currently unknown. It has been suggested that satellite cells express properly glycosylated α-DG (31) and that myoblasts/MPCs also express properly glycosylated α-DG although its signals are relatively weak compared with those of myotubes (32,33). The basement membrane of skeletal muscle contains DG ligands, laminins and perlecan. It is well established that interactions between the extracellular matrix and cell-surface receptors are involved in cell survival signaling (34). Since it has been proposed that DG-ligand interactions are also involved in cellular signaling mechanisms such as survival and apoptosis pathways (35,36), it is possible that loss of α-DG glycosylation may affect survival signaling regulated by α-DG-basement membrane interaction. In addition, since the Myf5-fukutin-cKO myofibers showed an earlier reduction of fukutin compared with those of MCK-fukutin-cKO, there is a possibility that earlier loss of α-DG glycosylation in myofibers affects disease progression and severity. For example, the absence of α-DG glycosylation during postnatal/juvenile muscle growth and development may have a high impact on muscle degeneration and/or dystrophic pathology in later stages.

As for other muscular dystrophy models with defects in the dystrophin–glycoprotein complex, impaired muscle regeneration has also been reported in *MORE-DG null* mice, in which the DG gene (*Dag1*) is ablated in all cells in the embryo (31), and in older (>1 year) dystrophin-deficient mdx and sarcoglycan-deficient mice (31). The regeneration defects in our Myf5-fukutin-cKO mice appeared at a relatively young age (~3 months), but at this age, Myf5-fukutin-cKO mice already show severe dystrophic pathology. The pathological environment may interfere with efficient muscle regeneration, resulting in decreased regeneration activity as the disease progresses (37). A very recent study also suggested that alterations to the basal lamina microenvironment perturb regeneration potential in dystroglycanopathy (38). Moreover, our data suggest that multiple cycles of degeneration/regeneration may also affect MPC viability, which is consistent with a previous study showing that the progressive exhaustion of functional muscle satellite cells is associated with severe dystrophic phenotype (39). It appears that these disease environments and impaired MPC viability caused by loss of fukutin-dependent modification additively deteriorate regeneration activity, eventually leading to severe and rapid progressive pathology. Together, we conclude that defects in MPC activity contribute to the severe pathology of dystroglycanopathy and propose that dystroglycanopathy is a regeneration-defective disorder.

We observed that some muscle specimens from 16-week-old Myf5-fukutin-cKO mice showed mild phenotype, which was consequently supported by the presence of functionally glycosylated α-DG. Beedle *et al.* (30) also reported that fukutin deletion resulted in moderate to severe muscular dystrophy using Myf5-Cre mice. Because phenotypic variation in our Myf5-fukutin-cKO colony was rarely seen before 12 weeks of age, the variation may be secondary to disease progression. The less phenotypic variation in our colonies could also be due to the number of backcross on C57BL/6 (backcross: more than seven). We speculate that during frequent cycles of muscle degeneration/regeneration, Myf5-independent or less-expressed myogenic cells (40) may be activated and differentiated into myofibers in which the *fukutin* gene escaped Cre-mediated recombination.

Many cases of dystroglycanopathy show the most severe skeletal muscle phenotype, and the severe/typical dystroglycanopathy patients end their short lives without ever standing or walking. Although an increasing number of patients are being diagnosed with dystroglycanopathy worldwide, there have been no therapeutic studies on dystroglycanopathy models after the disease progresses. In this study, for the first time, we succeeded in ameliorating the disease severity in dystroglycanopathy mouse models based on the pathomechanism. It is of importance that limited rescue of fukutin protein in myofibers of Myf5-fukutin-cKO muscles, which have MPC defects, ameliorated the severe phenotype. These data suggest that even after functional and/or substantial loss of MPC occurs, prevention of disease-causing defects in myofibers is a probable therapeutic strategy for muscular dystrophy. Moreover, it is noteworthy that therapeutic effects of the exogenous *fukutin* gene were achieved with relatively lower AAV titers than those used in other gene therapies for structural proteins such as dystrophin and sarcoglycans (41–43). Our results showed that titers that were ~2 orders of magnitude less than those required in previous studies were sufficient to produce a therapeutic effect in Myf5-fukutin-cKO mice. This is consistent with our previous study, which suggested that only a little amount of fukutin is necessary to prevent muscular dystrophy (29). Most dystroglycanopathy genes are identified as glycosyltransferases (11,12,15) or hypothesized to have enzyme-like properties, suggesting that a small amount of exogenous gene would be sufficient for producing therapeutic effects. A small dose of AAV vectors could lower the chances of adverse effects such as immune responses in human (44). In addition, the cDNA sizes of fukutin as well as other dystroglycanopathy genes are suitable for AAV vectors. Taken together, we propose that gene transfer is a promising therapeutic strategy for the amelioration of the severe skeletal muscle pathology of dystroglycanopathy. Human dystroglycanopathy is frequently accompanied by brain and, often, cardiac disorders (1,45). The efficacy of AAV delivery to these affected tissues, timing of administration and therapeutic effects in other fukutin-cKO models should be examined in the future. Although therapeutic interventions that rescue the developmental defects of dystroglycanopathy (such as anomalies in brain structure) are difficult at present, amelioration of the muscle phenotype would be highly beneficial to patients and their families. For example, such treatment might improve patients' physical abilities and postpone the need for respiratory interventions until much later in the course of the disease. Increased physical activity could positively influence both mental development and social interactions. Overall, this study may facilitate future clinical translational research in the field of dystroglycanopathy treatment.

MATERIALS AND METHODS

Generation of fukutin cKO mice

Construction of the targeted allele, establishment of targeted embryonic stem (ES) cells and generation of the chimera

and F1 mice were carried out by Unitech Co. (Kashiwa, Japan). Briefly, exon 2 of mouse *fukutin* was flanked by two loxP sequences (Supplementary Material, Fig. S1A). An Flp recognition target-flanked neo-cassette was inserted upstream of exon 2. The targeting vector was electroporated into C57BL/6 mouse ES cells. Positive clones were selected, and homologous recombination was confirmed by Southern blotting (Supplementary Material, Fig. S1B). The targeted ES cells were injected into blastocysts (BALB/c), and then, chimera mice were bred with C57BL/6 mice to generate founder mice. The founder mice were crossed with the FLPe transgenic mice, producing heterozygous flox mice without the neo-cassette ($fukutin^{lox/+}$). The heterozygous flox mice were intercrossed to obtain homozygous flox mice ($fukutin^{lox/lox}$).

MCK-Cre mice (22) [$MCK\text{-}Cre^{Tg}(+)$, backcrossed for at least 10 generations to C57BL/6] and Myf5-Cre knock-in mice (23) [$Myf5\text{-}Cre^{KI}(+)$] were obtained from The Jackson Laboratory. Myf5-Cre mice were backcrossed for more than six generations to C57BL/6 before crossing with $fukutin^{lox/lox}$ mice. The heterozygous $fukutin^{lox/+}$ carrying MCK-Cre [$fukutin^{lox/+}$:$MCK\text{-}Cre^{Tg}(+)$] or Myf5-Cre [$fukutin^{lox/+}$:$Myf5\text{-}Cre^{KI}(+)$] were then bred with $fukutin^{lox/lox}$ mice to obtain cKO mice. Using this breeding strategy, we obtained the following four genotypes (Supplementary Material, Fig. S1C): for MCK-fukutin-cKO line—[$fukutin^{lox/lox}$:$MCK\text{-}Cre^{Tg}(-)$] (used as WT control), [$fukutin^{lox/lox}$:$MCK\text{-}Cre^{Tg}(+)$] (used as cKO), [$fukutin^{lox/+}$:$MCK\text{-}Cre^{Tg}(+)$] (used as heterozygous control, HET) and [$fukutin^{lox/+}$:$MCK\text{-}Cre^{Tg}(-)$]; and for the Myf5-fukutin-cKO line—[$fukutin^{lox/lox}$:$Myf5\text{-}Cre^{KI}(-)$] (used as WT), [$fukutin^{lox/lox}$:$Myf5\text{-}Cre^{KI}(+)$] (used as cKO), [$fukutin^{lox/+}$:$Myf5\text{-}Cre^{KI}(+)$] (used as HET) and [$fukutin^{lox/+}$:$Myf5\text{-}Cre^{KI}(-)$]. Alternatively, we crossed cKO mice with $fukutin^{lox/lox}$ mice to obtain two genotypes: WT and cKO (Supplementary Material, Fig. S1C). Genotyping was performed using PCR (Supplementary Material, Fig. S1D). Primer sequences and PCR conditions are available on request. Mice were maintained in accordance with the animal care guidelines of Unitech Co. Ltd., Osaka University and Kobe University.

Antibodies

Antibodies used in western blots and immunofluorescence were as follows: mouse monoclonal antibody 8D5 against β-DG (Novacastra); mouse monoclonal antibody IIH6 against glycosylated α-DG (Millipore); goat polyclonal antibody against the C-terminal domain of the α-DG polypeptide (AP-074G-C) (29); goat polyclonal anti-fukutin antibody (106G2) and rabbit polyclonal anti-fukutin antibody (RY213) (5); rat anti-laminin antibody 4H8-2 (Alexis Biochemicals); rabbit polyclonal anti-collagen I antibody (AbD Serotec); and mouse monoclonal anti-embryonic myosin antibody (The Developmental Studies Hybridoma Bank, University of Iowa). A rat monoclonal antibody against the α-DG core protein (3D7-7) was generated using the recombinant α-DG-Fc fusion protein (46); hybridoma clones were selected for reactivity to the C-terminal domain of the α-DG polypeptide. To reduce high background staining of IIH6 in severely affected skeletal muscle sections, commercial IIH6 was labeled with biotin. The IIH6-IgM fractions were prepared from ascites, using protein L-beads (Pierce) and then biotinylated (EZ-Link Micro Sulfo-NHS-Biotinylation Kit; Pierce) according to the manufacturer's instructions.

Preparation of fukutin and DG

Endogenous fukutin was enriched by immunoprecipitation using polyclonal goat anti-fukutin antibody (106G2) from the skeletal muscle lysates. The immunoprecipitated materials were subjected to western blotting using polyclonal rabbit anti-fukutin antibody (RY213). DG from solubilized skeletal muscle was enriched with wheat germ agglutinin-agarose beads (Vector Laboratories) as previously described (29).

Histological and immunofluorescence analyses

For histological and immunofluorescence staining, cryosections (7 μm thick) were prepared. For H&E staining, sections were stained for 2 min in hematoxylin, 1 min in eosin, and then dehydrated with ethanol and xylene. The slides are washed with 0.5% glacial acetic acid, dehydrated and then mounted. For immunofluorescence analysis, sections were treated with cold ethanol/acetic acid (1:1) for 1 min, blocked with 5% goat serum in MOM mouse Ig blocking reagent (Vector Laboratories) at room temperature for 1 h, and then incubated overnight with primary antibodies diluted in MOM diluent (Vector Laboratories) at 4°C. The slides were washed with phosphate-buffered saline (PBS) and incubated with Alexa Fluor 488-conjugated or Alexa Fluor 555-conjugated secondary antibodies (Molecular Probes) at room temperature for 30 min. Sections were observed by fluorescence microscopy (Leica DMR, Leica Microsystems and BZ9000, Keyence). Quantification of the number of Pax7-positive cells and the population of Pax7/MyoD-double-positive cells was performed as previously described (47).

Quantitative and statistical analysis

For quantitative evaluation of muscle pathology, the proportion of myofibers with centrally located nuclei in at least 1000 fibers for each individual was counted. For the evaluation of connective tissue infiltration, the immunofluorescence signal of collagen I was quantitatively measured using the ImageJ software. For the assessment of myofiber size variation, areas of individual myofibers on transverse sections were measured using the ImageJ software. Data represent means with SEM, and P-values ≤ 0.05 were considered statistically significant (Mann–Whitney U test).

Preparation and culture of MPCs

Mononuclear cells from uninjured limb muscles were prepared using 0.2% collagenase type II (Worthington Biochemical) as previously described (28,48). Approximately $3-5 \times 10^6$ or $3-9 \times 10^6$ mononuclear cells from young mice (2-week-old) or adult mice (4-month-old), respectively, were subjected to MPC isolation experiments. Mononuclear cells derived from the skeletal muscles were stained with FITC-conjugated anti-CD31 (Pecam1, Mouse Genome Informatics), anti-CD45

(Ptprc, Mouse Genome Informatics), phycoerythrin-conjugated anti-Sca1 (Ly6a, Mouse Genome Informatics) and biotinylated SM/C-2.6 antibodies (28). Cells were then incubated with 1:400 streptavidin–allophycocyanin (BD Biosciences) on ice for 30 min and resuspended in PBS containing 2% fetal calf serum (FCS) and 2 μg/ml propidium iodide (PI). Cell sorting was performed using an FACS Aria II flow cytometer (BD Immunocytometry Systems). Debris and dead cells were excluded by forward scatter, side scatter and PI gating. Data were collected using the FACSDiva software (BD Biosciences). Myogenic cells from the regenerating muscles were also highly enriched in the SM/C-2.6(+) CD31(−) CD45(−) Sca1(−) cell fraction.

Freshly isolated myogenic cells were cultured in a growth medium of high-glucose Dulbecco's modified Eagle's medium (DMEM-HG; Sigma-Aldrich) containing 20% FCS, 2.5 ng/ml basic fibroblast growth factor (FGF2; PeproTech) and penicillin (100 U/ml)–streptomycin (100 μg/ml) (Gibco BRL) on culture dishes coated with Matrigel (BD Biosciences). Differentiation was induced in a differentiation medium containing DMEM-HG, 5% horse serum and penicillin–streptomycin for 3–4 days. Quantitative analysis for cell proliferation was performed as described previously (47). Fusion index was estimated as the ratio of nuclei in the myotubes to all the nuclei in more than four independent microscopy fields.

CTX experiments

CTX (30 μM; purified from the venom of the snake *Naja nigricollis*; Latoxan) was injected intramuscularly (for young mice, 30 μl to the tibialis anterior and 70 μl to the calf; for adult mice, 50 μl to the tibialis anterior and 100 μl to the calf). Mock injections used only saline solution. The injected muscles were examined 5 or 14 days after the injection. Five days after the injection, areas of individual embryonic myosin-positive fibers were measured (>300 fibers randomly chosen from 5–10 regions per toxin-challenged muscle) in each genotype. Fourteen days after the injection, fiber size variation was quantitatively evaluated by measuring individual laminin-positive fibers (>300 fibers) in each genotype.

AAV gene transfer

To generate *fukutin*-encoding AAV9 vector, the complete open reading frame of mouse *fukutin* gene was cloned into the pAAV-IRES-hrGFP vector (49). The recombinant *fukutin*-encoding AAV9 vector was produced as described previously (49). AAV vectors were injected intramuscularly into the calf and tibialis anterior (at 1 week to the tibialis anterior, $0.8–1.6 \times 10^9$ vector genome in saline solution; at 1 week to the calf or at 8 weeks to the tibialis anterior, $2–4 \times 10^9$ vector genome; and at 8 weeks to the calf, $4–8 \times 10^9$ vector genome). For tail vein injections and intraperitoneal injections, $\sim 2 \times 10^{10}$ and $\sim 1 \times 10^{10}$ vector genome was used, respectively.

Miscellaneous

For western blotting, the proteins were separated using 4–15% linear gradient SDS–PAGE (Bio-Rad). Gels were transferred to polyvinylidene fluoride membrane (Millipore). Blots were developed by horseradish peroxidase-enhanced chemiluminescence (Supersignal West Pico, Pierce; or ECL Plus, GE Healthcare). Laminin-binding activity was determined by the laminin overlay assay as described previously (29). Serum CK activity was measured using the CPK kit (WAKO). For Evans blue dye uptake, Evans blue dye (10 mg/ml in saline) was intraperitoneally injected (100 μl/10 g of body weight). After 5 h, the mice were made to exercise on a downhill (15°) treadmill for 60 min (MK-680S, Muromachi Kikai). Twenty-four hours after the exercise, frozen tissue samples were prepared. Serum was prepared before 24 h and after 2 h of the exercise. Grip strength was measured for 10 consecutive trials for each mouse, using a strength meter (Ohara Ika Sangyo Co. Ltd, Tokyo), and 20% of the top and the bottom values were excluded to obtain the mean value.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We would like to thank the past and present members of T.T.'s laboratory for fruitful discussions and scientific contributions. We also thank Hiromi Hayashita-Kinoh for providing technical support.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Ministry of Health, Labor and Welfare of Japan [Intramural Research Grant for Neurological and Psychiatric Disorders of National Center of Neurology and Psychiatry (23B-5)], the Ministry of Education, Culture, Sports, Science and Technology of Japan [a Grant-in-Aid for Scientific Research (A) 23249049 to T.T., a Grant-in-Aid for Young Scientists (A) 24687017 to M.K. and a Grant-in-Aid for Scientific Research on Innovative Areas (No. 23110002, Deciphering Sugar Chain-based Signals Regulating Integrative Neuronal Functions) 24110508 to M.K.], a Senri Life Science Foundation grant to M.K., a Takeda Science Foundation grant to M.K. and a Naito Foundation grant to M.K.

REFERENCES

1. Godfrey, C., Foley, A.R., Clement, E. and Muntoni, F. (2011) Dystroglycanopathies: coming into focus. *Curr. Opin. Genet. Dev.*, **21**, 278–285.
2. Hayashi, Y.K., Ogawa, M., Tagawa, K., Noguchi, S., Ishihara, T., Nonaka, I. and Arahata, K. (2001) Selective deficiency of alpha-dystroglycan in Fukuyama-type congenital muscular dystrophy. *Neurology*, **57**, 115–121.
3. Michele, D.E., Barresi, R., Kanagawa, M., Saito, F., Cohn, R.D., Satz, J.S., Dollar, J., Nishino, I., Kelley, R.I., Somer, H. *et al.* (2002) Post-translational disruption of dystroglycan-ligand interactions in congenital muscular dystrophies. *Nature*, **418**, 417–422.
4. Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., Nomura, Y., Segawa, M., Yoshioka, M., Saito, K., Osawa, M. *et al.* (1998) An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature*, **394**, 388–392.

5. Taniguchi-Ikeda, M., Kobayashi, K., Kanagawa, M., Yu, C.C., Mori, K., Oda, T., Kuga, A., Kurahashi, H., Akman, H.O., DiMauro, S. *et al.* (2011) Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature*, **478**, 127–131.
6. 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.
7. Godfrey, C., Clement, E., Mein, R., Brockington, M., Smith, J., Talim, B., Straub, V., Robb, S., Quinlivan, R., Feng, L. *et al.* (2007) Refining genotype phenotype correlations in muscular dystrophies with defective glycosylation of dystroglycan. *Brain*, **130**, 2725–2735.
8. Tachikawa, M., Kanagawa, M., Yu, C.C., Kobayashi, K. and Toda, T. (2012) Mislocalization of fukutin protein by disease-causing missense mutations can be rescued with treatments directed at folding amelioration. *J. Biol. Chem.*, **287**, 8398–8406.
9. Wells, L. (2013) The O-mannosylation pathway: glycosyltransferases and proteins implicated in congenital muscular dystrophy. *J. Biol. Chem.*, **288**, 6930–6935.
10. Buysse, K., Riemersma, M., Powell, G., van Reeuwijk, J., Chitayat, D., Roscioli, T., Kamsteeg, E.J., van den Elzen, C., van Beusekom, E., Blaser, S. *et al.* (2013) Missense mutations in β-1,3-*N*-acetylglucosaminyltransferase 1 (B3GNT1) cause Walker-Warburg syndrome. *Hum. Mol. Genet.*, **22**, 1746–1754.
11. 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: coexpression of POMT1 and POMT2 required for enzymatic activity. *Proc. Natl Acad. Sci. USA*, **101**, 500–505.
12. Yoshida, A., Kobayashi, K., Manya, H., Taniguchi, K., Kano, H., Mizuno, M., Inazu, T., Mitsuhashi, H., Takahashi, S., Takeuchi, M. *et al.* (2001) Muscular dystrophy and neuronal migration disorder caused by mutations in a glycosyltransferase, POMGnT1. *Dev. Cell*, **1**, 717–724.
13. 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 alpha-dystroglycan is required for laminin binding. *Science*, **327**, 88–92.
14. Kuga, A., Kanagawa, M., Sudo, A., Chan, Y.M., Tajiri, M., Manya, H., Kikkawa, Y., Nomizu, M., Kobayashi, K., Endo, T. *et al.* (2012) Absence of post-phosphoryl modification in dystroglycanopathy mouse models and wild-type tissues expressing non-laminin binding form of α-dystroglycan. *J. Biol. Chem.*, **287**, 9560–9567.
15. 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.
16. Ibraghimov-Beskrovnaya, O., Ervasti, J.M., Leveille, C.J., Slaughter, C.A., Sernett, S.W. and Campbell, K.P. (1992) Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature*, **355**, 696–702.
17. Barresi, R. and Campbell, K.P. (2006) Dystroglycan: from biosynthesis to pathogenesis of human disease. *J. Cell Sci.*, **119**, 199–207.
18. Taniguchi, M., Kurahashi, H., Noguchi, S., Fukudome, T., Okinaga, T., Tsukahara, T., Tajima, Y., Ozono, K., Nishino, I., Nonaka, I. and Toda, T. (2006) Aberrant neuromuscular junctions and delayed terminal muscle fiber maturation in alpha-dystroglycanopathies. *Hum. Mol. Genet.*, **15**, 1279–1289.
19. Hara, Y., Balci-Hayta, B., Yoshida-Moriguchi, T., Kanagawa, M., Beltrán-Valero de Bernabé, D., Gündeşli, H., Willer, T., Satz, J.S., Crawford, R.W., Burden, S.J. *et al.* (2011) A dystroglycan mutation associated with limb-girdle muscular dystrophy. *N. Engl. J. Med.*, **364**, 939–946.
20. Roscioli, T., Kamsteeg, E.J., Buysse, K., Maystadt, I., van Reeuwijk, J., van den Elzen, C., van Beusekom, E., Riemersma, M., Pfundt, R., Vissers, L.E. *et al.* (2012) Mutations in ISPD cause Walker-Warburg syndrome and defective glycosylation of α-dystroglycan. *Nat. Genet.*, **44**, 581–585.
21. Willer, T., Lee, H., Lommel, M., Yoshida-Moriguchi, T., de Bernabe, D.B., Venzke, D., Cirak, S., Schachter, H., Vajsar, J., Voit, T. *et al.* (2012) ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. *Nat. Genet.*, **44**, 575–580.
22. Brüning, J.C., Michael, M.D., Winnay, J.N., Hayashi, T., Hörsch, D., Accili, D., Goodyear, L.J. and Kahn, C.R. (1998) A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. *Mol. Cell*, **2**, 559–569.
23. Tallquist, M.D., Weismann, K.E., Hellström, M. and Soriano, P. (2000) Early myotome specification regulates PDGFA expression and axial skeleton development. *Development*, **127**, 5059–5070.
24. Lyons, G.E., Mühlebach, S., Moser, A., Masood, R., Paterson, B.M., Buckingham, M.E. and Perriard, J.C. (1991) Developmental regulation of creatine kinase gene expression by myogenic factors in embryonic mouse and chick skeletal muscle. *Development*, **113**, 1017–1029.
25. Trask, R.V. and Billadello, J.J. (1990) Tissue-specific distribution and developmental regulation of M and B creatine kinase mRNAs. *Biochim. Biophys. Acta*, **1049**, 182–188.
26. Vilquin, J.T., Brussee, V., Asselin, I., Kinoshita, I., Gingras, M. and Tremblay, J.P. (1998) Evidence of mdx mouse skeletal muscle fragility in vivo by eccentric running exercise. *Muscle Nerve*, **21**, 567–576.
27. Durbeej, M., Sawatzki, S.M., Barresi, R., Schmainda, K.M., Allamand, V., Michele, D.E. and Campbell, K.P. (2003) Gene transfer establishes primacy of striated vs. smooth muscle sarcoglycan complex in limb-girdle muscular dystrophy. *Proc. Natl Acad. Sci. USA*, **100**, 8910–8915.
28. Fukada, S., Higuchi, S., Segawa, M., Koda, K., Yamamoto, Y., Tsujikawa, K., Kohama, Y., Uezumi, A., Imamura, M., Miyagoe-Suzuki, Y. *et al.* (2004) Purification and cell-surface marker characterization of quiescent satellite cells from murine skeletal muscle by a novel monoclonal antibody. *Exp. Cell Res.*, **296**, 245–255.
29. Kanagawa, M., Nishimoto, A., Chiyonobu, T., Takeda, S., Miyagoe-Suzuki, Y., Wang, F., Fujikake, N., Taniguchi, M., Lu, Z., Tachikawa, M. *et al.* (2009) Residual laminin-binding activity and enhanced dystroglycan glycosylation by LARGE in novel model mice to dystroglycanopathy. *Hum. Mol. Genet.*, **18**, 621–631.
30. Beedle, A.M., Turner, A.J., Saito, Y., Lueck, J.D., Foltz, S.J., Fortunato, M.J., Nienaber, P.M. and Campbell, K.P. (2012) Mouse fukutin deletion impairs dystroglycan processing and recapitulates muscular dystrophy. *J. Clin. Invest.*, **122**, 3330–3342.
31. Cohn, R.D., Henry, M.D., Michele, D.E., Barresi, R., Saito, F., Moore, S.A., Flanagan, J.D., Skwarchuk, M.W., Robbins, M.E., Mendell, J.R., Williamson, R.A. and Campbell, K.P. (2002) Disruption of DAG1 in differentiated skeletal muscle reveals a role for dystroglycan in muscle regeneration. *Cell*, **110**, 639–648.
32. Barresi, R., Michele, D.E., Kanagawa, M., Harper, H.A., Dovico, S.A., Satz, J.S., Moore, S.A., Zhang, W., Schachter, H., Dumanski, J.P. *et al.* (2004) LARGE can functionally bypass alpha-dystroglycan glycosylation defects in distinct congenital muscular dystrophies. *Nat. Med.*, **10**, 696–703.
33. Hara, Y., Kanagawa, M., Kunz, S., Yoshida-Moriguchi, T., Satz, J.S., Kobayashi, Y.M., Zhu, Z., Burden, S.J., Oldstone, M.B. and Campbell, K.P. (2011) Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. *Proc. Natl Acad. Sci. USA*, **108**, 17426–17431.
34. Gilmore, A.P. (2005) Anoikis. *Cell Death Differ.*, **12**, 1473–1477.
35. Langenbach, K.J. and Rando, T.A. (2002) Inhibition of dystroglycan binding to laminin disrupts the PI3K/AKT pathway and survival signaling in muscle cells. *Muscle Nerve*, **26**, 644–653.
36. Munoz, J., Zhou, Y. and Jarrett, H.W. (2010) LG4-5 domains of laminin-211 binds alpha-dystroglycan to allow myotube attachment and prevent anoikis. *J. Cell. Physiol.*, **222**, 111–119.
37. Morgan, J.E. and Zammit, P.S. (2010) Direct effects of the pathogenic mutation on satellite cell function in muscular dystrophy. *Exp. Cell Res.*, **316**, 3100–3108.
38. Ross, J., Benn, A., Jonuschies, J., Boldrin, L., Muntoni, F., Hewitt, J.E., Brown, S.C. and Morgan, J.E. (2012) Defects in glycosylation impair satellite stem cell function and niche composition in the muscles of the dystrophic Large(myd) mouse. *Stem Cells*, doi: 10.1002/stem.1197.
39. Sacco, A., Mourkioti, F., Tran, R., Choi, J., Llewellyn, M., Kraft, P., Shkreli, M., Delp, S., Pomerantz, J.H., Artandi, S.E. and Blau, H.M. (2010) Short telomeres and stem cell exhaustion model Duchenne muscular dystrophy in mdx/mTR mice. *Cell*, **143**, 1059–1071.
40. Gensch, N., Borchardt, T., Schneider, A., Riethmacher, D. and Braun, T. (2008) Different autonomous myogenic cell populations revealed by ablation of Myf5-expressing cells during mouse embryogenesis. *Development*, **135**, 1597–1604.
41. Gregorevic, P., Allen, J.M., Minami, E., Blankinship, M.J., Haraguchi, M., Meuse, L., Finn, E., Adams, M.E., Froehner, S.C., Murry, C.E. and Chamberlain, J.S. (2006) rAAV6-Microdystrophin preserves muscle

function and extends lifespan in severely dystrophic mice. *Nat. Med.*, **12**, 787–789.
42. Yoshimura, M., Sakamoto, M., Ikemoto, M., Mochizuki, Y., Yuasa, K., Miyagoe-Suzuki, Y. and Takeda, S. (2004) AAV vector-mediated microdystrophin expression in a relatively small percentage of mdx myofibers improved the mdx phenotype. *Mol. Ther.*, **10**, 821–828.
43. Pacak, C.A., Walter, G.A., Gaidosh, G., Bryant, N., Lewis, M.A., Germain, S., Mah, C.S., Campbell, K.P. and Byrne, B.J. (2007) Long-term skeletal muscle protection after gene transfer in a mouse model of LGMD-2D. *Mol. Ther.*, **15**, 1775–1781.
44. Nathwani, A.C., Tuddenham, E.G., Rangarajan, S., Rosales, C., McIntosh, J., Linch, D.C., Chowdary, P., Riddell, A., Pie, A.J., Harrington, C. *et al.* (2011) Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N. Engl. J. Med.*, **365**, 2357–2365.
45. Pane, M., Messina, S., Vasco, G., Foley, A.R., Morandi, L., Pegoraro, E., Mongini, T., D'Amico, A., Bianco, F., Lombardo, M.E. *et al.* (2012) Respiratory and cardiac function in congenital muscular dystrophies with alpha dystroglycan deficiency. *Neuromuscul. Disord.*, **22**, 685–689.
46. Kanagawa, M., Omori, Y., Sato, S., Kobayashi, K., Miyagoe-Suzuki, Y., Takeda, S., Endo, T., Furukawa, T. and Toda, T. (2010) Post-translational maturation of dystroglycan is necessary for pikachurin binding and ribbon synaptic localization. *J. Biol. Chem.*, **285**, 31208–31216.
47. Fukada, S., Yamaguchi, M., Kokubo, H., Ogawa, R., Uezumi, A., Yoneda, T., Matev, M.M., Motohashi, N., Ito, T., Zolkiewska, A. *et al.* (2011) Hesr1 and Hesr3 are essential to generate undifferentiated quiescent satellite cells and to maintain satellite cell numbers. *Development*, **138**, 4609–4619.
48. Segawa, M., Fukada, S., Yamamoto, Y., Yahagi, H., Kanematsu, M., Sato, M., Ito, T., Uezumi, A., Hayashi, S., Miyagoe-Suzuki, Y. *et al.* (2008) Suppression of macrophage functions impairs skeletal muscle regeneration with severe fibrosis. *Exp. Cell Res.*, **314**, 3232–3244.
49. Shin, J.H., Nitahara-Kasahara, Y., Hayashita-Kinoh, H., Ohshima-Hosoyama, S., Kinoshita, K., Chiyo, T., Okada, H., Okada, T. and Takeda, S. (2011) Improvement of cardiac fibrosis in dystrophic mice by rAAV9-mediated microdystrophin transduction. *Gene Ther.*, **18**, 910–919.

OPEN ACCESS Freely available online

PLOS | ONE

Genome-Wide DNA Methylation and Gene Expression Analyses of Monozygotic Twins Discordant for Intelligence Levels

Chih-Chieh Yu[1☯], Mari Furukawa[1☯], Kazuhiro Kobayashi[1], Chizuru Shikishima[2], Pei-Chieng Cha[1], Jun Sese[3], Hiroko Sugawara[4], Kazuya Iwamoto[5], Tadafumi Kato[4], Juko Ando[6], Tatsushi Toda[1]*

1 Division of Neurology/Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe University, Kobe, Japan, **2** Keio Advance Research Centers, Keio University, Tokyo, Japan, **3** Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan, **4** Laboratory for Molecular Dynamics of Mental Disorders, RIKEN Brain Science Institute, Saitama, Japan, **5** Department of Molecular Psychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, **6** Faculty of Letters, Keio University, Tokyo, Japan

Abstract

Human intelligence, as measured by intelligence quotient (IQ) tests, demonstrates one of the highest heritabilities among human quantitative traits. Nevertheless, studies to identify quantitative trait loci responsible for intelligence face challenges because of the small effect sizes of individual genes. Phenotypically discordant monozygotic (MZ) twins provide a feasible way to minimize the effects of irrelevant genetic and environmental factors, and should yield more interpretable results by finding epigenetic or gene expression differences between twins. Here we conducted array-based genome-wide DNA methylation and gene expression analyses using 17 pairs of healthy MZ twins discordant intelligently. *ARHGAP18*, related to Rho GTPase, was identified in pair-wise methylation status analysis and validated via direct bisulfite sequencing and quantitative RT-PCR. To perform expression profile analysis, gene set enrichment analysis (GSEA) between the groups of twins with higher IQ and their co-twins revealed up-regulated expression of several ribosome-related genes and DNA replication-related genes in the group with higher IQ. To focus more on individual pairs, we conducted pair-wise GSEA and leading edge analysis, which indicated up-regulated expression of several ion channel-related genes in twins with lower IQ. Our findings implied that these groups of genes may be related to IQ and should shed light on the mechanism underlying human intelligence.

Citation: Yu C-C, Furukawa M, Kobayashi K, Shikishima C, Cha P-C, et al. (2012) Genome-Wide DNA Methylation and Gene Expression Analyses of Monozygotic Twins Discordant for Intelligence Levels. PLoS ONE 7(10): e47081. doi:10.1371/journal.pone.0047081

Editor: Valerie W. Hu, The George Washington University, United States of America

Received February 9, 2012; **Accepted** September 11, 2012; **Published** October 17, 2012

Funding: This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas (22129006 to TT) from the Ministry of Education, Culture, Sports, Science and Technology of Japan; by Grants-in-Aid for Scientific Research (S) (21223002 to JA and TT); by Scientific Research (B) (20390099 to TT); and by Challenging Exploratory Research (23650136 to KK) from the Japan Society for the Promotion of Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: toda@med.kobe-u.ac.jp

☯ These authors contributed equally to this work.

Introduction

Individual differences in cognitive abilities have long been an intriguing phenomenon to both lay people and scientists. Differences in intelligence, as measured by IQ tests, appear to remain relatively stable from childhood to late life [1,2]. Further, the fact that intelligence, in the general population, has a normal distribution and long-term constancy allows for the assumption of the nature of IQ as being, at least partly, hereditary with quantitative trait features. In effect, the heritability of intelligence is estimated to be somewhere between 30% to over 80% in classical MZ twin versus dizygotic twin studies [3–5], marking it one of the highest among human quantitative traits.

Despite the significant role that genes are supposed to play in deciding an individual's cognitive abilities, progress to identify intelligence-related genes in healthy adults is not as promising [6,7], and contrasts the increasing list of some 300 genes associated with mental retardation [8]. One possible explanation for the lack of replicated genetic findings in normal-range intelligence is the small effect size of each gene. The fact that genome-wide association studies in the scale of thousands of subjects identified no specific genetic variants associated with human intelligence implies that very large sample sizes are necessary to detect individual loci [9].

As underpowered studies face challenges in the attempt to identifying small effect quantitative trait loci, twin research might provide an alternative. Twin studies serve more than a means to estimate heritability of the aforementioned complex traits; they also present an important resource to evaluate quantitative trait loci. There are accumulating evidences that long thought to be genetically identical MZ twins manifest variations in copy number [10] or point mutations limited to one twin [11]; however, these genomic discordances have failed to explain all phenotypic discrepancies [12]. Differences in phenotypes between MZ twins can possibly be attributed to environmental factors, as well as epigenetic variants, which refer to the gene expression-related modifications that occur without altering DNA sequences.

Epigenetic regulatory mechanisms have been reported to be associated with a number of biological phenotypes, including intelligence [13]. Among all known epigenetic mechanisms, DNA methylation has been the most extensively studied. Specifically, such studies have observed patterns of negative correlation between promoter region methylation and gene expression [14]. Additionally, studies based on phenotypically discordant MZ twins have revealed that those who are as best matched for genetics, gender, age, prenatal influences, and shared environment as nature could provide, are considered to possess the potential to detect epigenetic and transcriptomic differences, although research has yet to yield exclusive results [12,15].

In the present study, we recruited 17 healthy MZ twin pairs who manifested discordance for intelligence between co-twins (i.e., more than 1 standard deviation (SD)). Regarding that MZ twins are identical in genetic composition, the IQ difference could be associated with environmental factors, via epigenetic mechanisms regulations. By analyzing array-based genome-wide DNA methylation and gene expression profiles, a novel list of genes with functions related to protein synthesis, DNA helicase activities, and ion channels was generated. To our knowledge, this is the first study tested for epigenetic and expression differences between phenotypically normal, yet discordant, MZ twins via genome-wide approaches.

Results

General characteristics of participants

This study is a part of the Keio Twin Study project [16]. We have collected 240 MZ twin pairs with IQ scores of both siblings tested (Fig.S1). Seventeen MZ twin pairs (5 male pairs and 12 female pairs) aged 25.1±2.5 years (range, 21–31 years), with IQ scores of normal range yet manifesting significant between co-twins differences formed the present study sample (Table 1). Mean IQ score of all 34 participants was 100.91±13.32 (range, 67–139), while the mean difference between co-twins was 20.76 points (range, 15–45). No documented psychological or physiological conditions were noted at the time of recruitment. Since there is no standard definition of discordance in MZ twins IQ scores, we adopted 15-point-difference as the principle inclusion criterion. Fifteen-point is not only one standard deviation of IQ scores in general populations, but it is also the average IQ difference for genetically unrelated individuals sharing family environments, whereas identical twins differ by only about 6 IQ points on average [17].

27 genes identified by screening for epigenetically regulated candidate loci

We analyzed the methylation profiles of 25,500 human promoters utilizing methylated DNA enriched genomic DNA derived from peripheral blood cells. We used MAT (model-based analysis of tiling arrays) program [18] to identify genes that significantly differed between co-twins pairwisely. With a significance threshold set to $p<10^{-6}$, a total of 27 genes were recognized in 13 of the 17 twin pairs; however, none was shared in plural pairs (Table S1). In addition, a moderate positive correlation of 0.417 (Spearman's rank correlation coefficient, $p=0.048$) was found between the number of genes that epigenetically differed and pair-wise IQ differences (Fig. S2). On the other hand, after applying the log signal ratio (of the higher IQ twins to the lower IQ co-twins) to one-class t-test with the same significance threshold across all 17 twin pairs, we identified no locus manifesting significantly different methylation status.

Validation of *ARHGAP18* by bisulfite sequencing and quantitative RT-PCR

We performed a sodium bisulfite analysis to confirm the methylation status at the 27 gene loci putatively identified as being differentially methylated. After sequencing at least 30 clones for each locus, statistically significant difference in methylation status between co-twins was validated on two genes, *ARHGAP18* (Rho GTPase activating protein 18, $p=5.12806\times10^{-8}$; chi-square test) and *OR4D10* (olfactory receptor 4D10, $p=0.0234658$; chi-square test) (Fig. 1A and B).

After confirmation of methylation status, we correlated the data with their expression levels using total RNA derived from lymphoblast cell lines by qRT-PCR. The observed reduction in DNA methylation status of *ARHGAP18* was correlated with its increased expression level in the subject with lower IQ scores of the twin pair from which the gene was identified (2.04 fold, $p=0.024767$; Mann-Whitney U test) (Fig. 1C, left graph). Increased transcription levels were observed after a 3-day treatment of the demethylating agent 5-aza-2-deoxycytidine (5-azadC), and the difference between the twins was eliminated, suggesting that the expression is regulated by the methylation of the promoter (Fig. 1C, right graph). Meanwhile, no expression of *OR4D10* could be detected in the lymphoblast cell lines.

No gene manifesting statistically significant differences between the group of twins with higher IQ and that of their co-twins after the expression array analysis

Apart from the epigenetic approach described above, we used an expression microarray analysis to directly compare the genome-wide gene expression profiles of the higher IQ twins versus their lower IQ co-twins. At first, principal component analysis (PCA) was performed to facilitate visualization of the relationships between groups (Fig. S3). As a result, the two groups compositing twins with higher or lower IQ scores could not be readily distinguished. Likewise, the dendrograms, produced by hierarchical clustering, also failed to demonstrate differences in general expression patterns between these two phenotypic groups (Fig. S4). Next, ANOVA analysis was applied to identify whether there were differentially expressed genes between these two groups; however, no gene met the criteria of a FDR (false discovery rate)-adjusted p of 0.05. A one-class t-test analysis with multiple sample correction was conducted across all log ratios. Similarly, no significantly differentially expressed gene was identified.

Candidate genes list resulted from direct pair-wise comparison of the expression array data

Next, we applied a direct pair-wise comparison to focus on the genes up-regulated in the same tendency. Fold-change values of the expression levels of all genes were first calculated for each twin pair, from which genes with a fold-change value more than 2 were included (Dataset S1). We then generated a list of candidate genes by picking up those replicated in most twin pairs (Table 2). *UCHL1* (ubiquitin carboxyl-terminal esterase L1), along with the other 7 genes, were found to have higher expression levels in the higher IQ twins of at least 4 pairs, while 6 genes were up-regulated in some lower IQ twins.

Identification of 3 genes with borderline significance by grouping samples according to individual gene expression level

To further increase the possibility of identifying candidate genes, we performed an analysis based on individual gene expression level. After dividing the twin siblings of each pair into

Table 1. Background data for participants.

			IQ Scores* (Points)		
Twin Pair ID	Age (year)	Gender	Twin A	Twin B	IQ score differences
1	31	F	99	82	17
2	30	F	110	91	19
3	22	M	82	112	−30
4	20	F	100	78	22
5	29	F	90	108	−18
6	20	F	86	104	−18
7	24	F	93	76	17
8	25	F	126	106	20
9	21	M	97	80	17
10	27	F	123	104	19
11	22	F	93	108	−15
12	23	M	97	117	−20
13	27	M	121	139	−18
14	23	M	126	108	18
15	26	F	112	67	45
16	26	F	108	90	18
17	24	F	110	88	22

*IQ scores calculated after participants took the full version of Kyodai Nx15- test.

doi:10.1371/journal.pone.0047081.t001

higher and lower expression groups according to the expression level of every gene, a paired t-test was carried out to determine if there was a significant difference between the mean IQ scores of the two groups. Whereas not a single gene reached the corrected cutoff p of 10^{-6}, 3 genes, *RFK* (riboflavin kinase), *RPL12* (ribosomal protein L12), and *RMRP* (RNA component of mitochondrial RNA processing endoribonuclease), manifested borderline significance (Fig. 2). The twins manifesting up-regulated expression level of *RFK* showed the tendency to have lower IQ scores than their co-twins, while *RPL12* and *RMRP* might likely to contribute to higher intelligence.

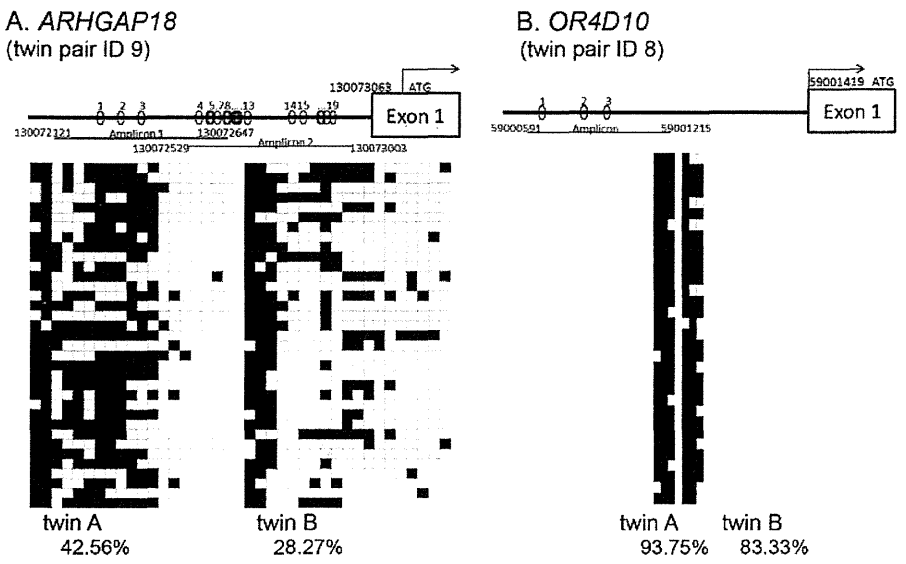
Identification of 4 differentially expressed gene sets by GSEA

It remained possible that functionally-related genes might have important gene expression changes in a set-wise matter without any individual transcript meeting the criteria of significance. Using GSEA [19] under the cutoff FDR q-value of <0.25, we denoted 1 and 4 up-regulated gene sets from KEGG and Gene Ontology (GO) database respectively in the group of higher IQ twins, while 1 gene set each from KEGG and Reactome pathway database was found to be up-regulated in the group of lower IQ twins (Table S2 and Fig. 3). From the results employing GO database, the leading edge analysis revealed 8 genes (*MRPS35*, *MRPL23*, *MRPL52*, *MRPL41*, *MRPL12*, *MRPS15*, *MRPS22*, and *MRPL55*) with core enrichment in the gene sets "Organellar Ribosome," "Ribosomal Subunit," and "Mitochondrial Ribosome", whereas the fourth gene set "ATP-dependent DNA Helicase Activity" was comprised of 7 other genes (*XRCC5*, *XRCC6*, *DHX9*, *PIF1*, *G3BP1*, *RUVBL2*, and *CHD4*) with core enrichment. By utilizing Reactome database, all 6 genes from the gene set "Reactome CREB phosphorylation through the activation of CAMKII" were enriched.

A different approach was carried out to the same end. Specifically, we performed a GSEA on each twin pair. Depending on the twin pairs analyzed and pathway databases used, as many as 284 gene sets were found to be significantly different between the siblings (Dataset S2, S3, S4, S5). Gene sets replicated in most twin pairs were listed. In general, pathways related to DNA replication, ribosomes, and proteases were found in higher IQ twins of most twin pairs, while cell signaling associated ones tended to be up-regulated in lower IQ twins (Table 3). In order to focus on those genes which effectively contributed to the enrichment of each given gene set, we first generated up- and down-regulated leading edge subsets for each twin pair, and then extracted those genes that were nominated most often across plural twin pairs (Table 4). Up-regulation of *IGF1* was found in the higher IQ twins of 4 pairs, whereas potassium channel-coding genes *KCNE2* and *KCNQ3*, along with an acetylcholine receptor-coding gene *CHRNA2* manifested higher expression levels in some lower IQ twins. Among all the candidates, *IGF1* was selected for further analysis for its important role in growth and development [20]. We performed bisulfite sequencing of the promoter regions for the 4 twin pairs manifesting an up-regulated expression level in the higher IQ siblings. However, no significant differences were noted in the methylation status of the 2 promoter domains (P1 and P2) between the siblings (Table S3).

Discussion

Given that the concept of general cognitive ability, designated as *g*, has been widely accepted to depict a near-universal positive covariation among diverse measures of cognitive abilities, naming even one genetic locus that is reliably related with normal-range intelligence remains challenging [21]. IQ is easy to quantify and compare among different individuals. Although not conclusively, the substantial *g*-loading for IQ [22] justifies its role to represent the general intelligence levels. Benefiting from the extraordinary similarities in genomic constitution and environmental factors,



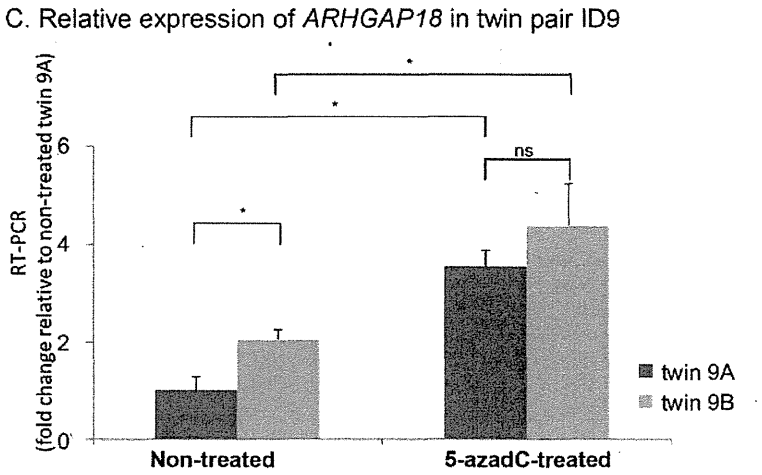


Figure 1. DNA methylation status of the 5′-regions of *ARHGAP18* and *OR4D10* analyzed by bisulfite sequencing along with the quantitative RT-PCR analysis for *ARHGAP18* expression. A, B. DNA methylation status analyzed by bisulfite sequencing. Schematic representation (top) for the relative position of CpGs within amplified regions and methylation profiling by bisulfite sequencing (bottom). The numbers at the ends of amplicons indicate the genome coordinates relative to the NCBI Build 36 genome assembly. Of all 27 loci identified by screening for epigenetically regulated candidate genes, A. *ARHGAP18* and B. *OR4D10* were validated by direct bisulfite sequencing. At least 30 clones were sequenced for each locus. Open squares indicate unmethylated CpG nucleotides and closed squares indicate methylated ones. Rows indicate the methylation status of each colony sequenced, while columns indicate the positions of CpG nucleotides.The percentages below refer to the ratio of CpG methylation. C. qRT-PCR analysis of *ARHGAP18* mRNA relative to *GAPDH* in the twin pair ID 9. *ARHGAP18* expression in twin 9A untreated with 5-azadC was normalized to 1. Error bars indicate ± SD. *P*-value was calculated using Mann-Whitney U test with asterisks indicating statistical significance ($p<0.05$). ns: not significant.
doi:10.1371/journal.pone.0047081.g001

studies based on discordant monozygotic twins, even with limited sample size (as small as 20–50 twin pairs), are capable of uncovering phenotype-associated epigenetic changes independent of underlying sequence variance [23]. By successfully recruiting 17 pairs of identical twins discordant for intelligence levels, we shall have a modest power in the identification of intelligence-related epigenetic differences. To our knowledge, this is the first genome-wide methylation and gene expression study administering the characteristics of monozygosity to access the epigenetic and expression changes for a quantitative trait.

Researchers have reported that patterns of epigenetic modifications in MZ twins diverge as they age [24]. Provided that all 17 twin pairs in this study were in early adulthood, it might not be surprising that only few loci revealed significant differences in methylation status. Of the 27 candidate loci nominated by promoter-arrays-based methylation analyses, bisulfite sequencing successfully validated only 2 genes. One explanation of the discrepancy between these two methods is that we adopted a less stringent criterion of p-value (10^{-6}, instead of a Boferroni corrected p-value of 10^{-8} considering that GeneChip Human

Table 2. List of genes having the same tendency of expression level in most twin pairs.

Up-regulated in co-twins with	Gene	Shared by twin pair ID
Higher IQ scores	*GTSF1*	1,2,7,11,14,15,16
	AK3L1	1,3,9,11,17
	PRKCH	2,9,11,14,17
	CDRT1	2,9,10,11,14
	LRIG3	1,2,9,15
	VSIG6	2,10,11,13
	SNORA20	3,6,10,12
	UCHL1	8,9,11,15
Lower IQ scores	*CD96*	1,2,4,9,14
	CXCL10	1,6,9,11
	CDC42BPA	1,2,6,9
	CXCR4	1,2,6,14
	EPS8	2,9,10,16
	FAM169A	3,9,11,15

doi:10.1371/journal.pone.0047081.t002

Promoter 1.0 Array contains 4.6 million probes). As a result, we were able to detect more candidate loci from the practically congruent MZ twin samples, only at the expense of precision rate. The technical limitations of MethylMiner collection might also contribute to false positives. After bisulfite sequencing and qRT-PCR validation, we identified one candidate gene, *ARHGAP18*, which encodes one of the Rho GTPase-activating proteins (GAPs) that modulate cell proliferation, migration, intercellular adhesion, cytokinesis, proliferation differentiation, and apoptosis [25]. Mutations in a handful of Rho-linked genes were documented to be associated with X-linked mental retardation [8], by which the importance of GAP activity in normal neuronal functions was proposed [26]. Notwithstanding the identification of *ARHGAP18* in a genome-wide association study for schizophrenia [27], it had not been previously connected to cognitive abilities until the present study.

We were not able to identify a single gene that displayed significantly different expression level between the group of twins with higher IQ scores and their co-twins. From the list of candidate genes generated by direct pair-wise comparison, UCHL1 is a brain-specific de-ubiquitinating enzyme. While the substrates are still unknown, loss of its enzyme activity has been reported in neurological diseases such as Alzheimer's disease and Parkinson's disease [28]. In a different approach, *RFK*, *RPL12*, and *RMRP* showed borderline significance. *RFK* encodes riboflavin kinase, an essential enzyme to form flavin mononucleotide, is important in a wide range of biological metabolisms [29]. *RPL12* encodes a ribosomal protein of the 60S subunit, while *RMRP* encodes the RNA component of mitochondrial RNA processing endoribonuclease. Although none of these 3 genes had ever been connected with cognitive functions, it remains possible that their biophysical characteristics might become more pronounced in cells having as high a metabolic rate as neurons.

In the gene set based approach, GSEA of between-group and between co-twin comparisons revealed several mitochondrial ribosomal protein-coding genes. Mitochondria, which are responsible for most of the energy requirement for cellular metabolism, have their own translation system for the 13 proteins essential for oxidative phosphorylation in mammals. All 78 human mitochondrial ribosome proteins are translation products of nuclear genes, of which some were identified as candidate genes for several congenital diseases [30]. No exclusive conclusion about the connection of mitochondrial ribosomal function and cognitive ability could be drawn before being further validated. Nevertheless, we hypothesized that, for the highly differentiated and high energy-demanding central nervous system, essential proteins for

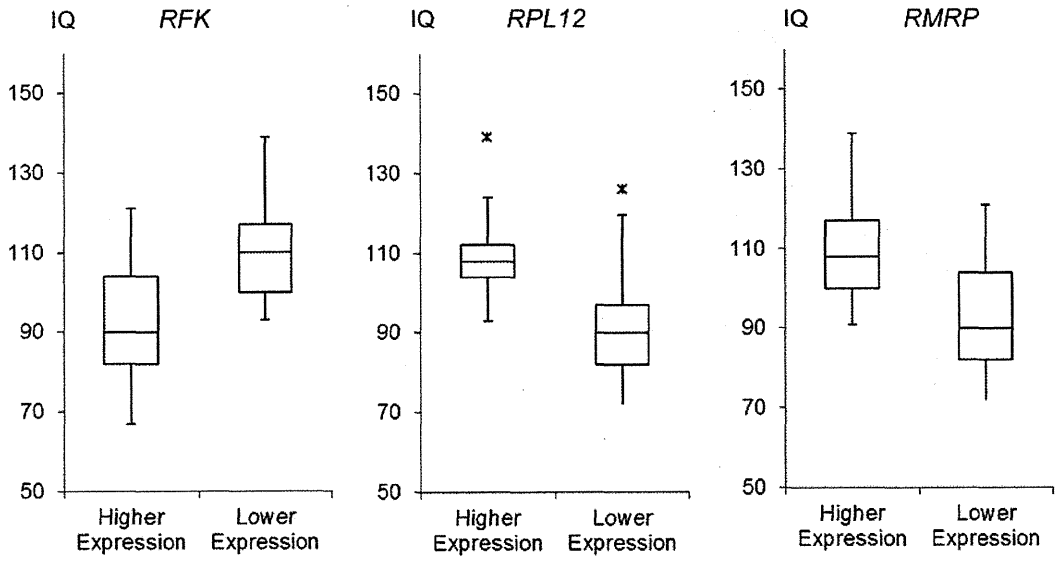


Figure 2. Three genes manifesting borderline significance by grouping samples according to individual gene expression level. IQ scores in subjects were grouped according to the relatively higher (Higher Expression) or lower (Lower Expression) expression levels within each twin pair. Three genes, *RFK*, *RPL12*, and *RMRP*, manifested the smallest p yet failed to meet the cutoff value of $p = 10^{-6}$. Data are presented as box plots (minimum, 25% quartile, median, 75% quartile, maximum). The red asterisks indicate maximum outliners.
doi:10.1371/journal.pone.0047081.g002

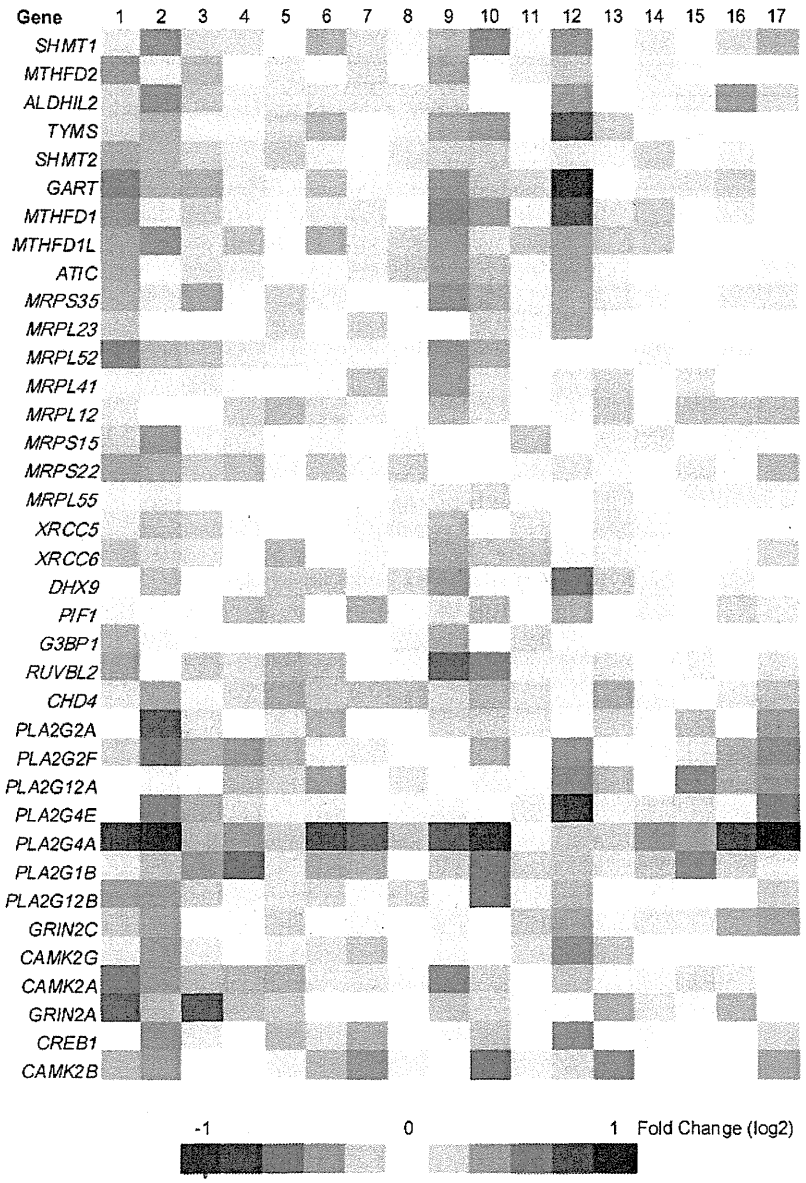


Figure 3. Heat map of the 37 genes with core enrichment from up-regulated genes sets. GSEA analysis was carried out to identify if any pre-defined gene set showing different expression levels between the group of higher IQ twins and the group of their lower IQ co-twins. Gene set databases BioCarta, KEGG, Reactome, and Gene Ontology were applied separative. Gene sets meeting the cutoff FDR *q*-value of 0.25 were subjected to leading edge analysis to determine the genes with core enrichment. The result was a list of 37 genes and we generated a heat map accordingly. Red and blue cells signify genes that were either up- or down-regulated, respectively, after the expression levels of the twins with higher IQ scores compared to their co-twins. The scale represents fold changes in log2 values, according to the color map at the bottom of the figure. The general tendency of higher expression levels in twins with higher IQ scores from the gene *SHMT1* to *CHD4*, and their lower expression levels from the gene *PLA2G2A* to *CAMK2B*, was visualized.
doi:10.1371/journal.pone.0047081.g003

mitoribosome function might play a role in the maintenance of neuronal biological processes.

Apart from mitochondrial ribosomal protein-related gene sets, "ATP-dependent DNA Helicase Activity" from the GO database was also found. DNA helicases are molecular motor proteins that use nucleoside 5'-triphosphate hydrolysis as a source of energy to open energetically stable duplex DNA into single strands. As such, they are essential in almost all aspects of cellular DNA machinery including DNA replication, repair, recombination, and transcription [31]. Of which, *XRCC5* and *XRCC6* encode the two subunits of the Ku protein, which plays an important role in the repair of double-stranded DNA breaks and telomere protection [32]. In neurodegenerative diseases, such as Alzheimer's disease, where cellular damage due to oxidative stress is proposed to contribute to pathophysiology, reduced Ku protein expression and its DNA binding activity have been thought to be involved [33]. Of the remaining five helicases denoted, *G3BP1* was demonstrated to play an essential part in proper embryonic growth and neonatal

Table 3. Summary of the most commonly shared gene sets showing the same tendency (up-regulated in the higher IQ twin or in the lower IQ twin) according to the pair-wise GSEA test.

Up-regulated in	Gene sets database	Gene set	Shared by twin pair ID
Higher IQ twins	BioCarta	MCM pathway	1,9,12,17
	KEGG	Proteasome	1,3,9,10,17
		Oxidative phosphorylation	1,9,10,12,16
		Ribosome	1,6,9,10,12,16
		DNA replication	1,3,8,9,10,12,17
		Valine leucine and isoleucine degradation	1,3,10,12,17
		Mismatch repair	1,3,9,12,17
		Peroxisome	1,9,10,12,17
	Reactome	RNA polymerase I promoter opening	1,3,9,11,12,13,16,17
		Electrotransport chain	1,3,9,10,12,13,16,17
		Packaging telomere ends	1,3,9,11,12,16,17
		Telomere maintenance	1,3,9,11,12,16,17
	Gene Ontology	*Structural constituent of ribosome*	9,10,12,17
		Microbody	1,9,12,17
		Peroxisome	1,9,12,17
		S-phase of mitotic cell cycle	1,8,15,17
		Small conjugating protein specific protease activity	1,9,15,17
Lower IQ twins	BioCarta	n/a	n/a
	KEGG	Neuroactive ligand receptor interaction	1,3,16,17
		Taste transduction	1,3,9,13,17
	Reactome	Cell-cell adhesion systems	1,3,9,17
	Gene Ontology	Anion transport	1,5,16,17
		Cation channel activity	1,9,16,17
		Cell-cell signaling	1,9,16,17
		Collagen	1,9,16,17
		Extracellular matrix part	1,9,16,17
		Extracellular region part	1,9,16,17
		G-protein coupled receptor protein signaling pathway	1,9,16,17
		Gated channel activity	1,9,16,17
		Intercellular junction	6,9,16,17
		Metal ion transmembrane transporter activity	1,9,16,17
		Second messenger mediated signal	1,9,16,17

n/a indicates no gene set was shared by at least 4 twin pairs.

doi:10.1371/journal.pone.0047081.t003

survival [34]. Although the direct associations between these seven genes and cognitive abilities have not been depicted, mutations in a list of DNA repair-related genes have already been reported to cause mental retardation [8]. As such, one hypothesis we proposed here is that the up-regulated expression of these helicases might provide better protection from oxidative damages and, thus, improve neuronal function and survival, which could bring forth higher levels of intelligence (or in other words, less compromised) as a phenotype.

On the other hand, the pair-wise GSEA, along with leading edge analysis, identified *CHRNA2* which encodes the α2 subunit of nicotinic acetylcholine receptors (nAChRs). Initially related to nicotine dependence, the role of nAChRs in cognitive performance has gained attention because nicotine is considered a powerful enhancer of cognitive capabilities [35] via the interaction of nicotine and nAChRs [36].

Additionally, two potassium voltage-gated channel-coding genes, *KCNE2* and *KCNQ3*, were identified. Voltage-gated ion channels possess diverse functions, include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, and smooth muscle contraction. By assembling with KCNQ2 or KCNQ5, KCNQ3 forms the M channel, a slow activating and deactivating potassium channel that plays a critical role in the regulation of neuronal excitability [37]. In addition to being identified as one cause for a dominantly inherited form of human generalized epilepsy, called benign familial neonatal convulsions, the electrogenic characteristics of KCNQ/M channels have importance in controlling intrinsic firing patterns of principal hippocampal neurons, thus, further modulating hippocampal learning and memory [38]. Of note, rats treated with Linopirdine, an M channel-specific inhibitor,

Table 4. Up-regulated genes with core enrichment shared by co-twins of multiple pairs.

Up-regulated in	Gene Sets database	Gene	Shared by twin pair ID
Higher IQ twins	BioCarta	*CASP6*	1,10,12,17
		CCNE1	1,9,12,17
		IGF1	1,2,7,17
		IL1a	2,7,8,15
		LTA	2,8,15,17
		C7	2,3,8,15
		IL6	2,3,7,15
		STAT4	3,8,10,17
	KEGG	*MOM6*	1,3,8,9,12,17
		RPA2	1,3,8,9,12,17
		POLA2	1,3,8,9,12
		PRIM2	1,3,8,9,12
		MOM2	1,3,8,9,12
		LTA	2,3,10,15,16
		SDHB	1,3,9,10,16
		POLE2	1,3,8,9,17
		COX7C	1,9,10,12,16
		COX8A	1,9,10,12,16
		NDUFB7	1,9,10,12,16
		NDUFA8	1,9,10,12,16
		NDUFA7	1,9,10,12,16
		SSBP1	1,8,9,12,17
	Reactome	*NUP93*	1,3,9,10,11,12,17
		POLE2	1,3,9,11,16,17
		POLR2L	1,3,9,11,12,16
		POLR2G	1,3,9,12,16,17
		PSMD14	1,3,9,11,12,17
		PSMC6	1,3,9,11,12,17
		RPA2	1,3,9,11,12,17
		NUP85	1.3.9.10.11.12
		NUP37	1.3.9.11.12.17
		NUP205	1.3.9.11.12.17
		CDC6	1.3.9.11.12.17
		CDC25A	1.3.9.11.12.17
		ORC4L	1.3.9.11.12.17
		KMCM6	1.3.9.11.12.17
		PRIM1	1.3.9,11,16,17
	Gene Ontology	n/a	n/a
Lower IQ Twins	BioCarta	n/a	n/a
	KEGG	*EP300*	1,3,6,12,13
		ACTN1	1,6,12,13,14
		PLOB2	1,3,8,13
		PIK30G	1,6,8,13
		CREBBP	1,3,12,13
		MAPK9	1,2,6,13
		LEF1	1,3,6,12
		PIK3R3	2,6,8,13
		AKR1O4	3,8,16,17
		PARD3	6,8,12,16
	Reactome	*MNAT1*	2,6,7,14
		RFC2	2,6,7,14
		PRIM1	2,6,7,14
		PRIM2	2,6,7,14
	Gene Ontology	*CHRNA2*	1,9,16,17
		KCNE2	1,9,16,17
		KCNQ3	1,9,16,17
		INHBA	1,9,14,17
		SLC34A3	1,5,16,17

n/a, no gene was shared by at least 4 twin pairs.

doi:10.1371/journal.pone.0047081.t004

demonstrated improved performance in various tests of learning and memory [39].

Identified by utilizing BioCarta pathway database, *IGF1* manifested up-regulation in higher IQ twins. The insulin-like growth factor (IGF) system is important in growth and development. While the exact mechanism remains unknown, the growth hormone (GH) and IGF-1 axis has been reported to play a role in the reduction of cognitive functions in aging population and patients with GH deficiency [20]. Methylation status of its promoter regions was studied, yet no difference was discerned between the twins. It is possible that *IGF1* is under some other epigenetic regulation, considering the actual mechanisms responsible for the cell type-specific expression patterns of this gene remain to be elucidated [40].

Since intelligence is a complex trait associated with many genes of small effect [41,42], it was not surprising that we failed to identify a single gene manifesting prevailing expression changes across all 17 twin pairs. We also noticed that by microarrays none of the candidate genes identified by methylation analyses was listed in the results of expression studies. A reason of the discrepancy of these two methods may be the lack of comprehensive accession to all known epigenetic regulations. DNA methylation at CpG sites across promoter regions has been deeply studied, while a number of other epigenetic regulatory mechanisms are also found to modulate gene expression [22]. Furthermore, ever since the genome-wide, single-base human DNA methylome mapping became possible, the correlation of methylation status of gene bodies and expression levels has been gaining attention [43]. It has been documented that DNA methylation of gene bodies is associated with gene activity. Therefore, it is possible that the intelligence-related expression profiles were subjected to this novel epigenetic regulation.

The present study has its conceptual and technical limitations. Conceptually, we hypothesized that changes in methylation status and expression levels could be captured in genomic DNA and total RNA extracted from whole blood and derivative lymphoblastic cell lines, respectively. Given the inability to probe DNA methylation status or gene expression in the human brain, except in postmortem studies, human blood is commonly used in transcriptional studies of various diseases including psychiatric disorders [44]. Although there is still no consensus regarding blood-based gene expression profiles as good surrogates for addressing neuroscientific research, the moderate correlation

between transcripts in whole blood and the central nervous system makes it an accessible alternative [45,46]. Several studies utilizing similar strategies to study twin pairs discordant for psychiatric disorders through comparisons of CpG islands methylation of peripheral blood cells or lymphoblast cell lines had detected a number of disease-associated epigenetic changes [47–49]. Moreover, it has been documented that the methylation changes of large-scale domains are linked to cell-specific differentiation [50]. Several functional gene sets we found (i.e., "cation channel activity", "cell-cell signaling", "extracellular region part", "G-protein coupled receptor protein signaling pathway", and "gated channel activity") were observed within neuronal highly methylated domains. The association between expression difference observed in lymphoblast cell lines and neuron-specific methylation patterns implies that these candidate gene sets are more likely to reflect the true differences in brains. Technical limitations of this study include low fold-change differences in expression levels manifested between co-twins. We did not carry out qRT-PCR for the genes identified by GSEA, considering that differences of less than 1.5-fold are thought to be beyond the limit of reproducibility [51].

Reverse causality should be considered in epigenetic studies, considering all known epigenetic marks are influenced by environmental exposures including diet, smoking, alcohol consumption, stress, or physical activities [52], It might be plausible that the changes we observed in this study resulted from the divergent lifestyle choices by subjects with different levels of intelligence, and not that these epigenetic changes caused the twins to differ.

Here, we presented the first study that used genome-wide epigenetic and transcriptomic profiling to identify epigenetic changes related to the discordance between MZ twins with normal-range intelligence. A list of new candidate genes possibly related to cognitive abilities was generated while further replications and functional analysis remain necessary.

Materials and Methods

Ethics statement

This study was conducted according to the principles expressed in the Declaration of Helsinki. Attendance was voluntary, and signed informed consent including information on genetic analyses was obtained from all participants. The Ethical Committees of Kobe University Graduate School of Medicine and Keio University Faculty of Letters approved study protocols.

Samples

A sub-sample of 326 twin pairs of twins from the Keio Twin Project were invited to Keio University, where the Kyodai Nx15-, one of the most often used group intelligence tests in Japan, was applied. The zygosity of participants was diagnosed by 15 polymorphic STR loci (AmpF, STR identifier kit, Applied Biosystems). Among the 240 pairs to have monozygosity, 34 MZ twin pairs who manifested differences in IQ score of more than 15 points between co-twins. One pair was excluded for a lower-than-normal IQ score (52 points). From the17 of the remaining 33 twin pairs, who agreed to participate in the study, peripheral blood was drawn and B-lymphoblastoid cell lines were established. For the treatment of 5-azadC (WAKO), a daily aliquot of 5 mM stock solution was added to flasks and thoroughly resuspended (final concentration of 1 mM). Cells were harvested after 3 days from the start of treatment.

Nucleic acids extraction

For human promoter microarrays and bisulfite genomic sequencing, genomic DNA was extracted from the blood via established methods. For gene expression microarrays and quantitative RT-PCR, total RNA was isolated using RNeasy Plus Mini kit (QIAGEN) from B-lymphoblastoid cell lines.

DNA methylation profiling

One microgram of genomic DNA was sonicated and subjected to methylated DNA enrichment using the MethylMiner methylated DNA enrichment kit (Invitrogen) as per the manufacturer's instructions. The methylated DNA fragments, amplified by the GenomePlex WGA reamplification kit 3 (SIGMA) and supplemented with dUTP, were further purified using the QIAquick PCR purification kit (QIAGEN).

According to Affymetrix's chromatin immunoprecipitation assay protocol, enriched methylated DNA was hybridized to GeneChip Human Promoter 1.0R arrays (Affymetrix), which comprised a coverage of over 25,500 human promoter regions.

AGCC (Affymetrix GeneChip Command Console)-format CEL files were first created, and then converted to GCOS (GeneChip Operating Software, Affymetrix)-format CEL files. For the pairwise analyses, paired CEL files were imported into MAT software to specify candidate regions (approximately 600 base pairs in length) with significantly different probe intensities between co-twins ($p<10^{-6}$).

To detect candidate loci across all 17 twin pairs, we utilized Partek Genomic Suite 6.5 software (Partek) to import the CEL files, and have the data converted to $\log_2$ values after normalized by the RMA (Robust Multichip Averaging) algorithm. After the signal from each probe for the higher-IQ sibling was subtracted from that of the lower-IQ co-twin across all probes, one-class *t*-test with statistical parameters set at $p<10^{-6}$ was carried out to detect significant regions.

Bisulfite sequencing

Genomic DNA was bisulfite-treated using the Methylcode bisulfite conversion kit (Invitrogen) as per the manufacturer's instructions. Amplification was performed with Takara LA Taq polymerase, with converted DNA-specific primers that were designed using MethPrimer. The amplicons were cloned into vectors using a TOPO TA cloning kit (Invitrogen). We performed direct sequencing of the plasmid DNA that was isolated using PI-200 auto-plasmid-isolator (KURABO) via the ABI 3730xl sequencing system (Applied Biosystems).

Quantitative RT-PCR

Two micrograms of the total RNA was subjected to reverse transcription using the SuperScript III first-strand synthesis system for RT-PCR (Invitrogen). Gene target amplifications, using Takara SYBR premix Ex Taq, were performed in triplicate in a matter of a serial 10 fold dilution. Housekeeping gene *GAPDH* served as the internal control gene. Mann-Whitney U test was performed to compare the relative expression levels between co-twins.

Gene expression profiling

We processed 300 nanograms of total RNA using the Ambion WT expression kit and the Affymetrix GeneChip WT terminal labeling kit according to the manufacturers' recommended methods. Hybridization and scanning of GeneChip Human Gene 1.0 ST arrays (Affymetrix), which comprised of more than 28,000 gene-level probe sets, were performed as per the manufacturer's

instructions. Partek Genomic Suite 6.5 software was used to import AGCC-format CEL files and normalize the data according to the RMA algorithm. ANOVA with an FDR-adjusted p set to 0.05 was used to determine those probe sets that were significantly different between the groups of twins with a higher IQ and their lower IQ co-twins. A one-class t-test analysis with multiple sample correction was conducted across all $\log_2$ ratios (higher-IQ twin/lower-IQ co-twin) for all 17 twin pairs. We also carried out pair-wise comparison for the expression array data and then included genes with a fold-change value more than 2. Genes replicated in the same tendency (up-regulated in the higher IQ twins, or up-regulated in the lower IQ twins) in most pairs were listed.

In another approach, the twins of each pair were categorized into higher and lower expression groups according to the expression level of every individual gene. A paired t-test was carried out to compare the mean IQ scores of the two groups. Corrected p of 10^{-6} was applied as the cutoff to define being positive.

GSEA was performed for functionally related genes across a spectrum of gene sets of C2 curated gene sets including BioCarta gene sets (217 gene sets), KEGG gene sets (186 gene sets), and Reactome gene sets (430 gene sets), and C5 GO gene sets (1,454 gene sets) separately. Pre-ranked gene lists, including lists with up-regulated/down-regulated genes in the group of twins with higher IQ scores sorted according to the p calculated by a between-group ANOVA test, and lists for each twin pair with genes sorted by between-sibling fold-change values, were constructed for the analyses. Gene sets with FDR q-value<0.25 after 1,000 permutation cycles were considered significantly enriched. Lists of leading edge subset genes, the cores of gene sets that account for the enrichment signal, were then generated. To create the list of enriched genes shared by plural twin pairs, the upper 100 enriched genes of each pair were included in the test.

Supporting Information

Dataset S1 Pair-wise comparison of expression array data. Genes with a fold-change value >2 were included. The positive value of fold-change designates up-regulation in the higher IQ twin, while the negative value designates the other way round. n/a indicates no gene matched the fold-change cutoff value.
(XLS)

Dataset S2 Pair-wise GSEA results using BioCarta database. Gene sets with a FDR q-value<0.25 were included. n/a indicates no gene set matched the FDR cutoff value.
(XLS)

Dataset S3 Pair-wise GSEA results using KEGG pathway database. Gene sets with a FDR q-value<0.25 were included. n/a indicates no gene set matched the FDR cutoff value.
(XLS)

Dataset S4 Pair-wise GSEA results using Reactome database. Gene sets with a FDR q-value<0.25 were included. n/a indicates no gene set matched the FDR cutoff value.
(XLS)

Dataset S5 Pair-wise GSEA results using GO database. Gene sets with a FDR q-value<0.25 were included. n/a indicates no gene set matched the FDR cutoff value.
(XLS)

Figure S1 Scatterplot of the 240 MZ twin pairs IQ scores. This diagram provides an overview of the IQ distribution for all 240 MZ twins from the Keio Twin Study. Each circle identifies one twin pair with its x-coordinate and y-coordinate representing, respectively, the IQ score of Twin A and Twin B. With the black line standing for regression, the correlation coefficient of 0.72 suggests the similarities between twins. Circles located outside of the space between two blue lines indicate twin pairs manifesting between-sibling IQ differences larger than 15 points and were considered to be recruited, while the red arrowhead points to one pair being excluded as a possible subject for a lower-than-normal IQ score.
(TIF)

Figure S2 Positive correlation between IQ scores differences and the number of loci different in methylation status. The positive correlation between the number of loci with significant differences in methylation patterns and the differences of IQ scores of the twin pairs was visualized. Twin pairs with larger differences in IQ scores tended to have more loci identified by screening for epigenetically regulated genes.
(TIF)

Figure S3 PCA results for the expression profiles of 17 twin pairs. Principle components ranked from the highest variance are named accordingly as PC 1st and PC 2nd. The PCA projection maps these two components data to 2 dimensions for visualization. In the scatter plots, each point represents a sample. The color of the symbol represents the relative IQ scores with red as higher and blue as lower. The number on each symbol indicates the twin pair ID. In contrast to the similarity between co-twins from each pair, no apparent gathering pattern could be recognized by the relative IQ scores.
(TIF)

Figure S4 Clustering analysis for the expression profiles of 17 twin pairs. Clustering analysis for the list of 644 genes manifesting more than a 1.1-fold change between the groups of twins with higher IQ scores and the groups of their co-twins under ANOVA analysis was performed. The number below each symbol indicates the twin pair ID. The color red and blue of the symbol indicate respectively the higher IQ and the lower IQ twin of each twin pair. The scale represents fold changes, according to the color map at the bottom of the figure. No apparent clustering could be recognized.
(TIF)

Table S1 Summary of candidate loci with methylation changes identified by promoter DNA methylation patterns and their bisulfite sequencing result.
(DOC)

Table S2 Up-regulated gene sets in the group of twins identified by GSEA (FDR q-value<0.25).
(DOC)

Table S3 Methylation profiling by bisulfite sequencing for IGF1 (Chr2:101335584□31398508) in Twin Pair ID 7.
(DOC)

Acknowledgments

We deeply thank all the study subjects. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE33478. This research was supported by JST, CREST.

Author Contributions

Conceived and designed the experiments: CCY MF KK TT. Performed the experiments: CCY MF. Analyzed the data: CCY MF. Contributed reagents/materials/analysis tools: CS PCC JS HS KI TK JA. Wrote the paper: CCY MF KK TT.

References

1. Deary I, Whalley LJ, Lemmon H, Crawford JR, Starr JM (2000) The stability of individual differences in mental ability from childhood to old age: follow-up of the 1932 scottish mental survey. Intelligence 28:49–55.
2. Deary IJ, Yang J, Davies G, Harris SE, Tenesa A, et al. (2012) Genetic contributions to stability and change in intelligence from childhood to old age. Nature 482:212–215.
3. Wright M, Geus ED, Ando J, Luciano M, Posthuma D, et al. (2001) Genetics of cognition: outline of a collaborative twin study. Twin Res 4:48–56.
4. Jacobs N, Gestel SV, Derom C, Thiery E, Vernon P, et al. (2001) Heritability estimates of intelligence in twins: effect of chorion type. Behav Genet 31:209–217.
5. Haworth CM, Wright MJ, Martin NW, Martin NG, Boomsma DI, et al. (2010) A twin study of the genetics of high cognitive ability selected from 11,000 twin pairs in six studies from four countries. Behav Genet 39:359–370.
6. Payton A (2009) The impact of genetic research on our understanding of normal cognitive ageing: 1995 to 2009. Neuropsychol Rev 19:451–477.
7. Vinkhuyzen AA, van der Sluis S, Posthuma D (2011) Life events moderate variation in cognitive ability (g) in adults. Mol Psychiatry 16:4–6.
8. Inlow JK, Restifo LL (2004) Molecular and comparative genetics of mental retardation. Genetics 166:835–881.
9. Davies G, Tenesa A, Payton A, Yang J, Harris SE, et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 16: 996–1005.
10. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet 82:763–771.
11. Biousse V, Brown MD, Newman NJ, Allen JC, Rosenfeld J, et al. (1997) De novo 14484 mitochondrial DNA mutation in monozygotic twins discordant for Leber's hereditary optic neuropathy. Neurology 49:1136–1138.
12. Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, et al. (2010) Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 464:1351–1356.
13. Haggarty P, Hoad G, Harris SE, Starr JM, Fox HC, et al. (2010) Human intelligence and polymorphisms in the DNA methyltransferase genes involved in epigenetic marking. PLoS One 5:e11329.
14. Holliday R (2006) Epigenetics: A Historical Overview. Epigenetics 1:76–80.
15. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, et al. (2010) Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res 20:170–179.
16. Shikishima C, Ando J, Ono Y, Toda T, Yoshimura K (2006) Registry of adolescent and young adult twins in the Tokyo area. Twin Res Hum Genet 9:811–816.
17. Plomin R, Defries JC (1980) Genetics and intelligence: recent data. Intelligence 4:15–24.
18. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, et al. (2006) Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA 103:12457–12462.
19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102:15545–15550.
20. van Nieuwpoort IC, Drent ML (2008) Cognition in the adults with childhood-onset GH deficiency. Eur J Endocrinol 159 Suppl 1:S53–57.
21. Deary IJ, Penke L, Johnson W (2010) The neuroscience of human intelligence differences. Nat Rev Neurosci 11:201–211.
22. Shikishima C, Hiraishi K, Yamagata S, Sugimoto Y, Takemura R, et al. (2009) Is *g* an entity? A Japanese twin study using syllogisms and intelligence tests. Intelligence 37:256–267.
23. Bell JT, Spector TD (2011) A twin approach to unraveling epigenetics. Trends Genet 27:116–125.
24. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, et al. (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci USA 102:10604–10609.
25. Maeda M, Hasegawa H, Hyodo T, Ito S, Asano E, et al. (2011) ARHGAP18, a GTPase-activating protein for Rho A, controls cell shape, spreading, and motility. Mol Biol Cell 22:3840–3852.
26. Govek EE, Newey SE, Akerman CJ, Cross JR, Van der Veken L, et al. (2004) The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis. Nat Neurosci 7:364–372.
27. Potkin SG, Macciardi F, Guffanti G, Fallon JH, Wang Q, et al. (2010) Identifying gene regulatory networks in schizophrenia. NeuroImage 53:839–847.
28. Nishikawa K, Li H, Kawamura R, Osaka H, Wang YL, et al. (2003) Alterations of structure and hydrolase activity of parkinsonism-associated human ubiquitin carboxyl-terminal hydrolase L1 variants. Biochem Biophys Res Commun 304:176–183.
29. Karthikeyan S, Zhou Q, Mseeh F, Grishin NV, Osterman AL, et al. (2003) Crystal structure of human riboflavin kinase reveals a beta barrel fold and a novel active site arch. Structure 11:265–273
30. O'Brien TW, O'Brien BJ, Norman RA (2005) Nuclear MRP genes and mitochondrial disease. Gene 354:147–151.
31. Tuteja N (2003) Plant DNA helicases: the long unwinding road. J Exp Bot 54:2201–2214.
32. Fisher TS, Zakian VA (2005) Ku: a multifunctional protein involved in telomere maintenance. DNA Repair (Amst) 4:1215–1226.
33. Davydov V, Hansen LA, Shackelford DA (2003) Is DNA repair compromised in Alzheimer's disease? Neurobiol Aging 24:953–968.
34. Zekri L, Chebli K, Tourrière H, Nielsen FC, Hansen TV, et al. (2005) Control of fetal growth and neonatal survival by the RasGAP-associated endoribonuclease G3BP. Mol Cell Biol 25:8703–8716.
35. Thiel CM, Fink GR (2008) Effects of the cholinergic agonist nicotine on reorienting of visual spatial attention and top-down attentional control. Neuroscience 152:381–390.
36. Nakauchi S, Brennan RJ, Boulter J, Sumikawa K (2007) Nicotine gates long-term potentiation in the hippocampal CA1 region via the activation of alpha2* nicotinic ACh receptors. Eur J Neurosci 25:2666–2681.
37. Brown BS, Yu SP (2000) Modulation and genetic identification of the M channel. Prog Biophys Mol Biol 73:135–166.
38. Peters HC, Hu H, Pongs O, Storm JF, Isbrandt D (2005) Conditional transgenic suppression of M channels in mouse brain reveals functions in neuronal excitability, resonance and behavior. Nat Neurosci 8:51–60.
39. Buxton A, Callan OA, Blatt EJ, Wong EH, Fontana DJ (1994) Cholinergic agents and delay-dependent performance in the rat. Pharmacol Biochem Behav 49:1067–1073.
40. Rodriguez S, Gaunt TR, Day INM (2007) Molecular genetics of human growth hormone, insulin-like growth factors and their pathways in common disease. Hum Genet 122:1–21.
41. Butcher LM, Meaburn E, Knight J, Sham PC, Schalkwyk LC, et al. (2005) SNPs, microarrays and pooled DNA: identification of four loci associated with mild mental impairment in a sample of 6000 children. Hum Mol Genet 14:1315–1325.
42. Bilder RM, Sabb FW, Cannon TD, London ED, Jentsch JD, et al. (2009) Phenomics: the systematic study of phenotypes on a genome-wide scale. Neuroscience 164:30–42.
43. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:351–322.
44. Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sásik R, et al. (2005) Comparative gene expression analysis of blood and brain provides concurrent validation of *SELENBP1* up-regulation in schizophrenia. Proc Natl Acad Sci U S A 102:15533–15538.
45. Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, et al. (2010) Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11:589.
46. Sullivan PF, Fan C, Perou CM (2006) Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet 141B:261–268.
47. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, et al. (2011) Disease-associated epigenetic changes in monozygotic twin discordant for schizophrenia and bipolar disorder. Hum Mol Genet 20:4786–4796.
48. Nguyen AT, Rauch TA, Pfeifer GP, Hu VW (2010) Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, *RORA*, whose protein product is reduced in autistic brain. FASEB J 24:3036–3051.
49. Sugawara H, Iwamoto K, Bundo M, Ueda J, Miyauchi T, et al. (2011) Hypermethylation of serotonin transporter gene in bipolar disorder detected by epigenome analysis of discordant monozygotic twins. Transl Psychiatry 1:e24.
50. Schroeder DI, Lott P, Korf I, Laselle JM (2011) Large-scale methylation domains mark a functional subset of neuronally expressed genes. Genome Res 21:1583–1591.
51. Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, et al. (2005) Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate? BMC Genomics 6:59.
52. Mathers JC, Strathdee G, Relton CL (2010) Induction of epigenetic alterations by dietary and other environmental factors. Adv Genet 71:3–39.

Large-scale replication and heterogeneity in Parkinson disease genetic loci

Manu Sharma, PhD
John P.A. Ioannidis, MD
Jan O. Aasly, MD
Grazia Annesi, MD
Alexis Brice, MD
Christine Van Broeckhoven, PhD
Lars Bertram, MD
Maria Bozi, MD
David Crosiers, MD
Carl Clarke, MD
Maurizio Facheris, MD
Matthew Farrer, PhD
Gaetan Garraux, MD
Suzana Gispert, MD
Georg Auburger, MD
Carles Vilariño-Güell, PhD
Georgios M. Hadjigeorgiou, MD
Andrew A. Hicks, PhD
Nobutaka Hattori, MD, PhD
Beom Jeon, MD
Suzanne Lesage, PhD
Christina M. Lill, MD
Juei-Jueng Lin, MD
Timothy Lynch, MD
Peter Lichtner, PhD
Anthony E. Lang, MD
Vincent Mok, MD
Barbara Jasinska-Myga, MD
George D. Mellick, PhD
Karen E. Morrison, MD
Grzegorz Opala, MD
Peter P. Pramstaller, MD
Irene Pichler, PhD
Sung Sup Park, MD
Aldo Quattrone, MD
Ekaterina Rogaeva, PhD
Owen A. Ross, PhD
Leonidas Stefanis, MD
Joanne D. Stockton, MD
Wataru Satake, MD, PhD
Peter A. Silburn, MD
Jessie Theuns, PhD
Eng-King Tan, MD
Tatsushi Toda, MD, PhD
Hiroyuki Tomiyama, MD, PhD
Ryan J. Uitti, MD
Karin Wirdefeldt, MD, PhD
Zbigniew Wszolek, MD
Georgia Xiromerisiou, MD
Kuo-Chu Yueh, MD
Yi Zhao, MD
Thomas Gasser, MD
Demetrius Maraganore, MD
Rejko Krüger, MD
On behalf of the GEO-PD Consortium

Correspondence & reprint requests to Dr. Sharma: manu.sharma@uni-tuebingen.de

Editorial, page 619

Supplemental data at www.neurology.org

Supplemental Data

ABSTRACT

Objective: Eleven genetic loci have reached genome-wide significance in a recent meta-analysis of genome-wide association studies in Parkinson disease (PD) based on populations of Caucasian descent. The extent to which these genetic effects are consistent across different populations is unknown.

Methods: Investigators from the Genetic Epidemiology of Parkinson's Disease Consortium were invited to participate in the study. A total of 11 SNPs were genotyped in 8,750 cases and 8,955 controls. Fixed as well as random effects models were used to provide the summary risk estimates for these variants. We evaluated between-study heterogeneity and heterogeneity between populations of different ancestry.

Results: In the overall analysis, single nucleotide polymorphisms (SNPs) in 9 loci showed significant associations with protective per-allele odds ratios of 0.78–0.87 (*LAMP3*, *BST1*, and *MAPT*) and susceptibility per-allele odds ratios of 1.14–1.43 (*STK39*, *GAK*, *SNCA*, *LRRK2*, *SYT11*, and *HIP1R*). For 5 of the 9 replicated SNPs there was nominally significant between-site heterogeneity in the effect sizes (I^2 estimates ranged from 39% to 48%). Subgroup analysis by ethnicity showed significantly stronger effects for the *BST1* (rs11724635) in Asian vs Caucasian populations and similar effects for *SNCA*, *LRRK2*, *LAMP3*, *HIP1R*, and *STK39* in Asian and Caucasian populations, while *MAPT* rs2942168 and SYT11 rs34372695 were monomorphic in the Asian population, highlighting the role of population-specific heterogeneity in PD.

Conclusion: Our study allows insight to understand the distribution of newly identified genetic factors contributing to PD and shows that large-scale evaluation in diverse populations is important to understand the role of population-specific heterogeneity. ***Neurology*® 2012;79:659–667**

GLOSSARY

CI = confidence interval; **GEO-PD** = Genetic Epidemiology of Parkinson's Disease; **GWAS** = genome-wide association studies; **HWE** = Hardy-Weinberg equilibrium; **MALDI-TOF** = matrix-assisted laser desorption/ionization time-of-flight; **MSA** = multiple system atrophy; **OR** = odds ratio; **PD** = Parkinson disease; **SNP** = single nucleotide polymorphism.

Genome-wide association studies (GWAS) have provided tangible gains in understanding the genetic architecture of complex diseases,[1,2] including Parkinson disease (PD).[3] Several GWAS have been conducted in PD in Caucasian populations and only 1 in the Asian population.[3–11] Consistent and reproducible association signals were confirmed in α-synuclein (*SNCA*), leucine-rich repeat kinase 2 (*LRRK2*), and microtubule-associated protein tau (*MAPT*), thus underscoring the importance of these 3 genes in the pathophysiology of the common sporadic forms of PD.[3–10,12] In addition to that, different studies have provided some evidence for an association for *BST1*, *GAK*, and *HLA-DRB5* with PD.[6–9,13]

A recently published GWAS meta-analysis in PD increased the number of identified PD genetic loci to 11.[14] This study reported significant between-study heterogeneity for some of the 11 genetic loci[14] even though data were restricted to Caucasian descent populations.

It is important to establish whether the 11 genetic loci that have been postulated to be associated with PD are replicated when tested with direct genotyping in a larger spectrum of diverse populations. The consistency or lack thereof of the genetic effects of these genetic

Author affiliations are provided at the end of the article.
Coinvestigators are listed on the *Neurology*® Web site at www.neurology.org.
Go to Neurology.org for full disclosures. Disclosures deemed relevant by the authors, if any, are provided at the end of this article.

variants across different populations may help to determine whether they represent genuine loci for PD susceptibility and whether they can be used for risk prediction across these diverse populations.[15] To gain further insight into genetic factors contributing to PD across different populations and define the implications of between-population heterogeneity, we performed a large-scale replication study within the GEO-PD consortium.

METHODS

Consortium. Investigators from the Genetic Epidemiology of PD (GEO-PD) Consortium were invited to participate in this study. A total of 21 sites representing 19 countries from 4 continents agreed to contribute DNA samples and clinical data for a total of 17,705 individuals (8,750 cases and 8,955 controls). Healthy individuals matched for age and gender served as controls. They underwent neurologic examination and were excluded from the study whenever there was clinical evidence for any extrapyramidal disorder.

Genotyping. We selected 1 SNP per each gene locus, exactly as they were proposed by the recently published GWAS meta-analysis.[14] Genotyping was performed by a central genotyping core (Department of Human Genetics, Helmholtz Zentrum, Munich). Each site provided 100–200 ng of DNA to the laboratory core. In total 11 SNPs located in and around the genes encoding *SYT11*, *ACMSD*, *STK39*, *LAMP3*, *GAK*, *BST1*, *SNCA*, *HLA-DRB5*, *LRRK2*, *HIP1R*, and *MAPT* were genotyped. The genotyping core was blinded to case-control status of each site. Genotyping was performed using a matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry on a MassArray system (Sequenom, San Diego, CA). Cleaned extension products were analyzed by a mass spectrometer (Bruker Daltronik, USA) and peaks were identified using the MassArray Typer 4.0.2.5 software (Sequenom). Assays were designed by the AssayDesigner software 4.0 (Sequenom) with the default parameters for the iPLEX Gold chemistry and the Human GenoTyping Tools ProxSNP and PreXTEND (Sequenom). All variants were genotyped in 1 multiplex assay. An experienced investigator blinded to case or control status of the samples visually checked genotype clustering. The average call rate of the variants was >97%.

In order to further enrich the samples of Asian ancestry populations, we also included GWAS data from a Japanese population (988 cases, 2,521 controls).[6] We used r^2 threshold of 0.8–1.0 to select proxy SNPs from the Japanese GWAS. Using this threshold, we were able to capture only 3 SNPs from *BST1*, *SNCA*, and *LRRK2* genes.

Standard protocol approvals, registrations, and patient consents. The local Ethics Committee approved the study. All participants signed an informed consent.

Analysis. An exact test was used to assess whether the genotype distributions for each SNP deviated from Hardy-Weinberg equilibrium (HWE) among controls; each site was tested separately and deviation from HWE was considered significant at <0.01. We excluded data from sites where the missing rate was >5%. For our analysis, we adhered to the same allele coding as in the previous GWAS meta-analysis.[14]

For consistency effect estimates based on minor vs major allele contrast were computed. We used an additive model adjusted for age and gender to obtain effect estimates. Results were then synthesized using fixed and random effects models. Fixed effect models assume that the genetic effect is the same in populations from different sites and that observed differences are due to chance alone. For associations showing between-study heterogeneity, fixed effect estimates yield narrower confidence intervals (CIs) and smaller p values as compared to random effects models, which incorporate between-study heterogeneity.[16–18] Fixed effects analysis tests the null hypothesis of no association in all studied populations that are analyzed. Routinely, this assumption is used in GWAS settings to increase the power of meta-analysis to detect associations that may exist in some (at least 1) population. However, in presence of heterogeneity the effects may differ substantially in different populations and not all populations may show a genetic effect for the variant of interest. Random effects models allow the genetic effects might be different due to genuine heterogeneity that may exist across different sites. Random effects calculations take into account the estimated between-study heterogeneity. We used the inverse variance method for fixed effects models. Cochran Q test of homogeneity and the I^2 metric were used to evaluate the between-site heterogeneity. The Q statistics follows χ^2-based distribution with $k - 1$ degrees of freedom (k = number of studies). I^2 is estimated by the ratio (Q-df)/Q, where df is degrees of freedom. The I^2 metric ranges from 0% to 100% and measures the proportion of variability that is beyond chance. Typically estimates of I^2 <25% are considered to reflect little or no heterogeneity, 25%–50% moderate heterogeneity, 50%–75% large heterogeneity, and >75% very large heterogeneity. It should be acknowledged that I^2 can have large uncertainty in its estimation especially for variants with low minor allele frequency. Therefore, we also estimated the 95% CI of I^2.[17]

The overall main analysis considered all sites and populations irrespective of ancestry. Then, we separately analyzed Caucasian and Asian sites and we compared the genetic effects in these 2 major ancestry groups.

The SNPs evaluated in the recently published GWAS meta-analysis are common with minor allele frequencies varying from 13% to 46%,[14] except for SNP, rs34372695 (SYT11) where the minor allele frequency is 2%. Therefore, based on minor allele frequency and effect estimates obtained in the GWAS meta-analysis,[14] power calculations showed that our study would have at least 99% power to detect an allele-based odds ratio (OR) of 1.2 for minor allele frequencies of 10% or higher for $\alpha = 0.05$. Based on genome-wide significance level ($\alpha = 5 \times 10^{-8}$), our study would have 43% power to detect an allele-based OR of 1.2 for minor allele frequency of 10%, but it would be 99% for same minor allele frequency and an OR of 1.4. Power would be only 69% for a minor allele frequency of 2% and OR of 1.2, but it would be 99% for the same minor allele frequency of 2% and an OR of 1.5.

Meta-analyses were performed using STATA 9.0 (Stata Corp., College Station, TX) and Review Manager 4.2.7. p Values are 2-tailed.

RESULTS

Characteristics of sites and overall database. Twenty-one sites contributed a total of 8,750 cases and 8,955 controls. Characteristics of all participating sites are shown in table 1. Most sites contributed participants of Caucasian ancestry ($n = 16$); 5 sites (counting also the GWAS performed in the Jap-

Table 1 Description of datasets contributed by each study site

Site	Country	No.	Case	Control	Male (%)	Female (%)	Mean AAO	Mean age at study	Diagnostic criteria
Annesi	Italy	394	197	197	204 (51.7)	190 (48.2)	61.5	63.7	UKPDBB
Brice[a]	France	505	272	233	302 (59.8)	203 (40.1)	47.6	57.8	UKPDBB
Bozi	Greece	222	114	108	107 (48.1)	115 (51.8)	69.9	74.5	UKPDBB
Wszolek	US	1,518	692	826	794 (52.3)	724 (47.6)	64.4	71.7	UKPDBB
Garraux	Belgium	82	68	14	45 (54.8)	37 (45.1)	62.1	69.6	UKPDBB
Hadjigeorgiou	Greece	714	357	357	379 (53.0)	335 (46.9)	63.4	63.7	UKPDBB
Jeon	Korea	749	408	341	314 (41.9)	435 (58.0)	57.6		UKPDBB
Opala	Poland	629	352	277	340 (54.0)	288 (45.7)	50.2	68.1	UKPDBB
Lynch	Ireland	740	368	372	340 (45.9)	400 (54.0)	50.5	70.7	UKPDBB
Lin	Taiwan	320	160	160	160 (50)	160 (50)	62.0	70.8	UKPDBB
Facheris	Italy	181	114	67	86 (47.5)	95 (52.4)	63.0		UKPDBB
Maraganore	US	1,024	801	223	600 (58.5)	361 (35.3)	59	74.7	Bower
Mellick	Australia	2,024	1,012	1,012	1,042 (51.4)	981 (48.4)	59	72.2	Bower
Morrison[a]	England	1,120	766	354	606 (54.1)	514 (45.8)	66.1		UKPDBB
Mok	China	436	260	176	264 (60.5)	170 (38.9)		63.5	UKPDBB
Aasly	Norway	1,278	656	622	721 (56.4)	557 (43.5)	58.8	72.9	UKPDBB
Wirdefeldt	Sweden	299	83	216	147 (49.1)	152 (50.8)	65.8	71.4	Gelb
Van Broeckhoven	Belgium	1,010	501	509	500 (49.5)	509 (50.3)	60.5	66.3	Pals/Gelb
Rogaeva	Canada	560	387	173	303 (54.1)	257 (45.8)	49.7	64.2	UKPDBB
Tan	Singapore	391	194	197	244 (62.4)	147 (37.5)	59.7	54.0	UKPDBB
Toda	Japan	3,509	988	2,521	1,844 (52.6)	1,665 (47.4)	58.7	66.0	UKPDBB
Total		17,705	8,750	8,955			59.5	67.6	

Abbreviations: AAO = age at onset; GWAS = genome-wide association studies; UKPDBB = UK Parkinson's Disease Brain Bank.
[a] Also included in the previously published GWAS.[7,8]

anese population[6]) included participants of Asian ancestry. We excluded 1 site with 114 cases and 67 controls from the analysis due to a plate layout error. The median age at onset was 59 years and median age at examination was 67 years.

We observed that for one site, effect estimates for all SNPs were "inverse" as compared to other Caucasian sites. Allele flipping for one particular site in the same Caucasian descent might reflect error in sampling ascertainment and is unlikely to reflect genuine effects.[19] This site (n = 181) was therefore excluded from further analyses. Overall, genotype call rates were >97%. The genotype distribution for each SNP in the controls of each site showed no departure from HWE, except for rs6599388 (*GAK*) in samples from 4 Asian sites. We therefore excluded this SNP (rs6599388) from analyses in the Asian population.

Overall data synthesis. We observed consistent and reproducible associations for *SNCA*, *LRRK2*, *MAPT*, *BST1*, *GAK*, *STK39*, *SYT11*, *LAMP3*, and *HIP1R* loci but not for *ACMSD* (rs10928513) or *HLA-DRB5* (rs3129882) where the per-allele OR was very close to the null (1.02 and 0.95, respectively) and statistically nonsignificant (table 2). Thus we provide unequivocal support for the involvement of these newly identified genetic loci in the pathogenesis of PD.

Summary effect estimates were generally comparable with the previous GWAS meta-analysis results (table 2), although effect estimates in this study were stronger for STK39 and somewhat weaker for *LRRK2* compared to the previous GWAS meta-analysis.[14] Exclusion of 1,625 samples that overlap with the previously published GWAS did not change any of the estimates (table e-1 on the *Neurology*® Web site at www.neurology.org). The protective per-allele OR ranged from 0.78 to 0.87 (*LAMP3*, *BST1*, and *MAPT*) and the susceptibility per-allele OR ranged from 1.14 to 1.43 (*STK39*, *GAK*, *SNCA*, *LRRK2*, *SYT11*, and *HIP1R*). Cochran Q statistics were nominally significant for *STK39*, *LAMP3*, *BST1*, and *SNCA* with I^2 estimates ranging from 39% to 48%. The heterogeneity reflected primarily differences in the magnitude of the effect sizes across different sizes, while the direction of the effect was consistent in all sites, with rare exceptions.

Table 2 Overall analysis irrespective of ethnicity and influence of between-study heterogeneity

Gene	SNP	Q test p value	I^2 (95% CI)	Odds ratio (95% CI) by random effects	Fixed effects p value	Random effects p value	IPDGC study odds ratio (fixed effects p values)
ACMSD	rs10928513	0.71	0 (0–54)	1.02 (0.96–1.08)	0.479	0.479	1.07 (0.003)
STK39	rs2102808	0.02	46 (0–68)	1.21 (1.08–1.35)	0.0001	0.001	1.12 (0.0016)
LAMP3	rs11711441	0.04	39 (0–64)	0.85 (0.77–0.94)	3.01×10^{-5}	0.002	0.87 (6.92×10^{-5})
GAK	rs6599388	0.06	39 (0–64)	1.19 (1.10–1.28)	3×10^{-6}	0.001	1.14 (7.46×10^{-8})
HLA-DRB5	rs3129882	0.23	18 (0–53)	0.95 (0.90–1.01)	0.12	0.15	0.80 (9.3×10^{-8})
BST1	rs11724635	0.02	43 (0–65)	0.87 (0.83–0.91)	2.01×10^{-6}	0.00001	0.87 (2.43×10^{-9})
SNCA	rs356219	0.02	48 (0–69)	1.30 (1.21–1.40)	4.23×10^{-23}	4.23×10^{-23}	1.27 (4.23×10^{-23})
SYT11	rs34372695	0.38	6 (0–50)	1.43 (1.15–1.78)	0.001	0.001	1.44 (1.18×10^{-6})
LRRK2	rs1491942	0.75	0 (0–50)	1.14 (1.07–1.21)	1.06×10^{-8}	1.06×10^{-8}	1.30 (1.06×10^{-8})
HIP1R	rs10847864	0.90	0 (0–44)	1.15 (1.09–1.21)	9.06×10^{-7}	9.06×10^{-7}	1.13 (9.06×10^{-7})
MAPT	rs2942168	0.14	29 (0–62)	0.78 (0.71–0.85)	1.37×10^{-13}	1.37×10^{-13}	0.80 (1.37×10^{-13})

Abbreviations: CI = confidence interval; IPDGC = International Parkinson disease Genomics Consortium; SNP = single nucleotide polymorphism.

Analysis including only Caucasian sites. Restricting the analysis to Caucasian sites only resulted in per-allele ORs that ranged from 0.78 to 0.90 for the 3 replicated protective loci (*BST1*, *LAMP3*, and *MAPT*) and from 1.14 to 1.43 for the 6 replicated susceptibility loci (*STK39*, *GAK*, *SNCA*, *LRRK2*, *SYT11*, and *HIP1R*), while *ACMSD* and *HLA-DRB5* still had no significant effect (table 3 and figure e-1).

Summary effect estimates were generally comparable to those of the previous GWAS meta-analysis,[14] except for modest differences in STK39 and LRRK2 effect sizes, as noted above also for the overall analysis. There was nominally significant heterogeneity only for SNCA and LAMP3 (I^2 estimates 51% and 46%, respectively), but this reflected primarily differences in the magnitude of the effect size estimates rather than direction of effects across sites (figure e-1).

Analysis including only Asian sites. In the Asian series, not only the *SYT11* SNP, but also the *ACMSD* and *MAPT* SNPs were monomorphic (table e-2). Summary effect estimates for the remaining SNPs are shown in table 4. We again observed consistent nominally significant evidence of association for all loci except for *STK39* (which still had an effect size estimate consistent with what was seen in the overall analysis) and *HLA-DRB5* (which had a point estimate very close to the null), which still had an effect size estimate consistent with what was seen in the overall analysis. Results were generally consistent across sites, with the exception of *STK39* that showed very large heterogeneity ($I^2 = 73\%$) (figure e-2).

Comparison of effect size estimates. Five gene loci (*HIP1R*, *LAMP3*, *LRRK2*, *SNCA*, and *STK39*)

Table 3 Caucasian specific effect estimates and influence of between-study heterogeneity

Gene	SNP	Q test p value	I^2 (95% CI)	Odds ratio (95% CI) by random effects	Fixed effects p value	Random effects p value
ACMSD	rs10928513	0.71	0 (0–54)	1.02 (0.96–1.08)	0.479	0.479
STK39	rs2102808	0.10	37 (0–65)	1.21 (1.08–1.35)	0.000	0.001
LAMP3	rs11711441	0.02	46 (0–70)	0.86 (0.76–0.97)	0.001	0.025
GAK	rs6599388	0.27	17 (0–57)	1.14 (1.06–1.23)	1.01×10^{-4}	0.001
HLA-DRB5	rs3129882	0.10	32 (0–63)	0.95 (0.88–1.02)	0.11	0.16
BST1	rs11724635	0.14	29 (0–62)	0.90 (0.85–0.95)	0.001	0.003
SNCA	rs356219	0.01	51 (0–72)	1.30 (1.19–1.42)	4.23×10^{-23}	4.23×10^{-23}
SYT11	rs34372695	0.38	6 (0–50)	1.43 (1.15–1.78)	0.001	0.001
LRRK2	rs1491942	0.80	0 (0–47)	1.15 (1.07–1.23)	1.06×10^{-8}	1.06×10^{-8}
HIP1R	rs10847864	0.90	0 (0–47)	1.15 (1.08–1.22)	9.06×10^{-7}	9.06×10^{-7}
MAPT	rs2942168	0.14	29 (0–62)	0.78 (0.71–0.85)	1.37×10^{-13}	1.37×10^{-13}

Abbreviations: CI = confidence interval; SNP = single nucleotide polymorphism.