

Genome-wide association study of glucocerebrosidase activity modifiers.

Emma N. Somerville^{a,b}, Lynne Krohn^{a,b}, Konstantin Senkevich^{a,c}, Eric Yu^{a,b}, Jamil Ahmad^{a,c}, Farnaz Asayesh^{a,c}, Jennifer A. Ruskey^{a,c}, Dan Spiegelman^{a,c}, Stanley Fahn^d, Cheryl Waters^d, S. Pablo Sardi^e, Roy N. Alcalay^{f,g}, Ziv Gan-Or^{a,b,c}

Affiliations:

^aThe Neuro (Montréal Neurological Institute-Hospital), McGill University, Montréal, QC, Canada

^bDepartment of Human Genetics, McGill University, Montréal, QC, Canada

^cDepartment of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada

^dDepartment of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA

^eRare and Neurological Diseases Therapeutic Area, Sanofi, Cambridge, MA, USA

^fDepartment of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA

^gTaub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA

Corresponding author:

Ziv Gan-Or

Department of Neurology and Neurosurgery

McGill University

1033 Pine Avenue West, Ludmer Pavilion, Room 312, Montréal, QC, Canada, H3A 1A1

Email: ziv.gan-or@mcgill.ca

Telephone: +1-514-398-5845

Keywords: genome wide association study, GWAS, Parkinson's disease, glucocerebrosidase, GAA, GBA1

Abstract

One of the most common genetic risk factors for Parkinson's disease (PD) are variants in *GBA1*, which encodes the lysosomal enzyme glucocerebrosidase (GCase). GCase deficiency has been associated with an increased PD risk, but not all individuals with low GCase activity are carriers of *GBA1* mutations, suggesting other factors may be acting as modifiers. We aimed to discover common variants associated with GCase activity, as well as replicate previously reported associations, by performing a genome-wide association study using two independent cohorts: a Columbia University cohort consisting of 697 PD cases and 347 controls and the Parkinson's Progression Markers Initiative (PPMI) cohort consisting of 357 PD cases and 163 controls. As expected, *GBA1* variants have the strongest association with decreased activity, led by p.N370S (beta = -4.36, se = 0.32, p = 5.05e-43). We also identify a novel association in the *GAA* locus (encoding for acid alpha-glucosidase, beta = -0.96, se = 0.17, p = 5.23e-09) that may be the result of an interaction between GCase and acid alpha-glucosidase based on various interaction analyses. Lastly, we show that several PD-risk loci are potentially associated with GCase activity. Further research will be needed to replicate and validate our findings and to uncover the functional connection between acid alpha-glucosidase and GCase.

Keywords: Genome-wide association study, glucocerebrosidase, Parkinson's disease, *GBA1*, lysosomal metabolism.

1. Introduction

Parkinson's disease (PD) is a complex neurodegenerative disorder, characterized by accumulation of alpha synuclein in Lewy bodies and a progressive loss of dopaminergic neurons in the substantia nigra [1]. Although the exact mechanisms remain unknown, there is a clear role for genetic factors in PD. One gene of particular interest is *GBA1*. In addition to being one of the most common genetic risk factors associated with PD, it is also associated with a faster rate of motor and non-motor progression [2].

GBA1 encodes for glucocerebrosidase (GCase), a lysosomal hydrolase whose biallelic deficiency causes the lysosomal storage disorder (LSD) known as Gaucher disease [3]. Heterozygous *GBA1* variants associated with reduced GCase activity are important risk factors for both PD and dementia with Lewy bodies (DLB), and a substantial portion of patients with these disorders have reduced GCase activity albeit not carrying *GBA1* variants [4, 5]. This observation suggests that other factors, genetic or environmental, may modify GCase activity. One such factor may be *TMEM175* variants, which have been associated with reduced GCase activity in patient data [6] and in cell models [7]. Additionally, variants in *LRRK2* have also been linked to modified GCase activity, although the directionality of the effect is still unclear [4, 8-10].

Understanding the genetic variation influencing GCase activity could be informative in several ways. First, it can provide a better understanding of the mechanisms underlying the role GCase plays in PD. Furthermore, it would allow for proper adjustments or stratification in future research and clinical trials. In the present study, we aimed to identify genetic factors associated with GCase activity by performing a genome-wide association study (GWAS) including a total

of 1054 PD cases and 510 healthy controls. A secondary analysis was conducted to identify PD-associated variants that may act as modifiers of GCase activity.

2. Methods

2.1 Study Population

The study population consisted of two separate cohorts with available genotype and GCase activity data: 1) a cohort of 697 PD cases and 347 controls collected from Columbia University in New York, and 2) a cohort of 357 PD cases and 163 controls from the Parkinson's Progression Markers Initiative (PPMI). Cohort demographics can be found in Table 1. Both cohorts have been previously described [4, 9]. All subjects were of European descent, confirmed with principal component analysis. Informed consent forms were signed by all participants prior to entering their respective studies and the study protocol was approved by the institutional review boards.

2.2 Enzyme Activity

Enzymatic activity for GCase, acid alpha-glucosidase (GAA), acid sphingomyelinase (ASM), alpha-galactosidase A (GLA), and galactosylceramidase (GALC) was measured from dried blood spots in the Columbia cohort, the protocol for which has been previously described [11, 12]. To summarize, enzyme activity was measured by Sanofi laboratories using liquid chromatography-tandem mass spectrometry (LC-MS/MS) from dried blood spots, using a multiplex assay. The dried blood spots were incubated in a reaction cocktail with substrates for the enzymes and buffer. Enzyme activity measurements were determined by calculating the amount of product obtained after the incubation, under the assumption that this is a direct indication of the activity of the enzymes. Lysosomal enzyme activity for the PPMI cohort was measured from frozen whole blood, which was slowly thawed and the processed similarly to the Columbia cohort. GCase outliers were identified as those with activity measurements lying

outside +/- 1.5 times the inter quartile range (IQR), and were subsequently removed in both cohorts.

2.3 Genome-Wide Association Study

Genotyping was performed with the OmniExpress GWAS array for the Columbia cohort and the NeuroX array for the PPMI cohort, according to manufacturer's protocols (Illumina Inc.). Quality control for individual and variant data was completed as previously described (<https://github.com/neurogenetics/GWAS-pipeline>). In brief, samples that were heterozygosity outliers (inclusion criteria of $-0.15 \leq F \leq 0.15$), call rate outliers (missingness $> 95\%$), had mismatched genetic and reported sex, or were identified as European ancestry outliers based on HapMap3 principal component analysis (PCA) in plink v1.9 were removed [13]. Additionally, we removed samples with relatedness closer than 3rd degree relatives ($\text{pihat} > 0.125$). Individual SNPs were excluded on the basis of variant missingness ($> 95\%$), differences in missingness between cases and controls ($p < 1e-04$), haplotype missingness ($p < 1e-04$), and deviation from Hardy-Weinberg equilibrium in controls ($p < 1e-04$). Imputation was performed on filtered data with the Michigan Imputation Server using the Haplotype Reference Consortium reference panel r1.1 2016 and default settings [14]. Linear regressions of GCase activity were performed in plink v1.9 using hard-call variants ($R^2 > 0.8$) and a minor allele frequency (MAF) threshold of > 0.01 . We manually added the PD-relevant variant *GBA1* p.N370S to the Columbia analyses, as well as *GBA1* p.T369M and *LRRK2* p.G2019S after filtering, due to all being under the minor allele threshold. The Columbia cohort included adjustments for age, sex, disease status, Ashkenazi Jewish ancestry, *LRRK2* p.G2019S status, and lysosomal enzyme activities (GAA, GLA, GALC, and ASM) to examine the isolated effect on GCase activity. The p.G2019S mutation was

included as a covariate due it being a known GCase modifier [4, 8, 9]. Lysosomal enzymes were adjusted for because they have been shown to correlate with one another in a previous study [15]. The PPMI analysis included the same adjustments, with the addition of white blood cell count (WBC) as previously suggested [9]. Fixed-effect meta-analyses of PPMI and Columbia cohorts were performed with METAL in plink v1.9. Principal components (PCs) were calculated using PCA with plink v1.9 and the top 10 PCs were included as covariates for each cohort. Conditional and joint analyses were performed with GTCA-COJO to identify independent variants after adjusting for lead SNPs [16]. ANOVA, linear regressions with interaction term, and interaction plots were created in R v4.3.1 [17]. Linkage disequilibrium Manhattan plots were created using LocusZoom [18].

2.4 Data Statement

All code for analyses used in this project can be found at https://github.com/gan-orlab/GCase_GWAS. PPMI data used for this study can be obtained by qualified researchers upon completion of a data access application (<https://www.ppmi-info.org/access-data-specimens/download-data>). Columbia cohort data can be obtained by request. The GCase GWAS summary statistics can be found on the GWAS catalog (<https://www.ebi.ac.uk/gwas/>).

3. Results

3.1 *GAA* and *GBAI* loci are associated with GCase activity.

After quality control and imputation, a total of 556 cases, 284 controls and 13327381 variants in the Columbia cohort and 328 cases, 144 controls and 941882 variants in the PPMI cohort were available for analysis. Mean GCase activity was similar between Columbia (mean = 11.1 umol/l/h, sd = +/- 3.14) and PPMI (mean = 11.63 umol/l/h, sd = +/- 2.59).

We performed a GWAS to identify potential associations between common variants and GCase activity in both cohorts separately, followed by a meta-analysis. We looked at the genomic inflation factors and QQ plots of each analysis to test for systematic bias, and found them to be acceptable (Columbia $\lambda = 0.99$, PPMI $\lambda = 1.04$, Supplementary Figure 1). The strongest associated locus in Columbia was in the *GBAI* locus, as expected (Figure 1A). This signal is driven by p.N370S (rs76763715, beta = -4.21, se = 0.35, p = 4.55e-31), which is in linkage disequilibrium (LD) with the lead SNP in the locus (rs745550122, beta = -4.46, se = 0.35, p = 6.61e-34, $R^2 = 0.78$, $D' = 0.91$). A second signal can be seen in chromosome 17 in the *GAA* locus, with the strongest associated variants comprising a high-LD region (Supplementary Figure 2) and the lead SNP having a negative direction of effect (beta = -0.96, se = 0.17, p = 7.55e-9). Applying conditional and joint analyses revealed no secondary independent associations in either signal. The *GBAI* signal was replicated in the PPMI cohort, with p.N370S once again driving the association (Figure 1B, beta = -5.05, se = 0.76, p = 1.25e-10). The *GAA* locus does surpass nominal significance in PPMI, although the lead SNP has an opposite direction of effect compared to the lead SNP from Columbia (Table 2). The meta-analysis likewise showed associations in the *GBAI* and *GAA* loci, but did not result in any additional associations (Figure 1C). The *GBAI* variants p.T369M (rs75548401) and p.E326K (rs2230288)

were also associated with GCase activity, although not at GWAS significance (Table 3). Both variants had similar effect sizes, directions, and p-values across both analyses.

Interestingly, analysis in the Columbia cohort without adjusting for acid alpha-glucosidase activity eliminates any association from the *GAA* locus (Figure 2A). Adding adjustments for common PD-relevant *GBA1* variants p.N370S, p.E326K and p.T369M results in both the *GBA1* and *GAA* peaks dropping below GWAS-level significance (Figure 2B). To better understand if these outcomes could be due to an interaction between the acid alpha-glucosidase and GCase enzymes, we constructed an interaction plot of enzyme activities by genotype of the top *GAA* SNP that was common in both cohorts (Figure 2B). The genotype of the lead SNP in the *GAA* locus (17:78061141:T:G) appears to impact the interaction of GCase and acid alpha-glucosidase enzyme activities in Columbia, with correlations becoming progressively weaker in an additive manner with the addition of the T allele (Figure 3). This trend is not as clear in controls, which could be due to the lower number of samples in this group. The effect is also not apparent in PPMI, reflective of the lack of association in this cohort. To further investigate the presence of an interaction, both an ANOVA and a linear regression using an interaction term were performed. The ANOVA used the absolute difference between GCase and acid alpha-glucosidase activity, which did significantly vary by the genotype status of the lead *GAA* variant ($df = 2$, Sum Sq = 130, Mean Sq = 64.75, $F = 13.62$, $p = 1.52e-06$). The linear regression was performed using the same covariates as the previous analysis, with the addition of a *GAA**GCase interaction term. The results indicated that the T/T genotype status was significant ($\beta = -0.25$, $se = 0.13$, $p = 0.048$), supporting the presence of an interaction effect between *GAA* genotype status, acid alpha-glucosidase activity, and GCase activity.

To examine if the *GAA* locus has any role in PD, we investigated the association of this region with PD risk and various measurements of progression using GWAS summary statistics from previous studies. We identified 26, 4, 5, 79, and 6 variants that were nominally associated in the GWASs of PD risk, age at onset, UPDRS III, MoCA, and MMSE scores, respectively (Supplementary Table 1), although no variants passed Bonferroni correction for multiple comparisons [19-21]. There were no *GAA* locus variants even nominally associated in the GWASs of motor, cognitive, and composite progression [22]. Of the associated variants, a total of 8, 2, 33, and 1 variants in the GWASs of PD risk, age at onset, MoCA, and MMSE scores were also nominally associated with GCase activity in the present meta-analysis.

3.2 PD-related variants are associated with GCase activity.

We investigated PD-related variants nominated from the largest PD GWAS to date to determine if any were associated with GCase activity in our cohorts (Nalls et al., 2019). There were 17, 7 and 14 loci below nominal significance in the main GWAS for Columbia, PPMI and the meta-analysis, respectively (Table 4). Of these, only p.N370S (rs76763715) and rs35749011 (in strong LD with p.E326K) in the *GBA1* locus, as well as rs13117519 in the *ANK2* locus passed multiple testing correction in the meta-analysis. We then attempted to replicate previously reported associations in *TMEM175* and *LRRK2* with GCase activity, by investigating these loci in a simplified meta-analysis using only age, sex, disease status, 10 PCs, and PD-relevant *GBA1* SNPs including p.N370S, p.E326K, and p.T369M as covariates (Supplementary Figure 3). There were 3 intronic variants in *LRRK2* that passed nominal significance (rs116911375, beta = -1.7, se = 0.79, p = 0.03; rs191242488, beta = -1.041, se = 0.52, p = 0.044; 12:40668909:T:G, beta = -1.041, se = 0.52, p = 0.044), and no variants in *TMEM175*. No variants passed multiple testing correction. The nominally significant variants do not appear to be in LD with p.G2019S or

p.M1646T. Although these variants are not associated with GCase activity in our analysis, p.G2019S and p.M1646T do show a consistent positive direction of effect with what has been demonstrated in previous research (Table 5) [4, 9, 10]. The difference seen in our results compared to the nominal association of p.M1646T found in a very similar analysis in these cohorts performed by Sosero et al. [10] is due to the removal of GCase outliers in our study. With the inclusion of outliers, we also see an association of this variant in our meta-analysis ($b = 1.16$, $se = 0.38$, $p = 0.002$, rs35303786).

4. Discussion

In the present GWAS, we identified variants in the *GAA* locus as potential modifiers of GCase activity, in addition to the known associations with *GBA1* variants. *GAA* encodes acid alpha-glucosidase, and deficiency in this enzyme leads to the LSD Pompe disease [23]. There is no known association between Pompe disease and PD, and the variants in the *GAA* locus are not associated with PD risk, age at onset, or various measures of progression. This is an important observation, since it may suggest that changes in GCase activity alone are insufficient to cause PD, or that the effects of *GAA* variants on GCase activity are too small to have a meaningful clinical effect. It was hypothesized that the mechanism underlying *GBA1*-associated PD is an imbalance, or disturbance in the flux within the lysosomal glycosphingolipid metabolism pathway, rather than reduced GCase activity on its own [24]. This is supported by the identification of other genes in this pathway that are associated with PD, such as *SMPD1*, *GALC*, *ARSA* and *ASAHI* [12, 25-32]. Previous work has also demonstrated that lysosomal enzymes are highly correlated, especially for GCase, acid alpha-glucosidase, and alpha galactosidase A which is encoded by *GLA* [15]. Taken into consideration with the interaction effect observed in the present study, there seems to be mounting evidence in favor of this hypothesis. The mechanism behind the interaction of GCase and acid alpha-glucosidase is still unknown. The products of GCase are lipids and glucose, while acid alpha-glucosidase hydrolyzes glycogen into glucose [33]. Our observation of genotype-dependent correlation between GCase and acid alpha-glucosidase activities may therefore indicate that the *GAA* locus variants associated with GCase activity could maintain equilibrium in this pathway, as they affect both GCase and alpha-glucosidase activities. This hypothesis requires additional genetic, enzymatic, and functional studies.

We further examined the association between known PD-associated loci and GCase activity. As to be expected, we found three PD variants in the *GBA1* locus to be associated with GCase activity. We also found an additional association with the *ANK2* locus. Phosphorylation in the *ANK2* region has been proposed to play a role in PD neurodegeneration through the inhibition of organelle autophagy, but its connection to GCase activity is not clear from the current literature and will need further research [34]. There are also multiple notable associations with ties to GCase activity that may have only failed multiple testing correction due to sample size. One such locus is *GALC*, encoding the enzyme galactosylceramidase which, similarly to GCase, breaks down large sphingolipids into lipids and ceramide. The association of *GALC* variants with GCase activity may suggest that a disturbance in lysosomal metabolism could be a causal mechanism in PD, which is also supported by previous research which found galactosylceramide and GCase to be correlated [15]. We also found nominal associations with multiple *SNCA* variants, complimenting previous associations found between levels of alpha synuclein aggregates and GCase activity [35-37]. A third interesting locus that fell just shy of the multiple testing correction threshold was *FYN*. This gene encodes a kinase that has been shown to phosphorylate alpha-synuclein, which is a key step towards its aggregation into Lewy bodies [38]. We did not replicate previously reported associations with *LRRK2* or *TMEM175* variants, which is likely due to additional adjustments in our study and the removal of GCase outliers compared to previous studies. Studies in larger cohorts will be required to confirm the associations of these PD loci.

This study has several limitations. The first limitation is a lack of power. Our discovery and replication analyses utilized the largest cohorts with enzyme activity measurements that we could find at the time of performing the study and allowed us to complete the first GWAS of

GCase activity, but only enabled us to discover common variants with moderate effect sizes or rare variants with large effect sizes. Due to this, there are likely more modulators of GCase activity yet to be uncovered. A second limitation is the use of only individuals of European and Ashkenazi Jewish ancestry, as there was insufficient data to carry out the study in other populations. Next, our study had minor differences in age and notable differences in sex, particularly in the Columbia cohort, between cases and controls. Both age and sex were adjusted for in all analyses to account for these differences. An additional limitation is the method of enzyme measurement used. Lysosomal activity was measured using dried blood spots, which is not guaranteed to capture the true enzyme activity within a functioning lysosomal environment. These results will strengthen if replicated in future studies that use lysosomal-specific methods of enzyme activity, in cells more representative of disease pathogenesis.

In conclusion, we found a novel potential association in the *GAA* locus associated with GCase activity, which may represent an interaction effect between GCase and acid alpha-glucosidase and could be indicative of a homeostatic relationship between lysosomal enzymes. We also support previously suggested connections of multiple PD-related genes with GCase activity, such as *GALC* and *SNCA*, as well as identifying a novel potential association with *ANK2*. These findings could be significant for improving our understanding of how GCase deficiency is related to overall lysosomal dysfunction, and how GCase functions in relation to other lysosomal enzymes. Additionally, associated loci can be taken into account in future research and clinical trials to help control for genetic influences on GCase activity. Due to the novelty of our results and lack of confident association in our replication cohort, further genetic and functional research will be required to validate our findings.

5. Competing Interest

Z.G.O received consultancy fees from Lysosomal Therapeutics Inc. (LTI), Idorsia, Prevail Therapeutics, Ono Therapeutics, Denali, Handl Therapeutics, Neuron23, Bial Biotech, Bial, UCB, Capsida, Vanqua bio, Congruence Therapeutics, Takeda, Jazz Guidepoint, Lighthouse and Deerfield.

6. Acknowledgements

This study was financially supported through grants from the Michael J. Fox Foundation (MJFF), the Canadian Consortium on Neurodegeneration in Aging (CCNA), and the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives initiative (HBHL). Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/access-data-specimens/download-data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including [list the full names of all of the PPMI funding partners found on the PPMI Website]. We would like to thank the research participants for contributing to this study. We thank Meron Teferra for her assistance. Z.G.O. is supported by the Fonds de recherche du Québec—Santé (FRQS) Chercheurs-boursiers award and is a William Dawson Scholar. KS is supported by a post-doctoral fellowship from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives initiative (HBHL). The Columbia cohort was funded by the National Institutes of Health (K02NS080915, and UL1 TR000040, formerly the National Center for Research Resources, Grant Number UL1 RR024156).

7. Tables

Table 1. Cohort demographics for individuals with available genotype and GCase data.

Cohort	N Cases	N Controls	Status	Males (%)	Females (%)	Mean Age (±SD)
Columbia	697	347	cases	65.1	34.9	65.74 (11.05)
			controls	38.3	61.7	64.25 (10.05)
PPMI	357	163	cases	66.7	33.3	61.85 (9.49)
			controls	67.5	32.5	60.79 (11.34)

N, number; SD, standard deviation; PPMI, Parkinson's Progression Markers Initiative.

Table 2. Comparison of lead *GAA* locus variants in Columbia and PPMI cohorts.

SNP	Cohort	beta	se	p-value
17:78061141:T:G	Columbia	-1.009	0.18	1.15e-08
	PPMI	-0.29	0.19	0.14
17:78096483:A:C	Columbia	0.39	0.162	0.017
	PPMI	0.60	0.18	7.44e-4

SNP, single nucleotide polymorphism; se, standard error; PPMI, Parkinson's Progression Markers Initiative.

Table 3. *GBA1* p.T369M and p.E326K association statistics for Columbia and PPMI cohorts.

Variant	Cohort	beta	se	p-value
p.T369M	Columbia	-3.28	0.78	2.79e-05
	PPMI	-2.91	0.68	2.11e-05
	Meta-analysis	-3.067	0.51	1.787e-09
p.E326K	Columbia	-1.62	0.59	0.0063
	PPMI	-1.43	0.5	0.0047
	Meta-analysis	-1.51	0.38	8.027e-05

se, standard error; PPMI, Parkinson's Progression Markers Initiative.

Table 4. Parkinson's disease risk SNPs associated with changes in GCase activity.

Cohort	SNP	rsID	Beta	SE	MAF	P	Locus
Columbia	1:155205634:T:C	rs76763715	-4.21	0.35	0.041	4.55E-31	<i>GBA1</i>
	20:6006041:C:T	rs77351827	0.63	0.19	0.16	0.00083	<i>CRSL1</i>
	6:112243291:A:G	rs997368	-0.49	0.16	0.24	0.0017	<i>FYN</i>
	4:114369065:C:T	rs13117519	0.53	0.19	0.16	0.006	<i>ANK2</i>
	4:90666041:C:T	rs356203	0.37	0.14	0.41	0.008	<i>SNCA</i>
	4:90626111:G:A	rs356182	0.36	0.14	0.37	0.011	<i>SNCA</i>
	1:155135036:G:A	rs114138760	-1.61	0.64	0.012	0.012	<i>GBA1</i>
	21:38852361:G:A	rs2248244	0.39	0.16	0.26	0.012	<i>DYRK1A</i>
	4:90636630:G:A	rs5019538	0.33	0.14	0.33	0.022	<i>SNCA</i>
	17:43933307:T:C	rs7221167	-0.3	0.14	0.4	0.029	<i>MAPT</i>
	14:37989270:T:C	rs12147950	-0.3	0.14	0.43	0.032	<i>MIPOL1</i>
	4:90607126:C:G	rs356228	0.28	0.13	0.46	0.034	<i>SNCA</i>
	18:31304318:G:T	rs1941685	-0.29	0.14	0.47	0.037	<i>ASXL3</i>
	3:28705690:T:C	rs6808178	0.3	0.14	0.36	0.039	<i>LINC00693</i>
	13:97865021:T:C	rs4771268	-0.32	0.16	0.28	0.044	<i>MBNL2</i>
	14:95194760:G:C	rs4905237	0.38	0.19	0.15	0.047	<i>SERPINA13P</i>
PPMI	1:155205634:T:C	rs76763715	-5.048	0.76	0.0087	1.25E-10	<i>GBA1</i>
	1:155135036:G:A	rs35749011	-1.43	0.5	0.02	0.0047	<i>GBA1</i>
	1:154898185:G:C	rs114138760	-1.61	0.62	0.014	0.0097	<i>GBA1</i>
	4:114369065:C:T	rs13117519	0.48	0.19	0.19	0.01	<i>ANK2</i>
	1:171719769:C:T	rs11578699	0.47	0.19	0.19	0.012	<i>VAMP4</i>
	14:88464264:G:T	rs979812	-0.48	0.2	0.45	0.018	<i>GALC</i>
	6:32578772:C:A	rs504594	-0.46	0.22	0.14	0.035	<i>HLA-DRB1</i>
Meta	1:155205634:T:C	rs76763715	-4.36	0.32	0.036	5.05E-43	<i>GBA1</i>
	1:155135036:G:A	rs35749011	-1.50	0.39	0.017	0.00014	<i>GBA1</i>
	4:114369065:C:T	rs13117519	0.50	0.13	0.17	0.00016	<i>ANK2</i>
	20:6006041:C:T	rs77351827	0.45	0.14	0.15	0.0013	<i>CRSL1</i>
	4:90666041:C:T	rs356203	0.31	0.1	0.41	0.0028	<i>SNCA</i>
	4:90636630:G:A	rs5019538	0.3	0.11	0.33	0.0041	<i>SNCA</i>
	6:112243291:A:G	rs997368	-0.34	0.12	0.22	0.0053	<i>FYN</i>
	14:88464264:G:T	rs979812	-0.34	0.13	0.44	0.0068	<i>GALC</i>
	4:90626111:G:A	rs356182	0.2784	0.1037	0.3811	0.007286	<i>SNCA</i>
	1:154898185:G:C	rs114138760	-1.606	0.618	0.0138	0.009358	<i>GBA1</i>
	21:38852361:G:A	rs2248244	0.3915	0.1556	0.2634	0.01187	<i>DYRK1A</i>
	18:31304318:G:T	rs1941685	-0.2872	0.1376	0.4749	0.03687	<i>ASXL3</i>
	4:90607126:C:G	rs356228	0.2054	0.1005	0.4607	0.04097	<i>SNCA</i>
	2:135438789:G:A	rs4954162	0.2489	0.1246	0.2193	0.04585	<i>TMEM163</i>
	14:95194760:G:C	rs4905237	0.3794	0.1904	0.1481	0.0463	<i>SERPINA13P</i>

SNP, single nucleotide polymorphism; se, standard error; MAF, minor allele frequency; PPMI, Parkinson's Progression Markers Initiative.

Table 5. *LRRK2* p.G2019S and p.M1646T association statistics for Columbia and PPMI cohorts.

Variant	Cohort	beta	se	p-value
p.G2019S	Columbia	0.64	0.4	0.11
	PPMI	-0.053	1.12	0.96
	Meta-analysis	0.56	0.56	0.14
p.M1646T	Columbia	0.41	0.45	0.36
	PPMI	0.82	0.54	0.13
	Meta-analysis	0.58	0.58	0.094

se, standard error; PPMI, Parkinson's Progression Markers Initiative.

Figures

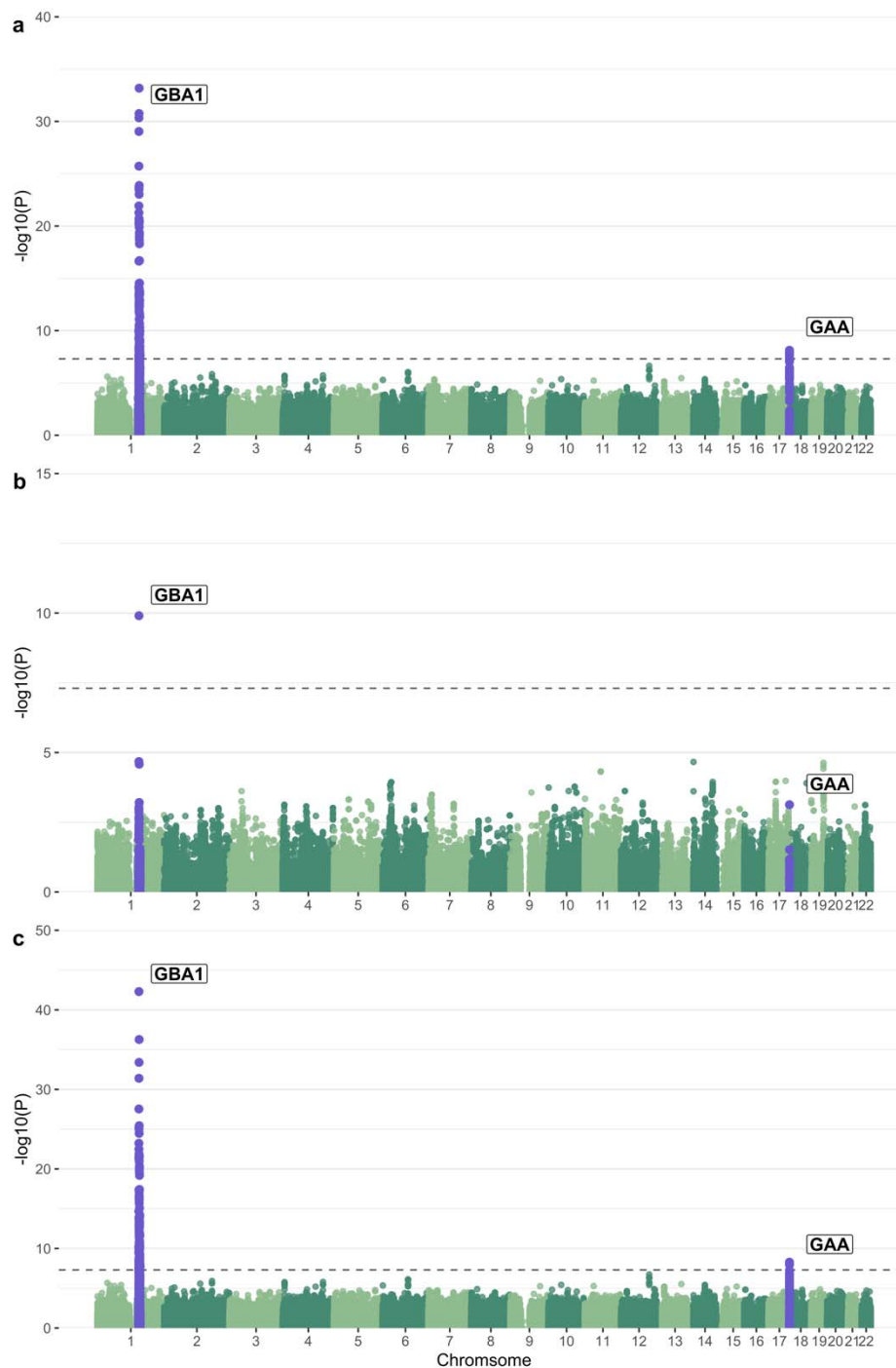


Figure 1. Manhattan plot of log adjusted p-values at each genomic position in a) the Columbia cohort with adjustments for age, sex, disease status, Ashkenazi Jewish status, *LRRK2* p.G2019S, ASM activity, GAA activity, GLA activity, GALC activity, and the top 10 PCs, b) the PPMI cohort with adjustments for age, sex, disease status, *LRRK2* G2019S genotype, ASM activity, GAA activity, GLA activity, GALC activity, white blood cell count, and the top 10 PCs, and c) the meta-analysis of these two analyses.

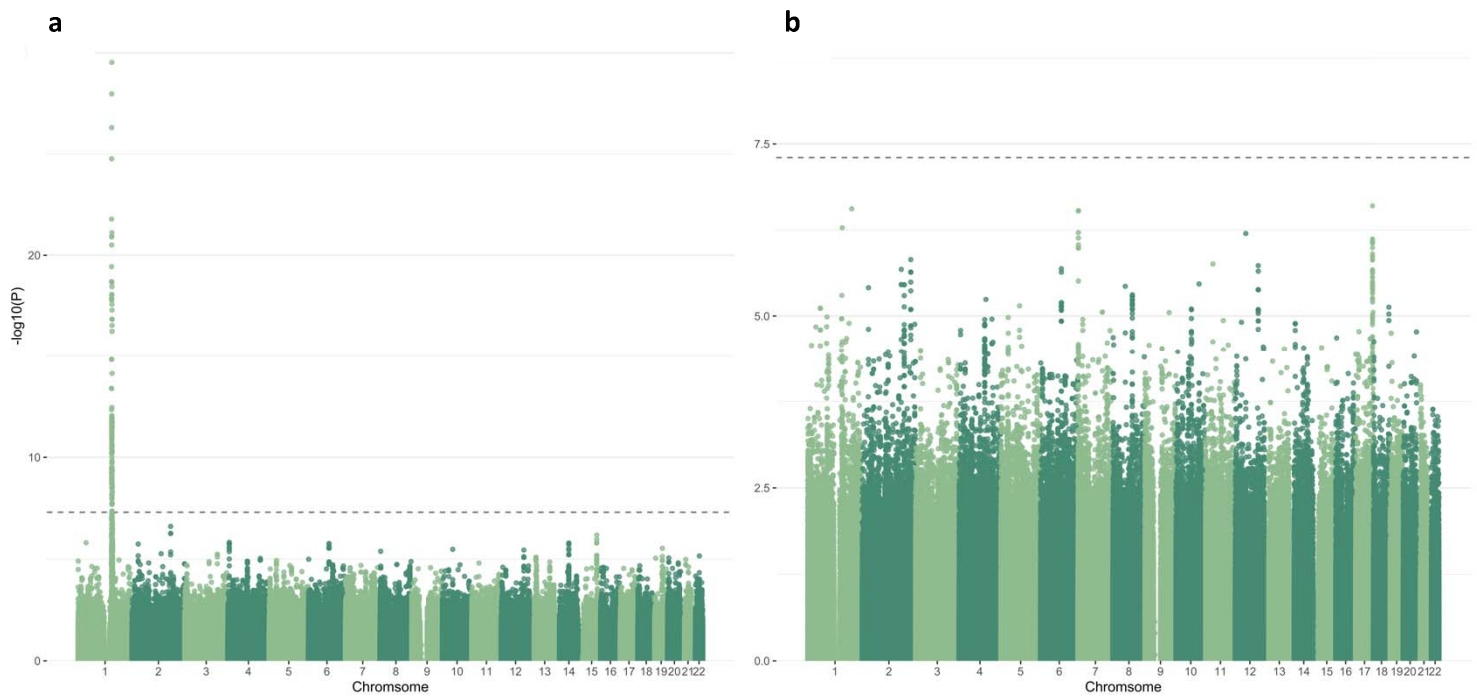


Figure 2. Manhattan plot of log adjusted p-values at each genomic position in the Columbia cohort with adjustments for a) age, sex, disease status, Ashkenazi Jewish status, *LRRK2* p.G2019S, ASM activity, GLA activity, GALC activity, and the top 10 PCs, and b) age, sex, disease status, Ashkenazi Jewish status, *LRRK2* p.G2019S, *GBA1* p.N370S, *GBA1* p.T369M, *GBA1* p.E326K, ASM activity, GAA activity, GLA activity, GALC activity, and the top 10 PCs.

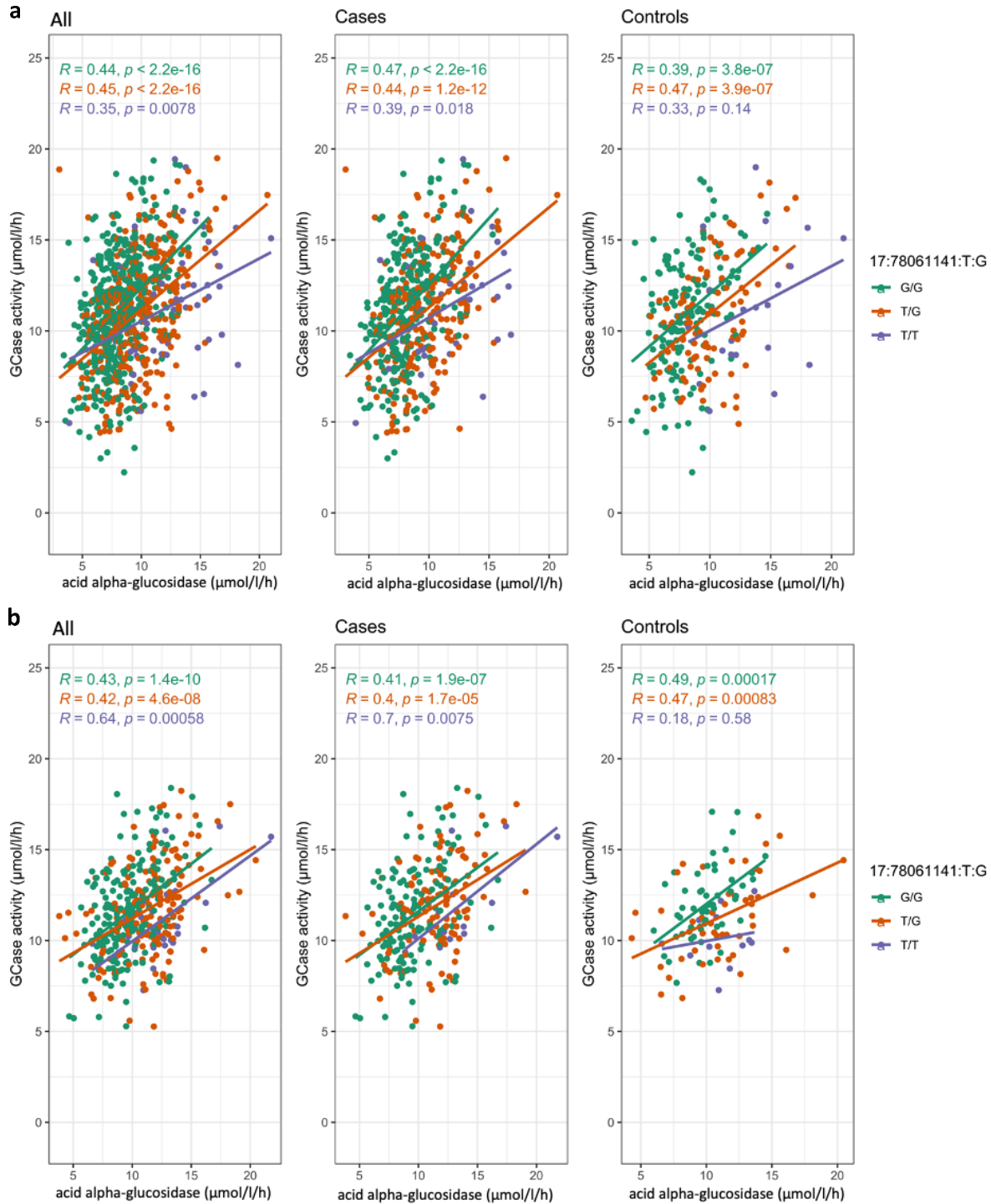


Figure 3. Interaction plots of GCase and acid alpha-glucosidase activity colored by 17:78061141:T:G genotype in a) Columbia and b) PPMI

References

- [1] Armstrong, M.J. and M.S. Okun, *Diagnosis and Treatment of Parkinson Disease: A Review*. JAMA, 2020. **323**(6): p. 548-560.
- [2] Gan-Or, Z., C. Liong, and R.N. Alcalay, *GBA-Associated Parkinson's Disease and Other Synucleinopathies*. Curr Neurol Neurosci Rep, 2018. **18**(8): p. 44.
- [3] Hruska, K.S., M.E. LaMarca, C.R. Scott, and E. Sidransky, *Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA)*. Hum Mutat, 2008. **29**(5): p. 567-83.
- [4] Alcalay, R.N., et al., *Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations*. Brain, 2015. **138**(Pt 9): p. 2648-58.
- [5] Moors, T.E., et al., *Characterization of Brain Lysosomal Activities in GBA-Related and Sporadic Parkinson's Disease and Dementia with Lewy Bodies*. Mol Neurobiol, 2019. **56**(2): p. 1344-1355.
- [6] Krohn, L., et al., *Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies*. Ann Neurol, 2020. **87**(1): p. 139-153.
- [7] Jinn, S., et al., *TMEM175 deficiency impairs lysosomal and mitochondrial function and increases alpha-synuclein aggregation*. Proc Natl Acad Sci U S A, 2017. **114**(9): p. 2389-2394.
- [8] Ysselstein, D., et al., *LRRK2 kinase activity regulates lysosomal glucocerebrosidase in neurons derived from Parkinson's disease patients*. Nat Commun, 2019. **10**(1): p. 5570.
- [9] Alcalay, R.N., et al., *Longitudinal Measurements of Glucocerebrosidase activity in Parkinson's patients*. Ann Clin Transl Neurol, 2020. **7**(10): p. 1816-1830.
- [10] Sosero, Y.L., et al., *LRRK2 p.M1646T is associated with glucocerebrosidase activity and with Parkinson's disease*. Neurobiol Aging, 2021. **103**: p. 142 e1-142 e5.
- [11] Reuser, A.J., et al., *The use of dried blood spot samples in the diagnosis of lysosomal storage disorders--current status and perspectives*. Mol Genet Metab, 2011. **104**(1-2): p. 144-8.
- [12] Senkevich, K., et al., *GALC variants affect galactosylceramidase enzymatic activity and risk of Parkinson's disease*. Brain, 2023. **146**(5): p. 1859-1872.
- [13] Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses*. Am J Hum Genet, 2007. **81**(3): p. 559-75.
- [14] Das, S., et al., *Next-generation genotype imputation service and methods*. Nat Genet, 2016. **48**(10): p. 1284-1287.
- [15] Alcalay, R.N., et al., *Alpha galactosidase A activity in Parkinson's disease*. Neurobiol Dis, 2018. **112**: p. 85-90.
- [16] Yang, J., et al., *Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits*. Nat Genet, 2012. **44**(4): p. 369-75, S1-3.
- [17] R Development Core Team, *R: A language and environment for statistical computing*. 2023, R Foundation for Statistical Computing: Vienna, Austria.
- [18] Pruim, R.J., et al., *LocusZoom: regional visualization of genome-wide association scan results*. Bioinformatics, 2010. **26**(18): p. 2336-7.

- [19] Blauwendraat, C., et al., *Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and alpha-synuclein mechanisms*. *Mov Disord*, 2019. **34**(6): p. 866-875.
- [20] Iwaki, H., et al., *Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts*. *Mov Disord*, 2019. **34**(12): p. 1839-1850.
- [21] Nalls, M.A., et al., *Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies*. *Lancet Neurol*, 2019. **18**(12): p. 1091-1102.
- [22] Tan, M.M.X., et al., *Genome-Wide Association Studies of Cognitive and Motor Progression in Parkinson's Disease*. *Mov Disord*, 2021. **36**(2): p. 424-433.
- [23] Stevens, D., S. Milani-Nejad, and T. Mozaffar, *Pompe Disease: a Clinical, Diagnostic, and Therapeutic Overview*. *Curr Treat Options Neurol*, 2022. **24**(11): p. 573-588.
- [24] Machtel, R., et al., *From Lysosomal Storage Disorders to Parkinson's Disease - Challenges and Opportunities*. *J Mol Biol*, 2023. **435**(12): p. 167932.
- [25] Alcalay, R.N., et al., *SMPD1 mutations, activity, and alpha-synuclein accumulation in Parkinson's disease*. *Mov Disord*, 2019. **34**(4): p. 526-535.
- [26] Gan-Or, Z., et al., *The p.L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease*. *Neurology*, 2013. **80**(17): p. 1606-10.
- [27] Senkevich, K., U. Rudakou, and Z. Gan-Or, *New therapeutic approaches to Parkinson's disease targeting GBA, LRRK2 and Parkin*. *Neuropharmacology*, 2022. **202**: p. 108822.
- [28] Robak, L.A., et al., *Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease*. *Brain*, 2017. **140**(12): p. 3191-3203.
- [29] Makarious, M.B., et al., *Large-scale rare variant burden testing in Parkinson's disease*. *Brain*, 2023.
- [30] Pan, H.X., et al., *Evaluating the role of ARSA in Chinese patients with Parkinson's disease*. *Neurobiol Aging*, 2022. **109**: p. 269-272.
- [31] Deng, S., et al., *Systematic Genetic Analysis of the SMPD1 Gene in Chinese Patients with Parkinson's Disease*. *Mol Neurobiol*, 2016. **53**(7): p. 5025-9.
- [32] Mao, C.Y., et al., *SMPD1 variants in Chinese Han patients with sporadic Parkinson's disease*. *Parkinsonism Relat Disord*, 2017. **34**: p. 59-61.
- [33] Roig-Zamboni, V., et al., *Structure of human lysosomal acid alpha-glucosidase-a guide for the treatment of Pompe disease*. *Nat Commun*, 2017. **8**(1): p. 1111.
- [34] Arango, D., et al., *Understanding the Potential of Genome Editing in Parkinson's Disease*. *Int J Mol Sci*, 2021. **22**(17).
- [35] Tayebi, N., et al., *Glucocerebrosidase haploinsufficiency in A53T alpha-synuclein mice impacts disease onset and course*. *Mol Genet Metab*, 2017. **122**(4): p. 198-208.
- [36] Munoz, S.S., et al., *The interplay between Glucocerebrosidase, alpha-synuclein and lipids in human models of Parkinson's disease*. *Biophys Chem*, 2021. **273**: p. 106534.
- [37] Fishbein, I., Y.M. Kuo, B.I. Giasson, and R.L. Nussbaum, *Augmentation of phenotype in a transgenic Parkinson mouse heterozygous for a Gaucher mutation*. *Brain*, 2014. **137**(Pt 12): p. 3235-47.
- [38] Nakamura, T., H. Yamashita, T. Takahashi, and S. Nakamura, *Activated Fyn phosphorylates alpha-synuclein at tyrosine residue 125*. *Biochem Biophys Res Commun*, 2001. **280**(4): p. 1085-92.

