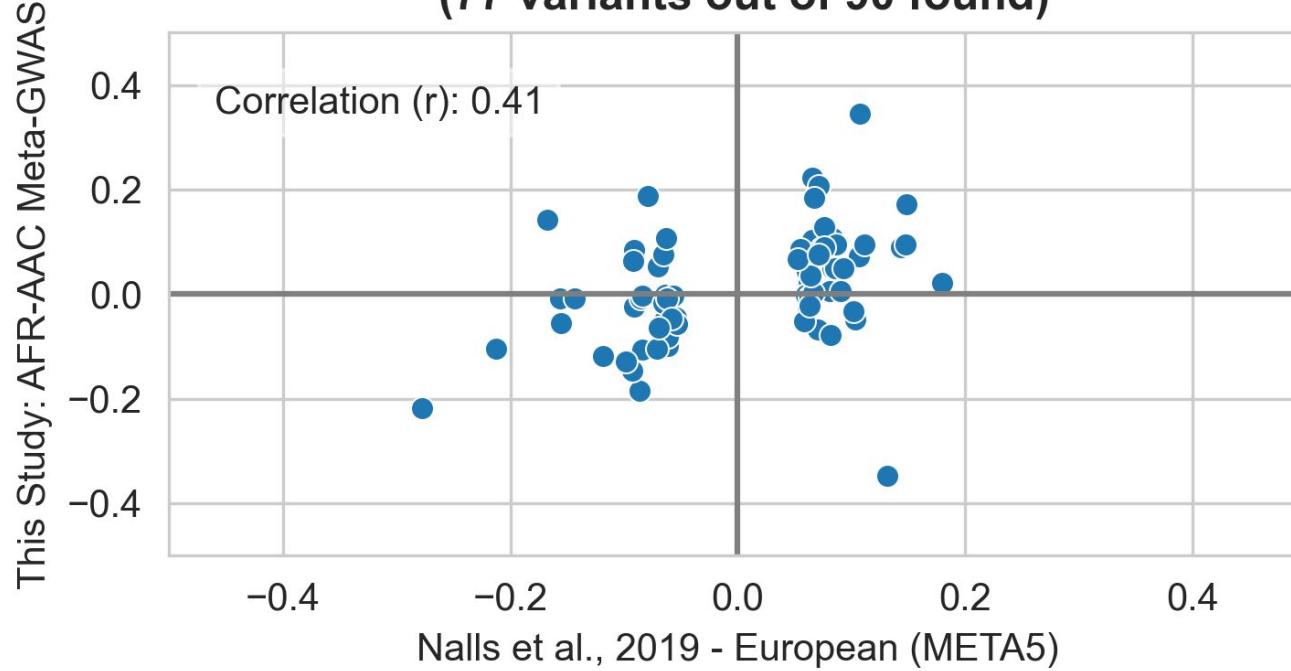
**Figure 19:** GCase activity analyses performed on *GBA1* - rs3115534-GG, rs3115534-GT, and rs3115534-TT carriers

Samples with Parkinson's disease were pulled from the 14 samples, including a type I Gaucher disease (GD) patient as a positive control. **A)** A total of 9 samples with Parkinson's disease were aggregated by rs3115534 genotype and average activity. Values are represented by mean and standard deviation. **B)** All 9 samples with Parkinson's were run in quadruplicate (Welch Two Sample t-test: GG versus GT;  $t = -3.189$ ,  $df = 7.3002$ ,  $p\text{-value} = 0.0.01446$ ; GG versus TT;  $t = -2.8158$ ,  $df = 13.003$ ,  $p\text{-value} = 0.01458$ ; GT versus TT;  $t = 1.7509$ ,  $df = 9.7545$ ,  $p\text{-value} = 0.1113$ ). A total of two carriers (one heterozygous for *GBA1* p.I320S and one heterozygous for *GBA1* p.T75del) were removed from further analyses.



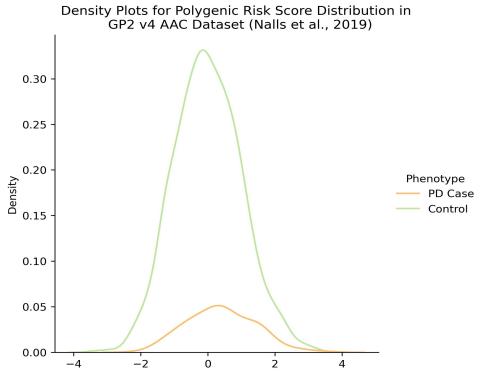
**Figure 20:** LocusZoom plots of *GBA* in AFR/AAC (A), EUR (B), EAS (C), AMR (D) populations  
*Chapter 4*

## Beta-Beta Plot: META5 vs AFR-AAC Meta-GWAS (77 variants out of 90 found)

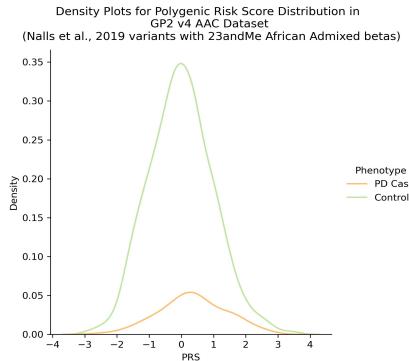


**Figure 21:** Beta-beta plot comparison of African versus African Admixed estimates for PD known risk loci identified in Europeans  
Chapter 4

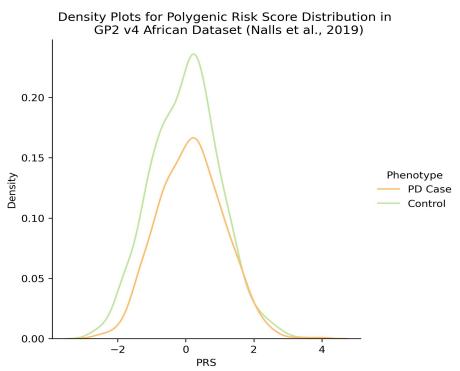
A



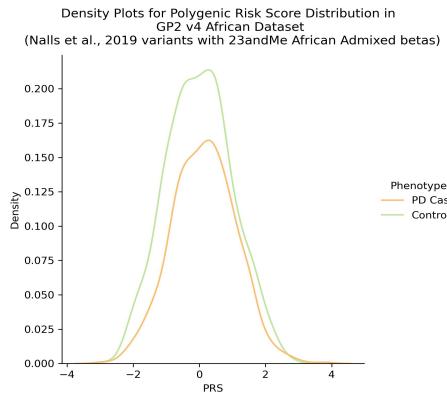
B



C



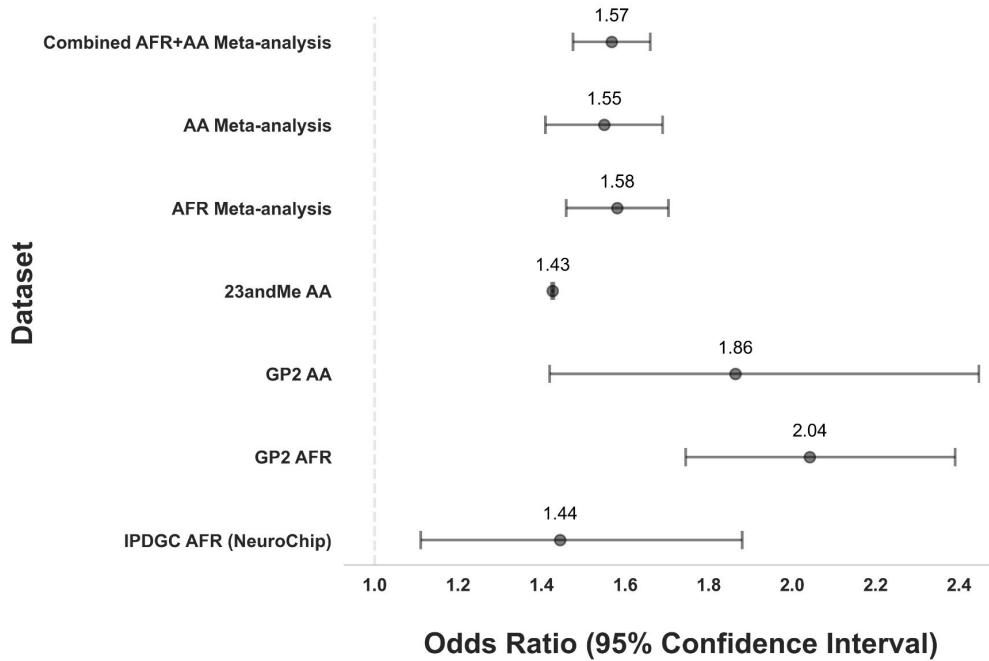
D



**Figure 22:** Density plots showing polygenic risk score distributions in the African and African Admixed individuals using the 90 Parkinson's disease risk loci

A) Nalls et al., 2019 reference estimates on African admixed individual level data; B) 23andMe African admixed reference estimates on African admixed individual level data  
C) Nalls et al., 2019 reference estimates on African individual level data ; D) 23andMe African admixed reference estimates on African individual level data

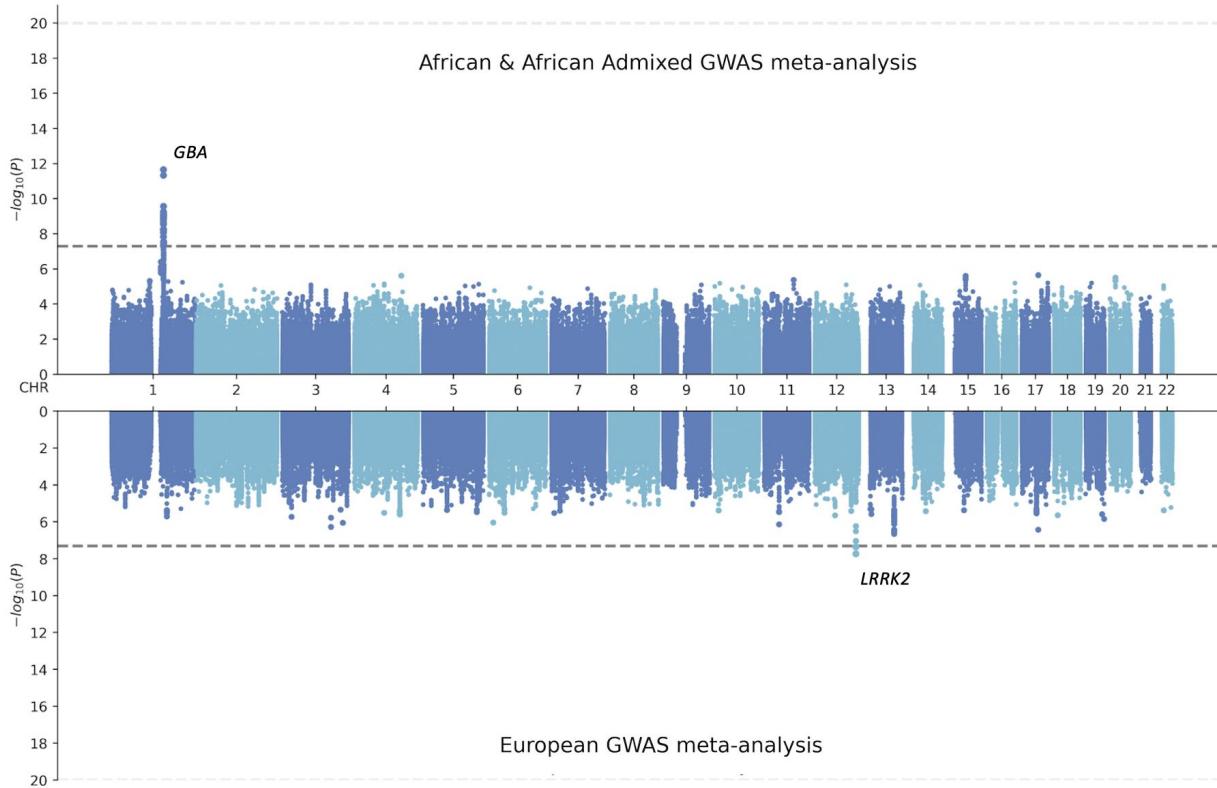
## Odds Ratio Analysis of rs3115534-G in Cohorts of Study



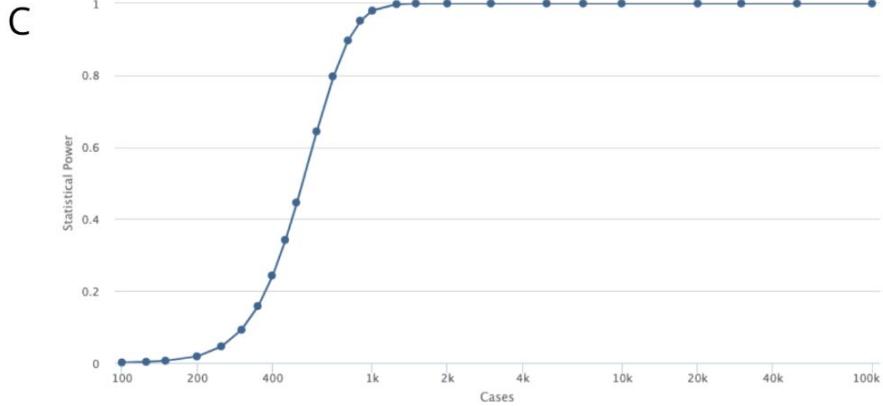
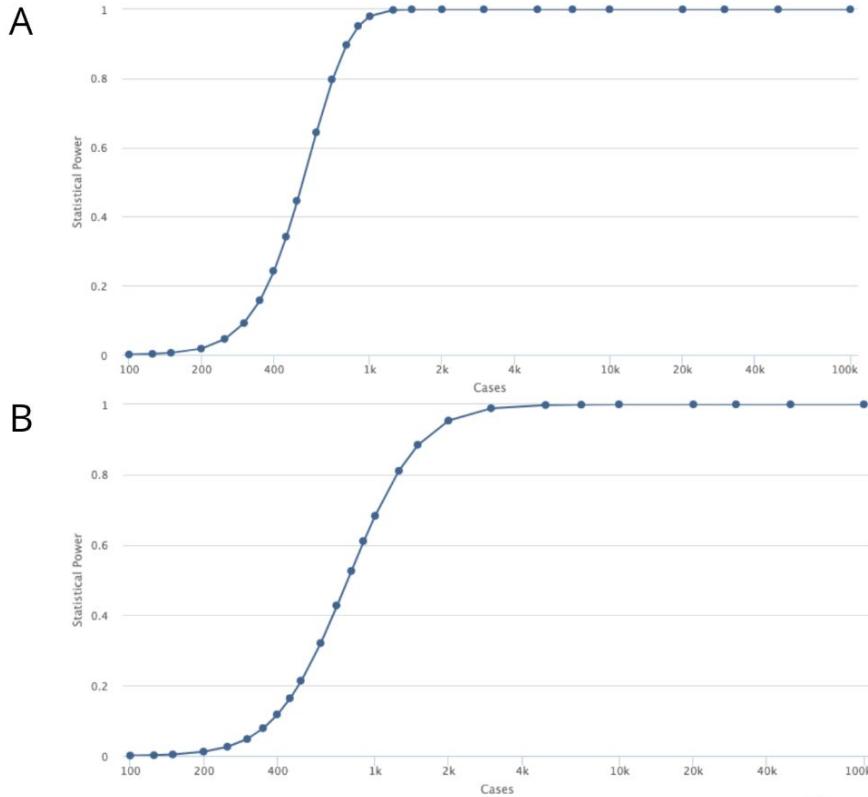
**Figure 23:** Comparative Odds Ratio Analysis Across Different Cohorts for rs3115534-G Variant in Parkinson's Disease Studies

AFR: African; AA: African admixed; IPDGC: International Parkinson's Disease Genomics Consortium

Chapter 4



**Figure 24:** Miami Plot comparing European versus African and African admixed GWAS meta-analysis assessing Parkinson's disease risk that are similarly randomly sampled ( $n_{\text{cases}} = 1,200$ ;  $n_{\text{controls}} = 2,445$ ). Randomly sampled 1,200 cases and 2,445 controls from European, African, and African admixed individuals for this analysis to illustrate that at this scale we are only powered to detect SNPs with the greatest association to increased disease susceptibility. The current European meta-analysis has 90 independent genome-wide significant risk signals in 78 genomic regions identified by Nalls and colleagues.



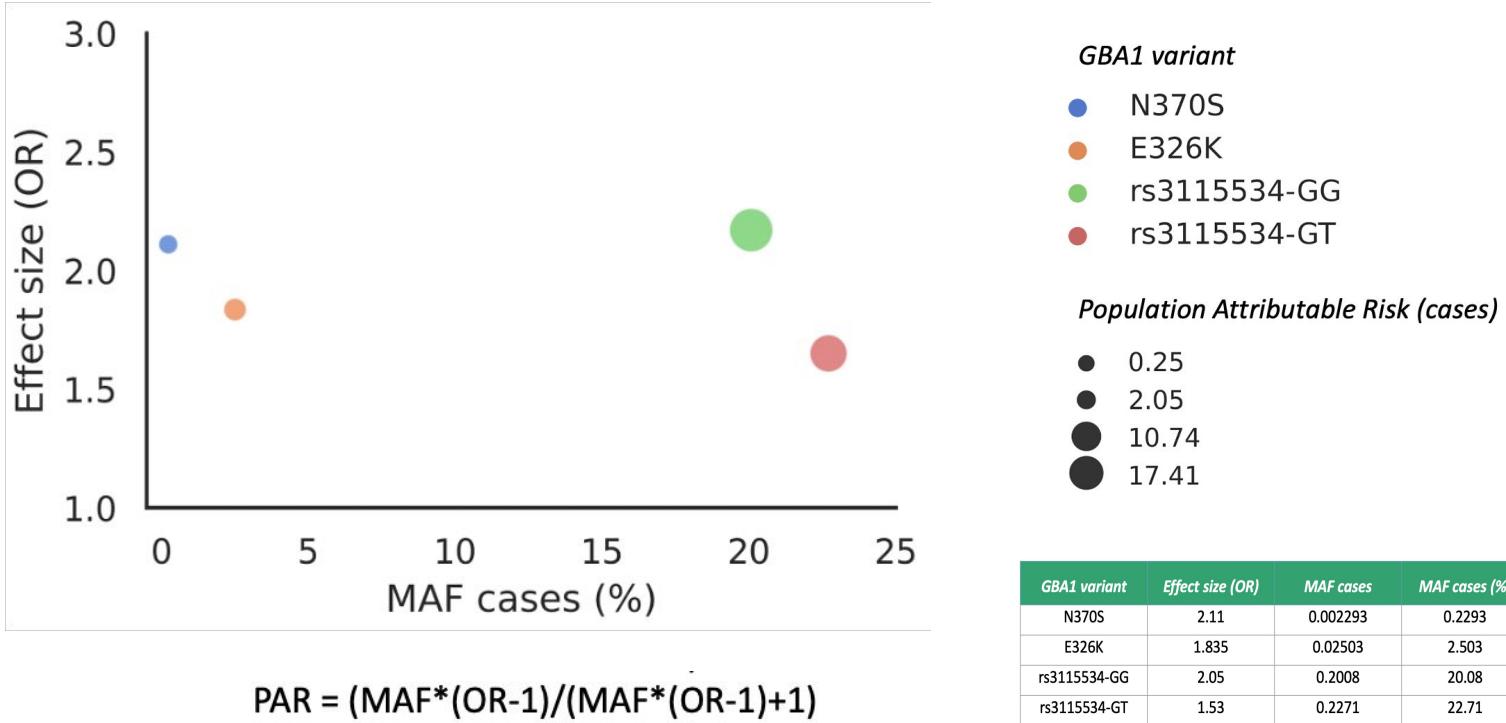
	Expected power for a one-stage study
<b>African Admixed</b> (477 cases; >100K controls)	0.404
<b>African</b> (1,015 cases; 1,296 controls)	0.693
<b>African and African Admixed</b> (1,488 cases; >100K controls)	1

**Figure 25: Power calculations**

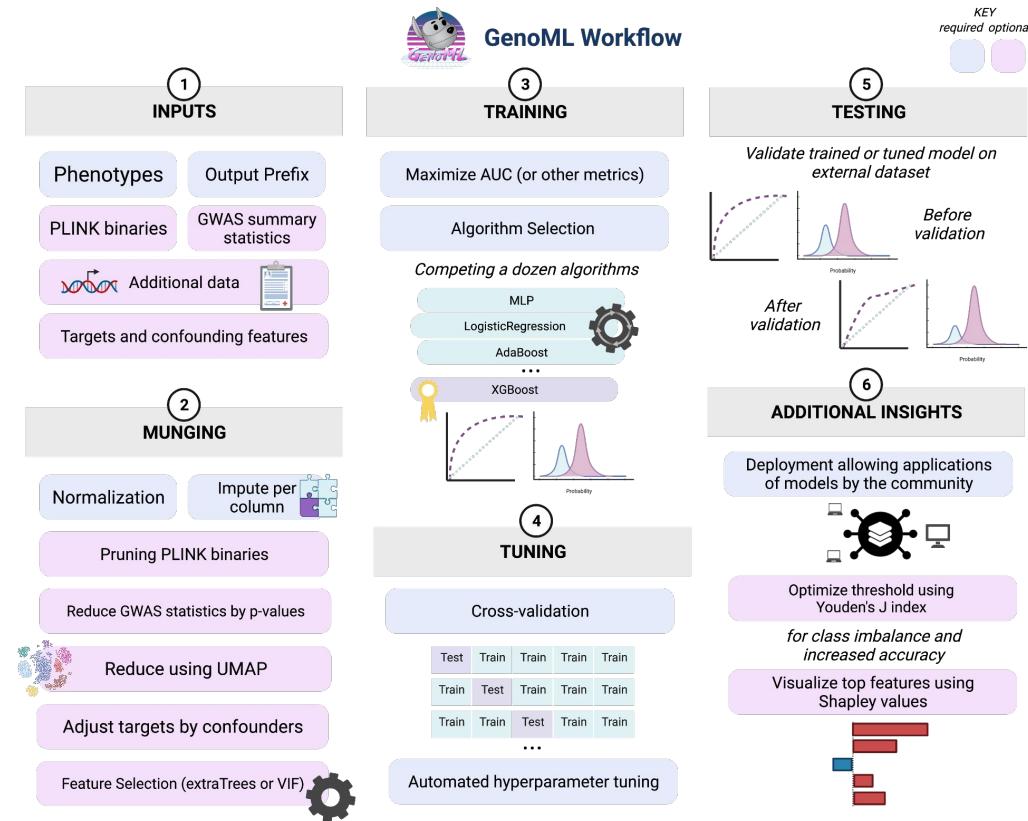
Calculated using the Genetic Association Study (GAS) power calculator provided by the University of Michigan under an additive model

Odds ratio of 1.58 for rs3115534; average of 19.5% disease allele frequency, and a 0.6% estimated prevalence in the African populations (according to Global, regional, and national burden of Parkinson's disease [Lancet Neurology 2018])

(A) African admixed (477 cases; >100K controls); (B) African (1,015 cases; 1,296 controls); (C) African and African admixed (1,488 cases; >100K controls)

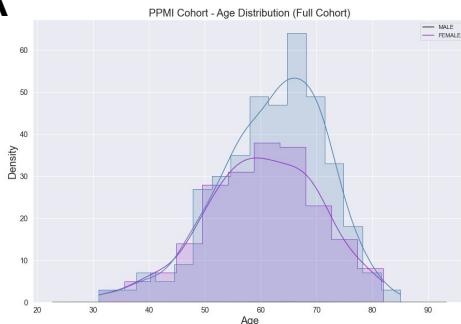
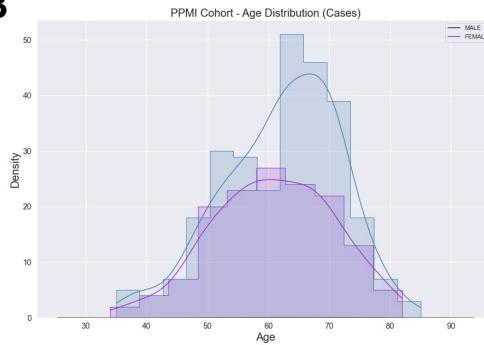
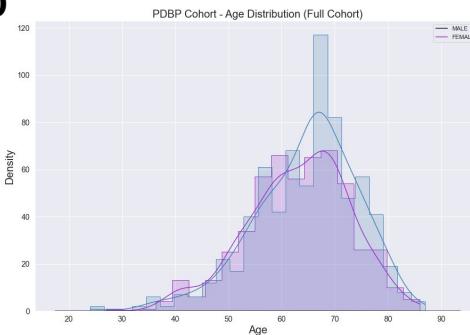
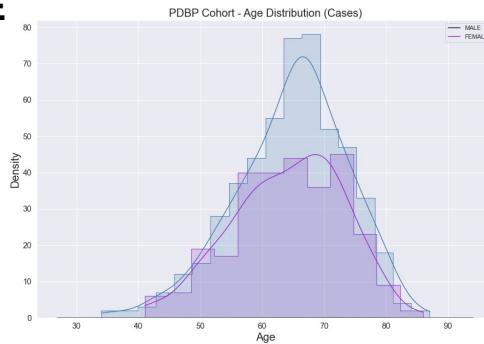
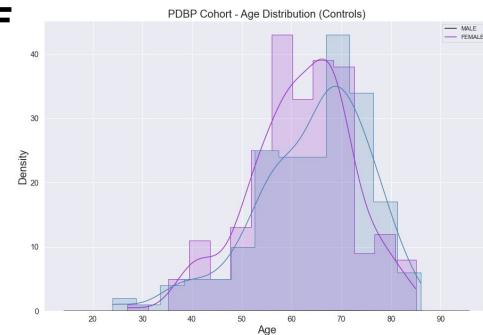


**Figure 26:** Population attributable risk comparison for *GBA1* known coding variants in the EUR population versus the novel *GBA1* intronic variant in the AFR population



**Figure 27:** GenoML Workflow

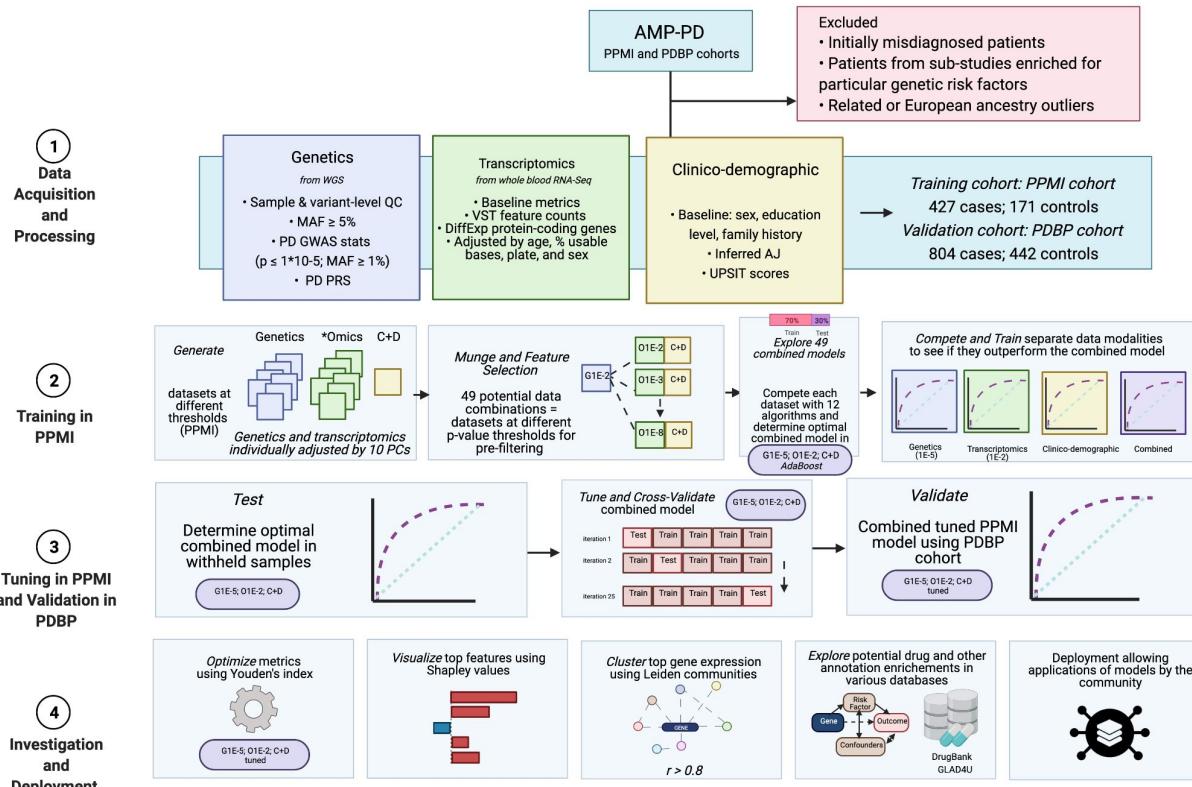
All current functionality in the GenoML open-source Python package. Blue indicates required inputs and features, while purple indicates optional inputs and features. GWAS: Genome-wide association study; AUC: area under the curve; UMAP: Uniform manifold approximation and projection; VIF: variance inflation factor; MLP: multi-layer perceptron networks; XGBoost: an extreme gradient boosting.

**A****B****C****D****E****F**

**Figure 28: PPMI and PDBP Age Distribution**

PDBP: Parkinson's Disease Biomarker Program; PPMI: Parkinson's Progression Markers Initiative

Chapter 5

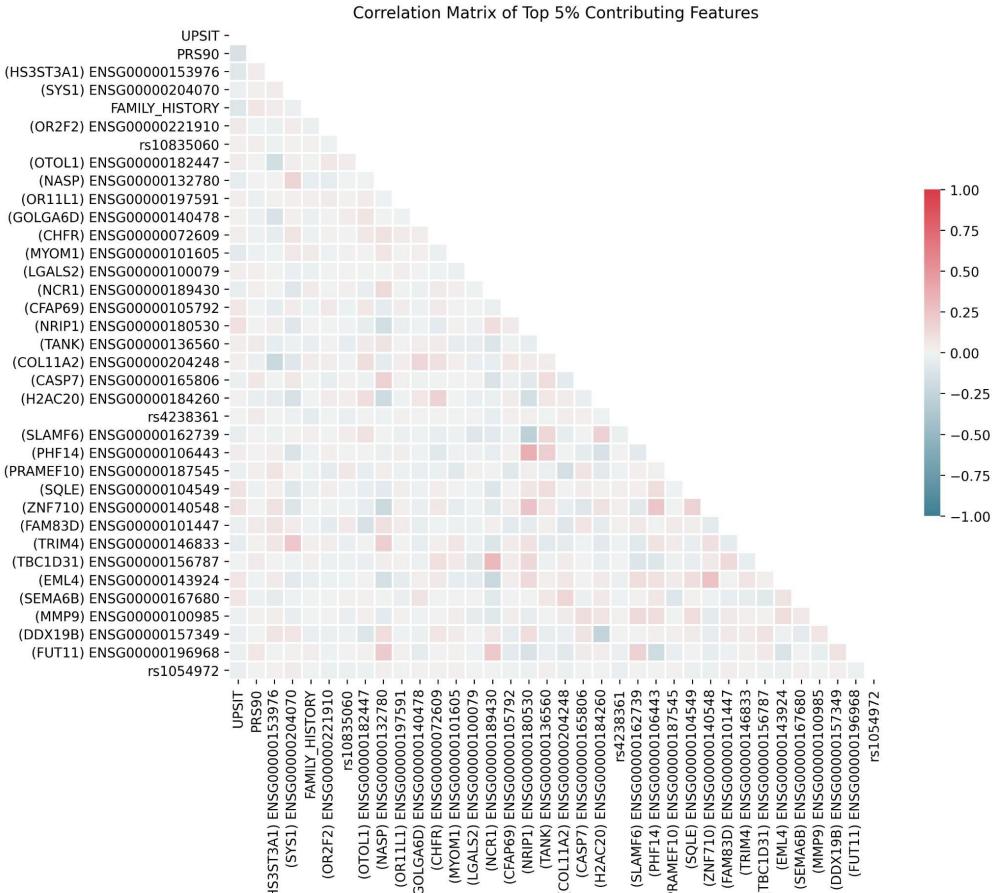


**Figure 29: Workflow and Data Summary for Multi-omics of PD Risk Prediction using ML Study**

Scientific notation in the workflow diagram denotes minimum p-values from reference GWAS or differential expression studies as a pre-screen for feature inclusion. Blue indicates subsets of genetics data (also denoted as "G"), green indicates subsets of transcriptomics data (also denoted as "\*omics or "O"), yellow indicates clinico-demographic data (also denoted as C+D), and purple indicates combined data modalities.

PD: Parkinson's Disease; AMP-PD: Accelerating Medicines Partnership in Parkinson's Disease; PPMI: Parkinson's Progression Marker Initiative; PDBP: Parkinson's Disease Biomarker Program; WGS: whole genome sequencing;

GWAS: Genome-wide association study; QC: quality control; MAF: minor allele frequency; PRS: polygenic risk score

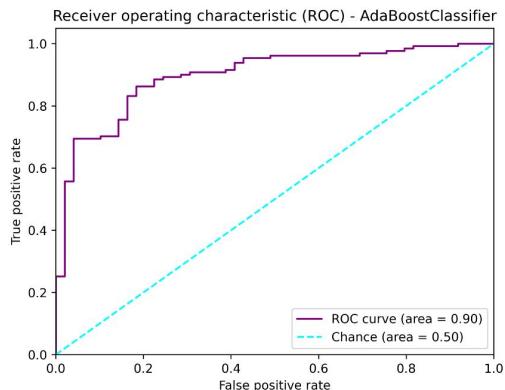


**Figure 30:** Correlation matrix of top 5% contributing features in ML model

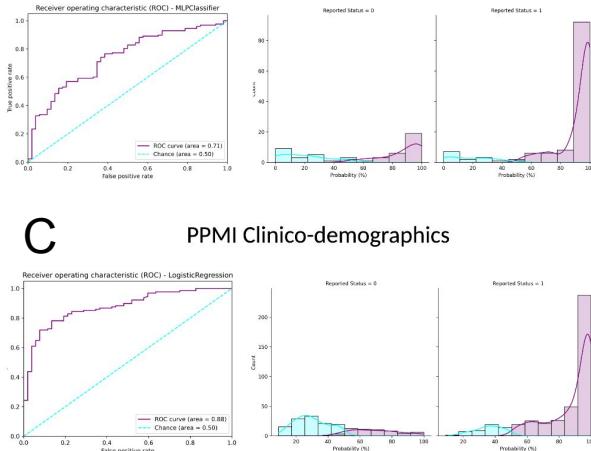
Chapter 5

**A**

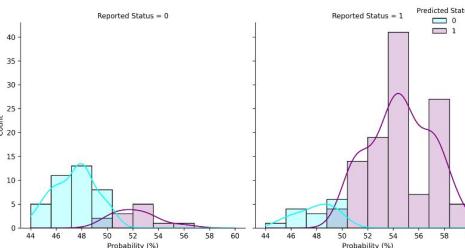
## PPMI - Combined \*Omics

**B**

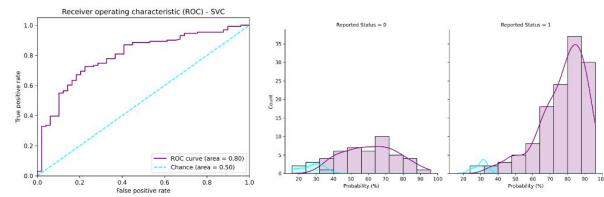
## PPMI Genetics (p-value threshold = 1E-5)

**C**

## PPMI Clinico-demographics

**D**

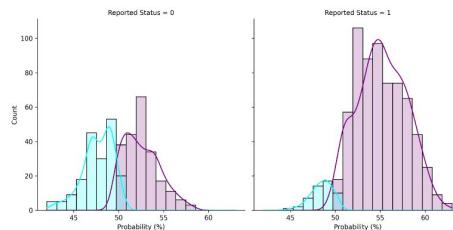
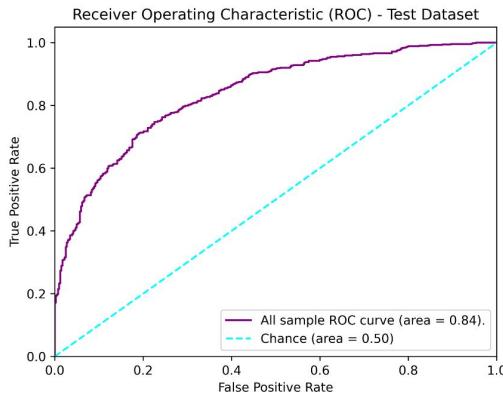
## PPMI Transcriptomics (p-value threshold = 1E-2)



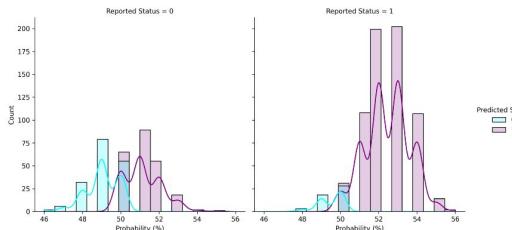
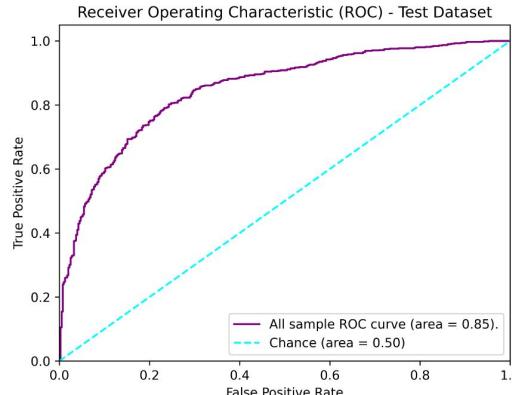
**Figure 31** Receiver operating characteristic curves and case probability density plots in withheld training samples at default thresholds comparing performance metrics in different data modalities from the PPMI dataset.

P values mentioned indicate the threshold of significance used per data type, except for the inclusion of all clinico-demographic features. A) PPMI combined \*omics dataset (genetics p-value threshold = 1E-5, transcriptomics p-value threshold = 1E-2, and clinico-demographic information); B) PPMI genetics only dataset (p-value threshold = 1E-5); C) PPMI clinico-demographics only dataset; D) PPMI transcriptomics only dataset (p-value threshold = 1E-2). Note that x-axis limits may vary as some models produce less extreme probability distributions than others inherently based on fit to the input data and the algorithm used.

**A** Combined \*Omics  
(Tested in PDBP after training the PPMI model)

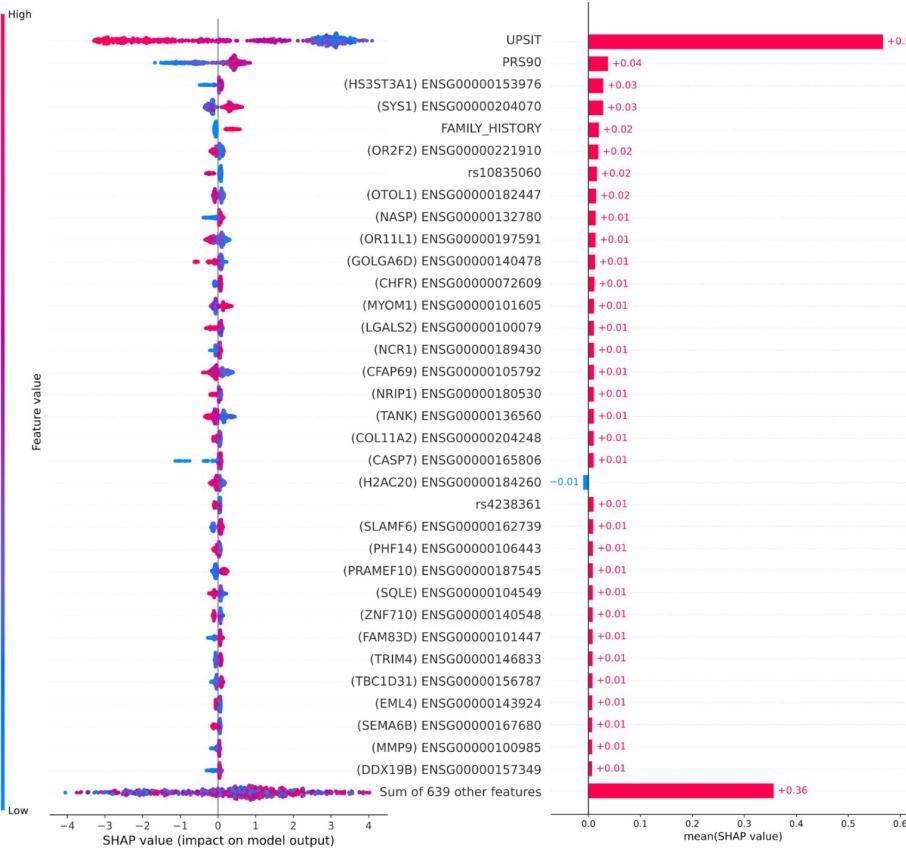


**B** Combined \*Omics  
(Tested in PDBP after tuning the PPMI model)



**Figure 32: Receiver operating characteristic and case probability density plots in the external dataset (PDBP) at validation for the trained and then tuned models at default thresholds.**

Probabilities are predicted case status (r1), so controls (a status of 0) skews towards more samples on the left, and positive PD cases (a status of 1) skews more samples on the right. A) Testing in PDBP the combined \*omics model (genetics p-value threshold = 1E-5, transcriptomics p-value threshold = 1E-2, and clinico-demographic information) developed in PPMI prior to tuning the hyperparameters of the model; B) Testing in PDBP the combined \*omics model (genetics p-value threshold = 1E-5, transcriptomics p-value threshold = 1E-2, and clinico-demographic information) developed in PPMI after tuning the hyperparameters of the model. PPMI: Parkinson's Progression Marker Initiative; PDBP: Parkinson's Disease Biomarker Program; ROC: Receiver Operating Characteristic curve



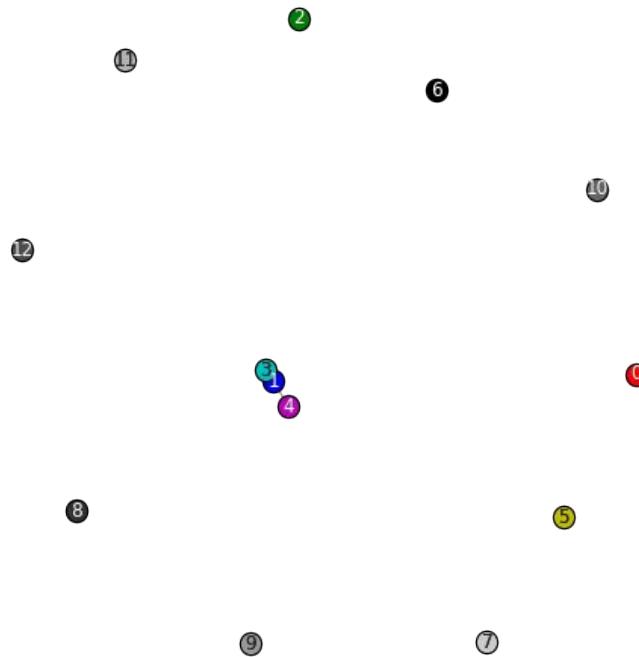
**Figure 33: Feature importance plots for top 5% of features in XGBoost surrogate of best combined \*omics model**

Plot on the left have lower values indicated by the color blue, while higher values are indicated in red compared to the baseline risk estimate. Plot on the right indicates directionality, with features predicting for cases indicated in red, while features better predicting controls are indicated in blue.

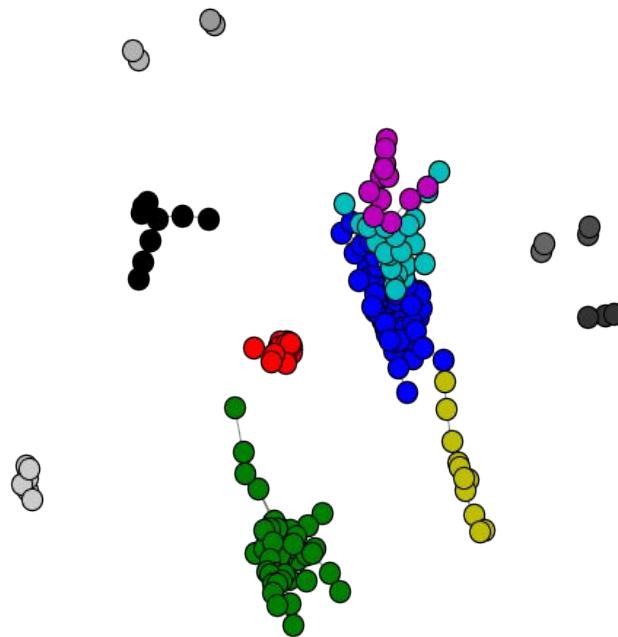
SHAP: Shapley values; UPSiT: University of Pennsylvania Smell Identification Test; PRS: Polygenic risk score

**A**

Macro-level View of Distance Between Communities

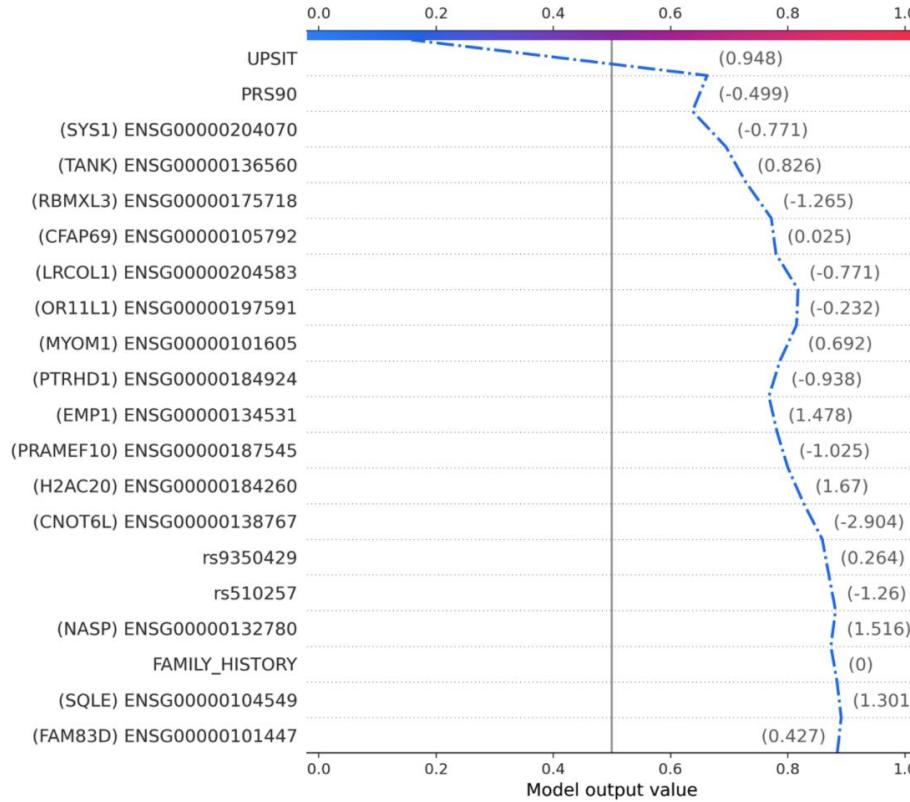
**B**

Micro-level View of Connectivity Within and Between Network Communities

**Figure 34: Network plot of nominated genes following best combined multi-model ML prediction model**

Panel A provides a macro-level view of the distance between communities. Panel B is a micro-level view of connectivity within and between network community modules. The colors of communities in Panel A correspond to those in panel B.

*Chapter 5*



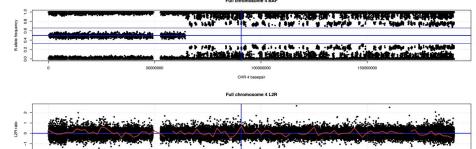
**Figure 35:** Misclassified case as a healthy control using the best multi-modal model  
Chapter 5

# Thesis Supplementary Figures

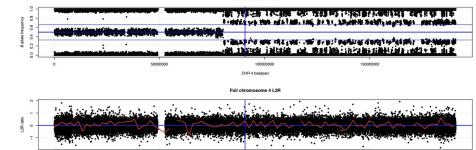
# Supplementary Figures

- **Supplementary Figure 1:** Complex *SNCA* genomic events of interest
- **Supplementary Figure 2:** Ten random “negative control” subjects for the UK biobank exome sequencing replication of *SNCA* alteration carriers
- **Supplementary Figure 3:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17807816:T:G)
- **Supplementary Figure 4:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17808270:G:T)
- **Supplementary Figure 5:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17808270:G:T)
- **Supplementary Figure 6:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17807919:G:T)
- **Supplementary Figure 7:** UK Biobank IGV Plot for Control (*B3GNT3* LoF; chr19:17807982:GC:G)
- **Supplementary Figure 8:** UK Biobank IGV Plot for PD Parent Proxy #1 (*B3GNT3* LoF; chr19:17808033:C:T)
- **Supplementary Figure 9:** UK Biobank IGV Plot for PD Parent Proxy #2 (*B3GNT3* LoF; chr19:17812105:C:CA)
- **Supplementary Figure 10:** *GBAP1* Duplication and MTX1, MTX1P1 Fusion vs Long Read Sequencing

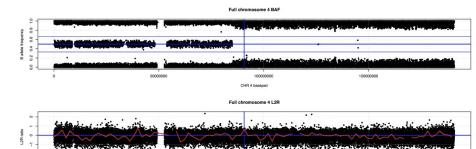
A) COMP1



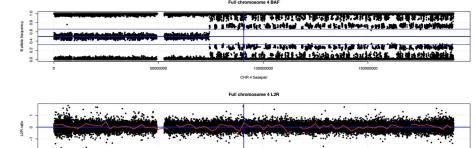
B) COMP2



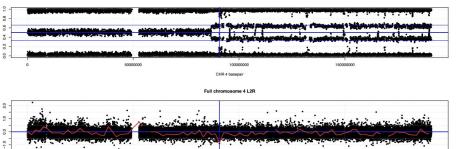
C) COMP3



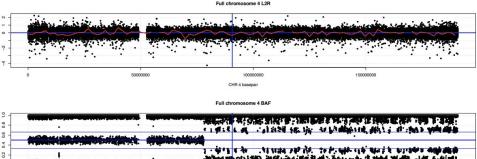
D) COMP4



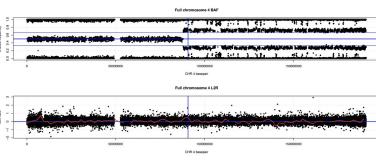
E) COMP5



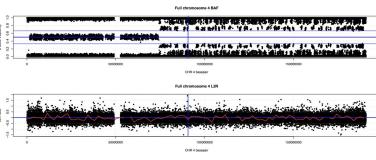
I) COMP9



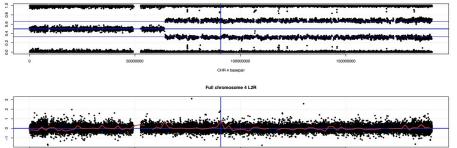
M) COMP13



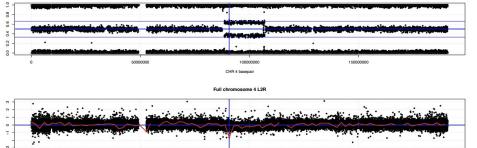
N) COMP14



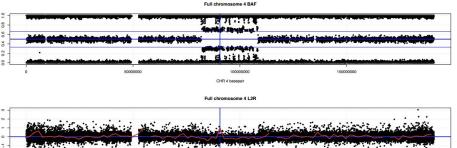
F) COMP6



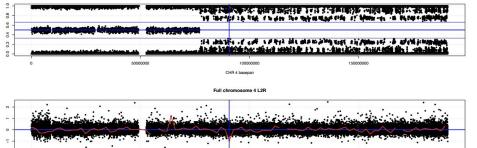
J) COMP10



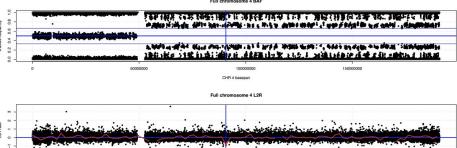
G) COMP7



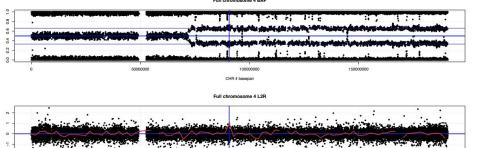
K) COMP11



H) COMP8

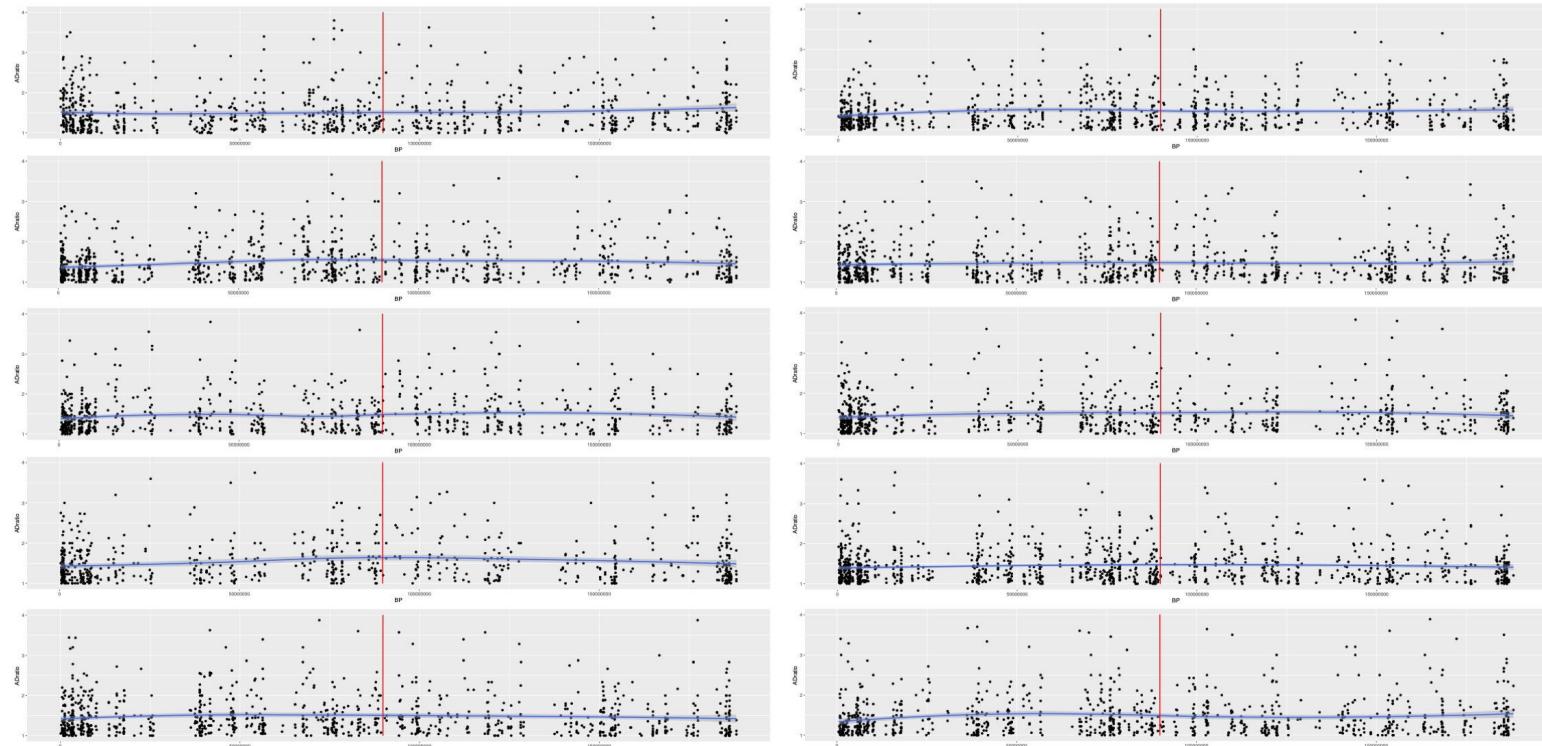


L) COMP12



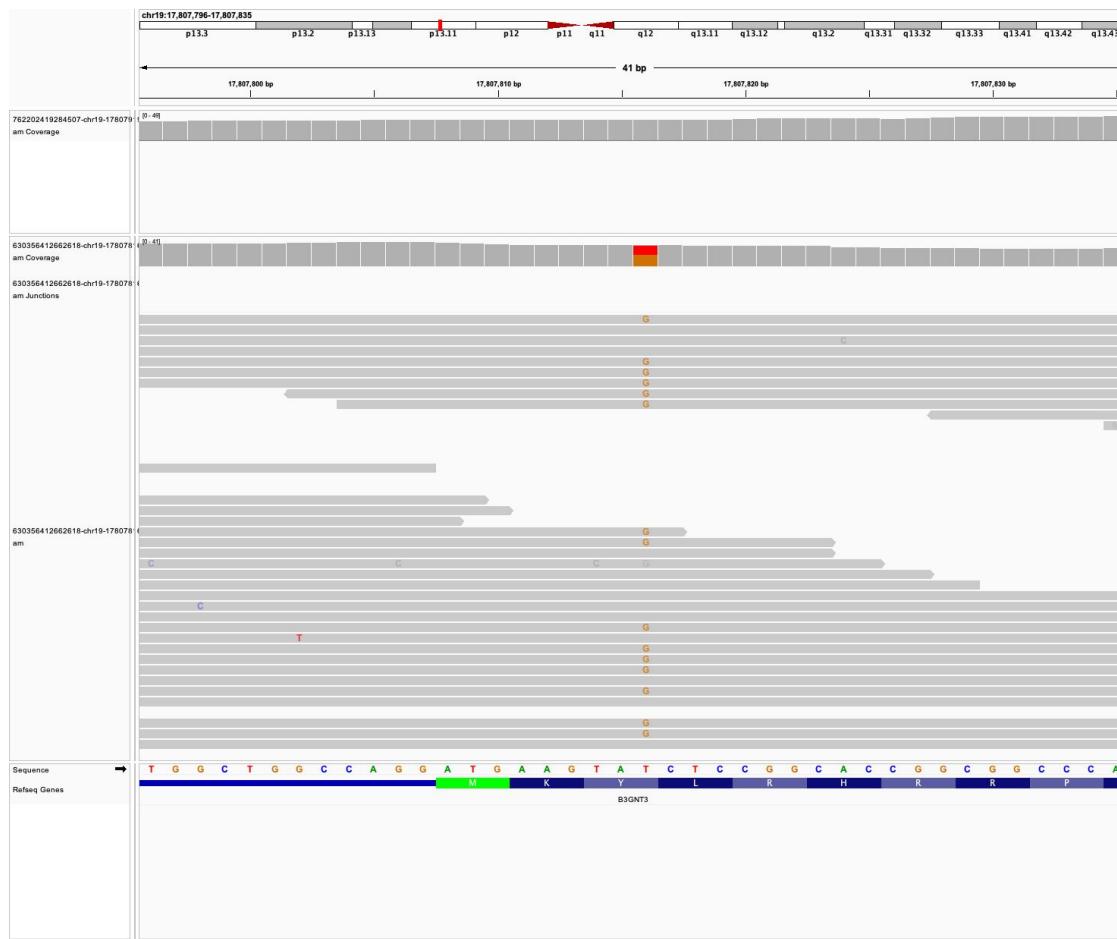
Supplementary Figure 1: Complex SNCA genomic events of interest

Chapter 2



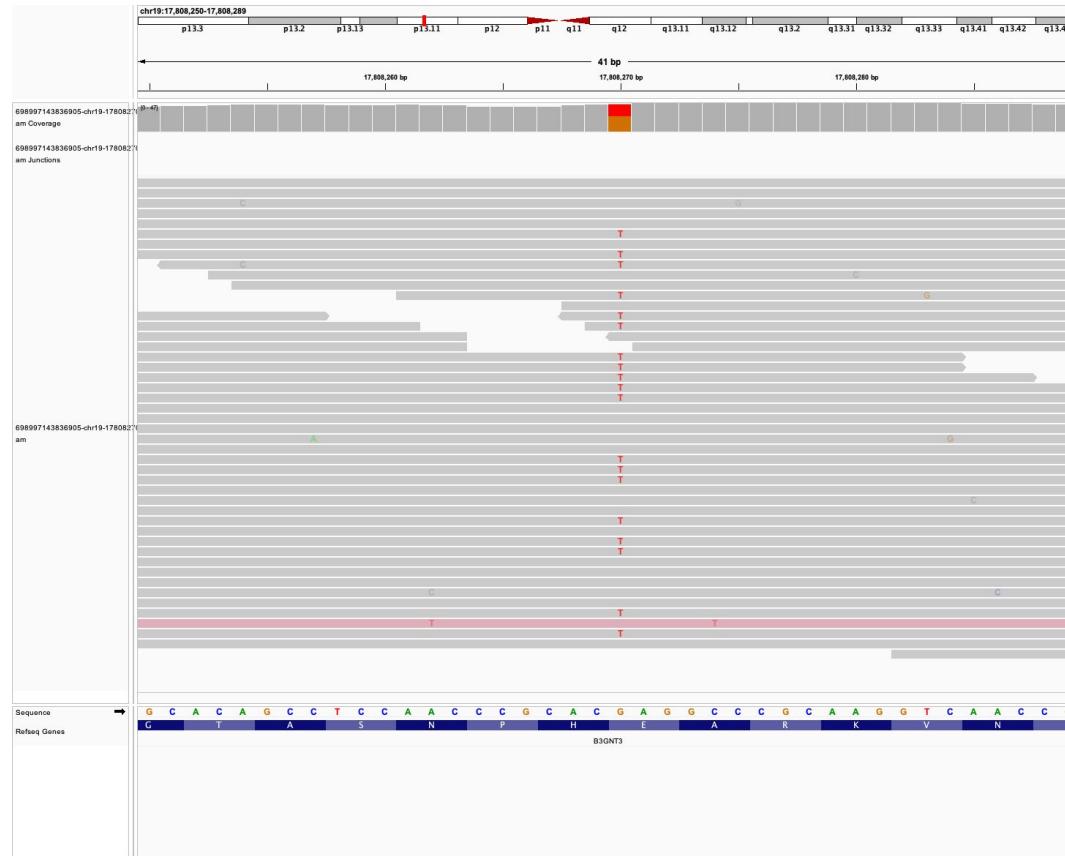
**Supplementary Figure 2:** Ten random “negative control” subjects for the UK biobank exome sequencing replication of *SNCA* alteration carriers. No differences are observed for allele depth ratios. Red line represents the *SNCA* gene body.

Chapter 2



**Supplementary Figure 3:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17807816:T>G)

Chapter 3

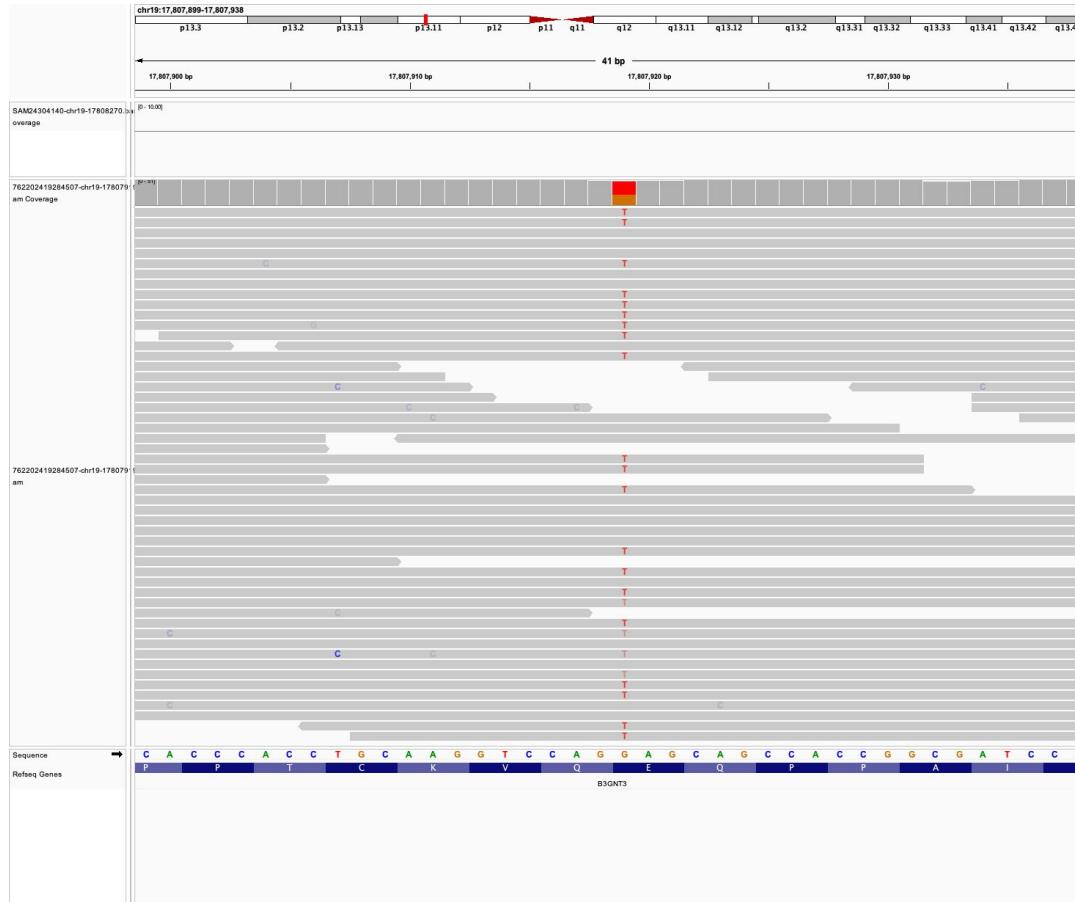


**Supplementary Figure 4:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17808270:G:T)



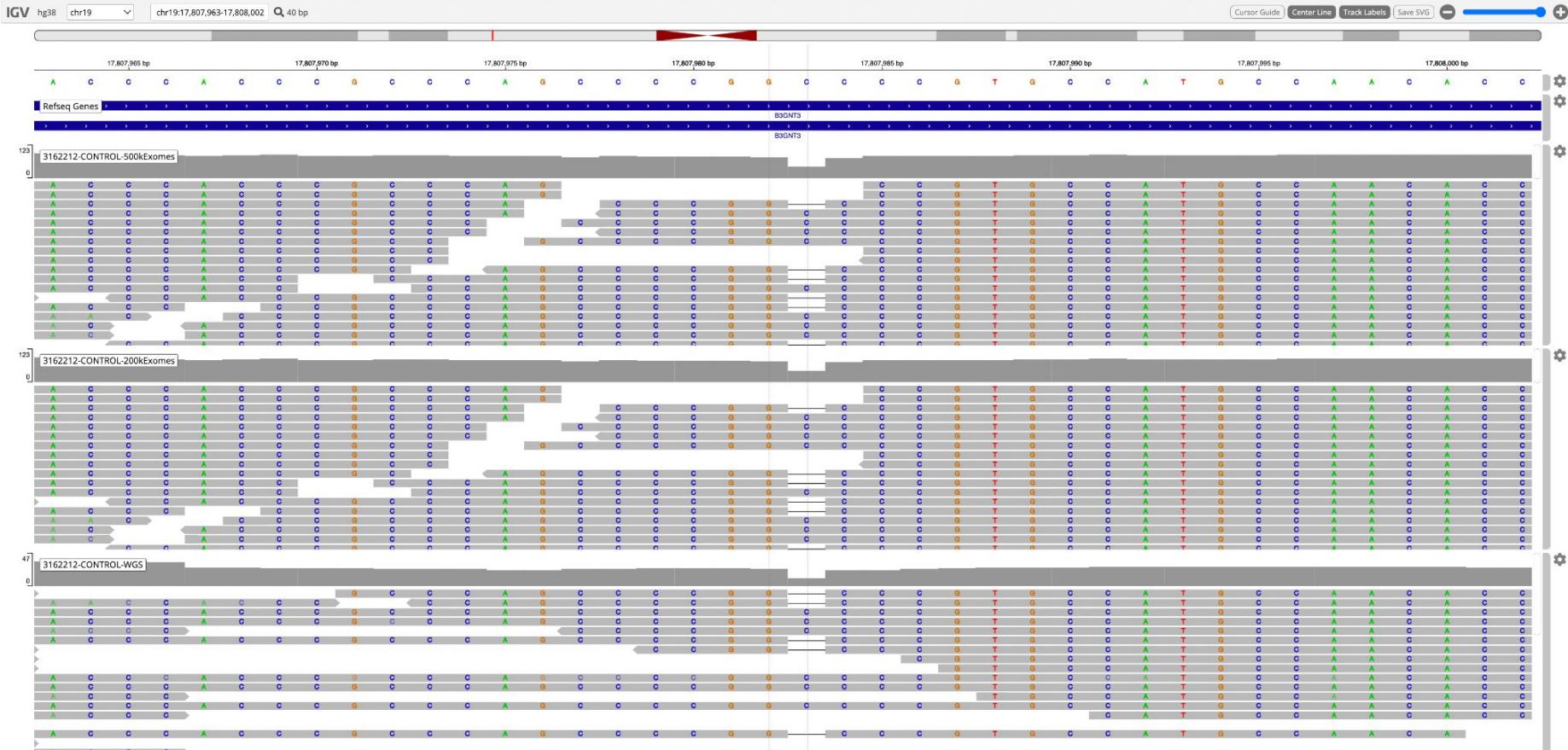
**Supplementary Figure 5:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17808270:G:T)

Chapter 3



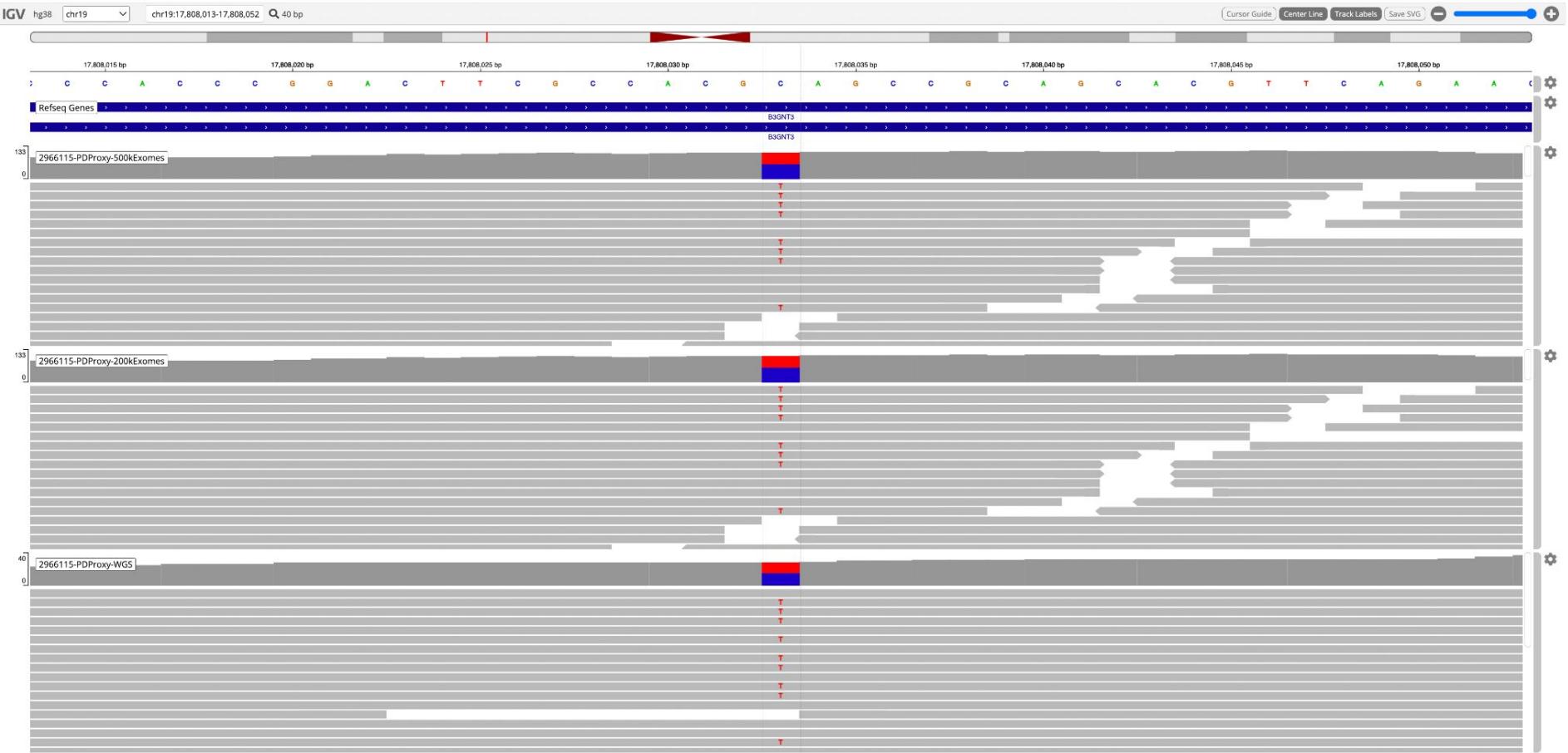
**Supplementary Figure 6:** Genentech IGV Plot (*B3GNT3* LoF; chr19:17807919:G:T)

Chapter 3



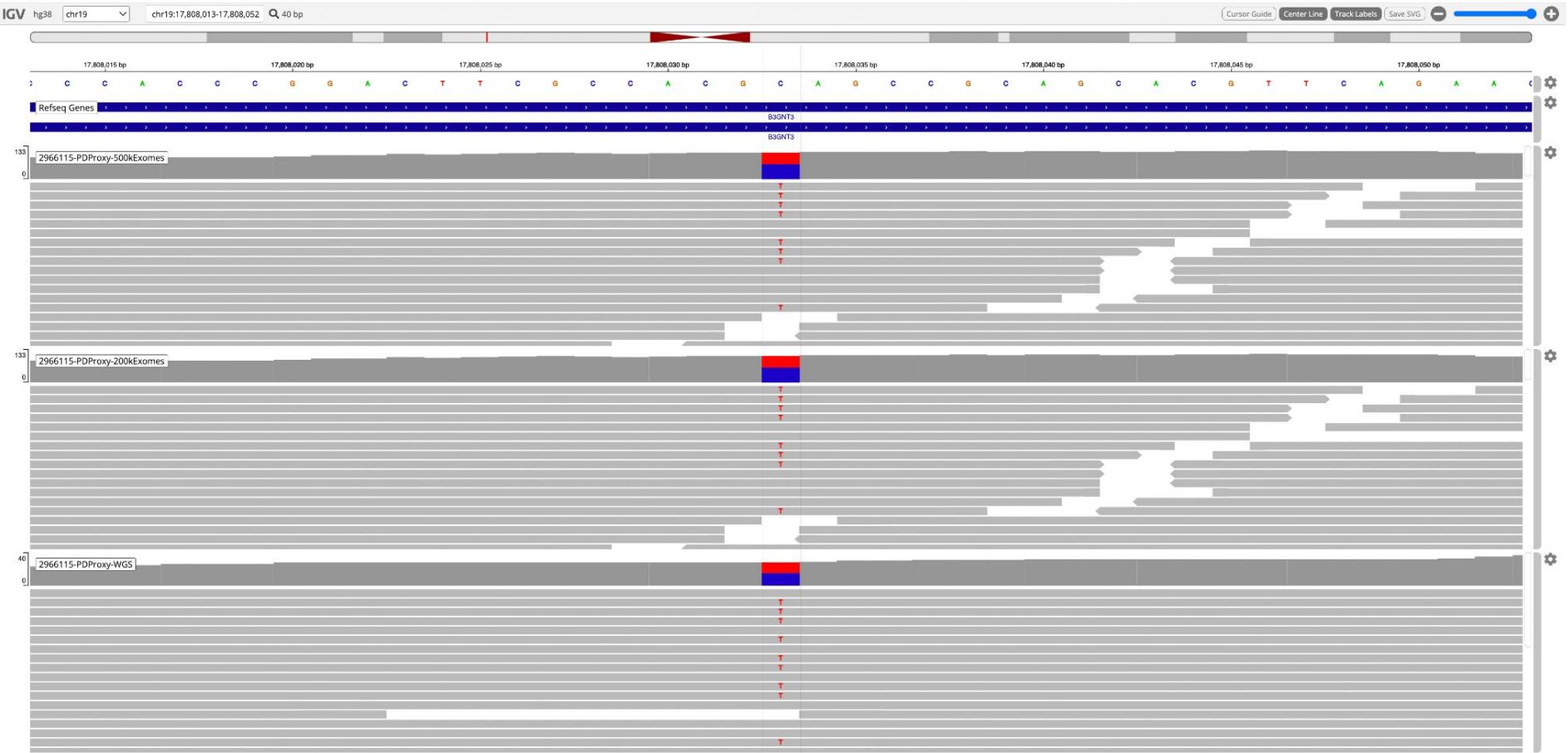
**Supplementary Figure 7:** UK Biobank IGV Plot for Control (*B3GNT3* LoF; chr19:17,807,982:GC>G)

Visualized using IGV via the UK Biobank DNAAnexus Research Analysis Platform. Each of the three tracks pictured above belong to one sample possessing a small deletion, with each track representing an independent data source. Track 1 visualizes the exomes from the latest release, which includes ~500K exomes (as of January 2023). Track 2 visualizes the exomes from the previous release, which includes ~200K exomes. Track 3 visualizes the whole genome sequences (as of January 2023). In summary, this visualization shows a small deletion, GC → G, in a neurologically healthy control where some of the reads are missing the second base pair, 'C'.



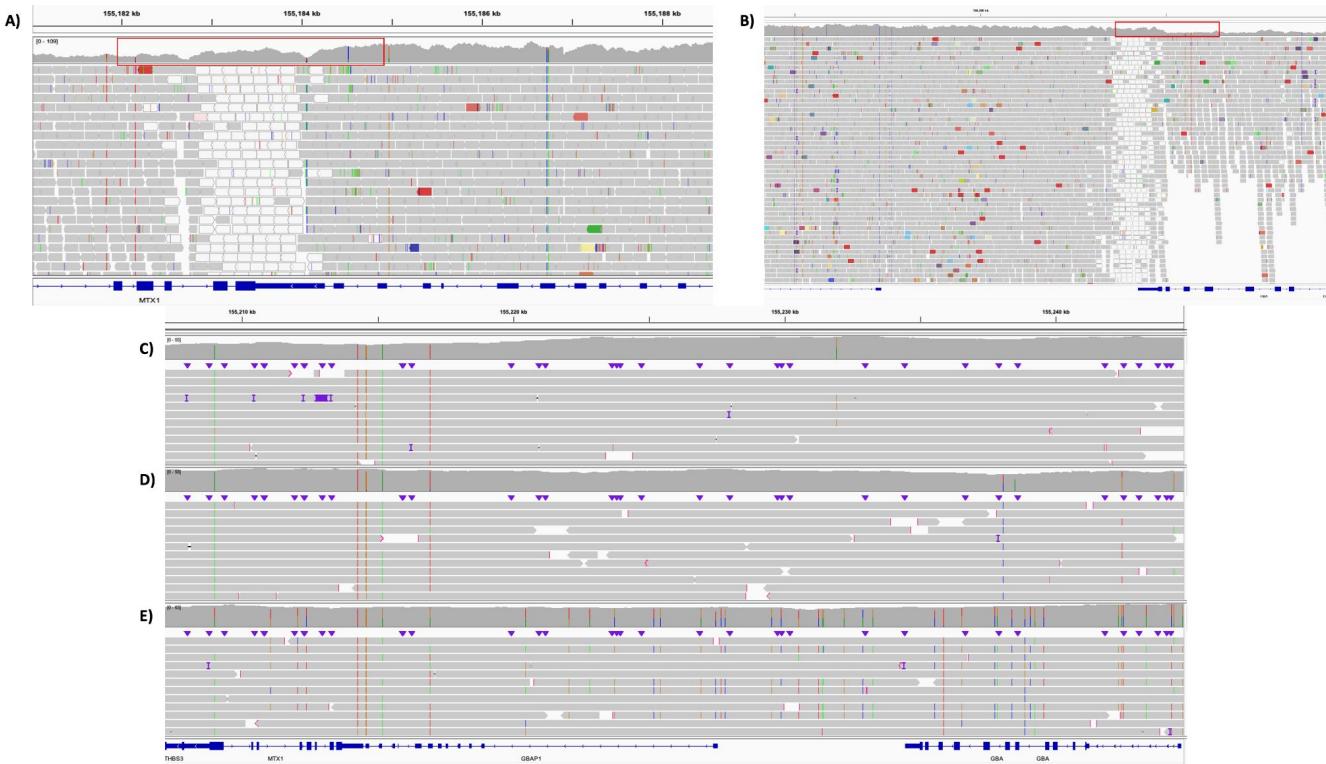
**Supplementary Figure 8:** UK Biobank IGV Plot for PD Parent Proxy #1 (*B3GNT3* LoF; chr19:17808033:C:T)

Visualized using IGV via the UK Biobank DNAnexus Research Analysis Platform. All tracks pictured above belong to one sample. Track 1 visualizes the exomes from the latest release, which includes ~500K exomes (as of January 2023). Track 2 visualizes the exomes from the previous release, which includes ~200K exomes. Track 3 visualizes the whole genome sequences (as of January 2023).



**Supplementary Figure 9:** UK Biobank IGV Plot for PD Parent Proxy #2 (*B3GNT3* LoF; chr19:17812105:C:CA)

Visualized using IGV via the UK Biobank DNAexus Research Analysis Platform. All tracks pictured above belong to one sample. Track 1 visualizes the exomes from the latest release, which includes ~500K exomes (as of January 2023). Track 2 visualizes the exomes from the previous release, which includes ~200K exomes. Track 3 visualizes the whole genome sequences (as of January 2023).



**Supplementary Figure 10: GBAP1 Duplication and MTX1, MTX1P1 Fusion vs Long Read Sequencing**

In 2000, Tayebi and colleagues identified a novel recombination occurring more frequently in African American patients and resulting in a duplication of glucosylceramidase beta pseudogene 1 (*GBAP1*) and fusion of metaxin (*MTX1*) and its pseudogene (*MTX1P1*) (PMID: 11129343, PMID: 11241475). Dr. Ellen Sidransky and Dr. Nahid Tayebi generously provided whole genome sequencing data (A,B) for a sample containing this recombination which was visualized on Broad Institute Integrative Genomics Viewer (2.15.4). We generated long read sequencing data for five rs3115534-GG carriers (5 PD cases), two rs3115534-GT carriers (2 PD cases), and 6 rs3115534-TT carriers (2 PD case; 4 PD controls). A handful of these samples are shown above for reference; C) rs3115534-GG (PD case), D) rs3115534-GG (PD case), E) rs3115534-TT (PD case). The long read sequencing data we generated was compared to the recombination found by Tayebi and colleagues. Coverage changes in A and B are highlighted in red and indicate the presence of a recombination. Our long read sequencing data failed to reveal any structural variants in this region.