The *GBA* variant E326K is associated with Parkinson's disease and explains a genome-wide association signal

Victoria Berge-Seidl [a,b], Lasse Pihlstrøm [b], Jodi Maple-Grødem [c,d], Lars Forsgren [e], Jan Linder [e], Jan Petter Larsen [f], Ole-Bjørn Tysnes [g], Mathias Toft [a*]

[a] Department of Neurology, Oslo University Hospital, Oslo, Norway

[b] Faculty of Medicine, University of Oslo, Oslo, Norway

[c] The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway

[d] The Centre for Organelle Research, University of Stavanger, Stavanger, Norway

[e] Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden

[f] Network for Medical Sciences, University of Stavanger, Stavanger, Norway

[g] Department of Neurology, Haukeland University Hospital, Bergen, Norway

Title character count: 92

Number of references: 16

Number of figures: 1

Word count abstract: 286

Word count paper: 1975

***Corresponding author:** Dr. Mathias Toft, Department of Neurology, Oslo University Hospital, P.O. Box 4950 Nydalen, N-0424 Oslo, Norway Tel: +47 99514189, Fax: +47 23074891 (mathias.toft@gmail.com).

Victoria Berge-Seidl berge.victoria@gmail.com

Lasse Pihlstrøm lasse.pihlstrom@gmail.com

Jodi Maple Grødem jodi.maple.grodem@sus.no

Lars Forsgren lars.forsgren@umu.se

Jan Linder jan.linder@umu.se

Jan Petter Larsen nevrologlarsen@gmail.com

Ole-Bjørn Tysnes ole-bjorn.tysnes@helse-bergen.no

Mathias Toft mathias.toft@gmail.com

Financial disclosure: The authors report no conflicts of interest concerning the research related to the manuscript.

Abstract

Objective: Coding variants in the *GBA* gene have been identified as the numerically most important genetic risk factors for Parkinson's disease (PD). In addition, genome-wide association studies (GWAS) have identified associations with PD in the *SYT11-GBA* region on chromosome 1q22, but the relationship to GBA coding variants have remained unclear. The aim of this study was to sequence the complete *GBA* gene in a clinical cohort and to investigate whether coding variants within the *GBA* gene may be driving reported association signals.

Methods: We analyzed high-throughput sequencing data of all coding exons of *GBA* in 366 patients with PD. The identified low-frequency coding variants were genotyped in three Scandinavian case-controls series (786 patients and 713 controls). Previously reported risk variants from two independent association signals within the *SYT11-GBA* locus on chromosome 1 were also genotyped in the same samples. We performed association analyses and evaluated linkage disequilibrium between the variants.

Results: We identified six rare mutations (1.6%) and two low-frequency coding variants in *GBA*. E326K (rs2230288) was significantly more frequent in PD patients compared to controls (OR 1.65, p=0.03). There was no clear association of T369M (rs75548401) with disease (OR 1.43, p=0.24). Genotyping the two GWAS hits rs35749011 and rs114138760 in the same sample set, we replicated the association between rs35749011 and disease status (OR 1.67, p=0.03), while rs114138760 was found to have similar allele frequencies in patients and controls. Analyses of the pairwise LD revealed that E326K and rs35749011 are in very high LD (r^2 0.95).

Conclusions: Our results confirm that the *GBA* variant E326K is a susceptibility allele for PD. The results suggest that E326K may fully account for the primary association signal observed at chromosome 1q22 in previous GWAS studies of PD.

Keywords: Parkinson's disease, glucocerebrosidase, E326K, T369M, synaptotagmin 11, GWAS

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder and is generally estimated to affect 1 % of people over 60 years of age, with increasing prevalence in higher age groups. PD is mainly a sporadic disease, but family and candidate gene studies have identified a number of genes related to PD pathogenesis [1]. There is particular interest in the *GBA* gene and its relationship to risk for PD. Homozygous *GBA* mutations cause the autosomal recessive lysosomal storage disorder Gaucher disease. However, heterozygous *GBA* mutations have been identified as the numerically most important genetic risk factors for PD, and 5-10% of PD patients have been reported to carry *GBA* mutations [2]. The two GBA coding variants E326K and T369M do not cause GD in the homozygous state and were initially considered to be benign polymorphisms. There is now increasing evidence in support of the variant E326K as a risk factor for PD, while the association between T369M and PD has been less clear.

Genome wide association studies (GWAS) have linked a number of risk loci to PD susceptibility [3]. Association signals emerging from GWAS typically involve dozens of gene variants in high linkage disequilibrium (LD) encompassing several genes. This complicates the identification of the functionally relevant variants within risk loci.

In PD, an early GWAS reported an intronic disease-associated polymorphism within the *SYT11* gene on chromosome 1q22 [4]. Later, a meta-analysis of several GWAS found an association between a coding variant in the *GBA* gene, E326K, and PD [5]. *GBA* is located about 650 kb from *SYT11*, within the same block of LD referred to as the *GBA-SYT11* locus. The largest and most recent meta-analysis of PD GWAS reported two independent associations within the *GBA-SYT11* locus, but the

relationship between the reported signals and *GBA* coding variants was not examined in detail [3].

The aim of this study was to investigate the frequency of *GBA* mutations in our population and if the two *GBA* variants E326K and T369M are associated with PD in a Scandinavian case-control series. We also wanted to assess to what degree coding *GBA* variants are linked to the *GBA-SYT11* association signals reported for PD.

2. Methods

2.1 Patients and controls

We included samples from three Scandinavian biobanks in our study. From Oslo University Hospital 486 patients (mean age at onset 56 years; SD 11 years) and 473 controls (mean age at inclusion 62 years; SD 11 years) were included. 173 patients (mean age at onset 66 years; SD 9 years) and 187 controls (mean age at inclusion 66 years; SD 9 years) originated from the ParkWest study. 127 patients (mean age at onset 68 years; SD 10 years) and 53 controls (mean age at inclusion 65 years; SD 7 years) were from the NYPUM study at Umeå University Hospital. All PD patients were examined by a neurologist and diagnosed according to the revised UKPDSBB criteria (Oslo and Umeå) or Gelb criteria (ParkWest). The majority of patients were screened for the *LRRK2* G2019S mutation, in addition a large subset of patients was also sequenced for genes causing Mendelian forms of PD (*SNCA*, *PRKN*, *PINK1*, *DJ-1*, *LRRK2*, and *VPS35*). Patients with pathogenic mutations in these genes were excluded from the study. Control subjects consist of spouses of patients, outpatients in primary care and healthy volunteers, all without neurological disease and known parkinsonism among first degree relatives. The study was

approved by the Regional Committee for Medical Research Ethics (Oslo, Norway). Sample and data collection at each study site was approved by local ethics committees. All participants gave written, informed consent.

2.2 Identification of GBA coding variants

To identify all coding variants in the *GBA* gene we analyzed sequencing data from 366 patients from the Oslo patient series. All coding exons of the *GBA* gene were part of a gene panel examined by targeted deep sequencing of DNA pools as described previously [6]. Putative variants were identified by bioinformatic analyses and individually validated by Sanger sequencing. Pools with a read depth below 80x at the relevant position were excluded from analysis of that specific variant. The *GBA* gene was amplified in distinct fragments. To avoid amplification of the pseudogene, we used primer sequences designed to DNA regions exclusive to the *GBA* gene. PCR products were sequenced with a selection of previously described sequencing primers (all primer sequences are available upon request). The conventional nomenclature for *GBA* alleles was used, excluding the 39-residue signal peptide. *In silico* prediction of deleteriousness of the identified variants was performed by the use of Combined Annotation Dependent Depletion (CADD) v1.3, a method integrating and combining multiple genome annotations [7].

2.3 Genotyping and statistical analyses

Two identified *GBA* variants, E326K (rs2230288) and T369M (rs75548401), were genotyped in all 786 cases and 713 controls. We also genotyped the primary risk SNP (rs35749011) and a second independent risk SNP (rs114138760) located within the *GBA-SYT11* locus identified by a recent meta-analysis of genome-wide association

studies [3]. Genotyping was performed by KASP and TaqMan SNP genotyping assays on a Viia7 instrument (Life Technologies, Foster City, CA, USA). The genotype call rate was above 98% for each individual variant. Statistical analyses were performed in PLINK (https://www.cog-genomics.org/plink/1.9/). We tested for Hardy-Weinberg equilibrium (HWE) in controls, observing no significant departure. We assessed the association between each single variant and disease status with Chi-square test and calculated odds ratio (OR). LD between *GBA* coding variants and GWAS risk SNPs were analyzed by using Haploview 4.2 software (https://www.broadinstitute.org/haploview/haploview).

3. Results

We identified two low-frequency coding variants in *GBA* (E326K and T369M) in the sequenced samples. Five additional coding variants and one potential splicing variant were identified by sequencing, each variant only occurring once. Only three of these variants have been described in Gaucher disease patients (N370S, R463C, IVS3+1G>A). The remaining three variants are to our knowledge novel and thus of unknown significance (V457A, G377D, W357R). The novel variants all have a CADD score above the suggested cutoff on deleteriousness. Information on the GBA variants identified by sequencing is summarized in Table 1.

Subsequent genotyping of the two low-frequency variants in all samples revealed that E326K (rs2230288) was significantly more frequent in PD patients compared to controls (OR 1.65, p=0.03). There was no clear association of T369M (rs75548401) with disease (OR 1.43, p=0.24). When genotyping the two meta-GWAS hits rs35749011 and rs114138760 in the same sample set we observed a significant association of rs35749011 in PD patients (OR 1.67, p=0.03), while

rs114138760 was found to have similar allele frequencies in patients and controls (OR 0.91, p=0.83) (Table 2).

The location of the *SYT11* and *GBA* genes, as well as top hit SNPs from previous GWAS are shown in Figure 1 a. Analyses of the pairwise LD between the four genotyped variants revealed that E326K and rs35749011 are in very high LD with a r^2 of 0.95 (D´=0.98) (Figure 1 b and c). Therefore, it is likely that E326K in *GBA* explains the association observed at rs35749011 in previous studies. The LD between T369M and rs114138760 was low, indicating that the secondary association signal reported by Nalls et al. is independent of this coding variant.

4. Discussion

Our results confirm that the *GBA* variant E326K is a susceptibility allele for PD. The frequency of E326K and T369M seem to be higher in our Scandinavian case-control series compared to other European populations. Our study was nevertheless underpowered to identify the previously reported association between PD and T369M. However, we note that the odds ratio was similar to that reported by a recent meta-analysis of T369M [8].

GBA mutations may cause a deficiency of the enzyme glucocerebrosidase (GCase) leading to an accumulation of glucocerebroside within lysosomes. Although E326K and T369M do not cause GD in the homozygous state, they have been shown to modify GCase activity. Studies expressing *GBA* constructs with E326K suggest that this polymorphism reduces enzyme activity [9]. An association between T369M and reduced enzyme activity has also been reported in carriers of this variant [10]. Such a modification of GCase activity may contribute to PD risk in concert with other risk variants/small biochemical alterations.

We found a low frequency of *GBA* mutations in our study, as only 6 of 366 (1.6%) carry known or novel rare mutations. The patients sequenced in our study are included from a tertiary care hospital, and a large proportion of these patients have been treated with deep brain stimulation (DBS). Cognitive impairment is an exclusion criterion when evaluating PD patients for DBS. We may have selected against carriers of *GBA* mutations since this group of PD patients have been reported to have an accelerated cognitive decline [11, 12]. The mutation frequency in a previous Norwegian study is low, indicating that GBA mutations may be rare in hospital-based studies from this population [13].

In this study *GBA* mutations were identified by analyses of data from a pooled sequencing experiment. We have previously reported a high sensitivity of this approach [6]. Furthermore, a high number of exons were Sanger sequenced to validate both rare mutations and low frequency variants, without identifying any additional mutations. We thus find it unlikely that the low frequency of *GBA* mutations should be caused by low sensitivity of our sequencing method.

Mutations in *GBA* play an important role in PD, as *GBA* mutation carriers have an increased disease risk, earlier age at onset, and faster progression. In addition to cognitive decline, various other nonmotor symptoms including REM sleep behavior disorder, hyposmia, and autonomic dysfunction seem to be more frequent [14]. Interestingly, it has recently been demonstrated that also the E326K variant predicts a more rapid progression of cognitive dysfunction and motor symptoms in patients with PD [15]. Thus, *GBA* variants influence the heterogeneity in symptom progression observed in PD. This observation may have important clinical implications, especially if *GBA*-specific treatment will become available.

Our results suggest that the low-frequency *GBA* variant E326K may fully account for the primary association signal observed at the chromosome 1 *SYT11-GBA* locus in previous GWAS of PD. This is in accordance with a previous report by Pankratz et al. where E326K reaches genome-wide significance [5]. Recent GWAS have not clearly reported the relationship between identified association signals and GBA variants, which could inform functional studies. *SYT11* has therefore been considered a potential PD-related gene, since a GWAS reported an intronic disease-associated polymorphism within this gene [4]. Further genetic evidence linking *SYT11* to PD has however been scarce. The largest and most recent meta-analysis of PD GWAS to date located the association signal in an intergenic region hundreds of kilobases away from *SYT11*, but still kept the gene in the naming of the locus. In an attempt to functionally characterize this locus, several studies of synaptotagmin (*SYT11)* and its role in PD pathogenesis have recently been performed [16]. We report very high linkage disequilibrium between E326K and the primary association signal, emphasizing *GBA* as the causal gene at the chromosome 1 *SYT11-GBA* locus. On the other hand, we found no evidence that the secondary signal at this locus was related to the coding *GBA* variant T369M. In the meta-analysis by Pankratz et al. the *GBA* mutation N370S is detected as a second independent signal at the *SYT11-GBA* locus [5]. We are not able to study this due to the very low frequency of N370S in our population.

Identifying the functionally relevant variants within disease risk loci identified by GWAS is important to understand the disease mechanisms involved in disease pathogenesis of PD. Most genetic risk variants fall outside coding regions and do not alter the amino acid sequence of proteins. Until recently, the functional characterization of risk-associated loci has been hindered by the limited annotation of

the human genome outside coding sequences. However, approaches to successfully characterize the functional nature of these loci are emerging. Future studies will hopefully lead to the identification of specific genes and pathways that could serve as actionable therapeutic targets.

Author contributions

The study was designed by VBS, with support and advice by LP and MT. LP, JPL, OBT, LF, JL and MT designed clinical studies and collected data. VBS and JMG carried out the genetic analyses. VBS performed statistical analyses and analyzed the data. VBS and MT drafted the manuscript. All the co-authors critically revised the manuscript for intellectual content and approved the final version for publication.

Acknowledgements

The authors thank the participating patients and control subjects. The authors acknowledge the contributions to data collection (Camilla Sætehaug), technical assistance (Sandra Pilar Henriksen) and supervision of personnel (Aina Haugen Rengmark) at Oslo University Hospital. All other personnel involved in the NYPUM and ParkWest studies are thanked for their contributions.

Funding

The study was funded by a grant from the South-Eastern Norway Regional Health Authority and the Norwegian Parkinson Research Fund. Mathias Toft is also supported by the Research Council of Norway. Lasse Pihlstrøm is supported by the Norwegian Health Association. The ParkWest study is currently supported by the

Western Norway Regional Health Authority (grant# 912049) and the Norwegian Parkinson Research Foundation. The funding organization had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] M.K. Lin, M.J. Farrer, Genetics and genomics of Parkinson's disease, Genome Med, 6 (2014) 48.

[2] E. Sidransky, M.A. Nalls, J.O. Aasly, J. Aharon-Peretz, G. Annesi, E.R. Barbosa, A. Bar-Shira, D. Berg, J. Bras, A. Brice, C.M. Chen, L.N. Clark, C. Condroyer, E.V. De Marco, A. Durr, M.J. Eblan, S. Fahn, M.J. Farrer, H.C. Fung, Z. Gan-Or, T. Gasser, R. Gershoni-Baruch, N. Giladi, A. Griffith, T. Gurevich, C. Januario, P. Kropp, A.E. Lang, G.J. Lee-Chen, S. Lesage, K. Marder, I.F. Mata, A. Mirelman, J. Mitsui, I. Mizuta, G. Nicoletti, C. Oliveira, R. Ottman, A. Orr-Urtreger, L.V. Pereira, A. Quattrone, E. Rogaeva, A. Rolfs, H. Rosenbaum, R. Rozenberg, A. Samii, T. Samaddar, C. Schulte, M. Sharma, A. Singleton, M. Spitz, E.K. Tan, N. Tayebi, T. Toda, A.R. Troiano, S. Tsuji, M. Wittstock, T.G. Wolfsberg, Y.R. Wu, C.P. Zabetian, Y. Zhao, S.G. Ziegler, Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease, The New England journal of medicine, 361 (2009) 1651-1661.

[3] M.A. Nalls, N. Pankratz, C.M. Lill, C.B. Do, D.G. Hernandez, M. Saad, A.L. DeStefano, E. Kara, J. Bras, M. Sharma, C. Schulte, M.F. Keller, S. Arepalli, C. Letson, C. Edsall, H. Stefansson, X. Liu, H. Pliner, J.H. Lee, R. Cheng, M.A. Ikram, J.P. Ioannidis, G.M. Hadjigeorgiou, J.C. Bis, M. Martinez, J.S. Perlmutter, A. Goate, K. Marder, B. Fiske, M. Sutherland, G. Xiromerisiou, R.H. Myers, L.N. Clark, K. Stefansson, J.A. Hardy, P. Heutink, H. Chen, N.W. Wood, H. Houlden, H. Payami, A. Brice, W.K. Scott, T. Gasser, L. Bertram, N. Eriksson, T. Foroud, A.B. Singleton, Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease, Nature genetics, 46 (2014) 989-993.

[4] M.A. Nalls, V. Plagnol, D.G. Hernandez, M. Sharma, U.M. Sheerin, M. Saad, J. Simon-Sanchez, C. Schulte, S. Lesage, S. Sveinbjornsdottir, K. Stefansson, M.

Martinez, J. Hardy, P. Heutink, A. Brice, T. Gasser, A.B. Singleton, N.W. Wood, Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies, Lancet, 377 (2011) 641-649.

[5] N. Pankratz, G.W. Beecham, A.L. DeStefano, T.M. Dawson, K.F. Doheny, S.A. Factor, T.H. Hamza, A.Y. Hung, B.T. Hyman, A.J. Ivinson, D. Krainc, J.C. Latourelle, L.N. Clark, K. Marder, E.R. Martin, R. Mayeux, O.A. Ross, C.R. Scherzer, D.K. Simon, C. Tanner, J.M. Vance, Z.K. Wszolek, C.P. Zabetian, R.H. Myers, H. Payami, W.K. Scott, T. Foroud, P.G. Consortium, Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2, Annals of neurology, 71 (2012) 370-384.

[6] L. Pihlstrom, A. Rengmark, K.A. Bjornara, M. Toft, Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease, Annals of human genetics, 78 (2014) 243-252.

[7] M. Kircher, D.M. Witten, P. Jain, B.J. O'Roak, G.M. Cooper, A general framework for estimating the relative pathogenicity of human genetic variants, 46 (2014) 310-315.

[8] V. Mallett, J.P. Ross, R.N. Alcalay, A. Ambalavanan, E. Sidransky, P.A. Dion, G.A. Rouleau, Z. Gan-Or, GBA p.T369M substitution in Parkinson disease: Polymorphism or association? A meta-analysis, Neurology. Genetics, 2 (2016) e104.

[9] E. Malini, S. Grossi, M. Deganuto, C. Rosano, R. Parini, S. Dominisini, R. Cariati, S. Zampieri, B. Bembi, M. Filocamo, A. Dardis, Functional analysis of 11 novel GBA alleles, European journal of human genetics : EJHG, 22 (2014) 511-516.

[10] R.N. Alcalay, O.A. Levy, C.C. Waters, S. Fahn, B. Ford, S.H. Kuo, P. Mazzoni, M.W. Pauciulo, W.C. Nichols, Z. Gan-Or, G.A. Rouleau, W.K. Chung, P. Wolf, P.

Oliva, J. Keutzer, K. Marder, X. Zhang, Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations, Brain : a journal of neurology, 138 (2015) 2648-2658.

[11] T. Oeda, A. Umemura, Y. Mori, S. Tomita, M. Kohsaka, K. Park, K. Inoue, H. Fujimura, H. Hasegawa, H. Sugiyama, H. Sawada, Impact of glucocerebrosidase mutations on motor and nonmotor complications in Parkinson's disease, Neurobiology of aging, 36 (2015) 3306-3313.

[12] K. Brockmann, K. Srulijes, S. Pflederer, A.K. Hauser, C. Schulte, W. Maetzler, T. Gasser, D. Berg, GBA-associated Parkinson's disease: Reduced survival and more rapid progression in a prospective longitudinal study, Movement disorders : official journal of the Movement Disorder Society, DOI 10.1002/mds.26071(2014).

[13] M. Toft, L. Pielsticker, O.A. Ross, J.O. Aasly, M.J. Farrer, Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian population, Neurology, 66 (2006) 415-417.

[14] S. Jesus, I. Huertas, I. Bernal-Bernal, M. Bonilla-Toribio, M.T. Caceres-Redondo, L. Vargas-Gonzalez, M. Gomez-Llamas, F. Carrillo, E. Calderon, M. Carballo, P. Gomez-Garre, P. Mir, GBA Variants Influence Motor and Non-Motor Features of Parkinson's Disease, PloS one, 11 (2016) e0167749.

[15] M.Y. Davis, C.O. Johnson, J.B. Leverenz, D. Weintraub, J.Q. Trojanowski, A. Chen-Plotkin, V.M. Van Deerlin, J.F. Quinn, K.A. Chung, A.L. Peterson-Hiller, L.S. Rosenthal, T.M. Dawson, M.S. Albert, J.G. Goldman, G.T. Stebbins, B. Bernard, Z.K. Wszolek, O.A. Ross, D.W. Dickson, D. Eidelberg, P.J. Mattis, M. Niethammer, D. Yearout, S.C. Hu, B.A. Cholerton, M. Smith, I.F. Mata, T.J. Montine, K.L. Edwards, C.P. Zabetian, Association of GBA Mutations and the E326K

Polymorphism With Motor and Cognitive Progression in Parkinson Disease, JAMA neurology, 73 (2016) 1217-1224.

[16] C.F. Bento, A. Ashkenazi, M. Jimenez-Sanchez, D.C. Rubinsztein, The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway, Nature communications, 7 (2016) 11803.