TABLE 1
Disease-causing missense mutations in the *fukutin* gene

Mutation	Other allele	Severity	References
A114T	T286 frame shift	Mild	10
G125S[a]	5370–5842 deletion (3'-UTR)	Severe	11
G125S[a]	F390 frame shift	Severe	11
M133T	3-kb insertion	Typical	28
A170E	Y371C	Typical	13
H172R	3-kb insertion	Typical	27
R179T	3-kb insertion	Mild	6
H186R	Homozygote	Severe	12
R246G	R47 nonsense	Mild	13
C250G	3-kb insertion	Typical	4
W305C	Homozygote	Typical	10
R307Q	Homozygote	Mild	13
R307Q	F390 frame shift	Mild	9
R307Q	N455 frame shift	Mild	9
Q358P	3-kb insertion	Mild	6
Y371C	3-kb insertion	Typical	5
Y371C	A170E	Typical	13

[a] G125S has been registered as a polymorphism (rs_34006675).

mixture, then directly boiled for 5 min in SDS-polyacrylamide gel loading buffer.

SDS-PAGE, Western Blotting, and Laminin Overlay Assay—Cell lysates were dissolved in SDS sample buffer and subjected to SDS-PAGE in 10% gels or 7.5% gels. Gels were transferred to polyvinylidene difluoride (PVDF) membranes (Millipore), probed with anti-FLAG, anti-α-DG core protein (AP-074G-C), anti-α-DG sugar chain (IIH6C4), or anti-β-DG antibodies, and then developed with horseradish peroxidase-conjugated secondary antibody (DAKO, Carpinteria, CA). Blots were processed using ECL plus Western blotting detection system (GE Healthcare) and exposed to Fuji RX-U x-ray film (Fuji Film, Kanagawa, Japan). The laminin overlay assay was performed according to the method of Michele *et al.* (14).

Additional methods are described under supplemental Experimental Procedure.

RESULTS

Mislocalization of Disease-causing Missense Mutant Fukutin Proteins—To examine the cellular location of fukutin proteins containing disease-causing missense mutations, we constructed expression vectors encoding wild-type or mutant fukutin proteins with a FLAG epitope at the C terminus. The missense mutants analyzed in this study have been identified inside and outside Japan (Table 1), and their clinical phenotypes vary from severe WWS-like to mild LGMD-type without mental retardation (Table 1 and Fig. 1). These constructs were transfected into C2C12 myoblast cells, and the cellular localizations of the expressed fukutin proteins were examined by immunofluorescence. Immunofluorescent signals indicated co-localization of the expressed wild-type fukutin with the Golgi apparatus marker GM130 (162/176 cells; Golgi + Golgi and around/total cells) (Fig. 2*A*, *merge*, and supplemental Table I). Nine of the 13 missense mutants (A114T (97/113), G125S (140/144), M133T (109/120), R179T (142/149), R246G (101/110), C250G (124/134), W305C (183/198), R307Q (131/134), and Q358P (174/182)) also co-localized with GM130 (supplemental Table I), indicating that these mutations do not affect the cellular location of fukutin protein. In contrast, the A170E (13/146), H172R (8/145), and H186R (6/141) mutants, as well as the previously reported Y371C (8/128) mutant (16), did not co-localize with GM130 (Fig. 2*A* and supplemental Table I), instead showing co-localization with the ER marker KDEL (Fig. 2*B*). These results indicated that A170E, H172R, H186R, and Y371C aberrantly localize to the ER.

Accumulation of Missense Fukutin Mutants in ER Caused by Impaired Transport to the Golgi—Accumulation of the four mutants in the ER might result from improper cellular trafficking. To determine whether the mutants are not properly transported from the ER to the Golgi or whether they are transported back to the ER after reaching the Golgi, we treated C2C12 cells expressing wild-type fukutin or the four mutants with nocodazole (an inhibitor for retrograde transport from the Golgi to the ER) or BFA (an inhibitor for anterograde transport from the ER to the Golgi). If the four mutants reached the Golgi and then were immediately transported back to the ER, the four mutant proteins should be detected in the Golgi after nocodazole treatment. Immunofluorescent signals indicating ER accumulation of the four mutants were observed following a 2-h incubation with 10 μg/ml nocodazole (Fig. 3*C–F*, *panels 4*). When cells expressing wild-type fukutin were incubated with 5 μg/ml BFA, the wild-type fukutin was detected in the ER (99/107; ER/total cells) (Fig. 3*B*, *panel 6*, and supplemental Table II), as seen in the cells expressing any of the four mutants. These data suggested that failure of proper transport via the anterograde pathway causes mislocalization of the four mutants to the ER.

Correction of Cellular Location of Mutant Fukutin Proteins by Low Temperature Culture—We hypothesized that mislocalization resulted from protein misfolding and therefore examined whether the localization of the four missense mutants could be corrected by folding amelioration. It has been reported that cell culture at low temperature can ameliorate folding and correct the subcellular localization of missense mutant proteins (31). In culture at 37 °C, the four missense fukutin mutants (A170E, H172R, H186R, and Y371C) were co-localized with KDEL (Fig. 4*A*). At low temperature (27.5 °C), the A170E (130/144; Golgi + Golgi and around/total cells), H172R (130/159), and Y371C (102/158) mutants preferentially co-localized with GM130 (Fig. 4*A* and supplemental Table I), indicating that the ER accumulation had decreased and proper localization to the Golgi was restored. Most of the expressed H186R mutant protein, however, remained in the ER (134/155; ER/total cells) at 27.5 °C; only a small proportion of this mutant shifted to the Golgi (Fig. 4*A* and supplemental Table I).

POMGnT1, which has been shown to interact with fukutin and localize to the Golgi (151/156; Golgi with fukutin/total cells) (supplemental Table III) (16), also mislocalized to the ER when co-expressed with the mutants A170E (89/103; ER/total cells), H172R (102/113), H186R (96/111), and Y371C (98/109) (Fig. 4*B*, 37 °C, and supplemental Table III). These results indicated that fukutin mislocalization also affects the cellular location of POMGnT1. Low temperature culture of cells expressing both POMGnT1 and any of the mutants A170E (158/173; Golgi with fukutin/total cells), H172R (103/129), or Y371C (126/192) restored POMGnT1 subcellular localization to the Golgi (Fig. 4*B*, 27.5 °C, and supplemental Table III). When expressed with the H186R mutant, the POMGnT1 localization shifted to the Golgi despite the majority of the H186R mutant remaining in

MSRINKNVVLALLTLTSSAFLLFQLYYYKHYLSTKNGAGLSKSKGSRIGFDSTQWRAVKKFIMLTSNQNV

PVFLIDPLILELINKNFEQVKNTSHGSTSQCKFFCVPRDFTAFALQYHLWKNEEGWFRIAENMGFQCLKI

ESKDPRLDGIDSLSGTEIPLHYICKLATHAIHLVVFHERSGNYLWHGHLRLKEHIDRKFVPFRKLQFGRY

PGAFDRPELQQVTVDGLEVLIPKDPMHFVEEVPHSRFIECRYKEARAFFQQYLDDNTVEAVAFRKSAKEL

LQLAAKTLNKLGVPFWLSSGTCLGWYRQCNIIPYSKDVDLGIFIQDYKSDIILAFQDAGLPLKHKFGKVE

DSLELSFQGKDDVKLDVFFFYEETDHMWNGGTQAKTGKKFKYLFPKFTLCWTEFVDMKVHVPCETLEYIE

ANYGKTWKIPVKTWDWKRSPPNVQPNGIWPISEWDEVIQLY

FIGURE 1. **Amino acid sequence of fukutin and location of missense mutations.** Amino acid changes are indicated along with severity of phenotype. *Blue*, mild form (CMD without mental retardation, LGMD, or cardiomyopathy). *Purple*, typical form (FCMD, CMD with mental retardation); *red*, severe form (WWS). *Single underline* indicates the transmembrane domain, the *asterisk* indicates an *N*-glycosylation site, and the *double underline* indicates the DxD motif.

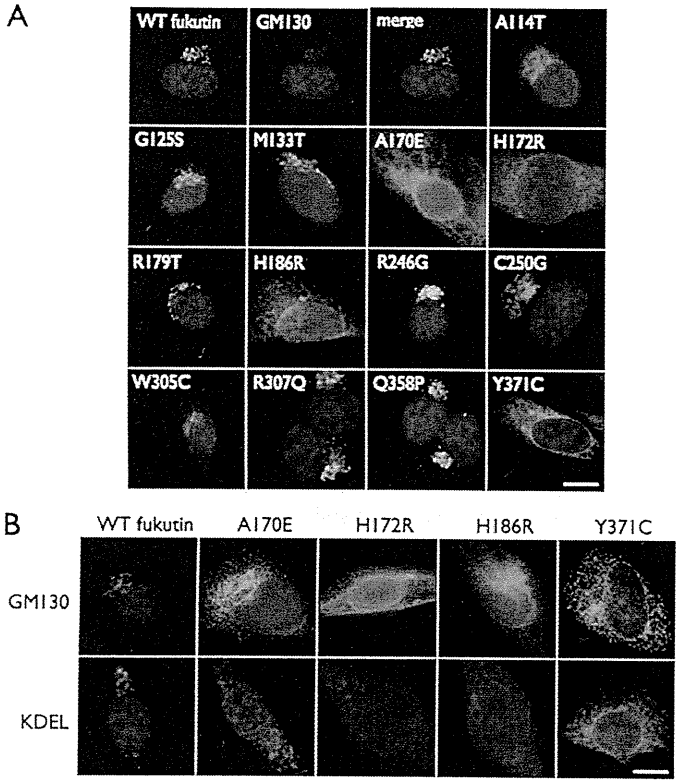


FIGURE 2. **Some disease-causing missense mutations cause abnormal cellular localization of fukutin.** *A*, localization of the FLAG-tagged wild-type or missense fukutin proteins in transfected C2C12 myoblast cultured cells. Cells were double-labeled with anti-FLAG (*green*) and a Golgi marker GM130 (*red*). No co-localization was seen for the missense fukutin mutants A170E, H172R, H186R, and Y371C. Wild-type (*WT fukutin*) and the other nine mutants co-localized with GM130. *B*, mislocalization of missense fukutin mutants to the ER. The fukutin missense mutants A170E, H172R, H186R, and Y371C (*green*) lost their localization to the Golgi (*red*) in transfected C2C12 cultured cells (*upper*). Co-localization of these mutants with the ER marker KDEL (*red*) was seen in the cultures (*lower*). *Blue*, DAPI, indicating the nucleus. *Scale bars*, 10 μm.

the ER (108/141; Golgi without fukutin/total cells) (Fig. 4*B* and supplemental Table III). These data showed that low temperature culture could correct mislocalization of the mutant fukutin proteins, but its effect may depend on the position of missense mutations within the protein.

Residual Function of Fukutin Missense Mutants to Restore α-Dystroglycan Glycosylation—C2C12 cells contain endogenous fukutin, making it difficult to evaluate the α-DG modification activity of exogenously expressed mutant fukutin proteins. Instead, we used *fukutin*-null mouse ES cells (FCMDp182 cell) (30) to examine whether the four missense mutant fukutin proteins retain the activity responsible for α-DG modification. We expressed the four mutants (A170E, H172R, H186R, and Y371C) in FCMDp182 cells (supplemental Fig. 1) and then examined the recovery of glycosylation and laminin binding activity of α-DG. In mock-transfected FCMDp182 cells cultured at either 37 °C or 27.5 °C, α-DG showed no detectable reactivity against the monoclonal IIH6C4 antibody, which recognizes the functionally glycosylated form of α-DG. We also observed hypoglycosylation of α-DG, as indicated by lower molecular mass, and no α-DG laminin binding activity in mock-transfected cells (Fig. 5, *lanes 2* and *10*). When wild-type fukutin was expressed in FCMDp182 cells, we observed restoration of the laminin binding activity and the IIH6C4 reactivity of α-DG, but at much weaker levels than for α-DG in the wild-type mouse ES cell line AB2.2 (*lanes 1* and *3*). This partial restoration and the faint amount of α-DG core protein bands around 100 kDa (the size of α-DG in the AB2.2 cells) may result from low transfection efficiency. Expression of each of the four missense mutants in the FCMDp182 cells restored the IIH6C4 reactivity and the laminin binding activity to levels comparable with those observed in the wild-type fukutin transfectants (*lanes 3–7*). Although these missense mutants preferentially localized to the ER at 37 °C, it is possible that small amounts of the expressed mutants reached the Golgi. To support this interpretation, a report has shown that the mislocalized mutant protein ΔF508 cystic fibrosis transmembrane conductance regulator, which is the most common mutation leading to cystic fibrosis, shows partial function when overexpressed (32). We have not detected restoration of the IIH6C4 reactivity and the laminin binding activity by expression of a mutant fukutin protein with direct substitutions in the DxD motif (D317A/V/D319A) (supplemental Fig. 2). Expression of FKRP, a member

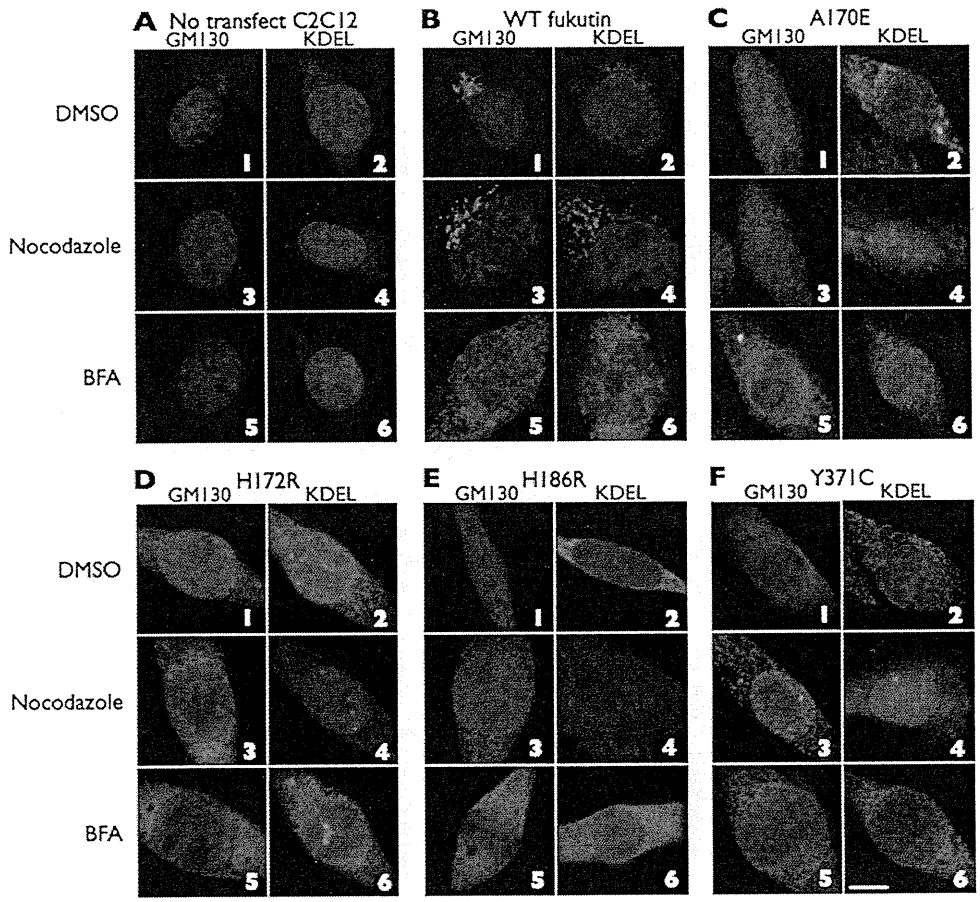


FIGURE 3. **Accumulation of the missense mutants in the ER could be caused by the failure of proper transport via the anterograde pathway.** C2C12 cells without (*A*) or with expression of wild-type fukutin (*B*) or the missense mutants (*C–F*) were incubated with nocodazole or BFA. Nocodazole treatment did not change the cellular localization of the expressed mutant fukutin proteins, whereas BFA treatment shifted wild-type fukutin to the ER. *Red*, GM130 or KDEL; *green*, expressed fukutin proteins; *blue*, DAPI. *Scale bar*, 10 μm.

of the fukutin protein family, in FCMDp182 cells showed no effect on α-DG glycosylation (*lane 8*). We also performed these experiments at low culture temperature (27.5 °C) and obtained similar results (Fig. 5, *lanes 9–16*). The reason for reduced efficiency of glycosylation recovery in low temperature culture is unclear, but the low temperature might affect enzymatic activities involved in the α-DG glycosylation pathway. Because the expression level of each mutant protein was different (Fig. 5, anti-FLAG), it was not possible to simply compare their residual activity. Importantly, however, all four mutants retained the activity responsible for α-DG modification.

Correction of Mislocalization of Missense Fukutin Mutants by Curcumin Treatment—Our results indicated that mislocalization of some missense mutants could cause disease-related abnormal glycosylation of α-DG. Therefore, we next searched for chemicals that could restore proper localization. Curcumin, a nontoxic natural constituent of turmeric spice, has shown the ability to correct misfolding and mislocalization of the ΔF508 mutant (33). We incubated C2C12 cells expressing the four mutants (A170E, H172R, H186R, or Y371C) with 10 μg/ml curcumin at 37 °C for 24 h. Among the four mutants, the A170E mutant showed the greatest benefit from curcumin treatment. In the absence of curcumin, approximately 90% of the cells showed ER localization of the A170E mutant (159/175; ER/total cells), and only a few cells (7/175; Golgi + Golgi and around cells) showed Golgi localization (Fig. 6). Curcumin treatment significantly decreased the ER mislocalization signal of the A170E mutant (69/156; ER/total ells) and increased the Golgi- and Golgi/around signals (72/156; Golgi + Golgi and around cells) ($p = 8.28 \times 10^{-20}$). Although not as dramatic as seen with A170E, the other mutants showed slight beneficial changes in their cellular distributions following curcumin treatment (Fig. 6*B* and supplemental Table II). We also examined glycerol, arginine, and 17-allylaminogeldanamycin for beneficial effects on localization of the mutants, but none was observed (data not shown).

DISCUSSION

In this report, we have demonstrated that some disease-causing missense fukutin mutants lost their Golgi localization in C2C12 cultured cells and that this mislocalization can be corrected by low temperature culture or curcumin treatment. We identified four missense mutants that localized abnormally to the ER. Nocodazole treatment did not alter their ER localization, and low temperature culture shifted three of the four mutants to the Golgi. It is generally recognized that cell culture at low temperature can ameliorate protein misfolding and correct abnormal cellular localization (31). Therefore, we presume that some missense mutants could not be transported to the Golgi via the anterograde pathway because of protein misfold-

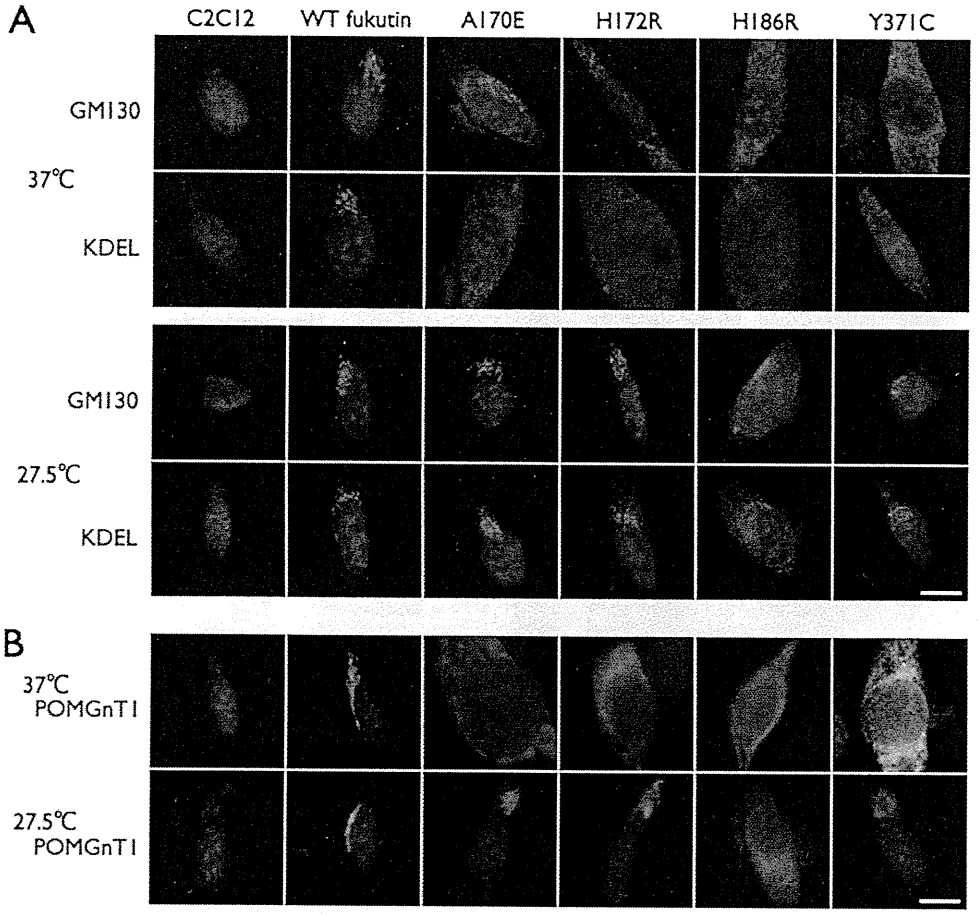


FIGURE 4. **Low temperature culture corrects mislocalization of mutant fukutin proteins.** *A*, effects of low temperature culture on the localization of the missense fukutin mutants. At 37 °C, the missense fukutin mutants (A170E, H172R, H186R, and Y371C, *green*) co-localized with KDEL (ER marker, *red*) in C2C12 cultured cells (*upper*). In contrast, at 27.5 °C, the missense fukutin mutants A170E, H172R, and Y371C lost their co-localization with KDEL and acquired co-localization with GM130 (Golgi marker, *red*) (*lower*). A large amount of the mutant H186R protein retained the ER localization in 27.5 °C culture. *B*, POMGnT1 localization when co-expressed with the missense mutants. POMGnT1 (*red*) mislocalized to the ER when co-expressed with the missense fukutin mutants (A170E, H172R, H186R, or Y371C, *green*) at 37 °C. However, at 27.5 °C, localization of POMGnT1 was restored to the Golgi even when co-expressed with mutant fukutin proteins. *Blue*, DAPI. *Scale bars*, 10 μm.

ing caused by amino acid substitutions. A large amount of the fourth mutant protein, H186R, retained the ER localization even under low temperature conditions (Fig. 4*A*). It has been reported that the patient with the homozygous H186R mutation is affected with WWS (12), which shows more severe pathological features than typical FCMD. The H186R mutation in the *fukutin* gene may affect protein folding severely enough that low temperature conditions cannot correct mislocalization. We reported previously that POMGnT1 interacts with fukutin and co-localizes to the Golgi (16). Our present data show that correction of the mislocalization of the three missense fukutin mutants by low temperature culture was accompanied by correction of POMGnT1 cellular localization (Fig. 4*B*). This POMGnT1 behavior is rational because the transmembrane region of fukutin, through which fukutin binds to POMGnT1 (16), remained intact in these missense fukutin mutants (Fig. 1, *single underline*). Immunoprecipitation experiments confirmed the interaction between each of the missense fukutin mutants and POMGnT1 (supplemental Fig. 3). Most of the H186R mutant remained in the ER at 27.5 °C, but POMGnT1 expressed with the H186R localized to the Golgi. The reason for this result is uncertain, but a possible explanation is that the H186R mutant may have a harmful (dominant negative-like) effect on POMGnT1 localization, and this effect may be ameliorated at low temperature.

The above-mentioned four mutants, which showed abnormal localization to the ER, were identified in patients presenting with a typical or a severe phenotype. However, several of the remaining nine mutants, which showed Golgi localization, were also identified in patients presenting with the typical or the severe phenotype (G125S, M133T, C250G, and W305C) (Table 1). Therefore, the typical or severe phenotypes seem not always to be related to abnormal cellular localization of mutated proteins. Given that FCMD is inherited in an autosomal recessive manner, disease-causing mutations in the *fukutin* gene must lead to loss of function of the fukutin protein. Although the exact function of fukutin is undetermined, these nine mutations may disrupt an important functional domain in the protein. For example, the W305C (10) and R307Q (9, 13) mutations are located in the DxD motif (Fig. 1, *double underline*), which is predicted to be involved in the modification of cell surface glycoproteins or glycolipids (17). These substitutions may disrupt the DxD motif and produce dysfunctional fukutin protein. Five mutations (A114T, R179T, R246G, R307Q, and Q358P) have been identified in patients presenting with mild phenotypes (congenital muscular dystrophy with no

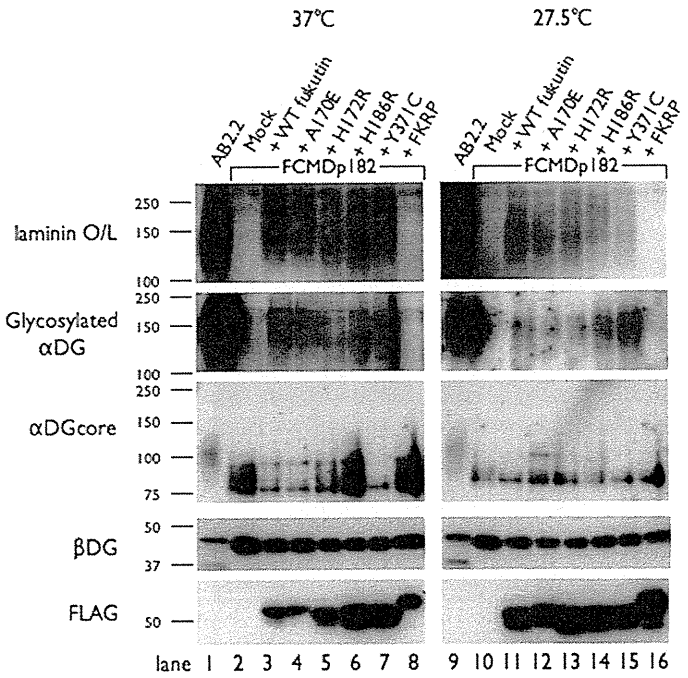
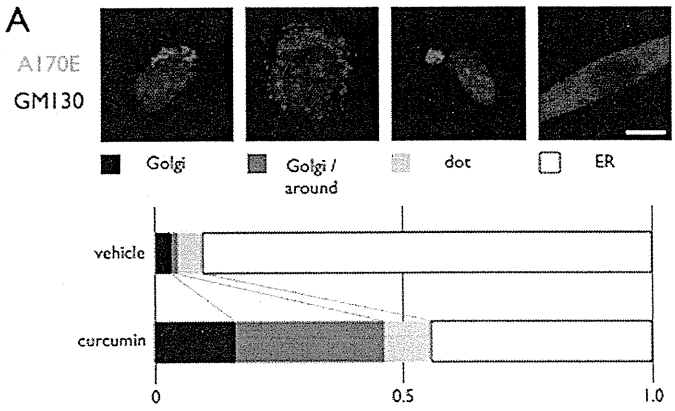


FIGURE 5. **Fukutin missense mutants retain the activity responsible for α-DG glycosylation.** Western blot analysis was performed to detect IIH6C4 (glycosylated α-DG), core α-DG, β-DG, and FLAG tag, and the laminin overlay assay. DG and fukutin proteins were prepared from wild-type mouse ES cells (AB2.2), fukutin-deficient mouse ES cells (FCMDp182), and fukutin- or FKRP-transfected FCMDp182 cells, and cultured at 37 °C and 27.5 °C. FCMDp182 cells expressing any of the mutant fukutin proteins showed levels of glycosylated α-DG signal (IIH6C4-reactivity) and laminin binding activity that were comparable with those observed in FCMDp182 cells expressing wild-type fukutin.



B

	A170E		H172R		H186R		Y371C	
	vehicle	curcumin	vehicle	curcumin	vehicle	curcumin	vehicle	curcumin
Golgi	6	25	1	0	0	0	0	0
Golgi / around	1	47	1	32	0	32	1	11
dot	9	15	1	5	1	5	11	11
ER	159	69	109	102	113	158	115	114
p value	8.28×10^{-20}		1.10×10^{-6}		0.005		0.02	

FIGURE 6. **Curcumin treatment partly corrects mislocalization of the missense fukutin mutants.** *A*, effects of curcumin treatment on the cellular localization of the missense A170E fukutin mutants. C2C12 cells expressing the missense fukutin mutant A170E were cultured in the presence or absence of 10 μg/ml curcumin. Cells were classified into four classes (Golgi localization, Golgi and around localization, dot localization, and ER localization), then counted and statistically analyzed using the χ^2 test. *Red*, GM130; *green*, expressed missense fukutin protein A170E; *blue*, DAPI. *Scale bar*, 10 μm. *B*, statistical analysis of effects of curcumin treatment on the cellular localization of the missense fukutin mutants A170E, H172R, H186R, and Y371C.

mental retardation, LGMD with no mental retardation, or cardiomyopathy) (6, 9, 10, 13). In the present study, these five mutant proteins were localized to the Golgi when expressed in C2C12 cells. In patients with R179T or Q358P mutations, α-DG shows residual reactivity against the monoclonal antibody IIH6C4, which recognizes functionally glycosylated α-DG (6), indicating that these fukutin mutations retain partial function in the DG maturation pathway.

It is of interest that some missense fukutin mutants retain α-DG glycosylation activity and that their mislocalization could be partly corrected by treatments directing at folding amelioration. These observations suggest that drugs capable of correcting the localization might have therapeutic benefits in patients who carry these missense fukutin mutants. Although approximately half of the cells expressing the A170E mutant retained the ER accumulation signals following curcumin treatment (Fig. 6), recent studies have indicated that even partial restoration of α-DG glycosylation can produce therapeutic effects (29). Efforts to identify more efficient folding amelioration reagents may lead to therapeutic strategies.

A large number of missense mutations have been identified in dystroglycanopathy. It has been reported that missense mutations in *POMT1* and *POMGnT1* compromise enzymatic activity in the gene products (18, 34, 35). Disease-causing missense mutations in *POMT2*, *LARGE*, and *FKRP* have been also identified (10, 21–23, 36, 37). Recently, Kawahara *et al.* have reported that expression of some disease-causing missense FKRP mutant proteins in FKRP knock-down zebrafish restores α-DG glycosylation, suggesting a residual FKRP function in these missense mutants (38); however, phenotype improvement depends on the location of the mutation. The FKRP missense mutations C318Y and A455D, which failed to improve the fish phenotype in the report from Kawahara *et al.* (38), were reported to show abnormal cellular localization when expressed in certain cell lines (39). On the other hand, using several different cell lines, Dolatshad *et al.* suggested that a reduced protein (putative enzymatic) function of FKRP rather than protein mislocalization is the primary mechanism of disease (39). Interestingly, Bao *et al.* have reported cells lacking FKRP transcripts but expressing IIH6C4-reactive α-DG (40). This finding may indicate a possibility of a FKRP-independent glycosylation pathway. Alternatively, only a subtle amount of FKRP, even below detectable level by RT-PCR, may be sufficient for α-DG glycosylation. This implies that some missense mutants can restore IIH6C4-reactive α-DG if only a little protein function remains. Of the increasing number of identified disease-causing missense mutations, some may alter the cellular location of the protein, which can be a direct cause of disease. Correction of cellular localization or folding amelioration may have a therapeutic benefit for dystroglycanopathies caused by missense mutations, although the finding from cell culture experiments must be interpreted cautiously when extrapolating to human disease.

In this study, we have used curcumin to correct mislocalization of missense fukutin mutants. It has been reported that curcumin can correct misfolding and mislocalization of cystic fibrosis transmembrane conductance regulator with the ΔF508 mutation (33). Curcumin has also been reported to have protective effects in neurodegenerative diseases by inhibiting protein misfolding and aggregation in Creutzfeldt-Jakob disease and Parkinson disease (41, 42). These studies indicate that curcumin or its derivatives might be new candidate compounds for protein-folding diseases. In addition, the use of pharmacological chaperones to stabilize or promote correct folding of mutant proteins has been shown as a potential therapeutic approach to phenylketonuria, in which more than 500 disease-causing mutations have been identified (43). Our results contribute to a deeper understanding of the molecular pathogenesis of fukutin-deficient muscular dystrophy and have led us to propose a novel therapeutic strategy directed at correction of cellular localization and/or folding amelioration of disease-causing missense mutant proteins.

Acknowledgments—We thank past and present members of the Dr. Toda's laboratory for fruitful discussions and scientific contributions and Dr. Jennifer Logan for help in editing the manuscript.

REFERENCES

1. Fukuyama, Y., Osawa, M., and Suzuki, H. (1981) Congenital progressive muscular dystrophy of the Fukuyama type: clinical, genetic and pathological considerations. *Brain Dev.* **3,** 1–29
2. Toda, T., Segawa, M., Nomura, Y., Nonaka, I., Masuda, K., Ishihara, T., Sakai, M., Tomita, I., Origuchi, Y., Ohno, K., Misugi, N., Sasaki, Y., Takada, K., Kawai, M., Otani, K., Murakami, T., Saito, K., Fukuyama, Y., Shimizu, T., Kanazawa, I., and Nakamura, Y. (1993) Localization of a gene for Fukuyama-type congenital muscular dystrophy to chromosome 9q31–33. *Nat. Genet.* **5,** 283–286
3. Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M., Hamano, K., Sakakihara, Y., Nonaka, I., Nakagome, Y., Kanazawa, I., Nakamura, Y., Tokunaga, K., and Toda, T. (1998) An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* **394,** 388–392
4. Kondo-Iida, E., Kobayashi, K., Watanabe, M., Sasaki, J., Kumagai, T., Koide, H., Saito, K., Osawa, M., Nakamura, Y., and Toda, T. (1999) Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). *Hum. Mol. Genet.* **8,** 2303–2309
5. Kobayashi, K., Sasaki, J., Kondo-Iida, E., Fukuda, Y., Kinoshita, M., Sunada, Y., Nakamura, Y., and Toda, T. (2001) Structural organization, complete genomic sequences and mutational analyses of the Fukuyama-type congenital muscular dystrophy gene, fukutin. *FEBS Lett.* **489,** 192–196
6. Murakami, T., Hayashi, Y. K., Noguchi, S., Ogawa, M., Nonaka, I., Tanabe, Y., Ogino, M., Takada, F., Eriguchi, M., Kotooka, N., Campbell, K. P., Osawa, M., and Nishino, I. (2006) Fukutin gene mutations cause dilated cardiomyopathy with minimal muscle weakness. *Ann. Neurol.* **60,** 597–602
7. Silan, F., Yoshioka, M., Kobayashi, K., Simsek, E., Tunc, M., Alper, M., Cam, M., Guven, A., Fukuda, Y., Kinoshita, M., Kocabay, K., and Toda, T. (2003) A new mutation of the fukutin gene in a non-Japanese patient. *Ann. Neurol.* **53,** 392–396
8. de Bernabé, D. B., van Bokhoven, H., van Beusekom, E., Van den Akker, W., Kant, S., Dobyns, W. B., Cormand, B., Currier, S., Hamel, B., Talim, B., Topaloglu, H., and Brunner, H. G. (2003) A homozygous nonsense mutation in the fukutin gene causes a Walker-Warburg syndrome phenotype. *J. Med. Genet.* **40,** 845–848
9. Godfrey, C., Escolar, D., Brockington, M., Clement, E. M., Mein, R., Jimenez-Mallebrera, C., Torelli, S., Feng, L., Brown, S. C., Sewry, C. A., Rutherford, M., Shapira, Y., Abbs, S., and Muntoni, F. (2006) Fukutin gene mutations in steroid-responsive limb girdle muscular dystrophy. *Ann. Neurol.* **60,** 603–610
10. Godfrey, C., Clement, E., Mein, R., Brockington, M., Smith, J., Talim, B., Straub, V., Robb, S., Quinlivan, R., Feng, L., Jimenez-Mallebrera, C., Mercuri, E., Manzur, A. Y., Kinali, M., Torelli, S., Brown, S. C., Sewry, C. A., Bushby, K., Topaloglu, H., North, K., Abbs, S., and Muntoni, F. (2007) Refining genotype phenotype correlations in muscular dystrophies with defective glycosylation of dystroglycan. *Brain* **130,** 2725–2735
11. Cotarelo, R. P., Valero, M. C., Prados, B., Peña, A., Rodríguez, L., Fano, O., Marco, J. J., Martínez-Frías, M. L., and Cruces, J. (2008) Two new patients bearing mutations in the fukutin gene confirm the relevance of this gene in Walker-Warburg syndrome. *Clin. Genet.* **73,** 139–145
12. Manzini, M. C., Gleason, D., Chang, B. S., Hill, R. S., Barry, B. J., Partlow, J. N., Poduri, A., Currier, S., Galvin-Parton, P., Shapiro, L. R., Schmidt, K., Davis, J. G., Basel-Vanagaite, L., Seidahmed, M. Z., Salih, M. A., Dobyns, W. B., and Walsh, C. A. (2008) Ethnically diverse causes of Walker-Warburg syndrome (WWS): FCMD mutations are a more common cause of WWS outside of the Middle East. *Hum. Mutat.* **29,** E231-E241
13. Vuillaumier-Barrot, S., Quijano-Roy, S., Bouchet-Seraphin, C., Maugenre, S., Peudenier, S., Van den Bergh, P., Marcorelles, P., Avila-Smirnow, D., Chelbi, M., Romero, N. B., Carlier, R. Y., Estournet, B., Guicheney, P., and Seta, N. (2009) Four Caucasian patients with mutations in the fukutin gene and variable clinical phenotype. *Neuromuscul. Disord.* **19,** 182–188
14. Michele, D. E., Barresi, R., Kanagawa, M., Saito, F., Cohn, R. D., Satz, J. S., Dollar, J., Nishino, I., Kelley, R. I., Somer, H., Straub, V., Mathews, K. D., Moore, S. A., and Campbell, K. P. (2002) Post-translational disruption of dystroglycan-ligand interactions in congenital muscular dystrophies. *Nature* **418,** 417–422
15. Matsumoto, H., Noguchi, S., Sugie, K., Ogawa, M., Murayama, K., Hayashi, Y. K., and Nishino, I. (2004) Subcellular localization of fukutin and fukutin-related protein in muscle cells. *J. Biochem.* **135,** 709–712
16. Xiong, H., Kobayashi, K., Tachikawa, M., Manya, H., Takeda, S., Chiyonobu, T., Fujikake, N., Wang, F., Nishimoto, A., Morris, G. E., Nagai, Y., Kanagawa, M., Endo, T., and Toda, T. (2006) Molecular interaction between fukutin and POMGnT1 in the glycosylation pathway of α-dystroglycan. *Biochem. Biophys. Res. Commun.* **350,** 935–941
17. Aravind, L., and Koonin, E. V. (1999) The fukutin protein family: predicted enzymes modifying cell surface molecules. *Curr. Biol.* **9,** R836–837
18. Yoshida, A., Kobayashi, K., Manya, H., Taniguchi, K., Kano, H., Mizuno, M., Inazu, T., Mitsuhashi, H., Takahashi, S., Takeuchi, M., Herrmann, R., Straub, V., Talim, B., Voit, T., Topaloglu, H., Toda, T., and Endo, T. (2001) Muscular dystrophy and neuronal migration disorder caused by mutations in a glycosyltransferase, POMGnT1. *Dev. Cell* **1,** 717–724
19. Beltrán-Valero de Bernabé, D., Currier, S., Steinbrecher, A., Celli, J., van Beusekom, E., van der Zwaag, B., Kayserili, H., Merlini, L., Chitayat, D., Dobyns, W. B., Cormand, B., Lehesjoki, A. E., Cruces, J., Voit, T., Walsh, C. A., van Bokhoven, H., and Brunner, H. G. (2002) Mutations in the *O*-mannosyltransferase gene POMT1 give rise to the severe neuronal migration disorder Walker-Warburg syndrome. *Am. J. Hum. Genet.* **71,** 1033–1043
20. van Reeuwijk, J., Janssen, M., van den Elzen, C., Beltran-Valero de Bernabé, D., Sabatelli, P., Merlini, L., Boon, M., Scheffer, H., Brockington, M., Muntoni, F., Huynen, M. A., Verrips, A., Walsh, C. A., Barth, P. G., Brunner, H. G., and van Bokhoven, H. (2005) POMT2 mutations cause α-dystroglycan hypoglycosylation and Walker-Warburg syndrome. *J. Med. Genet.* **42,** 907–912
21. Brockington, M., Blake, D. J., Prandini, P., Brown, S. C., Torelli, S., Benson, M. A., Ponting, C. P., Estournet, B., Romero, N. B., Mercuri, E., Voit, T., Sewry, C. A., Guicheney, P., and Muntoni, F. (2001) Mutations in the fukutin-related protein gene (FKRP) cause a form of congenital muscular dystrophy with secondary laminin α2 deficiency and abnormal glycosylation of α-dystroglycan. *Am. J. Hum. Genet.* **69,** 1198–1209
22. Brockington, M., Yuva, Y., Prandini, P., Brown, S. C., Torelli, S., Benson, M. A., Herrmann, R., Anderson, L. V., Bashir, R., Burgunder, J. M., Fallet, S., Romero, N., Fardeau, M., Straub, V., Storey, G., Pollitt, C., Richard, I.,

Sewry, C. A., Bushby, K., Voit, T., Blake, D. J., and Muntoni, F. (2001) Mutations in the fukutin-related protein gene (FKRP) identify limb girdle muscular dystrophy 2I as a milder allelic variant of congenital muscular dystrophy MDC1C. *Hum. Mol. Genet.* **10,** 2851–2859

23. Longman, C., Brockington, M., Torelli, S., Jimenez-Mallebrera, C., Kennedy, C., Khalil, N., Feng, L., Saran, R. K., Voit, T., Merlini, L., Sewry, C. A., Brown, S. C., and Muntoni, F. (2003) Mutations in the human LARGE gene cause MDC1D, a novel form of congenital muscular dystrophy with severe mental retardation and abnormal glycosylation of α-dystroglycan. *Hum. Mol. Genet.* **12,** 2853–2861
24. Manya, H., Chiba, A., Yoshida, A., Wang, X., Chiba, Y., Jigami, Y., Margolis, R. U., and Endo, T. (2004) Demonstration of mammalian protein *O*-mannosyltransferase activity: co-expression of POMT1 and POMT2 required for enzymatic activity. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 500–505
25. Inamori, K., Yoshida-Moriguchi, T., Hara, Y., Anderson, M. E., Yu, L., and Campbell, K. P. (2012) Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. *Science* **335,** 93–96
26. Yoshida-Moriguchi, T., Yu, L., Stalnaker, S. H., Davis, S., Kunz, S., Madson, M., Oldstone, M. B., Schachter, H., Wells, L., and Campbell, K. P. (2010) O-mannosyl phosphorylation of α-dystroglycan is required for laminin binding. *Science* **327,** 88–92
27. Matsumoto, H., Hayashi, Y. K., Kim, D. S., Ogawa, M., Murakami, T., Noguchi, S., Nonaka, I., Nakazawa, T., Matsuo, T., Futagami, S., Campbell, K. P., and Nishino, I. (2005) Congenital muscular dystrophy with glycosylation defects of α-dystroglycan in Japan. *Neuromuscul. Disord.* **15,** 342–348
28. Yoshioka, M., Higuchi, Y., Fujii, T., Aiba, H., and Toda, T. (2008) Seizure-genotype relationship in Fukuyama-type congenital muscular dystrophy. *Brain Dev.* **30,** 59–67
29. Kanagawa, M., Nishimoto, A., Chiyonobu, T., Takeda, S., Miyagoe-Suzuki, Y., Wang, F., Fujikake, N., Taniguchi, M., Lu, Z., Tachikawa, M., Nagai, Y., Tashiro, F., Miyazaki, J., Tajima, Y., Takeda, S., Endo, T., Kobayashi, K., Campbell, K. P., and Toda, T. (2009) Residual laminin binding activity and enhanced dystroglycan glycosylation by LARGE in novel model mice to dystroglycanopathy. *Hum. Mol. Genet.* **18,** 621–631
30. Takeda, S., Kondo, M., Sasaki, J., Kurahashi, H., Kano, H., Arai, K., Misaki, K., Fukui, T., Kobayashi, K., Tachikawa, M., Imamura, M., Nakamura, Y., Shimizu, T., Murakami, T., Sunada, Y., Fujikado, T., Matsumura, K., Terashima, T., and Toda, T. (2003) Fukutin is required for maintenance of muscle integrity, cortical histiogenesis, and normal eye development. *Hum. Mol. Genet.* **12,** 1449–1459
31. Denning, G. M., Anderson, M. P., Amara, J. F., Marshall, J., Smith, A. E., and Welsh, M. J. (1992) Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* **358,** 761–764
32. Cheng, S. H., Fang, S. L., Zabner, J., Marshall, J., Piraino, S., Schiavi, S. C., Jefferson, D. M., Welsh, M. J., and Smith, A. E. (1995) Functional activation of the cystic fibrosis trafficking mutant ΔF508-CFTR by overexpression. *Am. J. Physiol.* **268,** L615–624
33. Egan, M. E., Pearson, M., Weiner, S. A., Rajendran, V., Rubin, D., Glöckner-Pagel, J., Canny, S., Du, K., Lukacs, G. L., and Caplan, M. J. (2004) Curcumin, a major constituent of turmeric, corrects cystic fibrosis defects. *Science* **304,** 600–602
34. Manya, H., Sakai, K., Kobayashi, K., Taniguchi, K., Kawakita, M., Toda, T., and Endo, T. (2003) Loss-of-function of an *N*-acetylglucosaminyltransferase, POMGnT1, in muscle-eye-brain disease. *Biochem. Biophys. Res. Commun.* **306,** 93–97
35. Akasaka-Manya, K., Manya, H., and Endo, T. (2004) Mutations of the *POMT1* gene found in patients with Walker-Warburg syndrome lead to a defect of protein *O*-mannosylation. *Biochem. Biophys. Res. Commun.* **325,** 75–79
36. Brockington, M., Torelli, S., Prandini, P., Boito, C., Dolatshad, N. F., Longman, C., Brown, S. C., and Muntoni, F. (2005) Localization and functional analysis of the LARGE family of glycosyltransferases: significance for muscular dystrophy. *Hum. Mol. Genet.* **14,** 657–665
37. Keramaris-Vrantsis, E., Lu, P. J., Doran, T., Zillmer, A., Ashar, J., Esapa, C. T., Benson, M. A., Blake, D. J., Rosenfeld, J., and Lu, Q. L. (2007) Fukutin-related protein localizes to the Golgi apparatus and mutations lead to mislocalization in muscle *in vivo*. *Muscle Nerve* **36,** 455–465
38. Kawahara, G., Guyon, J. R., Nakamura, Y., and Kunkel, L. M. (2010) Zebrafish models for human FKRP muscular dystrophies. *Hum. Mol. Genet.* **19,** 623–633
39. Dolatshad, N. F., Brockington, M., Torelli, S., Skordis, L., Wever, U., Wells, D. J., Muntoni, F., and Brown, S. C. (2005) Mutated fukutin-related protein (FKRP) localizes as wild type in differentiated muscle cells. *Exp. Cell Res.* **309,** 370–378
40. Bao, X., Kobayashi, M., Hatakeyama, S., Angata, K., Gullberg, D., Nakayama, J., Fukuda, M. N., and Fukuda, M. (2009) Tumor suppressor function of laminin-binding α-dystroglycan requires a distinct β3-*N*-acetylglucosaminyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 12109–12114
41. Hafner-Bratkovic, I., Gaspersic, J., Smid, L. M., Bresjanac, M., and Jerala, R. (2008) Curcumin binds to the α-helical intermediate and to the amyloid form of prion protein: a new mechanism for the inhibition of PrP(Sc) accumulation. *J. Neurochem.* **104,** 1553–1564
42. Ono, K., and Yamada, M. (2006) Antioxidant compounds have potent anti-fibrillogenic and fibril-destabilizing effects for α-synuclein fibrils *in vitro*. *J. Neurochem.* **97,** 105–115
43. Pey, A. L., Ying, M., Cremades, N., Velazquez-Campoy, A., Scherer, T., Thöny, B., Sancho, J., and Martinez, A. (2008) Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. *J. Clin. Invest.* **118,** 2858–2867

OPEN ACCESS Freely available online PLoS GENETICS

Comprehensive Research Synopsis and Systematic Meta-Analyses in Parkinson's Disease Genetics: The PDGene Database

Christina M. Lill[1,2,3,4], Johannes T. Roehr[1,5], Matthew B. McQueen[6], Fotini K. Kavvoura[7,8,9], Sachin Bagade[2], Brit-Maren M. Schjeide[1], Leif M. Schjeide[1], Esther Meissner[1], Ute Zauft[1], Nicole C. Allen[2], Tian Liu[1,10], Marcel Schilling[1,5], Kari J. Anderson[11], Gary Beecham[12], Daniela Berg[13,14], Joanna M. Biernacka[11], Alexis Brice[15,16,17,18], Anita L. DeStefano[19,20], Chuong B. Do[21], Nicholas Eriksson[21], Stewart A. Factor[22], Matthew J. Farrer[23], Tatiana Foroud[24], Thomas Gasser[13,14], Taye Hamza[25], John A. Hardy[26], Peter Heutink[27], Erin M. Hill-Burns[25], Christine Klein[28], Jeanne C. Latourelle[19], Demetrius M. Maraganore[29], Eden R. Martin[12], Maria Martinez[30,31], Richard H. Myers[19], Michael A. Nalls[32], Nathan Pankratz[24], Haydeh Payami[25], Wataru Satake[33], William K. Scott[12], Manu Sharma[13,14], Andrew B. Singleton[32], Kari Stefansson[34], Tatsushi Toda[33], Joyce Y. Tung[21], Jeffery Vance[12], Nick W. Wood[35,36], Cyrus P. Zabetian[37], 23andMe, The Genetic Epidemiology of Parkinson's Disease (GEO-PD) Consortium¶, The International Parkinson's Disease Genomics Consortium (IPDGC)¶, The Parkinson's Disease GWAS Consortium¶, The Wellcome Trust Case Control Consortium 2 (WTCCC2)¶, Peter Young[4], Rudolph E. Tanzi[2], Muin J. Khoury[38], Frauke Zipp[3], Hans Lehrach[1], John P. A. Ioannidis[7,39,40,41], Lars Bertram[1,2]*

1 Neuropsychiatric Genetics Group, Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany, **2** Department of Neurology, Massachusetts General Hospital, Charlestown, Massachusetts, United States of America, **3** Department of Neurology, Medical Center of the Johannes Gutenberg-University, Mainz, Germany, **4** Department of Neurology, University Hospital, Münster, Germany, **5** Department of Mathematics and Computer Science, Free University, Berlin, Germany, **6** Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, United States of America, **7** Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, **8** Centre for Diabetes and Endocrinology, Royal Berkshire Hospital, Reading, United Kingdom, **9** Oxford Centre for Diabetes, Endocrinology, and Metabolism, Churchill Hospital, University of Oxford, Oxford, United Kingdom, **10** Max Planck Institute for Human Development, Berlin, Germany, **11** Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, United States of America, **12** John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida, United States of America, **13** Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany, **14** DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany, **15** INSERM, UMR_S975, Paris, France, **16** Université Pierre et Marie Curie-Paris, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, UMR-S975, Paris, France, **17** CNRS, UMR 7225, Paris, France, **18** AP-HP, Pitié-Salpêtrière Hospital, Department of Genetics and Cytogenetics, Paris, France, **19** Department of Neurology, Boston University School of Medicine, Boston University, Boston, Massachusetts, United States of America, **20** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **21** 23andMe, Mountain View, California, United States of America, **22** Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, United States of America, **23** Department of Medical Genetics, University of British Columbia, Vancouver, Canada, **24** Indiana University School of Medicine, Indianapolis, Indiana, United States of America, **25** New York State Department of Health Wadsworth Center, Albany, New York, United States of America, **26** Department of Molecular Neuroscience, UCL Institute of Neurology, University College London, London, United Kingdom, **27** Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre, Amsterdam, The Netherlands, **28** Section of Clinical and Molecular Neurogenetics, Department of Neurology, University of Lübeck, Lübeck, Germany, **29** Department of Neurology, NorthShore University Health System, Evanston, Illinois, United States of America, **30** INSERM UMR 1043, CPTP, Toulouse, France, **31** Paul Sabatier University, Toulouse, France, **32** Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, **33** Division of Neurology/Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe, Japan, **34** deCODE genetics, Reykjavik, Iceland, **35** UCL Genetics Institute, University College London, London, United Kingdom, **36** Department of Molecular Neuroscience, UCL Institute of Neurology, University College London, London, United Kingdom, **37** VA Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, Washington, United States of America, **38** Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, **39** Biomedical Research Institute, Foundation for Research and Technology–Hellas, Ioannina, Greece, **40** Center for Genetic Epidemiology and Modeling and Tufts Clinical and Translational Science Institute, Tufts University School of Medicine, Boston, Massachusetts, United States of America, **41** Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, United States of America

Abstract

More than 800 published genetic association studies have implicated dozens of potential risk loci in Parkinson's disease (PD). To facilitate the interpretation of these findings, we have created a dedicated online resource, PDGene, that comprehensively collects and meta-analyzes all published studies in the field. A systematic literature screen of ~27,000 articles yielded 828 eligible articles from which relevant data were extracted. In addition, individual-level data from three publicly available genome-wide association studies (GWAS) were obtained and subjected to genotype imputation and analysis. Overall, we performed meta-analyses on more than seven million polymorphisms originating either from GWAS datasets and/or from smaller scale PD association studies. Meta-analyses on 147 SNPs were supplemented by unpublished GWAS data from up to 16,452 PD cases and 48,810 controls. Eleven loci showed genome-wide significant ($P<5\times10^{-8}$) association with disease risk: *BST1*, *CCDC62/HIP1R*, *DGKQ/GAK*, *GBA*, *LRRK2*, *MAPT*, *MCCC1/LAMP3*, PARK16, *SNCA*, *STK39*, and *SYT11/RAB25*. In addition, we identified novel evidence for genome-wide significant association with a polymorphism in *ITGA8* (rs7077361, OR 0.88, $P=1.3\times10^{-8}$). All meta-analysis results are freely available on a dedicated online database (www.pdgene.org), which is cross-linked with a customized track on the UCSC Genome Browser. Our study provides an exhaustive and up-to-date summary of the status of PD genetics research that can be readily scaled to include the results of future large-scale genetics projects, including next-generation sequencing studies.

Citation: Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, et al. (2012) Comprehensive Research Synopsis and Systematic Meta-Analyses in Parkinson's Disease Genetics: The PDGene Database. PLoS Genet 8(3): e1002548. doi:10.1371/journal.pgen.1002548

Editor: Amanda J. Myers, University of Miami, United States of America

Received October 16, 2011; **Accepted** January 5, 2012; **Published** March 15, 2012

Funding: The main funding for this study was provided by the Michael J. Fox Foundation for Parkinson's Disease (MJFF) with additional financial support by the Cure Alzheimer's Fund (CAF), the National Alliance for Research on Schizophrenia and Depression (NARSAD), Prize4Life, and EMD Serono (all to L Bertram). CM Lill was supported by a fellowship from the Deutscher Akademischer Austauschdienst (DAAD) and Fidelity Biosciences Research Initiative (FBRI). L Bertram is also supported by the German Ministry for Education and Research (BMBF). JPA Ioannidis was supported through the Tufts Clinical and Translational Science Institute (Tufts CTSI) under funding from the National Institute of Health/National Center for Research Resources (UL1 RR025752). Points of view or opinions in this paper are those of the authors and do not necessarily represent the official position or policies of the Tufts CTSI. M Sharma was supported by the Michael J. Fox Foundation. The NeuroGenetics Research Consortium GWAS [15] was funded by the Edmond J. Safra Michael J. Fox Foundation Global Genetics Consortium Initiative and NIH R01 NS 036960. The work of the International Parkinson's Disease Genomics Consortium (IPDGC) was supported in part by the Intramural Research Programs of the National Institute on Aging, National Institute of Neurological Disorders and Stroke, National Institute of Environmental Health Sciences, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services: project numbers Z01 AG000949-02 and Z01-ES101986. In addition the work of the IPDGC was supported by the U.S. Department of Defense, award number W81XWH-09-2-0128. Portions of the work of the IPDGC utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). T Foroud received funds from the National Institutes of Health (R01CA141668 and R01NS37167). C Klein is the recipient of a career development award from the Volkswagen Foundation and from the Hermann and Lilly Schilling Foundation. DM Maraganore acknowledges active funding support from the National Institutes of Health (2R01 ES10751), Alnylam Pharmaceuticals, Medtronic, and NorthShore University Health System. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: CB Do, N Eriksson, and JY Tung are employed by 23andMe and own stock options in the company. MJ Farrer and Mayo Foundation received royalties from H.Lundbeck A/S and Isis Pharmaceuticals. In addition, MJ Farrer has received an honorarium for a seminar at Genzyme. T Gasser has received consultancy fees from Cephalon and Merck-Serono, grants from Novartis, payments for lectures including service on speakers' bureaus from Boehringer Ingelheim, Merck-Serono, UCB, and Valean, and holds patents NGFN2 and KASPP. JA Hardy has received consulting fees or honoraria from Eisai and his institute has received consulting fees or honoraria from Merck-Serono. DM Maraganore has received extramural research funding support from the National Institutes of Health (2R01 ES10751), the Michael J. Fox Foundation (Linked Efforts to Accelerate Parkinson Solutions Award, Edmond J. Safra Global Genetics Consortia Award), and from Alnylam Pharmaceuticals and Medtronic (observational studies of Parkinson's disease). DM Maraganore has also received intramural research funding support from the Mayo Clinic and from NorthShore University Health System. DM Maraganore filed a provisional patent for a method to predict Parkinson's disease. This provisional patent is unlicensed. He also filed a provisional patent for a method to treat neurodegenerative disorders. That provisional patent has been licensed to Alnylam Pharmaceuticals and DM Maraganore has received royalty payments in total of less than $20,000. K Stefansson has received grants from deCODE.

* E-mail: lbertram@molgen.mpg.de

¶ Memberships of the consortia are provided in Text S1.

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease with a prevalence of ~1% over 60 years of age [1]. Approximately 5–10% of the patients show an autosomal dominant or recessive mode of inheritance, and several causative genes have been identified, e.g. *SNCA*, *LRRK2*, *PARK2*, and *PINK1* (for review see ref. [2]). Recently, two other novel autosomal dominant PD genes, *VPS35* and *EIF4G1* [3–5], have been identified, the former via application of next-generation sequencing techniques. It can be anticipated that causal mutations in additional genes will emerge within the next years. However, the vast majority of patients suffer from non-Mendelian forms of PD, which are likely caused by the combined effects of genetic and environmental factors. In order to decipher the genetic architecture underlying PD susceptibility, more than 800 genetic association studies have been performed over the past 20 years. While early candidate gene studies and subsequent meta-analyses provided conclusive evidence showing that polymorphisms in *SNCA* [6] (encoding alpha-synuclein), *LRRK2* [7] (leucine-rich repeat kinase 2), *MAPT* [8] (microtubule-associated protein tau), and *GBA* [9] (acid beta-glucosidase) significantly impact PD susceptibility, most association studies in the field provided inconclusive or even conflicting results.

During the last few years, genome-wide association studies (GWAS) [10–19] have postulated additional PD loci. While the early GWAS and a GWAS-meta-analysis [20] were of limited sample sizes and yielded mostly inconsistent results, more recent studies have identified a number of loci that were independently confirmed in follow-up studies (e.g. *GAK*, *BST1*, and PARK16, see Table 1 for all proposed GWAS findings across GWAS publications). Very recently, a GWAS meta-analysis [21] implicated several other new putative PD loci which currently await further validation. Despite this progress, approximately 40% or more of the population-attributable risk probably remains unexplained by today's most promising PD loci [21]. To this end, genetic association studies remain one of the mainstays of PD genetics research. However, GWAS and other large-scale association studies typically only highlight the most promising results and often do not provide data on variants showing suggestive evidence for association, or previously implied variants that could not be confirmed in the GWAS setting. As a result, the cumulative genetic evidence in favor of or against association with certain variants in the PD field is becoming increasingly difficult to follow, evaluate and interpret. To address this problem, we have comprehensively collected, catalogued and systematically meta-analyzed the data from all genetic association studies published in the field of non-Mendelian PD, including GWAS, and made all results publicly available on a regularly updated online database, "PDGene" (http://www.pdgene.org).

Results

Database content

The results of this research synopsis are based on a freeze of the PDGene database content on March 31st 2011 (available upon request from the authors). At that time, PDGene included details on 828 individual studies across more than 50 different countries and six continents reporting on 3,382 polymorphisms in 890 genetic loci. Data for more than 2,000 SNPs were supplemented by results derived from up to three publicly available GWAS datasets [10,12,13] following extensive quality control and imputation. Ultimately, this procedure yielded a total of 867 polymorphisms across ~300 genetic loci that met our criteria for meta-analysis (see Methods). Additional independent GWAS data

Author Summary

The genetic basis of Parkinson's disease is complex, i.e. it is determined by a number of different disease-causing and disease-predisposing genes. Especially the latter have proven difficult to find, evidenced by more than 800 published genetic association studies, typically showing discrepant results. To facilitate the interpretation of this large and continuously increasing body of data, we have created a freely available online database ("PDGene": http://www.pdgene.org) which provides an exhaustive account of all published genetic association studies in PD. One particularly useful feature is the calculation and display of up-to-date summary statistics of published data for overlapping DNA sequence variants (polymorphisms). These meta-analyses revealed eleven gene loci that showed a statistically very significant ($P<5\times10^{-8}$; a.k.a. genome-wide significance) association with risk for PD: *BST1*, *CCDC62/HIP1R*, *DGKQ/GAK*, *GBA*, *LRRK2*, *MAPT*, *MCCC1/LAMP3*, PARK16, *SNCA*, *STK39*, *SYT11/RAB25*. In addition and purely by data-mining, we identified one novel PD susceptibility locus in a gene called *ITGA8* (rs7077361, $P=1.3\times10^{-8}$). We note that our continuously updated database represents the most comprehensive research synopsis of genetic association studies in PD to date. In addition to vastly facilitating the work of other PD geneticists, our approach may serve as a valuable example for other complex diseases.

for 147 SNPs yielding P values of ≤ 0.1 in these initial meta-analyses were provided by researchers of all remaining currently published Caucasian GWAS datasets [13,15–19,22]. Following the identification of genome-wide significant association with an intronic SNP (rs7077361) in *ITGA8* after addition of these data, we obtained additional data from the same GWAS datasets on ~1,400 SNPs in the chromosomal region encompassing *ITGA8* (chr10:15346353–15801533, hg18). Finally, independent replication data in Caucasian and Asian populations from the GEO-PD consortium [23] generated for ten recently described PD loci [21] were made available for inclusion. As a result, we were able to substantially increase the sample size (up to 16,452 PD cases and 48,810 controls) for a large number of some of the most promising PD loci. For instance, we were able to add data from up to 48,861 previously not analyzed combined cases and controls to meta-analyses of some of the recently proposed PD loci [21] (median sample size 14,896, see Table 2 and Table S1 for details). In addition to these focused analyses, PDGene displays meta-analysis results for more than seven million additional SNPs originating from up to three publicly available GWAS datasets [10,12,13]. The results are available online (e.g. as summarized in http://www.pdgene.org/largescalemeta.asp), where they are cross-linked to a customized and fully browsable track on the UCSC Genome Browser.

PDGene meta-analysis results

The PDGene meta-analyses of the 867 core polymorphisms were based on a median of 7,680 subjects (interquartile range 4,612–16,726). Additional meta-analyses were performed after stratification for Caucasian and Asian ancestry (for details on sample size and included ethnicities for individual meta-analyses see Table S1). In addition, we also performed random-effects meta-analyses across all three publicly available GWAS datasets [10,12,13] following genotype imputation using data from the International HapMap Consortium and 1000 Genomes Project. Ultimately this yielded 7,123,920 SNPs that could be meta-analyzed across at least two GWAS datasets (see Figure S1 for a quantile-to-quantile plot of the GWAS-only meta-analyses). All 867 core meta-analysis results are available online on PDGene as forest plots, summarizing the relative contributions of each dataset to the most current summary effect estimate, and in the form of cumulative plots, illustrating how summary ORs evolve over time. All meta-analysis results are plotted in Figure 1 (green dots) alongside the GWAS-only meta-analysis results (black and grey dots).

One-hundred-three meta-analyses across 12 genetic loci (*BST1*, *CCDC62/HIP1R*, *DGKQ/GAK*, *GBA*, *ITGA8*, *LRRK2*, *MAPT*, *MCCC1/LAMP3*, PARK16, *SNCA*, *STK39*, *SYT11/RAB25*) yielded summary ORs suggesting a genome-wide significant ($P\leq5\times10^{-8}$) increase or decrease in PD risk in all ethnicities and/or after stratification for ethnic ancestry (Table 2, Table S1, and Figure S2 [forest plots]). None of these loci contained more than one SNP independently associated at genome-wide significance (as judged by pair-wise linkage disequilibrium assessments using 'SNAP' and r^2-values of 0.2 as cut off http://www.broadinstitute.org/mpg/snap/). The majority of polymorphisms tested in the genome-wide significant loci do not show evidence for publication bias (Table S1). Finally, all genome-wide significant signals were robust against potential undetected sample overlap using a recently proposed procedure [24] (see Table S2 for more details). Combined sample sizes for all 12 loci were substantially larger here as compared to any previously published meta-analysis (Table S1), providing unequivocal evidence for an involvement of these loci in PD susceptibility. While power to detect genome-wide significance was excellent for most of these loci (>80% based on an OR of 1.15, and a minor allele frequency down to 0.05 using the Genetic Power Calculator, http://pngu.mgh.harvard.edu/~purcell/gpc/), power was less for a large number of other meta-analyses due to smaller sample sizes and allele frequencies (see Table S1 for details). Thus, no simple statistic can summarize the overall power of our study.

The above list includes an intronic polymorphism in *ITGA8* located on chromosome 10p13 for which we identified novel evidence for genome-wide association with PD risk (OR 0.88, $P=1.3\times10^{-8}$, $I^2=0$, see Table 2, and Figure 2). This SNP had previously been proposed to be associated with PD risk at sub-genome-wide significance by Simon-Sanchez et al [13]. After obtaining and meta-analyzing GWAS data from ~1,400 additional SNPs in this region derived from all Caucasians GWAS datasets [10,12,13,15–19,21,22], rs7077361 remained the most significantly associated SNP in this region (Figure S3).

In addition to using random-effects models, we also performed exploratory fixed-effect meta-analyses on all eligible polymorphisms. These analyses did not reveal genome-wide significant effect sizes for any additional locus, except *ACMSD/TMEM163* (most significant SNP rs6723108, OR 0.91, $P=1.3\times10^{-9}$, $I^2=46\%$ [95% CI 0–73%], Figure S4, panel 1) and *HLA* (most significant SNP chr6:32609909, OR 0.78, $P=8.8\times10^{-15}$, $I^2=84\%$ [95% CI 70–91%], Figure S4, panel 2), both of which were reported to be associated with PD risk at genome-wide significance in previous work [16,21]. In both instances, the lack of genome-wide significance in the random-effects models (Table S1) was due to relatively pronounced heterogeneity of effect estimates across studies. However, the heterogeneity across the 11 datasets in the *ACMSD/TMEM163* meta-analysis is almost entirely due to variance of effect size estimates in the same direction (see Figure S4, panel 1), making it likely that *ACMSD/TMEM163* represents a genuine PD risk locus. For the SNP tested in the *HLA* locus (chr6:32609909, Figure S4, panel 2), heterogeneity is more pronounced and more complex owing to ORs on either side of

Table 1. Overview of genome-wide association studies (GWAS) published in PD until March 31, 2011.

GWAS	Design GWAS (Follow-up)	Population GWAS (Follow-up)	# SNPs	# PD GWAS (Follow-up)	# CTRL GWAS (Follow-up)	"Featured" genetic loci
Maraganore, 2005 (ref. 9)	Family-based (case-control)	USA-LEAPS (USA)	198,345	443 (332)	443 (332)	*CDCP2, GALNT3, GWA_2q36.3, GWA_4q28.1, GWA_4q28.3, GWA_5p15.32, GWA_7p14.2, GWA_10q21.1, PASD1, PRDM2, SEMA5A*
Fung, 2006 (ref. 10)	Case-control (-)	USA-NINDS	408,803	267 (-)	270 (-)	*BRDG, DLG2, GLT25D2, GWA_4q35.2, GWA_7p12, GWA_10q11.21, GWA_11q11, GWA_16q23.1, GWA_22q13, NEGR1, ULK2, ZNF313*
Pankratz, 2009 (ref. 11)	Case-control (-)	USA-PROGENI/GenePD (-)	328,189	857 (-)	867 (-)	***DGKQ/GAK, GPRIN3, MAPT, SNCA***
Simon-Sanchez, 2009 (ref. 12)	Case-control (case-control)	USA-NINDS, Germany (USA, Germany, UK)	463,185	1,745 (3,452)	4,047 (4,756)	***LRRK2, MAPT,* PARK16, *SNCA***
Satake, 2009 (ref. 13)	Case-control (case-control)	Japan (Japan)	435,470	1,078 (993)	2,628 (15,753)	***BST1, LRRK2,* PARK16, *SNCA***
Edwards, 2010 (ref. 14)	Case-control (-)	USA-HIHG (-)	491,376	604 (-)	619 (-)	***MAPT, SNCA***
Hamza, 2010 (ref. 15)	Case-control (-)	USA-NGRC (-)	811,597	2,000 (-)	1,986 (-)	***GAK/DGKQ,* HLA locus, *MAPT, SNCA***
Spencer, 2011 (ref. 16)	Case-control (case-control)	UK-WTCCC2 (France)	1,733,533	1,705 (1,039)	5,175 (1,984)	***BST1, GAK/DGKQ, MAPT,* PARK16, *SNCA***
Saad, 2011 (ref. 17)	Case-control (case-control)	France (UK-WTCCC2, Australia)	492,929	1,039 (3,232)	1,984 (7,064)	***BST1,* GWA_12q24, *SNCA***
Simon-Sanchez, 2011 (ref. 18)	Case-control (case-control)	Netherlands	514,799	772 (-)	2024 (-)	***BST1,* HLA locus, *GAK/DGKQ, MAPT, SNCA***

The overview is based on content on the PDGene website (http://www.pdgene.org; current on March 31st, 2011). Studies are listed in order of publication date. '# PD GWAS' and '# CTRL GWAS' refers to sample sizes used in the initial GWAS datasets, whereas 'Follow-up' refers to the total number of replication samples where applicable. 'Featured genes' are those genes/loci that were declared as 'associated' in the original publication; note that criteria for declaring association varies across studies. Genetic loci in bold font denote genes showing genome-wide significant results ($P<5\times10^{-8}$) in the PDGene meta-analyses.
doi:10.1371/journal.pgen.1002548.t001

1. This could be due to a number of reasons, e.g. subtle and uncorrected population substructure and/or different LD patterns between the analyzed SNP and the actual functional variant(s) [16]. Thus, although the evidence is currently not as conclusive as for *ACMSD/TMEM163* it still appears quite possible that there is one or more PD association signals in the *HLA* region. Regardless of these considerations, additional data are needed to more firmly assess the role of both loci in contributing to PD susceptibility.

Ethnicity-specific meta-analysis results

SNCA, *LRRK2*, *BST1*, and PARK16 show evidence for genome-wide significance in meta-analyses restricted to Caucasian and Asian populations (Table 2). Furthermore, data obtained from the GEO-PD consortium [23] suggest that the effect estimates for some of the recently discovered PD loci (i.e. *CCDC62/HIP1R*, *MCC1*, and *STK39*) [21] may be comparable in Caucasian and Asian populations (Table S1), although additional datasets are needed to establish genome-wide significance in populations of Asian-descent for these loci. Conversely, only insufficient data are currently available to assess the effect sizes of *GAK* and *SYT11/RAB25* on PD risk in Asians: *GAK* rs6599388 violated Hardy-Weinberg equilibrium in Asian datasets from the GEO-PD consortium and was thus excluded from further analyses on that ethnic group [23]. *SYT11/RAB25* chr1:154105678 was excluded from all analyses due to technical reasons in the study by the GEO-PD consortium [23]. Moreover, none of the reported *SYT11/RAB25* and *GAK* SNPs from the recent GWAS meta-analysis [21] were captured directly or by proxy (with an $r^2\geq0.8$) in the Japanese GWAS dataset [14,23]. Finally, Asian-descent populations cannot be appropriately assessed for PD association with the *MAPT*-H1/H2 haplotype, rs10928513 in *ACMSD*, and rs7077361 in *ITGA8* owing to monomorphicity at these sites [14,23].

Evaluating the credibility of significant associations

To estimate the epidemiologic credibility of associations with polymorphisms showing sub-genome-wide significant association with PD ($P>5\times10^{-8}$), we applied two "credibility" measures for each such result. First, we calculated Bayes factors (BF, expressed here as $\log_{10}$-values, "logBF") assuming an average non-null odds ratio of 1.15, as approximation of a typical "complex disease effect size", and a spike and smear prior distribution of effects [25]. Our

Table 2. Genome-wide significant summary meta-analysis results of the PDGene database in populations of Caucasian and Asian decent.

Caucasian ethnicity

Locus	Polymorphism	Location (hg18)	MAF	Allele contrast	N datasets	N samples	OR (95% CI)	*P*-value	I^2 (95% CI)	HuGENet	BF
GBA	N370S	chr1:153451576	0.01	G vs. A	15	44,851	3.51 (2.55–4.83)	1.44×10^{-14}	38 (0–66)	A	6.6
SYT11/RAB25	chr1:154105678	chr1:154105678	0.02	T vs. C	6	17,300	1.73 (1.48–2.02)	2.35×10^{-12}	0 (0–52)	B*	8.2
PARK16	rs947211	chr1:204019288	0.23	A vs. G	12	69,262	0.91 (0.88–0.94)	8.00×10^{-10}	0 (0–66)	A	6.8
STK39	rs2390669	chr2:168800188	0.13	C vs. A	14	35,159	1.19 (1.12–1.25)	1.37×10^{-09}	18 (0–56)	A	4.9*
MCCC1/LAMP3	rs11711441	chr3:184303969	0.14	A vs. G	25	46,502	0.86 (0.82–0.91)	9.20×10^{-10}	18 (0–50)	A	6.8
DGKQ	rs11248060	chr4:954359	0.12	T vs. C	10	57,716	1.21 (1.15–1.27)	3.04×10^{-12}	11 (0–52)	A	9.2
BST1	rs11724635	chr4:15346199	0.43	C vs. A	26	46,586	0.88 (0.84–0.91)	1.87×10^{-10}	43 (10–64)	A	7.5
SNCA	rs356219	chr4:90856624	0.41	G vs. A	31	79,494	1.29 (1.25–1.33)	6.06×10^{-65}	16 (0–46)	A	61.0
ITGA8	rs7077361	chr10:15601549	0.12	C vs. T	11	61,036	0.88 (0.84–0.92)	1.51×10^{-08}	0 (0–55)	A	5.7
LRRK2	rs1491942	chr12:38907075	0.21	G vs. C	21	34,123	1.17 (1.13–1.22)	6.44×10^{-15}	0 (0–38)	A	11.8
CCDC62/HIP1R	rs10847864	chr12:121892551	0.39	T vs. G	23	38,367	1.15 (1.11–1.18)	4.37×10^{-17}	0 (0–35)	A	14.4
MAPT/STH	H1H2	chr17:42131818–41149582	0.20	H2 vs. H1	37	50,389	0.78 (0.75–0.80)	7.97×10^{-52}	0 (0–29)	A	48.1

Asian ethnicity

Locus	Polymorphism	Location (hg18)	MAF	Allele contrast	N datasets	N samples	OR (95% CI)	*P*-value	I^2 (95% CI)	HuGENet	BF
PARK16	rs823156	chr1:204031263	0.17	G vs. A	5	22,870	0.74 (0.68–0.81)	2.09×10^{-12}	0 (0–58)	A	9.2
BST1	rs4538475	chr4:15347035	0.38	G vs. A	3	20,393	0.80 (0.75–0.86)	9.53×10^{-10}	0 (-)	A	6.8
SNCA	rs6532194	chr4:90999925	0.40	T vs. C	5	22,844	1.29 (1.20–1.39)	4.91×10^{-11}	31 (0–74)	A	8.0
LRRK2	rs34778348	chr12:39043595	0.04	A vs. G	13	10,441	2.23 (1.89–2.63)	2.97×10^{-21}	0 (0–53)	B*	15.2

Whenever multiple polymorphisms showed genome-wide significant association in the same locus, only the variant with the smallest *P*-value is listed here. Note that, overall, 103 PDGene meta-analyses results across the 12 loci listed above yield genome-wide significant evidence for association with PD. For a complete list of these as well as the non-genome-wide significant meta-analysis results performed for the datafreeze, see Table S1. MAF = minor allele frequency in cases and controls combined; N = Number, OR = Odds Ratio; CI = confidence interval; I^2 = estimate of percentage of between-study heterogeneity that is beyond chance. BF = Bayes factor. *Note that additional polymorphisms in these loci showing genome-wide significant association with PD yield are graded with "strong epidemiologic credibility" (grade A) according to the HuGENet criteria [26,27], and a Bayes Factor >5 [25], respectively (see Table S1 for more details).

doi:10.1371/journal.pgen.1002548.t002

second assessment was based on the Human Genome Epidemiology Network's (HuGENet) interim criteria for the assessment of cumulative epidemiologic evidence in genetic association studies [26,27]. The results of these analyses are summarized in Table S1.

There was strong epidemiologic support in both assessments for all loci showing genome-wide significant association. This included several additional polymorphisms in these same loci that only showed sub-genome-wide significant association. However, there was no additional sub-genome-wide significantly associated locus that received unequivocally strong support from both credibility assessments (Table S1). In this list, the strongest support was assigned to SNP chr6:32588205 in the *HLA* locus receiving the best possible grade in the HuGENet criteria (grade A), but more moderate support in the Bayesian analyses (logBF = 4.4). However, the relevance of this assessment needs to be evaluated as the underlying analysis was only based on four GWAS datasets.

Discussion

The PDGene database represents a comprehensive, regularly updated and freely available online research synopsis of genetic association studies in PD. Detailed summaries of the most compelling findings are provided within an easy-to-use, dedicated online framework, displaying forest plots, cumulative meta-analyses, and an up-to-date ranking of "Top Results". To allow comparison of PDGene results with association findings from other complex diseases and to facilitate their interpretation with respect to functional genetics data, all meta-analysis results have been ported as a customized track onto the UCSC Genome Browser. This will also allow for a integration and visualization [28] of association results from large-scale resequencing data (e.g. from whole-exome or whole-genome studies) into PDGene once these become available.

To the best of our knowledge, our study represents the most comprehensive research synopsis in the field of PD genetics. In addition, it represents the first disease-specific genetic database that allows a systematic and exhaustive inclusion of GWAS data, and may serve as a model for similar databases in other complex genetic diseases. Owing to our multi-pronged data retrieval and analysis protocol we were able to perform meta-analyses on the vast majority of PD risk-gene candidates, including those "featured" as top association results in all published GWAS. In particular, this includes the five novel loci recently featured in the recent GWAS meta-analysis [21]. Through collaboration with other PD genetics laboratories we obtained independent summary data for these and 142 additional SNPs, substantially extending the hitherto available evidence. Taken together, our analyses provide unequivocal evidence that *BST1*, *CCDC62/HIP1R*, *DGKQ/GAK*, *GBA*, *ITGA8*, *LRRK2*, *MAPT*, *MCCC1/LAMP3*, PARK16, *SNCA*, *STK39*, *SYT11/RAB25* represent genuine PD risk loci, while the role of several other loci (e.g. *ACMSD/*

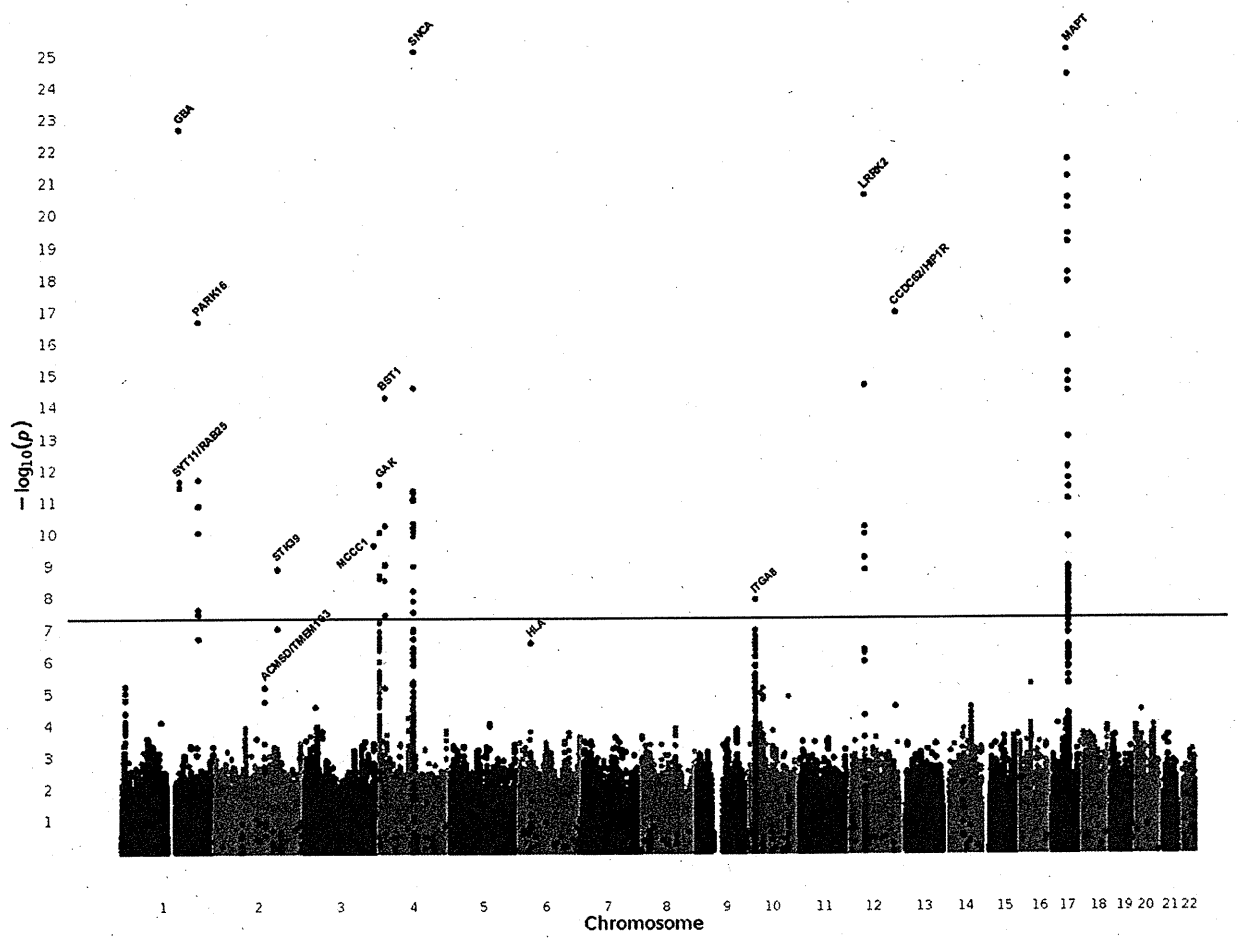


Figure 1. Manhattan plot of all meta-analysis results performed in PDGene. This summary combines association results from 7,123,986 random-effects meta-analyses based on the March 31st 2011 datafreeze of the PDGene database. Results are plotted as $-\log_{10}$ *P*-values (y-axis) against physical chromosomal location (x-axis). Black and grey dots indicate results originating exclusively from the three fully publicly available GWAS datasets [10,12,13] (see Methods), while green dots are based on a combination of smaller scale studies, supplemented by GWAS datasets (where applicable). Gene annotations are provided for genes highlighted in the main text.
doi:10.1371/journal.pgen.1002548.g001

TMEM163, and the *HLA* locus) remains to be determined. The unpublished data aggregated here from various PD genetics groups for selected candidate genes represents the first step towards a systematic meta-analysis across the full GWAS datasets from the same populations. Once completed, the results of this "mega" meta-analysis will be posted on the PDGene database, allowing users to browse the complete results via the customized genome browser track already in place.

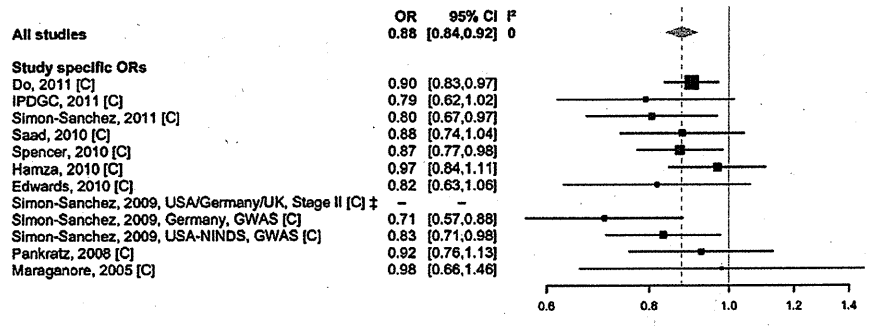


Figure 2. Forest plot of the meta-analysis of rs7077361 in *ITGA8*. Study-specific allelic odds ratios (ORs, black squares) and 95% confidence intervals (CIs, lines) were calculated for each included dataset. The summary OR and CI was calculated using the DerSimonian Laird random-effects model (grey diamond) [31]. C = Caucasian ancestry.
doi:10.1371/journal.pgen.1002548.g002

Of particular interest are loci with unusually large effect sizes. While most loci in PDGene have only small effects on PD risk (with ORs ranging from 1.10 to 1.35, which are typical for complex diseases), for some loci much larger ORs were estimated (i.e. *GBA* [OR 3.51 in Caucasians], *LRRK2* [OR 2.23 in Asians], and *SYT11/RAB25* [OR 1.73 in Caucasians], see Table 2). The risk-allele frequencies at these polymorphisms are typically rather small (i.e. below 0.05), resulting in low population attributable risks for these loci (for the above mentioned loci individually less than 2%).

Interestingly, the meta-analysis results of *GBA* N370S as well as the *LRRK2* rs34778348 are solely based on candidate-gene approaches since these SNPs are not on any of the current GWAS arrays or imputation reference panels. Thus, even in the "GWAS era" smaller-scale, non-GWAS but "focused" genetic studies, will likely continue to play an important role. This is also true when it comes to providing independent replication of proposed disease associations and/or when validating imputation-derived results by direct genotyping in sufficiently sized datasets. PDGene systematically concatenates all these different types of data into one database framework, vastly facilitating an assessment of the overall evidence for any given SNP or locus.

The strength of our approach is further exemplified by the identification of genome-wide significant association between disease risk and a SNP in *ITGA8*, which was not featured as a relevant PD gene in any previous study. *ITGA8* (encoding integrin alpha 8, a type-I transmembrane protein) is functionally interesting as it is expressed in brain [29], mediates cell-cell interactions and regulates neurite outgrowth of sensory and motor neurons [30]. Additional studies are needed to further assess the potential role of this gene in PD pathogenesis. Furthermore, PDGene shows that two additional loci, not highlighted by the recent GWAS meta-analysis [21], yield genome-wide signficiant results in the PDGene meta-analyses, i.e. PARK16, originally implicated as a PD susceptibility locus in an Asian GWAS [14] but not highlighted in the recent GWAS meta-analysis on Caucasian samples [21] and *GBA*, a gene that was found soley by candidate-gene approaches. Another strength of our study is that it combines genetic data from currently more than 50 different countries allowing a systematic assessment of genetic associations across populations of different ethnic descent. For instance, these analyses suggest that variants in *BST1*, *LRRK2*, the PARK16 locus, and *SNCA* show genome-wide significant association with PD risk in both Caucasian and Asian-descent samples. Furthermore, the recently described Caucasian GWAS loci *CCDC62/HIP1R*, *MCC1*, and *STK39* [21] also show similar effect size estimates in populations of Asian-descent [23]. PD association data originating from other ethnic groups are still relatively scarce. However, they could easily be added to the already existing data on the respective polymorphisms available on PDGene.

In summary, we have created a continuously updated online resource for genetic association studies in the field of PD. Synthesizing essentially all available data in the field led to the identification of *ITGA8* as a novel potential PD risk locus. Our quantitative approach to data integration across a multitude of different study designs can be readily scaled to include large-scale resequencing efforts that will emerge over the coming years, making the complex field of PD genetics accessible to a broad range of investigators.

Methods

Note that the following section only provides a brief summary of the methods applied to our study. A much more detailed description can be found in Text S1.

Literature searches

Inclusion criteria. For inclusion in PDGene, a study has to meet three criteria: 1) It must evaluate the association between a bi-allelic genetic polymorphism (minor allele frequency $\geq$0.01 in the healthy control population of at least one study) and Parkinson's disease (PD) risk in datasets comprised of both affected (defined as clinically and/or neuropathologically diagnosed "Parkinson's disease") and unaffected individuals; 2) it must be published in a peer-reviewed journal; 3) it must be published in English. For this manuscript, we also included data on ten SNPs generated in the GEO-PD Consortium datasets [14,23] and obtained data for the newly identified SNP rs7077361 in *ITGA8* from the Japanese GWAS dataset [14].

Exclusion criteria. In brief, genetic association data of the following studies were excluded from the meta-analyses (see Text S1 for details): family-based studies without available subject-level data (however, unrelated case-control data enriched for familial cases were not excluded), studies investigating only disease controls, multi-allelic polymorphisms, and studies of polymorphisms in mitochondrial DNA. We also excluded genetic data of apparently "poor" quality if discrepancies could not be resolved after contacting the study authors (e.g. inadequate genotyping/sequencing protocols or discrepancies in terms of allele names or frequencies when compared with public databases; more details can be found in Text S1).

Search strategies. Our literature searches until March 31st, 2011, yielded 27,210 articles, which were screened for eligibility using the title, abstract, or full-papers, as necessary. Additional screening of bibliographies in reviews, published meta-analyses, and original genetic association studies were also performed. Overall, full text versions of 1,534 articles were obtained. Following the inclusion and exclusion criteria outlined above, 828 articles were included in PDGene until March 31st 2011 (also see Figure 3).

Statistical analyses

Meta-analyses. Random-effects allelic meta-analyses [31] were performed if a minimum of four independent datasets existed per polymorphism. Summary odds ratios [ORs] and 95% confidence intervals [CIs] were calculated irrespective of ethnic descent as well as for distinct ethnic groups (i.e. Caucasians, and Asians) if sufficient data were available. In addition, we performed a number of sensitivity analyses (excluding the initial studies and datasets in which HWE was violated in control individuals), systematically assessed between-study heterogeneity (via I^2), and assessed the credibility of each at least nominally significant meta-analysis result by calculating Bayes factors (BF; here expressed as log10(BF) = "logBF") [25] and by determining a grading score developed by the Human Genome Epidemiology Network (HuGENet) [26,27].

Assessment of small-study bias/publication bias. This is of particular importance in meta-analyses of published association data and was carefully addressed here: First, we added *publicly* available GWAS data [10,12,13] to the vast majority of SNPs. Since these data are typically unbiased, this should decrease the potential for small-study bias/publication bias. Secondly, for 147 SNPs of the core PDGene meta-analyses that showed statistically suggestive results ($P \leq 0.1$), we obtained additional data from all currently published, but *not publicly* available GWAS datasets, further decreasing a potential impact of small-study bias/publication bias. Thirdly, we directly assessed the evidence for small study bias by applying a recently proposed regression test [32] on all nominally significant ($P<0.05$) meta-analysis results. The results of these analyses are fully displayed in Table S1.

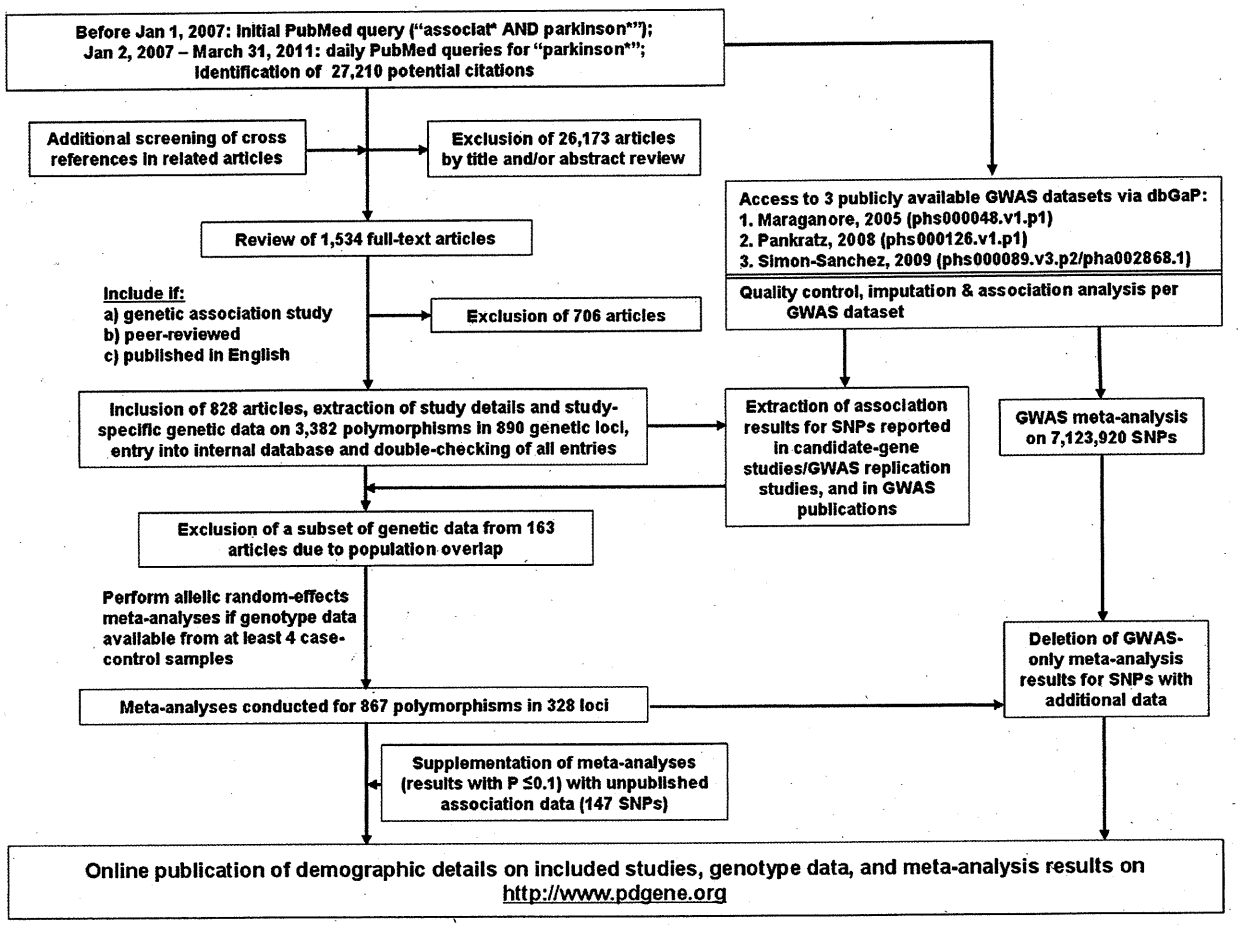


Figure 3. Flowchart of literature search, data extraction, and analysis strategies applied for PDGene.
doi:10.1371/journal.pgen.1002548.g003

GWAS-only meta-analyses. We obtained individual-level genotype data for all publicly available PD GWAS datasets from NCBI's "dbGAP" database (a total of three [10,12,13] at the time of the datafreeze, March 31st, 2011). Genotype data were cleaned using standard procedures, followed by imputation of untested genotypes (using reference panels from HapMap and the 1000 Genomes Project), and association analyses incorporating imputation uncertainty (case-control datasets only), age, sex, and population stratification. Overall, this procedure led to a total of 7,723,931 unique SNPs, 7,123,920 of which were present in at least two, and 711,271 in at least three datasets. Meta-analyses (either combining test-statistics and standard errors using random-effects models, or by combining P-values weighted by sample size, see Text S1 for more details) were performed on the 7,123,920 SNPs present in at least two of the GWAS datasets.

Online database

After completion of all data-management and analysis steps, all study-specific variables, genotype data (except for GWAS), and meta-analysis plots are posted on a dedicated, publicly available, online adaptation of the PDGene database using the same software and code as our databases for Alzheimer's disease [33] and schizophrenia [34]. The online database is hosted by the "Alzheimer Research Forum" and can be accessed via its own designated URL (http://www.pdgene.org).

Database code

The database software can easily be ported to other genetically complex diseases and will be made available on a collaborative basis to interested researchers upon request.

Supporting Information

Figure S1 QQ plots showing the distribution of expected versus observed P-values for the GWAS-only meta-analysis results. Analyses were performed using the METAL software (ref. [21] in Text S1). The excess of observed *P*-values (Figure S1, panel 1) is entirely due to association signals in the *SNCA*, *MAPT*, *LRRK2*, and *DGKQ/GAK* loci as can be seen in Figure S1, panel 2 that showcases the *P*-value distributions after removal of 18,622 SNPs in these regions (lambda = 1.007).
(TIF)

Figure S2 Forest plots of allelic meta-analyses for SNPs showing genome-wide significant association ($P<5\times10-8$) with PD susceptibility in the March 31st 2011 datafreeze. Study-specific allelic odds ratios (ORs, black squares) and 95% confidence intervals (CIs, lines) were calculated for each included dataset. The summary OR and CI was calculated using random-effects models (grey diamond). Whenever multiple polymorphisms showed genome-wide significant association in the same locus, only the variant with the smallest P-value is listed here for meta-

analysis results after stratification for Caucasian and Asian ancestries. For a complete list of meta-analyses performed for the datafreeze, see Table S1. Figure S1, panel 1-S1, panel 12 and S1, panel 13-S1, panel 16 display the SNP showing the most significant genome-wide association in datasets of Caucasian ancestry and Asian ancestry, respectively. Details and references of all included studies displayed here can be found on the PDGene database (http://www.pdgene.org). I2 = estimate of percentage of between-study heterogeneity that is beyond chance, "excl initial" = summary OR and 95%CI after meta-analysis after exclusion of the initial study, C = Caucasian ancestry, A = Asian ancestry, H = Hispanic descent, D = African descent, "•" = initial study (applies to candidate-gene studies), "†" = no data provided or data was not eligible for inclusion in meta-analysis, "‡" = study excluded due to overlap, "#" = HWE violation in controls ($P<0.05$, not applicable to quality-controlled GWAS datasets, see Text S1), "i" = SNP monomorphic in the respective dataset, "ø" = meta-analysis after excluding initial study not applicable.
(PDF)

Figure S3 Locus plot of the ITGA8 region on chromosome 10p13 (15346353–15801533 bp, hg18). The figure displays association results for ~1,400 SNPs in the *ITGA8* region including at least four independent datasets. SNPs are color-coded based on linkage disequilibrium (r^2) estimates from the CEU 1000G dataset (release June 2010). All LD estimates refer to the most significantly associated SNP rs7077361. SNPs color-coded in grey indicate missing LD estimates in the CEU dataset. Recombination rates were estimated based on the CEU dataset, and are displayed as blue line in the background. Gene annotations are based on RefSeq and the UCSC Genome browser. Locus plots were generated using the LocusZoom Stand-alone package (http://genome.sph.umich.edu/wiki/LocusZoom_Standalone).
(TIF)

Figure S4 Forest plots of fixed-effect meta-analyses for SNP rs6723108 in the ACMSD/TMEM163 locus and chr6:32609909 in the HLA locus. Symbols are the same as for Figure S2 (see above).
(TIF)

Table S1 Overview of all 867 polymorphisms meta-analyzed in the March 31st 2011 datafreeze using random-effects allelic models. Random-effects allelic meta-analyses were performed on polymorphisms for which four or more independent datasets were available. Meta-analyses after stratification for different ethnic descent were performed if at least three independent datasets were available in the respective stratum (applicable only to samples of European and Asian descent). Each nominally significant meta-analysis result ($P<0.05$) was graded according to the HuGENet interim criteria. For details on how these criteria are applied, see Text S1. Meta-analysis results in this table are ordered by genomic location. OR = Odds Ratio, CI = confidence interval, N minor = number of minor alleles, Ethnicities: C = Caucasian, A = Asian, D = African Descent, H = Hispanic, O = Other/Mixed, Low OR = OR<1.15 or ≥0.87, respectively, F = loss of significance in the respective meta-analysis after exclusion of the first study, HWE = loss of significance after excluding studies violating HWE ($P<0.05$), Regr = evidence for small-study/publication bias using a modified regression test (see Text S1), A = Grade A ('strong' epidemiologic credibility), B = Grade B ('modest' epidemiologic credibility), C = Grade C ('weak' epidemiologic credibility), logBF = Bayes Factor (see Text S1). "*" denotes SNPs that have been supplemented by additional data after the datafreeze (in total this applies to 147 SNPs, see Text S1 for the description of included datasets).
(XLS)

Table S2 Investigation of the extent of statistical inflation assuming sample overlaps of 1%, 5%, and 10% across cases and controls in datasets originating from the same countries. Hypothetical sample overlap across datasets was assumed between different candidate-gene/replication studies and between candidate-gene/replication studies and GWAS datasets if they originated from the same country. These analyses were performed applying random-effects models and adding the sum of weighted co-variances of overlapping datasets to the overall study variance (see ref. [24] in the main text). Note that the assumption of undetected overlapping samples does not apply (and was therefore not modeled here) to overlap between individual GWAS as duplicate samples in these datasets were removed prior to meta-analysis. It also does not apply to independent datasets used in the same publication where duplicate samples had been removed by the authors prior to analysis and publication. We emphasize that this table describes hypothetical scenarios, because the geographical origin of each study had been investigated extensively and potentially overlapping datasets had been excluded as part of PDGene's data inclusion protocol. Thus, the extent of overlap across geographically distinct datasets within the same countries is reduced to accidental recruitment of the same subjects more than once in different datasets throughout the respective countries, and can be expected to be less than ~1%. This estimate is based on data of the GEO-PD consortium, where sufficient data were centrally available of 6,072 subjects from 20 geographically distinct sites in 13 countries that had been investigated for potentially duplicate samples across sites, but no duplicate subjects (neither between not within countries) were identified when matching on ethnicity, birth, sex, and genotype. The investigation of overlap was not applicable here for Asian datasets, as they originated from different countries and/or were cleaned by the respective authors prior to publication.
(DOC)

Text S1 Supplementary material. This file includes supplementary methods and references as well as the list of members of the GWAS consortia, the GEO-PD Consortium, and consortia-specific acknowledgements.
(PDF)

Acknowledgments

23andMe acknowledges Elizabeth Dorfman, Amy K. Kiefer, Emily M. Drabant, Uta Francke, Joanna L. Mountain, David Hinds, and Anne Wojcicki from 23andMe, as well as Samuel M. Goldman, Caroline M. Tanner, and J. William Langston from the Parkinson's Institute, Sunnyvale, CA, USA. We also acknowledge the contribution of Mitsutoshi Yamamoto, Nobutaka Hattori, and Miho Murata for sample collection in the Japanese GWAS 1.0 [14]. We are grateful to the Alzheimer Research Forum—in particular to June Kinoshita, Colin Knep, Paula Noyes, and Gabrielle Ströbel—for hosting PDGene on their website. We also thank the many PD researchers who have kindly provided us with genotype data and helpful information beyond those included in the original publications. Finally, we would like to thank the many PD patients and control subjects who volunteered to participate in the individual studies.

Author Contributions

Conceived and designed the experiments: CM Lill, MB McQueen, JPA Ioannidis, L Bertram. Performed the experiments: CM Lill, JT Roehr, S Bagade, B-M Schjeide, E Meissner, U Zauft, NC Allen, KJ Anderson, G Beecham, D Berg, JM Biernacka, A Brice, AL DeStefano, CB Do, N

Eriksson, SA Factor, MJ Farrer, T Foroud, T Gasser, T Hamza, JA Hardy, P Heutink, C Klein, JC Latourelle, DM Maraganore, ER Martin, M Martinez, RH Myers, H Payami, WK Scott, M Sharma, AB Singleton, K Stefansson, T Toda, JY Tung, J Vance, NW Wood, CP Zabetian, 23andMe, GEO-PD, IPDGC, Parkinson's Disease GWAS, WTCC2. Analyzed the data: CM Lill, JT Roehr, MB McQueen, FK Kavvoura, L Bertram. Wrote the paper: CM Lill, JPA Ioannidis, L Bertram. Helped write the manuscript: E Meissner, MJ Farrer, T Foroud, T Gasser, C Klein, DM Maraganore, H Payami, AB Singleton, M Sharma, F Zipp, H Lehrach. Helped analyze the data: S Bagade, T Liu, M Schilling, CB Do, N Eriksson, T Hamza, EM Hill-Burns, MA Nalls, N Pankratz, W Satake, M Sharma. Interpretation of results: CM Lill, JPA Ioannidis, L Bertram. Study coordination: CM Lill, T Foroud, JA Hardy, H Payami, AB Singleton, P Young, RE Tanzi, MJ Khoury, F Zipp, H Lehrach, JPA Ioannidis, L Bertram. Literature searches and data entry: CM Lill, S Bagade, B-M Schjeide, E Meissner, U Zauft, N Allen.

References

1. de Lau LML, Breteler MMB (2006) Epidemiology of Parkinson's disease. Lancet Neurol 5: 525–535. doi:10.1016/S1474-4422(06)70471-9.
2. Hardy J, Lewis P, Revesz T, Lees A, Paisan-Ruiz C (2009) The genetics of Parkinson's syndromes: a critical review. Curr Opin Genet Dev 19: 254–265. doi:10.1016/j.gde.2009.03.008.
3. Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, et al. (2011) VPS35 mutations in Parkinson disease. Am J Hum Genet 89: 162–167. doi:10.1016/j.ajhg.2011.06.001.
4. Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, et al. (2011) A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet 89: 168–175. doi:10.1016/j.ajhg.2011.06.008.
5. Chartier-Harlin M-C, Dachsel JC, Vilariño-Güell C, Lincoln SJ, Leprêtre F, et al. (2011) Translation initiator EIF4G1 mutations in familial Parkinson disease. Am J Hum Genet 89: 398–406. doi:10.1016/j.ajhg.2011.08.009.
6. Maraganore DM, de Andrade M, Elbaz A, Farrer MJ, Ioannidis JP, et al. (2006) Collaborative analysis of alpha-synuclein gene promoter variability and Parkinson disease. JAMA 296: 661–670. doi:10.1001/jama.296.6.661.
7. Zabetian CP, Yamamoto M, Lopez AN, Ujike H, Mata IF, et al. (2009) LRRK2 mutations and risk variants in Japanese patients with Parkinson's disease. Mov Disord 24: 1034–1041. doi:10.1002/mds.22514.
8. Goris A, Williams-Gray CH, Clark GR, Foltynie T, Lewis SJG, et al. (2007) Tau and alpha-synuclein in susceptibility to, and dementia in, Parkinson's disease. Ann Neurol 62: 145–153. doi:10.1002/ana.21192.
9. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, et al. (2009) Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 361: 1651–1661. doi:10.1056/NEJMoa0901281.
10. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, et al. (2005) High-resolution whole-genome association study of Parkinson disease. Am J Hum Genet 77: 685–693. doi:10.1086/496902.
11. Fung H-C, Scholz S, Matarin M, Simón-Sánchez J, Hernandez D, et al. (2006) Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol 5: 911–916. doi:10.1016/S1474-4422(06)70578-6.
12. Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, et al. (2009) Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Hum Genet 124: 593–605. doi:10.1007/s00439-008-0582-9.
13. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, et al. (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet 41: 1308–1312. doi:10.1038/ng.487.
14. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, et al. (2009) Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet 41: 1303–1307. doi:10.1038/ng.485.
15. Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, et al. (2010) Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. Ann Hum Genet 74: 97–109. doi:10.1111/j.1469-1809.2009.00560.x.
16. Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, et al. (2010) Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet 42: 781–785. doi:10.1038/ng.642.
17. Spencer CCA, Plagnol V, Strange A, Gardner M, Paisan-Ruiz C, et al. (2011) Dissection of the genetics of Parkinson's disease identifies an additional association 5′ of SNCA and multiple associated haplotypes at 17q21. Hum Mol Genet 20: 345–353. doi:10.1093/hmg/ddq469.
18. Saad M, Lesage S, Saint-Pierre A, Corvol J-C, Zelenika D, et al. (2011) Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. Hum Mol Genet 20: 615–627. doi:10.1093/hmg/ddq497.
19. Simón-Sánchez J, van Hilten JJ, van de Warrenburg B, Post B, Berendse HW, et al. (2011) Genome-wide association study confirms extant PD risk loci among the Dutch. Eur J Hum Genet 19: 655–661. doi:10.1038/ejhg.2010.254.
20. Evangelou E, Maraganore DM, Ioannidis JPA (2007) Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. PLoS ONE 2: e196. doi:10.1371/journal.pone.0000196.
21. Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin U-M, et al. (2011) Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet 377: 641–649. doi:10.1016/S0140-6736(10)62345-8.
22. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, et al. (2011) Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. PLoS Genet 7: e1002141. doi:10.1371/journal.pgen.1002141.
23. Sharma M, Ioannidis JPA, Aasly JO, Annesi G, Brice A, et al. (n.d.) Large-scale replication and heterogeneity in Parkinson disease genetic loci. Neurology in press.
24. Lin D-Y, Sullivan PF (2009) Meta-analysis of genome-wide association studies with overlapping subjects. Am J Hum Genet 85: 862–872. doi:10.1016/j.ajhg.2009.11.001.
25. Ioannidis JPA (2008) Effect of formal statistical significance on the credibility of observational associations. Am J Epidemiol 168: 374–383; discussion 384–390. doi:10.1093/aje/kwn156.
26. Ioannidis JPA, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, et al. (2008) Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol 37: 120–132. doi:10.1093/ije/dym159.
27. Khoury MJ, Bertram L, Boffetta P, Butterworth AS, Chanock SJ, et al. (2009) Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases. Am J Epidemiol 170: 269–279. doi:10.1093/aje/kwp119.
28. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D (2010) BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26: 2204–2207. doi:10.1093/bioinformatics/btq351.
29. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, et al. (2007) A survey of genetic human cortical gene expression. Nat Genet 39: 1494–1499. doi:10.1038/ng.2007.16.
30. Varnum-Finney B, Venstrom K, Muller U, Kypta R, Backus C, et al. (1995) The integrin receptor alpha 8 beta 1 mediates interactions of embryonic chick motor and sensory neurons with tenascin-C. Neuron 14: 1213–1222.
31. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7: 177–188.
32. Harbord RM, Egger M, Sterne JAC (2006) A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med 25: 3443–3457. doi:10.1002/sim.2380.
33. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39: 17–23. doi:10.1038/ng1934.
34. Allen NC, Bagade S, McQueen MB, Ioannidis JPA, Kavvoura FK, et al. (2008) Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. Nat Genet 40: 827–834. doi:10.1038/ng.171.

New Technology

国際共同研究における Genome-Wide Association Study（GWAS）

Genome-wide association study (GWAS) by international collaboration

戸田　達史

Key Words：GWAS，メタ解析，SNP，HapMap計画

Abstract

dbSNPやHapMap計画などの情報基盤や数十万種のSNPを解析できる技術基盤の整備により，ゲノムワイド関連解析（GWAS）が2007年頃より実用的な戦略となり，疾患感受性遺伝子の発見が相次いだ。すでに165の疾患について，ゲノムワイド有意水準P< 5×10^{-8}をクリアする904個の論文が発表された。国際共同研究によりGWASメタ解析を行い，よりeffect sizeの小さなものも同定するという試みが行われだしており，「第２世代のGWAS」とも言われている。

■GWASの現状

2006年以降，common diseaseの感受性遺伝子の探索研究は新しい段階を迎えた。これをもたらしたおもな要因は二つの基盤整備である。まず情報基盤として，dbSNPやHapMap計画で代表されるように，ヒトゲノム全域にわたる膨大な多様性情報が集積されてきた。次に技術基盤として，数十万種のSNP（Single Nucleotide Polymorphism）を数千もの個体について並列解析できるプラットフォームが市販化された。

まずHapMap計画によると，日本人と白人は約25-30万個のタグSNP（これを調べれば連鎖不平衡で結ばれた近傍の多くのSNPの代表になる）で，ほぼ全ゲノムの遺伝子がカバーされる。つまり約30万個のタグSNPを患者と対照で調べれば，ほぼ全ゲノムの遺伝子を調べたことになる。そこで具体的には例えば患者1,000人，対照1,000人，計2,000人各人の50万個のSNPの遺伝子型を決定する。すなわちSNPチップとして2,000枚の実験を行う。それぞれSNP-1,,,,, SNP-500000ずつ，患者，対照におけるそれぞれのアレルの出現頻度を合計し，偏りがないかどうかの検定を行うのである。

これらを活用することによって，ゲノムワイド関連解析（GWAS: Genome-Wide Association Study）が実用的な戦略となり，2007年にはNature，Science誌などに立て続けに成果が発表されることとなった。その後のGWASによる疾患感受性遺伝子の発見ラッシュには目を見張るものがあり，すでに165の疾患について，ゲノムワイド有意水準P < 5×10^{-8}をクリアする904個の論文が発表されている（図）。

■GWASによる疾患感受性遺伝子発見

神経疾患のパーキンソン病を例にとると，我々のグループは患者の95%を占める孤発性PDのリスク遺伝子を同定するため，GWASを行い，PD発症に関わる2つの新しい遺伝子座*PARK16*, *BST1*を同定した。また，常染色体優性遺伝性PDの原因遺伝子*α-synuclein*, *LRRK2*の弧発性PDへの関与を証明した。国際共同研究として次に，我々と併行してヨーロッパ起源の集団のPDのGWAS研究を行っていたグループと，データを交換した。彼らの研究では，*α-synuclein*領域，*Tau*領域のみに，強い関連が検出されていた。そこで，我々の発見した*PARK16*・*BST1*・*LRRK2*の人数を増やして再現研究をしたところ，*PARK16*と*LRRK2*の関連は強く再現されたが，*BST1*は再現されなかった。逆に，我々は，彼らの検出した*Tau*の関連の再現を試みたが，我々の検体セットでは，再現されなかった。よって，*α-synuclein*・*PARK16*・*LRRK2*は2人種に共通のPDリスクであり，*Tau*・*BST1*はのリスク多型の影響は人種特異的であると考えた[1]。

Tatsushi Toda
神戸大学大学院医学研究科神経内科学／分子脳科学
Division of Neurology/Moleucular Brain Science, Kobe University Graduate School of Medicine

さらにパーキンソン病においては，さらなる国際共同研究として欧米の５つのグループがそれぞれ独立に行っていたGWASを合わせてメタ解析を行い（計患者5,333，対照12,019），有意なSNPをさらに患者7,053，対照9,007で再現実験を行い，ゲノムワイド有意水準P＜5×10^{-8}を超える遺伝子を従来の６個の他に，*ACMSD*, *STK39*, *LAMP3*, *SYT11*, *CCDC62*の５個を同定した[2]。

すべては網羅できないが他の例では，クローン病においては，６個のGWASを合わせてメタ解析を行い（計患者6,333，対照15,056），さらに有意なSNPを患者15,694，対照14,026で再現実験を行い，ゲノムワイド有意水準を超える遺伝子を新規に30個同定し，計71個となった[3]。

また２型糖尿病においては，８個のGWASを合わせてメタ解析を行い（計患者8,130，対照38,987），さらに有意なSNPを患者34,412，対照59,925で再現実験を行い，ゲノムワイド有意水準を超える遺伝子を新規に12個同定した[4]。

さらに驚くべきことに，GIANT Consortiumは国際共同研究として46個のGWASを合わせてメタ解析を行い（計133,653人），50,074人で再現実験を行い，ゲノムワイド有意水準を超える身長に関わる遺伝子を少なくとも180個同定した[5]。

GWASによって多数の疾患感受性遺伝子が同定されたものの，それらは遺伝要因全体の一部しか説明できないことから（missing heritability），このような圧倒的な数の試料を各地から集めてゲノムワイドメタ解析を行いよりeffect sizeの小さなものも同定するという試みが行われだしており，「第２世代のGWAS」とも言われている。

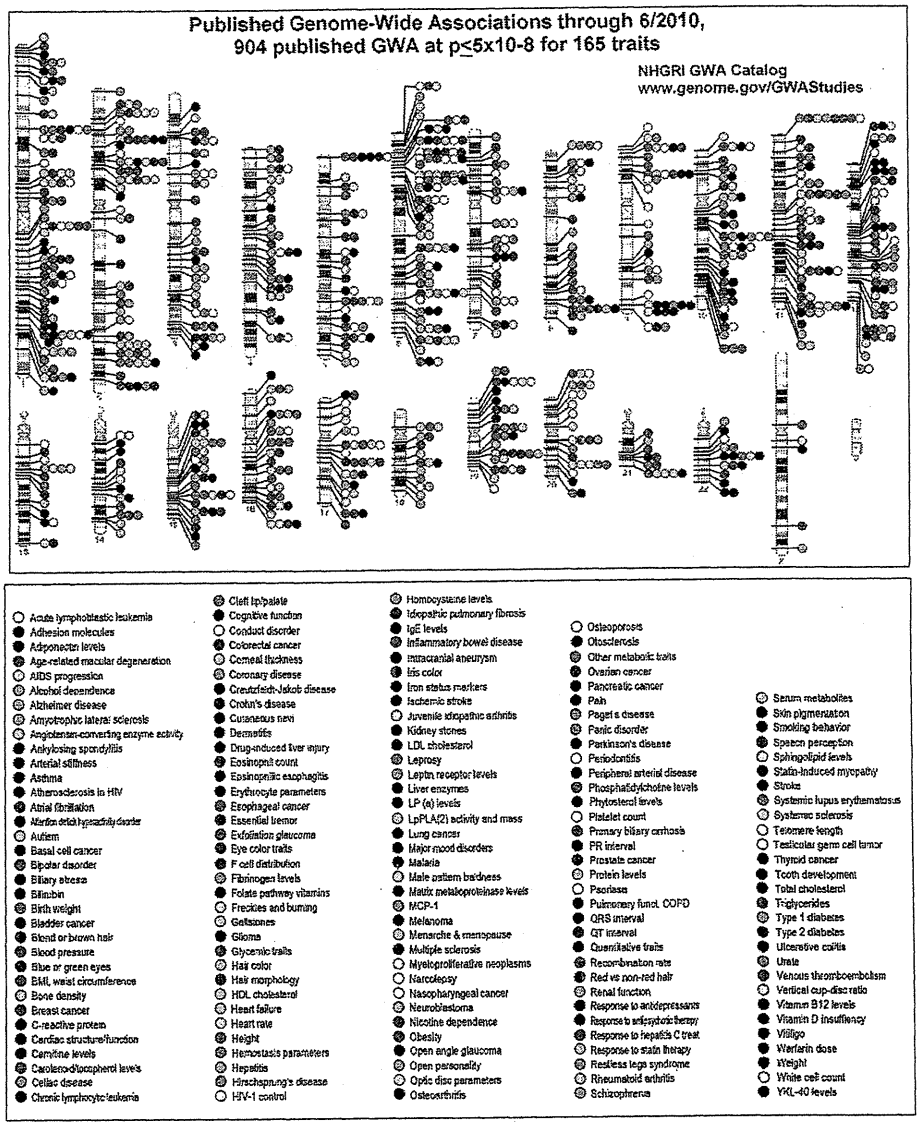


図　ゲノムワイド有意水準P＜5×10^{-8}をクリアする多因子疾患の遺伝子座

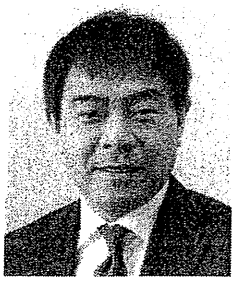
一方でCommon Disease-Multiple Rare Variant仮説による，頻度は低いが発症へのEffect sizeが大きいrare variantも，missing heritabilityとして重要な位置を占める。

文　献

1) Satake W, *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet. 2009; 41: 1303-7.
2) International Parkinson Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet. 2011; 377: 641-9.
3) Franke A, *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010; 42: 1118-25.
4) Voight BF, *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42: 579-89.
5) Lango Allen H, *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467: 832-8.

6 パーキンソン病の分子遺伝学—ゲノム関連解析研究—

■戸田 達史（とだ たつし）・佐竹 渉（さたけ わたる）
神戸大学大学院医学研究科　神経内科学



戸田　達史
1985年東京大学医学部卒業，神経内科入局。94年東京大学人類遺伝学助手。96年東京大学医科学研究所助教授。2000年大阪大学臨床遺伝学教授。2008年神戸大学神経内科学／分子脳科学教授。日本人類遺伝学会賞，日本神経学会賞，朝日賞，文部科学大臣表彰。研究テーマは神経内科学，神経疾患の分子遺伝学・治療。趣味はワイン，グルメ。

Key words：パーキンソン病，αシヌクレイン，PARK16，GWAS

Abstract
患者の95%を占める孤発性PDは多因子疾患である。孤発性PDのリスク遺伝子を同定するため，ゲノムワイド関連解析を行い，PD発症に関わる2つの新しい遺伝子座PARK16, BST1を同定した。また，常染色体優性遺伝性PDの原因遺伝子SNCA, LRRK2の弧発性PDへの関与を証明した。さらに，人種差が，PDの遺伝的不均一性に，寄与していることを示唆した。国際共同研究によりGWASメタ解析を行いより多くの感受性遺伝子を同定することが行われだしており，「第２世代のGWAS」とも言われている。さらなる遺伝子の解明が期待される。

はじめに

パーキンソン病（PD）症例の90%以上は孤発性発症であるが、5-10%は家族性（その一部はメンデル遺伝性）に発症する。メンデル遺伝性パーキンソン病（PD）家系の連鎖解析などから6つのメンデル遺伝性PD原因遺伝子（αシヌクレイン、パーキン、LRRK2遺伝子など）が明らかにされた。孤発性PD、メンデル遺伝性PDとも、一部共通の発症メカニズムが存在していると考えられ、それらを切り口にして孤発性PDの病態解明が進んでおり、ミトコンドリア障害、酸化ストレス障害の病態への関与に加え、新たにユビキチン・プロテアソーム系の機能低下，つまり蛋白分解異常からドパミン細胞死に至る経路の重要性が示された[1]。

一方dbSNPやHapMap計画などの情報基盤や数十万種のSNPを解析できる技術基盤の整備により，ゲノムワイド関連解析（GWAS）が2007年頃より実用的な戦略となり，現在多数の疾患でGWAS（ゲノムワイド関連解析）が行われており，疾患感受性遺伝子の発見があいついでいる。我々はパーキンソン病の大規模なGWASを行い，４つの感受性遺伝子を発見した[2]。本稿ではパーキンソン病のGWASについて述べる。

1. パーキンソン病は多因子遺伝性疾患

症例的には大多数（95%）の孤発性PDの原因は，現時点では不明であるが，環境因子と１つ１つは影響力の弱い遺伝因子（おそらく数10個）によってなり，その総和が，ある閾値を超えたとき発症するという多因子疾患であると考えられている。アイスランド国民を対象とした大規模な疫学的調査の結果が発表され，同胞再発危険率は6.7で，PD発症には遺伝因子が影響していることが示された。

Molecular genetics for Parkinson's disease -genomewide association study-：Tatsushi Toda, Wataru Satake, Division of Neurology, Kobe University Graduate School of Medicine, 7-5-1 Kusunoki-chou, Chuo-ku, Kobe 650-0017, Japan