Fine-Mapping, Gene Expression and Splicing Analysis of the Disease Associated LRRK2 Locus

Daniah Trabzuni1,2,9, Mina Ryten1,9, Warren Emmett3, Adaikalan Ramasamy4, Karl J. Lackner5, Tanja Zeller6, Robert Walker7, Colin Smith7, Patrick A. Lewis1,8, Adamantios Mamas1,9, Rohan de Silva1,9, Jana Vandrovcová1,5, International Parkinson Disease Genomics Consortium (IPDGC)9, Dena Hernandez10, Michael A. Nalls10, Manu Sharma11, Sophie Garnier12, Suzanne Lesage13, Javier Simon-Sanchez14, Thomas Gasser11, Peter Heutink14, Alexis Brice13, Andrew Singleton10, Huaibin Cai15, Eric Schadt16, Nicholas W. Wood1, Rina Bandopadhyay1,9, Michael E. Weale4, John Hardy1,9, Vincent Plagnol3

1 Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, United Kingdom, 2 Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 3 University College London Genetics Institute, University College London, London, United Kingdom, 4 Department of Medical and Molecular Genetics, King’s College London, Guy’s Hospital, London, United Kingdom, 5 Institute of Clinical Chemistry and Laboratory Medicine, University Medical Centre Mainz, Mainz, Germany, 6 University Heart Center Hamburg, Clinic for General and Interventional Cardiology, Hamburg, Germany, 7 MRC Sudden Death Brain Bank Project, University of Edinburgh, Department of Neuropathology, Edinburgh, Scotland, United Kingdom, 8 School of Pharmacy, University of Reading, Whiteknights, Reading, United Kingdom, 9 Reta Lila Weston Institute of Neurological Studies, London, United Kingdom, 10 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 11 Division of Neurodegenerative Disorders, Hertie Institute for Clinical Brain Research, University of Tubingen, Tubingen, Germany, 12 Pierre and Marie Curie University, Institut National de la Santé et de la Recherche Médicale UMR 937, Paris, France, 13 CRICM, University Pierre et Marie Curie, Institut National de la Santé et de la Recherche Médicale UMR 975, CNRS UMR 7225, Hospital Pitéï¿½-Salpêï¿½trière, Paris, France, 14 Department of Clinical Genetics, Section of Medical Genomics, VU University Medical Centre, Amsterdam, The Netherlands, 15 Unit of Transgenesis, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 16 Institute for Genomics and Multiscale Biology, Mount Sinai School of Medicine, New York, New York, United States of America

Abstract

Association studies have identified several signals at the LRRK2 locus for Parkinson’s disease (PD), Crohn’s disease (CD) and leprosy. However, little is known about the molecular mechanisms mediating these effects. To further characterize this locus, we fine-mapped the risk association in 5,802 PD and 5,556 controls using a dense genotyping array (ImmuChip). Using samples from 134 post-mortem control adult human brains (UK Human Brain Expression Consortium), where up to ten brain regions were available per individual, we studied the regional variation, splicing and regulation of LRRK2. We found convincing evidence for a common variant PD association located outside of the LRRK2 protein coding region (rs117762348, A>G, P=2.56×10⁻⁸, case/control MAF 0.083/0.074, odds ratio 0.86 for the minor allele with 95% confidence interval [0.80–0.91]). We show that mRNA expression levels are highest in cortical regions and lowest in cerebellum. We find an exon quantitative trait locus (eQTL) in brain samples that localizes to exons 32–33 and investigate the molecular basis of this eQTL using RNA-Seq data in n = 8 brain samples. The genotype underlying this eQTL is in strong linkage disequilibrium with the CD associated non-synonymous SNP rs3761863 (M2397T). We found two additional QTLs in liver and monocyte samples but none of these explained the common variant PD association at rs117762348. Our results characterize the LRRK2 locus, and highlight the importance and difficulties of fine-mapping and integration of multiple datasets to delineate pathogenic variants and thus develop an understanding of disease mechanisms.


Editor: Ted M. Dawson, Johns Hopkins, United States of America

Received September 15, 2012; Accepted June 23, 2013; Published August 13, 2013

Copyright: © 2013 Trabzuni et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Medical Research Council (MRC) through the MRC Sudden Death Brain Bank (C.S.) and by a Project Grant (G0901254 to J.H. and M.W.) and Training Fellowship (G0802462 to M.R.). D.T. was supported by the King Faisal Specialist Hospital and Research Centre, Saudi Arabia. P.A.L. is a Parkinson’s UK research fellow (fellowship F1002). A.-G. and C.M. are supported by a Michael J. Fox Foundation LRRK2 consortium grant to P.A.L. and R.B. This work was undertaken in part by the Wellcome Trust/MRC Joint Call in Neurodegeneration award (WT089698) to the UK Parkinson’s Disease Consortium (UKPDC) whose members are from the UCL Institute of Neurology, the University of Sheffield and the MRC Protein Phosphorylation Unit at the University of Dundee. This work was undertaken in part at UCLH/UCL who received a proportion of funding from the Department of Health’s NIHR Biomedical Research Centres funding scheme. This US work was undertaken in part by the Intramural Research Programs of the National Institute on Aging, National Institute of Neurological Disorders and Stroke, National Institute of Environmental Health Sciences, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services; project numbers Z01 AG000949-02 and Z01-ES101986. In addition this study was supported by the US Department of Defense, award number W81XWH-09-2-0128. Funding to support collection of a portion of the samples was obtained from the National Institutes of Health (grants NS057105 and RR024992), the American Parkinson Disease Association (APDA), Barnes Jewish Hospital Foundation, and the Greater St. Louis Chapter of the APDA. The KORA research platform (KORA: Cooperative Research in the Region of Augsburg; http://www.gsf.de/KORA) was initiated and financed by the Forschungszentrum für Umwelt und Gesundheit (GSF), which is funded by the German Federal Ministry of Education, Science, Research, and Technology and by the State of Bavaria. The study was additionally funded by the German National Genome Network (NGFNplus #01GS08134; German Ministry for Education and Research) and by the German Federal Ministry of Education and Research (BMBF) NGFN (01GQ0468) and in the frame of ERA-Net NEURON (01GW0908). This work was also supported by the Helmholtz Alliance Mental Health in an Ageing Society (HelMA, HA-215) funded by the Initiative and Networking Fund of the Helmholtz Association. The authors also want to thank the Hersenstichting Nederland (http://www.hersenstichting.nl), the Neuroscience Campus Amsterdam and the section of Medical genomics, the Primes Beatrix Fonds (http://www.primesbeatrixfonds.nl) for sponsoring this work. The authors have no conflicting financial, personal, or...
Introduction

The role of LRRK2 in human disease was first recognised in 2004 when dominant mutations in the LRRK2 gene were linked to Parkinson’s disease (PD). Rare genetic variants located in the LRRK2 gene contribute to a significant fraction of familial clustering of the disease [1]. In particular, heterozygous carriers of the non-synonymous change G2019S have an estimated PD lifetime risk close to 50% [2,3], which directly implicates the gene LRRK2 as causally implicated in PD aetiology. In addition, recent GWAS results suggest that common variants with a more modest effect on PD risk also exist at this locus [4]. Intriguingly, GWAS have implicated LRRK2 in the pathogenesis of Crohn’s disease (CD) and leprosy [5–7]. However, the mechanisms linking PD and the LRRK2 gene, and more generally the LRRK2 gene and human disease, remain largely unknown and are the focus of an intense research effort.

The LRRK2 gene spans 144 kb and is made up of 51 exons. It produces a 2,527 amino acid protein with multiple functional domains, including leucine-rich repeats (LRR), a GTPase domain (Ras of Complex proteins or ROC domain), a domain of unknown function termed the C-terminal of ROC (COR) domain, a kinase domain and a WD40 domain [8]. Coding changes causative for PD are located within the enzymatic core of LRRK2, namely the ROC-COR-Kinase triad of domains, and several of the mutations described in this region disrupt the enzymatic activities of the protein – strongly implicating the enzymatic function of LRRK2 in the pathogenesis of PD [9–11]. Existing data demonstrates that LRRK2 expression is not restricted to the human brain, but is also found in liver, kidney and thymus [12,13]. A number of studies have been undertaken to investigate the function, expression and cellular localization of LRRK2 in the human brain [14–17]. However, these findings have been based on relatively small numbers of individuals. Consequently, much of the biology of this gene remains unknown. In particular, very little is known about the pattern of LRRK2 gene expression across the brain, to what extent it is the subject of alternative splicing and, if this is the case, whether splicing is region-specific, and finally how the expression of LRRK2 is regulated – all information that is likely to be critical for deciphering how LRRK2 dysfunction results in disease.

A key area of research is the identification of the molecular mechanisms linking variants in the LRRK2 region with modified risk for PD, CD and leprosy [1,4,6,7,18]. We hypothesize that some of these effects are mediated by genetic control of the expression and/or splicing of LRRK2 mRNA. This hypothesis can be tested by combining disease GWAS results with expression datasets and by investigating whether these association signals are compatible with shared causal variants [19]. In addition to suggesting causal mechanisms, this analysis can point to tissue types involved in disease pathogenesis. Therefore, on the basis that PD is driven by pathophysiological processes resulting in the death of neuronal cell populations, there is considerable interest in dissecting the genetic basis of PD susceptibility at the LRRK2 locus by analysing these results in parallel with expression QTL (eQTL) studies conducted in multiple human tissues.

Results

Fine-mapping of the LRRK2 PD association signals

Using a GWAS of 5,333 cases and 12,019 controls, Nalls et al [4] identified a PD association in the LRRK2 chromosome region with a common variant (rs1491942, GWAS P = 5.23 × 10^{-8}, MAF 7%). To fine-map this locus and confirm that this result is independent of the previously described rare variant association at the nonSNP rs34637584/G2019S, we densely genotyped the LRRK2 region in the GWAS replication set of 5,802 PD cases and 5,556 controls using the IlluminaChip (2.3 typed SNPs per kb on average in the region defined by hg19 chr12:40,541,601-40,830,814, see Methods and [21]). We further increased marker density using 1,000 genomes based imputation (Methods and Table S2).

Direct ImmunoChip genotyping in the GWAS replication set confirmed the strong and previously reported association at the nonSNP rs34637584/G2019S (rs117762348, imputed A>G SNP with imputation r^{2}>0.99, P = 2.56 × 10^{-8}, UK control MAF 0.083, UK case MAF 0.074, OR 0.86 with 95% CI: 0.80–0.91 for the minor allele) which is located 5’ of LRRK2. rs117762348 is in moderate LD (D’=1, r^{2}=0.31, Table S3) with the initial GWAS SNP rs1491942 (combined GWAS and ImmunoChip data for rs117762348 is P=1.55×10^{-15}). While other SNPs showed a comparable level of association (Figure 1A), none of them were coding (P > 10^{-14} for all exonic SNPs, either typed or imputed). After conditioning on both rs117762348 and rs34637584 no other SNP remained significant (P > 10^{-5}). The PD association signal could therefore be summarized using a combination of rs34637584/G2019S and rs117762348.

CD meta-analysis [7] results indicate that at least two independent CD associations (Table 1) exist at the LRRK2 locus.
(lead SNPs rs11564258 and rs3761863). The low correlation coefficient ($r^2 = 0.05$, Table S3) between these two SNPs is consistent with independent associations. None of the CD SNPs is associated with PD in the immunoChip dataset (Table 1). In contrast, we observed suggestive evidence for PD association in the immunoChip dataset at the leprosy lead SNP rs1491938 ($P = 0.001$, OR 1.09, 95% CI: [1.04–1.15], Table 1). However, we found no support for this PD association in the GWAS dataset [4] ($P = 0.43$ for PD at rs1491938).

**Figure 1. Multiple association signals in the LRRK2 region chr12:40,351,601-40,830,814.** The x-axis shows the physical position (hg19 build) of each variant and the y-axis shows the $-\log_{10}(p)$ for association. Typed SNPs are shown in red and imputed SNPs in grey. (A) Fine-mapping of the PD association using case-control ImmunoChip genotyping. (B) Combined brain expression data across all brain regions and all exons of LRRK2 (C) Exon specific QTL identified in brain samples (combining expression data from ten brain regions). The expression probes showing this signal are located in exons 32 and 33 of LRRK2. (D) LRRK2 eQTL identified in 1,490 monocytes samples. (E) LRRK2 eQTL identified in 966 liver samples. doi:10.1371/journal.pone.0070724.g001
**Table 1.** Common variant associations in the LRRK2 region for PD, CD and leprosy.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Parkinson’s disease</th>
<th>Crohn, first signal</th>
<th>Crohn, second signal</th>
<th>Leprosy</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs117762348</td>
<td>rs11564258</td>
<td>rs3761863⁷</td>
<td>rs1491938</td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>chr12:40,597,612</td>
<td>chr12:40,792,300</td>
<td>chr12:40,758,652</td>
<td>chr12:40,645,630</td>
</tr>
<tr>
<td>Alleles (major &gt; minor)</td>
<td>A&gt;G</td>
<td>G&gt;A</td>
<td>C&gt;T</td>
<td>T&gt;C</td>
</tr>
<tr>
<td>MAF</td>
<td>0.076</td>
<td>0.034</td>
<td>0.32</td>
<td>0.396</td>
</tr>
<tr>
<td>Odds ratio for PD (95% CI) in the immunochip dataset</td>
<td>0.65 (0.80–0.91)</td>
<td>0.89 (0.75–1.05)</td>
<td>0.97 (0.91–1.02)</td>
<td>1.09 (1.04–1.15)</td>
</tr>
<tr>
<td>Odds ratio for minor allele (95% CI) for each of the respective disease</td>
<td>0.65 (0.80–0.91)</td>
<td>1.74 (1.55–1.95)</td>
<td>1.1 (1.05–1.15)</td>
<td>0.86 (0.80–0.92)</td>
</tr>
<tr>
<td>Other SNP in LD (r², Table S3)</td>
<td>rs1491942 (r²=0.3, PD GWAS SNP)</td>
<td>-</td>
<td>rs10784486 (r²=0.7, brain eQTL)</td>
<td>rs10784428 (r²=0.3, monocyte eQTL)</td>
</tr>
<tr>
<td>P-values for overlap with eQTL datasets (direction of effect for the minor allele)</td>
<td>P-value brain LRRK2 eQTL, exon 32–33 specific</td>
<td>NS</td>
<td>NS</td>
<td>1.35E-11 (+)</td>
</tr>
<tr>
<td>P-value LRRK2 eQTL liver</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>P-value monocyte LRRK2 eQTL</td>
<td>9.4E-4 (+)</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

The CD GWAS results indicate a minimum of two independent associations. For each disease association we list the P-values in the expression datasets. NS: not-significant (P>0.001). (a): See Figure S3. 

doi:10.1371/journal.pone.0070724.t001

**LRRK2 mRNA expression levels in different brain regions**

The genetic evidence for a regulatory (rather than protein coding) mechanism to explain this common variant association motivated further investigation of LRRK2 mRNA expression. To assess the pattern of gene expression in human brain, we generated expression data for n = 134 samples without diagnosed neuropathology [22].

To perform our exon eQTL analysis we combined our brain expression dataset separately quantified mRNA regional levels by Affymetrix GeneChip Human Exon 1.0 ST Arrays in ten brain regions of adult samples (for the majority of the n = 134 samples: 1,231 arrays overall): frontal cortex, occipital cortex (specifically primary visual cortex), temporal cortex, intralobular white matter, thalamus, putamen, substantia nigra, hippocampus, medulla (specifically inferior olivary nucleus) and cerebellum. No significant correlation was found between the data across all the exonic probes, we found no significant correlation between overall LRRK2 gene expression and genotypes (Figure 1B). However, we identified a brain eQTL implicating specifically exons 32 and 33 with rs10784486 (Table 2 and Figure 1C) across all brain regions (Figure S2). We also identified two additional and independent eQTLs (Table 2): one in monocytes (with rs10784428, Figure 1D) and one in liver (with rs11175518, Figure 1E and S1).

To perform our exon eQTL analysis we combined our brain expression dataset with liver ([23], n = 970, Methods) and monocytes (Gutenberg Health Study [24], n = 1,490) samples. All expression probes were carefully checked for variants located within them using the 1,000 Genomes dataset and the NHLBI exome sequencing project data (http://evs.gs.washington.edu/EVS/) and excluded if necessary. In brain, and after combining the data across all the exonic probes, we found no significant correlation between overall LRRK2 gene expression and genotypes (Figure 1B). However, we identified a brain eQTL implicating specifically exons 32 and 33 with rs10784486 (Table 2 and Figure 1C) across all brain regions (Figure S2). We also identified two additional and independent eQTLs (Table 2): one in monocytes (with rs10784428, Figure 1D) and one in liver (with rs11175518, Figure 1E and S1). For the liver and monocyte data, a single good quality exonic probe was present on the array. We therefore expect that these signals capture information about the expression of the overall LRRK2 gene, in contrast with the brain eQTL that was observed only for exons 32 and 33.

We found no significant evidence in the brain samples for an exon eQTL at the liver SNP rs11175518 (P = 0.054 in brain) or at the monocyte exon eQTL SNP (P = 0.073 in brain for rs10784428), although this may be a consequence of the smaller sample size of the brain expression study.
Figure 2. Regional variability in LRRK2 expression. (A) Box plot of mRNA expression levels for \textit{LRRK2} in 10 brain regions, based on microarray experiments and plotted on a log2 scale (y axis). Whiskers extend from the box to 1.5 times the inter-quartile range. (B) Box plot of mRNA expression levels for \textit{LRRK2} in 4 brain regions, based on QuantiGene experiments. Whiskers extend to the maximum and minimum values. Stars indicate significant differences in expression \textit{p}-value < 0.01, Wilcoxon signed rank testing). (C) Dot plot of mRNA expression levels for \textit{LRRK2} in 3 brain regions based on TaqMan Real Time PCR experiments. The expression levels were normalized to the geometric mean of 3 housekeeping genes. The graph shows higher expression in OCTX compared with other regions. Abbreviations: frontal cortex (FCTX), occipital cortex (specifically primary visual cortex, OCTX), temporal cortex (TCTX), intralobular white matter (WHMT), thalamus (THAL), putamen (PUTM), substantia nigra (SNIG), hippocampus (HIPP), medulla (specifically inferior olivary nucleus, MEDU) and cerebellum (CRBL).

doi:10.1371/journal.pone.0070724.g002
mechanism to mediate this secondary CD association signals. Denser fine-mapping data for CD is required to answer this question.

We also observed a weak association between the PD associated lead SNP rs117762348 and the \( LRRK2 \) eQTL in monocyte cells \( (P = 9.4 \times 10^{-4}, \text{Table 1}). \) However, while the lead SNPs for PD \( (\text{rs117762348}) \) and monocyte exon eQTL \( (\text{rs10784428}) \) are physically close (separated by 6,996 bp, Figure 1), these two SNPs are not in LD \( (r^2 = 0.05, \text{Table S3}). \) Accordingly, the strength of the monocyte exon eQTL association at the PD lead SNP is much weaker than the signal observed at the lead monocyte exon eQTL SNP \( (P = 9 \times 10^{-4} \text{ for rs117762348 in Table 1 versus } P = 1.3 \times 10^{-10} \text{ for rs10784428 in Table 2}). \) Hence, the data are not supportive of a shared causal variant for the PD association and the monocyte exon eQTL.

**Table 2.** List of identified \( LRRK2 \) eQTLs in three gene expression datasets: brain \( (n = 134) \), liver \( (n = 970) \) and monocytes \( (n = 1,490) \).

<table>
<thead>
<tr>
<th>Lead SNP</th>
<th>Position</th>
<th>Alleles</th>
<th>MAF</th>
<th>Probe location</th>
<th>( P )-value (direction of effect for minor allele)</th>
<th>( \text{PD association P-value (estimated OR for minor allele)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10784486*</td>
<td>chr12:40,677,029</td>
<td>C&gt;A</td>
<td>0.33</td>
<td>exons 32 and 33</td>
<td>2.24E-13 (−)</td>
<td>0.11 (0.95)</td>
</tr>
<tr>
<td>rs11175518*</td>
<td>chr12:40,580,318</td>
<td>C&gt;T</td>
<td>0.0709</td>
<td>exon 51</td>
<td>4.18E-21 (+)</td>
<td>0.015 (0.88)</td>
</tr>
<tr>
<td>rs10784428</td>
<td>chr12:40,604,608</td>
<td>C&gt;A</td>
<td>0.44</td>
<td>exon 50</td>
<td>1.3E-10 (−)</td>
<td>0.66 (0.99)</td>
</tr>
</tbody>
</table>

For the \( P \)-value computations in brain samples, \( LRRK2 \) expression values are averaged across all 10 brain regions. MAF: minor allele frequency. *: See Figure S2. #: See Figure S1.

doi:10.1371/journal.pone.0070724.t002

**Discussion**

Motivated by the role of \( LRRK2 \) in PD, we fine-mapped this locus in a large case-control collection and found firm support for the presence of a common non-coding variant PD association at this locus (MAF \( \sim 8\% \)). However, several SNPs show comparably high level of PD association and we cannot at this stage make a firm statement about the exact causal variant, but instead highlight a group of SNPs in high LD. The analysis of gene expression datasets highlighted the complexity of the genetic control of \( LRRK2 \), with at least three independent exon QTLs at this locus. The convincing exon QTL found in brain tissue involves exons 32 and 33 of \( LRRK2 \) but does not co-localize with the PD association. Similarly, the exon QTLs identified in liver and monocyte populations do not co-localize with the PD association.

The most recent meta-analysis results for CD indicate a secondary association for CD, for which the nsSNP rs3761863 is a leading associated variant \( (P = 3.0 \times 10^{-10}) \) \[7\]. rs3761863/M2397T is also strongly associated with the exon 32–33 brain exon eQTL lead SNP rs10784486 that we identified \( (P = 1.35 \times 10^{-12}, \text{Table 1}). \) The human intestine is heavily innervated and therefore brain tissue may not be irrelevant to CD, but it is also plausible that this exon/expression correlation is shared across other tissue types. Our data suggest a potential mechanism to mediate the CD association but additional CD fine-mapping will be required to test whether the CD and brain exon QTL datasets are fully consistent with a shared variant. We also note that a previous study has associated the same nsSNP rs3761863 with \( LRRK2 \) protein stability \[26\]. These results highlight several factors that complicate the follow-up of associated variants at the \( LRRK2 \) locus in all diseases and in particular PD: the existence of multiple \( LRRK2 \) isoforms, variability in mRNA expression and splicing across brain regions, and localisation of \( LRRK2 \) protein to both neuronal and non-neuronal cell types.

On the basis of both exon array and RT-PCR data we predict that alternative isoforms including the splicing of either one or both of exons 32–33 must exist. However, in our brain samples RNA-Seq data, we could not find a strong support for junction reads that support the splicing out of exons of 32 and/or 33. While we cannot fully exclude that unexpected artefacts generated a false positive result for that exon 32–33 specific brain eQTL, an alternative scenario would implicate a more complex set of splicing events, potentially involving more than exons 32–33. Analytical challenges, limited read depth and short length of RNA-Seq reads, may prevent us from characterising this event. The lower number of junctions in the exons 29–35 region of \( LRRK2 \) suggests that this...
may be the case, but further analysis will be required to confirm this hypothesis.

In summary, this study provides novel insights into \textit{LRRK2} expression, splicing and regulation with a potential link to the etiology of CD. It also highlights the relevance of imputation techniques to provide the dense coverage required to integrate multiple disease association and gene expression studies. The localization of the common variant PD association outside of the coding region of \textit{LRRK2} suggests that it is likely that the effect on disease risk is mediated by control of mRNA expression. However the tissue type where this effect might take place remains an open question. These data, together with our recent analysis of the MAPT locus, illustrate the complexities of defining precisely how risk loci contribute to disease and illustrate that there is as much work required to dissect a locus as to identify it in the first place [27]. We expect that future gene expression studies will increase the quality and quantity of data to maximize the power to dissect the genetic control of \textit{LRRK2} expression in all those diseases in which it is implicated.

\textbf{Materials and Methods}

\textbf{Case control collections}

Participating studies were genotyped using the ImmunoChip as part of a collaborative agreement with the ImmunoChip Consortium. Genotyping of the UK cases using the Immunochip was undertaken by the WTCCC2 at the Wellcome Trust Sanger Institute which also genotyped the UK control samples. The constituent studies comprising the IPDGC have been described in detail elsewhere [4]. All samples were of Caucasian origin from the following countries: UK (1,864 cases and 1,271 controls), USA (2,215 cases and 2,807 controls), Dutch (402 cases and 304 controls), German (712 cases and 1,153 controls) and French (363 cases and 267 controls).

\textbf{Genotyping}

4 ul of 50 ng/ul gDNA and whole genome amplified DNA samples were marked on the Illumina Infinium Omni1-Quad BeadChip (Illumina), which characterizes over 1 million SNPs across the genome, according to the manufacturer’s instructions. The BeadChips were scanned using an iScan (Illumina) with an AutoLoader (Illumina).

Brain and case control samples were also genotyped using the Immunochip, a custom genotyping array designed for the fine-mapping of auto-immune disorders. The Immunochip also contains 2,000 SNPs selected for replication of a Parkinson’s disease genome-wide association study performed by the International Parkinson’s Disease Genomics Consortium [4,21]. GenomeStudio v.1.8.X (Illumina Corp.) was used for analysing the data and generating SNP calls. SNP calls were created using the HumanOmni1-Quad_v1-0_B(C) cluster file provided by Illumina as a reference. Standard quality control checks carried out on both dataset include the removal of SNPs that are either labelled as copy number variations, labelled as indels, monomorphic in this data, have less than 95% genotyping rate across samples, deviate from Hardy-Weinberg Equilibrium at p<0.0001, had no genomic position information, had less than two heterozygotes or redundant with an existing SNP. After quality control, we had 813,859 SNPs from the Illumina Omi 1M chip and 137,456 from Immunochip and obtained a total of 905,943 after merging the two dataset and resolving for overlapping SNPs. All manipulation with genotyped data was performed in PLINK version 1.07 [28]. Next, we imputed un-typed SNPs from the autosomes and chromosome X using the MACH [29,30], minimac software (http://genome.sph.umich.edu/wiki/Minimac) and the 1000 Genomes Project (May 2011 Haplotype release) which is based on 381 individuals of European descent. For this paper, we restricted the imputed data to SNPs with imputation quality (R\textsuperscript{2}) greater than 0.5.

\textbf{QTL and disease association analysis}

We tested the association between each SNP and each expression profile using the R package snapStats (available as part of the suite of packages Bioconductor). We assumed an additive genetic model for each SNP (1 degree-of-freedom trend test) without additional covariates. Unless otherwise specified, computations using brain expression data use the averaged expression values for each sample across the ten regions surveyed.

\textbf{Liver and monocyte expression datasets}

The liver dataset has been previously described in [23]. It consists of n = 966 liver samples typed using a custom Affymetrix gene expression array, with DNA typed using the Illumina 660W platform. The unique \textit{LRRK2} probe is located in exon 51 of the canonical transcript. The monocyte Gutenberg Health Study dataset has been described in [24]. It consists of n = 1,372 monocyte samples, with expression data measured using the Illumina HT12 BeadChip expression array and DNA was typed using the Affymetrix 6.0 genotyping array. The unique \textit{LRRK2} probe is located in exon 50 of the canonical transcript.

\textbf{Human post-mortem brain tissue collection and dissection}

Brain and CNS tissue originating from 137 control individuals was collected by the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank, Edinburgh, UK [31], and the Sun Health Research Institute (SHRI) an affiliate of Sun Health Corporation, USA [32]. Samples originating from the MRC Sudden Death Brain and Tissue Bank were removed from whole brains as fresh tissue and anatomical regions of interest were sampled from brain coronal slices at autopsy and immediately flash frozen. In the case of samples originating from the SHRI, whole brains were removed as fresh tissue at autopsy and brain coronal slices were frozen. Anatomical regions of interest were sampled from brain coronal slices on dry ice. In all cases control status was confirmed by a consultant neuropathologist. A detailed description of the samples used in the study, tissue processing, pH determination, dissection, quality controls and rationale for covariate correction and statistical analysis of this data set is provided in Trabzuni et al., 2011 [22]. All samples had fully informed consent for retrieval and were authorized for ethically approved scientific investigation (Research Ethics Committee number 10/H0716/3, The national Hospital for Neurology and Neurosurgery & Institute of Neurology Joint Research Ethics Committee). Brain gene expression data are available within the GEO archive, accession number GSE46706.

\textbf{DNA extraction from brain samples}

DNA was extracted from human post-mortem brain tissues using Qiagen DNeasy kit (Qiagen, UK). The DNA quality was accessed using ethidium bromide stained agarose gel. The concentration and purity of each DNA sample was assessed using the NanoDrop ND-1000 Spectrophotometer V3.3.0. The concentration of each sample was calculated, together with the ratio of absorbance at 260 nm/280 nm and 260 nm/230 nm.
RNA isolation, processing, and microarray hybridization

Total RNA was isolated from human post-mortem brain tissues based on the Single-step method of RNA isolation [33] using the miRNAeasy 96 kit (Qiagen, UK). The quality of total RNA was evaluated by the 2100 Bioanalyzer (Agilent, UK) and RNA 6000 Nano Kit (Agilent) before processing with the Ambion® WT Expression Kit and Affymetrix GeneChip Whole Transcript Sense Target Labelling Assay, and hybridization to the Affymetrix Exon 1.0 ST Arrays following the manufacturers’ protocols. Hybridized arrays were scanned on an Affymetrix GeneChip® Scanner 3000 7G and visually inspected for hybridization artefacts. Further details regarding RNA isolation, quality control and processing are reported in [22].

Exon array data analysis

All arrays were pre-processed using RMA quantile normalisation with background correction and probe set summarisation with median polish according to the Robust Multi-array Average (RMA) [34] algorithm in Affymetrix Power Tools 1.14.3 (http://www.affymetrix.com/partners_programs/programs/developer/tools/power_tools.affx). After re-mapping the Affymetrix probe sets onto human genome build 19 (GRCh37) as documented in the NetAffx annotation file (HuEx-1.0-st-v2 Probeset Annotations, Release 354.31), we restricted analysis to 294,943 probe sets that: i) had gene annotation according, ii) did not target intronic regions and iii) contained at least two 25-mer probes with the following properties: a) unique hybridization to target sequence and b) did not contain SNPs as identified at 1% minor allele frequency from the 1,000 Genomes project (haplotype release May 2011). Since most exons are represented by only one probe set, we used the probe set signal intensity as a synonym of exon expression level, unless explicitly mentioned. Gene-level summary data for 26,684 transcripts was also generated by calculating the 10% trimmed mean of the expression values from the corresponding probe sets. In all types of analysis, the date of array hybridisation, brain bank mean of the expression values from the corresponding probe sets. The approach to the selection of reference genes is explained in previous studies [35,36]. In addition, a recent study confirms the efficiency of using this approach in selecting housekeeping genes to normalize in different tissues [37]. Further details regarding sample processing for this section are reported in Trabzuni et al., 2011 [22].

Array validation using direct RNA quantification with branched DNA, QuantiGene® 2.0 assay

Cerebellum, occipital cortex, putamen and white matter samples from 12 individuals were analysed using the QG platform for validation of exon array results. We focused on the target gene for validation, Leucine-rich repeat kinase 2 (LRRK2). We selected ribosomal protein, large, P0 (RPLP0) and ubiquitin C (UBC) as housekeeping genes to normalize the target genes as they showed relatively low variability in expression levels (i.e. low coefficient of variation) in all brain regions in our dataset. The approach to the selection of reference genes is explained in previous studies [35,36]. In addition, a recent study confirms the efficiency of using this approach in selecting housekeeping genes to normalize in different tissues [37]. Further details regarding sample processing for this section are reported in Trabzuni et al., 2011 [22].

Quantitative RT-PCR

Gene expression was quantified by TaqMan real-time PCR technique (Invitrogen, UK) from 30 individuals in cerebellum, occipital cortex and white matter. The LRRK2 specific assays which cover exon-exon boundary included (Hs00968193), (Hs00411194). Fluorescence was collected using the MXPro system (Agilent, UK). All runs were performed in triplicates and we normalized to a geometric mean of three housekeeping genes (PPIA, RPL0, and UBC). The relative expression values were calculated using the delta delta Ct method (ΔΔCt).

Semi-quantitative Reverse-Transcriptase RT-PCR

The validation for the array splicing events was done by using QIAGEN Long Range 2step RT-PCR kit (Qiagen, UK). All primers for this analysis were designed using the Primer3 software (fokker.wi.mit.edu/primer3/input.htm), and then they were BLAST searched against UCSC human. In-silico PCR tools. An aliquot of total RNA from 48 samples, originating from four brain regions (occipital cortex, substantia nigra, medulla and cerebellum) and 12 individuals were used as a subset for a further validation. cDNA was synthesis from 1–2 μg of total RNA using gene specific designed primers under the following conditions: Incubation at 42°C for 90 minutes, followed by enzyme activation at 85°C for 5 minutes. 2.5 μl of cDNA was used to perform the Semi-quantitative RT-PCR for the targeted exons under the following conditions: Initial activation at 93°C for 3 min, followed by denaturation at 93°C for 15 sec, annealing 59°C for 50 sec, elongation 68°C for 50 sec for 30–40 cycles. PCR products were run on a 2% agarose gel (Invitrogen, UK) and photographed using UV illumination to visualize GelRed staining. Images were inverted in Adobe Photoshop.

RNA-Seq data and alignment

cDNA was produced using 100 ng of total RNA and the Ovation RNA-seq system v2 (NuGEN, UK). The concentration of ds-cDNA was determined using the Qubit dsDNA BR assay kit (Invitrogen, UK). Fragmentation of the cDNA was done using a Covaris S220 to a fragment size of 100-500 bp with a median size of 250. DNA libraries were prepared using 1.2 μg cDNA with the TruSeq DNA LT Sample prep kit (illumina, UK). Library concentrations were determined using the KAPA library Quant Kit (KAPA biosystems) before being pooled into pools of three samples. Each pool was sequenced on a single lane of a Paired End HiSeq v3 Flow cell (illumina) using a read length of 101 bp for each read. Reads were aligned against the hg19 reference genome using STAR with default parameters. The STAR alignment used the default standard junctions data provided with the software with the addition of junctions 31–34 and 32–34. We then extracted the count data for all junctions (previously known and novel) from the output of STAR.

Supporting Information

Figure S1 LRRK2 exon 51 expression stratified by rs11175518 in 966 liver samples. (PDF)

Figure S2 LRRK2 exon 33 expression stratified by rs10784486 in 134 brain samples (all ten brain regions shown as well as combined mean across region). (PDF)

Figure S3 LRRK2 exon 33 expression stratified by rs376186 in 134 brain samples (all ten brain regions shown as well as combined mean across region). (PDF)

Table S1 Full list of IPDGc consortium members. (DOCX)

Table S2 Summary of imputability data and functional role of SNPs in the 479 kb long LRRK2 gene region as defined as chr12:40,351,601-40,830,814 (hg19). (DOCX)
Table S3 Summary of the pattern of linkage disequilibrium (correlation coefficient $r^2$) between the SNPs mentioned throughout the text.

(DOCX)

Table S4 Summary of existing eQTL studies performed in human control tissues and cells with relevance to the detection of exon eQTLs relevant to LRRK2.

(DOCX)

Table S5 PCR primers used to characterize the exome exon junctions.

(DOCX)

Table S6 Summary information for the 8 samples (4 individuals and 2 brain regions) with RNA-Seq data.

(DOCX)

Table S7 Number of sequencing reads spanning exon-exon junctions in the 8 samples with RNA-Seq data.

(DOCX)

References


Results are organised by: (top) canonical transcript, annotated junctions (middle) non-canonical transcript, annotated junctions (bottom) novel junctions.

CSV

Acknowledgments

We would like to thank AROS Applied Biotechnology AS company laboratories and Affymetrix for their valuable input.

Author Contributions

Conceived and designed the experiments: DT MR AR CS PL AM RD SL JS. Analyzed the data: DT MR AR PL RB MW JH VP SG TZ AR WE MN. Contributed reagents/materials/analysis tools: RW CS PL AM RD JV HC ES NW RB MW JH AS PH TG KL. AB. Wrote the paper: DT MR PL RB VP.

Supporting Information

Table S7 Number of sequencing reads spanning exon-exon junctions in the 8 samples with RNA-Seq data.