Coding variation in *GBA* explains the majority of the SYT11-GBA Parkinson's disease GWAS locus

Cornelis Blauwendraat PhD[1], Jose M. Bras PhD[2], Mike A. Nalls PhD[2,3], Patrick A. Lewis PhD[4,5], Dena G. Hernandez PhD[1], Andrew B. Singleton PhD[1,ca] for the International Parkinson's Disease Genomics Consortium

[1]Laboratory of Neurogenetics, National Institute on Aging, 35 Lincoln Drive, Bethesda, MD 20892, USA
[2]UK Dementia Research Institute at UCL and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square House, Queen Square, London WC1N 3BG, UK
[3]Founder/consultant with Data Tecnica International, Glen Echo, MD, 20812 USA
[4]School of Pharmacy, University of Reading, Reading, SL6 7DS, UK
[5]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square House, Queen Square, London WC1N 3BG, UK

Andrew B. Singleton PhD
Laboratory of Neurogenetics, NIA IRP NIH
35 Convent Drive, Bethesda, MD 20892, USA
[ca]singleta@mail.nih.gov

Running title: GBA risk for PD
Keywords: Genome wide association; Parkinson disease; GBA; risk allele; glucocerebrosidase

Letter format

A landmark discovery in the genetics of for Parkinson's disease (PD) was the determination that mutations in the gene encoding glucocerebrosidase (*GBA*) convey risk for PD and dementia with Lewy bodies[1, 2]. Mutations in *GBA* are also causing the autosomal recessive disorder Gaucher's disease (GD). GD is a lysosomal storage disorder which present with a variety of features, including parkinsonism. While it has been clear that some GD-causing *GBA* mutations such as p.N409S are risk factors for PD, there has been some controversy regarding *GBA* mutations that do not cause GD and the risk for PD. Resolving this issue and the effect of such mutations on e.g. glucocerebrosidase activity, will have important implications for our understanding of the pathobiology of PD.

We present here a large genetic assessment of *GBA* using NeuroX genotyping of 6,249 PD cases and 6,032 controls[3]. In total, we detected 11 *GBA* coding variants. No large deletions or duplications at the *GBA* locus were identified. As expected p.N409S and p.E365K were enriched in cases compared to controls (Table1). Unfortunately, another relatively common disease associated L483P could not be genotyped[1]. Interestingly, variant p.T408M, reported as uncertain pathological significance, was also significantly overrepresented in cases. In total the majority of identified *GBA* variants were more common in cases compared to controls, the exception being

p.K13R which is also reported as benign (Table1). Although some individuals were carrying multiple *GBA* mutations no high LD patterns were identified between coding variants (Supplementary Results).

In previous PD genome-wide association studies (GWAS), a locus approximately 150kb from *GBA* was implicated as a risk factor for PD (nominated as *SYT11-GBA*)[4]. Analysis of the non-coding variant rs71628662 previously implicated revealed a strong association for disease (Table1). Previous analyses suggested that this was independent of known GD causing mutations and the mechanism of action at this locus was therefore unclear. Conditioned analysis of rs71628662 on each of the identified coding variants revealed that the association at p.E365K completely accounted for the previously identified GWAS signal (conditional p.E365K rs71628662, p=0.6234). Within the current dataset the linkage disequilibrium between p.E365K and rs7168662 is very high (D'=0.987, r^2=0.84).

Previously, another non-coding variant (rs114138760) was also identified as independent signal [5]. In the NeuroX dataset, this variant also showed association with disease (p=0.0001853) and remained significant following adjustment for the primary GWAS variant (p=0.0001511). Additionally, this signal also remained significant following adjustment for p.E365K, p.N409S, and p.T408M (p>0.007146) and of the 15 rs114138760 risk variant carriers, none carry p.L483P

In summary, this work affirms that functional alleles within *GBA* (p.E365K and p.T408M) that are insufficient to cause GD variants are robustly associated with PD. Further we show that a previously reported GWAS signal at this locus can be largely explained by protein coding variability within *GBA.* These data are consistent with p.E365K being the functional effector allele underlying the original GWA locus (GBA-SYT11 locus), as suggested previously[6]. The results from this current study suggest that efforts directed at replacing or augmenting glucocerebrosidase activity already underway for PD may have a wider ranging applicability than just for individuals with GD associated mutations in glucocerebrosidase.

					Cases				Controls				ExAC (EUR)			
Amino Acid	**GRCh38 (bp)**	**rs number**	**Clinvar**	**HGMD Pathogenicity**	**M/M**	**M/W**	**W/W**	**MAF**	**M/M**	**M/W**	**W/W**	**MAF**	**M/M**	**MAF**	**OR**	**P-value**
K13R (K-27R)	155240707	rs150466109	Benign	DM?	0	0	6249	0	0	2	6030	0.00017	0	0.00021	NA	0.999
C62Y (C23Y)	155240008	rs145888253	na	na	0	1	6248	0.00008	0	0	6032	0	0	0.00001	NA	1
E150K (E111K)	155239622	none	na	DM?	0	1	6248	0.00008	0	0	6031	0	0	0	NA	0.9993
F255Y (F216Y)	155237576	rs74500255	Likely pathogenic	Pathogenic	0	2	6247	0.00016	0	1	6031	0.00008	0	0	NA	0.5827
F298L (F259L)	155237446	none	na	Pathogenic	0	1	6248	0.00008	0	0	5990	0	0	0	NA	0.9993
A348V (A309V)	155236426	rs78396650	Pathogenic	Pathogenic	0	5	6244	0.00040	0	0	6032	0	0	0	NA	0.9984
E365K (E326K)	155236376	rs2230288	Conflicting interpretations	DM?	4	256	5989	0.02158	0	121	5911	0.01013	7	0.01196	2.037	1.64E-09
T408M (T369M)	155236246	rs75548401	Conflicting interpretations	DM?	2	169	6078	0.01404	0	107	5925	0.00895	5	0.00976	1.53	0.001188
N409S (N370S)	155235843	rs76763715	Likely pathogenic	Pathogenic	0	87	6161	0.00701	0	41	5991	0.00341	1	0.00363	1.882	0.002014
D419N/Y (D380N/Y)	155235814	none	na	Pathogenic	0	2	6247	0.00016	0	0	6032	0	0	0	NA	0.9991
E427K (E388K)	155235790	rs149171124	Uncertain significance	DM?	0	10	6239	0.00080	0	5	6027	0.00041	0	0.00031	NA	0.2157
na	155390201	rs71628662	na	na	1	187	6059	0.01513	0	88	5935	0.007305	na	na	2.07	1.324E-07
na	154925709	rs114138760	na	na	1	166	6082	0.01344	0	107	5925	0.008869	na	na	1.52	0.0001853

Table 1: *GBA* variant frequencies in cases and controls. Indicated are uncorrected p values (Bonferroni correction for 0.05 lies at >0.0045) and odds ratios. Tests were carried out in plink using logistic regression with covariates age, sex, first 20 principal components based on genome wide genotype data, and sample provenance. Na = not applicable, bp = base-pair, W = wildtype, M =

mutant, MAF = minor allele frequency, EUR = European, ExAC = Exome Aggregation Consortium (http://exac.broadinstitute.org/), OR = odds ratio, HGMD = Human Gene Mutation Database.

Financial disclosure and conflict of interest
Drs Nalls, Blauwendraat, Lewis, Hernandez, Bras, and Singleton have nothing to disclose.

Author Roles
Dr Singleton was responsible for supervision of the study and initial drafting of the manuscript
Dr Blauwendraat was responsible for data analysis and revision of the manuscript
Dr Bras was responsible for data analysis and revision of the manuscript
Dr Nalls was responsible for data analysis and revision of the manuscript
Dr Hernandez was responsible for data acquisition and revision of the manuscript
Dr Lewis was responsible for revision of the manuscript

Financial Support
Dr Nalls' participation is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging, NIH, Bethesda, MD, USA. Dr Nalls also consults for Genoom Health, Illumina Inc, The Michael J. Fox Foundation for Parkinson's Research and University of California Healthcare.
Dr Singleton received financial support from the NIH, the Michael J Fox Foundation, and the Department of Defense.
Dr Hernandez received financial support from the NIH, and the Michael J Fox Foundation.
Dr Lewis received support from the Medical Research Council, the NIH, the Biotechnology and Biological Sciences Research Council, Parkinson's UK and the Diamond Light Source
Dr Bras received support from the Alzheimer's Society
Dr Blauwendraat received financial support from the NIH

1. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 2009;361(17):1651-1661.
2. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med 2004;351(19):1972-1977.
3. Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol Aging 2015;36(3):1605.e1607-1612.
4. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet 2014;46(9):989-993.
5. Mazzulli JR, Zunke F, Tsunemi T, et al. Activation of β-Glucocerebrosidase Reduces Pathological α-Synuclein and Restores Lysosomal Function in Parkinson's Patient Midbrain Neurons. J Neurosci 2016;36(29):7693-7706.
6. Berge-Seidl V, Pihlstrøm L, Maple-Grødem J, et al. The GBA variant E326K is associated with Parkinson's disease and explains a genome-wide association signal. Neurosci Lett 2017;658:48-52.