Computational analysis of the LRRK2 interactome

Claudia Manzoni, Paul Denny, Ruth C Lovering, Patrick A Lewis

LRRK2 was identified in 2004 as the causative protein product of the Parkinson's disease locus designated PARK8. In the decade since then, genetic studies have revealed at least 6 dominant mutations in LRRK2 linked to Parkinson's disease, alongside one associated with cancer. It is now well established that coding changes in LRRK2 are one of the most common causes of Parkinson's. Genome wide association studies (GWAs) have, more recently, reported single nucleotide polymorphisms (SNPs) around the LRRK2 locus to be associated with risk of developing sporadic Parkinson's disease and inflammatory bowel disorder. The functional research that has followed these genetic breakthroughs has generated an extensive literature regarding LRRK2 pathophysiology; however there is still no consensus as to the biological function of LRRK2. To provide insight into the aspects of cell biology that are consistently related to LRRK2 activity, we analysed the plethora of candidate LRRK2 interactors available through the BioGRID and IntAct data repositories. We then performed GO terms enrichment for the LRRK2 interactome. We found that, in two different enrichment portals, the LRRK2 interactome was associated with terms referring to transport, cellular organization, vesicles and the cytoskeleton. We also verified that 21 of the LRRK2 interactors are genetically linked to risk for Parkinson's disease or inflammatory bowel disorder. The implications of these findings are discussed, with particular regard to potential novel areas of investigation.

Computational analysis of the LRRK2 interactome

Claudia Manzoni[1, 2]*, Paul Denny[3], Ruth C. Lovering[3] and Patrick A. Lewis[1, 2, 4]

1. School of Pharmacy, University of Reading, Whiteknights, Reading, RG6 6AP, UK;

2. UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK;

3. UCL Institute of Cardiovascular Science, London, WC1E 6BT, UK.

4. Centre for Integrated Neuroscience and Neurodynamics, University of Reading, Whiteknights, Reading, RG6 6AP, UK

*To whom correspondence should be addressed: c.manzoni@reading.ac.uk

Introduction

LRRK2 is the most frequently mutated gene in familial Parkinson's disease (PD) (Paisán-Ruíz et al., 2004; Zimprich et al., 2004), and has also been identified as a risk locus for the sporadic form of the disease (Nalls et al., 2014). There are additional reports implicating allelic variants at the *LRRK2* locus with increased risk of developing inflammatory bowel disorder (IBD) (Jostins et al., 2012). *LRRK2* has also been linked to susceptibility to multibacillary Leprosy (Zhang et al., 2009), although reproducibility of the data is controversial (Marcinek et al., 2013; Zhang et al., 2011). Furthermore, genetic, as well as epidemiological data, supports a role for *LRRK2* in cancer for certain populations (Ruiz-Martínez et al., 2014; Inzelberg et al., 2012; Saunders-Pullman et al., 2010). The protein product of the *LRRK2* gene, Leucine Rich Repeat Kinase 2 (LRRK2) is a large enzyme hosting kinase and GTPase functions surrounded by protein-protein interactions domains (Greggio et al., 2008). It is now well recognized that there are complex regulatory events linking the two enzymatic activities within LRRK2. These include, although are likely not to be limited to, regulation of dimerization by GTP and/or GDP binding and auto-phosphorylation events (Taymans et al., 2011). The importance of LRRK2 in a number of human diseases has motivated many different groups to try and solve the conundrum of its pathophysiological activity. This has led to a large and increasing body of information emerging from functional studies of LRRK2. Despite the bulk of existing data, there is no consensus on the phosphorylation targets of the LRRK2 kinase domain, or with regard to the binding partners with which LRRK2 interacts. Many different functions have been reported to be associated with LRRK2, ranging from cytoskeleton organization to vesicle trafficking, from synaptic activities to autophagy, from mitochondria homeostasis to protein synthesis, involving multiple different signalling

33 cascades, such as the m-TOR and Wnt pathways (Paisán-Ruiz et al., 2013). None of these have been

34 categorically proven; the only reproducible data at the moment appears to be the interaction between

35 LRRK2 and 14-3-3 that has been shown to regulate LRRK2 localization in the cell (Dzamko et al.,

36 2010). The reasons for this volume of different and sometimes contradictory, literature may be the

37 result of variable reproducibility across the functional models available to study LRRK2; the not yet

38 proven, and sometimes axiomatic assumption, that the physiological function of LRRK2 can be

39 inferred *via* its pathology and *vice versa*; and finally the difficulties encountered studying an enzyme

40 possessing two potentially independent activities that may link to independent cellular functions

41 (Lewis & Manzoni, 2012).

42 In the past decade, a number of databases have been established to collect protein annotations from the

43 published literature and high-throughput datasets (Hermjakob et al., 2004; Stark and Tyres, 2003;

44 Ashburner et al., 2000). These repositories store gene and protein data in easily accessible,

45 standardized formats, allowing comparisons between datasets, ontological groups, genes, and

46 gene/protein families. In addition, they can be used to infer new data by deriving them from statistical

47 analysis of gene/protein sets (Orchard et al., 2012a).

48 For a protein such as LRRK2, where the volume of functional literature exceeds the capacity of

49 researchers to analyse on an *ad hoc* basis (>1200 PubMed entries as of 30th October 2014) and where

50 the range of techniques and models can cloud interpretation, a bioinformatics approach using freely

51 available, manually-curated protein-protein interactions (PPI) datasets provides a powerful tool to

52 develop a picture of LRRK2 in the cell environment and infer functional implications.

53 Herein we describe an *in silico* investigation of the LRRK2 interactome. Taking advantage of the

54 manually curated datasets stored in the BioGRID and IntAct repositories (Hermjakob et al., 2004;

55 Stark and Tyres, 2003), we prepared a filtered list of potential LRRK2 interactors, sorted them into

56 heterologous interactions (LRRK2-human protein), which were used for GO terms enrichment, and

57 into homologous interactions (LRRK2-LRRK2), analysed to shed light on LRRK2 dimerization.

58

59 **Materials and Methods**

60 **LRRK2 interactors: IntAct database**

61 http://www.ebi.ac.uk/intact/

62 The IntAct database (Kerrien et al., 2006; Hermjakob et al., 2004) was queried for LRRK2 (human,

63 Q5S007). The list of interactors was filtered to remove the spoke expanded co-complex (i.e.

64 annotations derived by an expansion algorithm sampling co-complexes data into binary interactions

65 and therefore listing proteins which may not interact directly with LRRK2). This reduces the number

66 of proteins not directly interacting with LRRK2, but does not remove all of these proteins. The MI-

67 TAB 2.7 file was downloaded (26th July 2014) and imported into an Excel spreadsheet. Filter #1

68 screened the taxid, to retain only human-human interactions (annotation removed for non-human taxid

69 or chemicals). Filter #2 was applied to remove multiple entries (in the same publication) when: i)

70 Interactor A was used as bait in the first experiment and as prey in the second one and interactor B *vice*

71 *versa*; ii) the same experiment was performed using differently tagged A and B interactors (e.g. GST

72 Vs FLAG Vs HA). To apply filter #2 the following cut-off was established: 2 entries with the same

73 "Publication Identifier(s)" and "#ID(s) interactor A and ID(s) interactor B", were considered duplicated

74 (i.e. one of them to be removed) if "Interaction detection method(s)" AND "Host organism(s)" were

75 identical.

76 As a general note, protein fragments can be retrieved with the same UniProtKB identifier as for the

77 entire protein (e.g. MAP1B catalytic domain (P46821-PRO_0000018605) has the same identifier as the

78 entire MAP1B protein (P46821)); protein isoforms are queried by the same UniProtKB identifier as for

79 the principal isoform (e.g. DVL2 isoform 2 (O14641-2) has the same identifier as DVL2 (O14641)).

80 For this reason no discrimination was made between protein fragments and the entire protein; nor

81 between different protein isoforms and the principal isoform.

82 **LRRK2 interactors: BioGrid database**

83 http://thebiogrid.org/

84 BioGrid database (Chatr-Aryamontri et al., 2012; Breitkreutz, Stark & Tyers, 2003) was queried for

85 LRRK2 (human, Q5S007). The list of interactions was downloaded (12th August 2014) as a TAB 2.0

86 file and imported into an Excel spreadsheet to be analysed. Filters #1 and #2 were applied, as described

87 above for the processing of the IntAct dataset. Filter #1 removed non-human interactors, filter #2

88 removed multiple entries with the same experimental setting.

89 **Enrichment Analysis**

90 GO enrichment analysis has been performed by the use of 2 different web-based applications.

91 WebGestalt (analysis on 20th August 2014, WebGestalt was last updated on 30th January 2013.) (Wang

92 et al., 2013; Zhang, Kirov & Snoddy, 2005) http://bioinfo.vanderbilt.edu/webgestalt/ was used with the

93 following parameters: Reference set for enrichment analysis: h.sapiens_genome, Enrichment: GO

94 analysis; Databases: biological process and cellular component. Analysis was performed using

95   hypergeometric statistics and Bonferroni's correction for multiple testing. Minimum number of genes

96   was set at 2 and significance level was set to retrieve the top10 hits. All the input proteins were

97   mapped. GO enrichment (supported by Panther 9.0, released 20 January 2014; analysis on 19th August

98   2014) (Mi, Muruganujan & Thomas, 2012) http://www.geneontology.org/page/go-enrichment-analysis

99   was run with the following parameters: Species: human, Ontology: biological process. The statistical

100 analysis included Bonferroni's correction for multiple testing. All the input proteins were mapped.

101 Since the enrichment in WebGestalt was for p-values between 10^{-9} and 10^{-7}, an arbitrary p-value

102 threshold was set at 10^{-7} as cut-off for the Panther enrichment.

103 **Comparison between LRRK2 interactome and GWAs hits.**

104 PD GWAs meta-analysis (Nalls et al., 2014) was use to extract SNPs associated with sporadic PD (22

105 SNPs were from discovery phase and replication, 2 SNPs were from replication of previous results (Do

106 et al., 2011; IPDGC & WTCCC2, 2011)). 163 IBD risk associated SNPs were extracted from Jostins *et*

107 *al*., 2012. The position of each positive SNP was retrieved in Entrez SNP using the GRCh38 genome

108 build; genes within an interval of +/- 200 Kbp were listed. The entire set of genes was then queried

109 against the LRRK2 interactome. Positive matches were extracted and the distance between the

110 associated SNP and the coding region was estimated.

111 **Software**

112 Data were stored and analyzed as Excel spreadsheet files. Graphs were prepared by the use of Prism

113 (GraphPad) and Cytoscape (2.6.2, freely available online at http://www.cytoscape.org/) (Saito et al.,

114 2012; Shannon et al., 2003).

115

116 **Results**

117 **LRRK2 interactors: IntAct and BioGRID merged dataset**

118 The IntAct database was queried for LRRK2 (human, Q5S007), and 542 unfiltered annotations were

119 downloaded. After applying filter #1 to include only human interactions and filter #2 to remove

120 multiple annotations with similar experimental details in the same publication (see materials and

121 methods), the total number of annotations dropped from 542 to 307. Of these 307 annotations, 278

122 described interactions between LRRK2 and different partners (heterologous interactions) and 29

123 annotations described LRRK2 self-interaction (homologous interaction).

124   The BioGRID database was similarly queried, and 260 unfiltered annotations were downloaded. After

125   applying filter #1 and filter #2 the total number of annotations dropped from 260 to 230; of which 194

126   were heterologous and 36 homologous interactions.

127   Detailed analysis of the PubMed IDs for the heterologous interactions in the IntAct and BioGRID

128   filtered datasets revealed that annotations came from 43 publications in IntAct and 42 publications in

129   BioGRID. However, only 22 publications were identical in both databases. Thus, considering a total of

130   63 annotated papers, 35% of them were annotated by both IntAct and BioGRID; 65% were contained

131   either in IntAct (21 publications) or BioGRID (20 publications). For this reason the IntAct and the

132   BioGRID datasets for LRRK2 heterologous interactions, after the previously described filtering, were

133   merged to prepare the final dataset for heterologous interactions only (hereafter referred to as "merged

134   dataset"). In the case of publications duplicated between IntAct and BioGRID only one set of

135   annotations was moved to the merge dataset. Discrepancies were found for some of the shared records;

136   in particular, the number of annotations was not necessarily consistent between IntAct and BioGRID,

137   this has been previously reported (Lehne & Schlitt, 2009) and is likely due to differences in the

138   technical classification of the experiments. To overcome this problem, in the case of shared

139   publications, the record containing the major number of annotations was selected to be moved into the

140   merged dataset; however, for 6 shared publications differences were as such that a merge of the two

141   records was necessary (details of these 6 problematic records and how they were merged are in

142   Supplementary Table1). The final, merged dataset for human LRRK2-human protein interactions

143   (heterologous interactions) consisted of 63 publications describing 422 annotations, captured *via*

144   different PPI detection methods (Fig.1), for a total of 269 LRRK2 binding partners (here after referred

145   to as the "complete" LRRK2 interactome).

146   The 422 annotations in the merged dataset were scored based on the following:

147   Low occurrence - the protein identifier was reported just once in the list of 422 annotations.

148   Medium-low occurrence - the protein identifier was reported in 1 publication but with 2 or more

149   experimental approaches.

150   Medium-high occurrence - the protein identifier was reported in 2 or more publications but with the

151   same experimental technique.

152   High occurrence - the protein identifier was reported in 2 or more publications with 2 or more

153   experimental approaches.

154 This ranking system allowed us to classify the interactors as follows: 207 interactors were annotated

155 only once; 24 LRRK2 interactors were reported in 1 publication only but with different methods; 13

156 LRRK2 interactors were reported in more than 1 publication but with the same detection method;

157 finally, 25 LRRK2 interactors were reported in more than 1 publication and with different experimental

158 techniques. Interactors annotated only once are detailed in Supplementary Table 2. We designated

159 interactors that were reported with 2 or more annotations to represent the "LRRK2 interactome", this

160 dataset contains 62 proteins (Fig. 2) and is the dataset that was used for enrichment analysis.

161 The 62 proteins in the LRRK2 interactome were segregated into protein family groups according to

162 their UniProtKB record "family&domains" (http://www.uniprot.org/uniprot/). 22 of the genes could be

163 assigned to 10 different families: heat shock protein 90 family; tubulin (alpha and beta) family;

164 argonaute family; small GTPase superfamily, Rho family; STE Ser/Thr protein kinase family; 14-3-3

165 family; TRAFAC class dynamin-like GTPase superfamily, dynamin/Fzo/YdjA family; DSH family,

166 small GTPase superfamily, Rab family and actin family (Fig.3).

167

168 **GO: "biological process" and "cellular component" enrichment analysis (WebGestalt)**

169 The online platform WebGestalt was used to conduct a GO term enrichment analysis for the 62

170 LRRK2 interactors described in Figure 2. The enrichment settings allowed for retrieval of the first

171 top10 hits. The enriched "GO biological process" terms identified were related to transport/localization

172 and cell organization, with a couple of terms supporting an involvement of the LRRK2 interactome in

173 regulating kinase activity (Fig. 4). For the enriched "GO cell component" terms, as expected, the

174 majority of the LRRK2 interactors were annotated as located in the cytoplasm or cytosol-associated

175 components. However, on a more specific level, the interactors were clustered to vesicles, cytoskeleton

176 and cell projections (Fig. 5). The enriched terms were sorted in GO term groups. Groups were

177 identified as cell organization, transport/localization, regulation of kinase activity, cytosol,

178 cytoskeleton, vesicles and cell projections (Supplementary Table 3). The WebGestalt enrichment

179 portal returns the break-down of proteins contributing toward the enrichment of each single term

180 (Supplementary Dataset 1). Therefore LRRK2 interactors were listed as associated to the groups for

181 which they contributed toward enrichment. The lists were then compared to calculate the percentage of

182 proteins in the LRRK2 interactome to contribute to each single GO term group and the percentage of

183 intersection between groups (Fig. 6). Nearly half of the LRRK2 interactors fall in the groups of

184 transport/localization, cytoskeleton and cell organization. Nearly 30% of them were associated with

185   vesicles and cell projections. As expected, each of the GO term groups presented with an intersection

186   of around 100% with the GO term cytosol group. Of more interest is that 50% of interactors associated

187   with the regulation of kinase activity were shared with vesicles, and 72% with transport/localization

188   thus suggesting a possible role of LRRK2 in controlling kinase activities related to vesicle/membranes

189   and transport processes. 79% of interactors associated with vesicles were also associated with

190   transport/localization, thus implying a possible role of LRRK2 in vesicle dynamics related to

191   intracellular transport/trafficking.

192

193   **GO: "biological process" enrichment analysis (Panther)**

194   The filtered list of 62 LRRK2 interactors, was also used for a second, independent enrichment analysis

195   using the GO Consortium term enrichment service, supported by Panther (Mi, Muruganujan &

196   Thomas, 2012). The analysis retrieved 150 "GO biological process" enriched terms. These terms were

197   manually divided into 13 groups: general terms, transport/localization, membrane processes, signalling,

198   regulation of catalysis, metabolism, catabolism, regulation of kinase activity, regulation of

199   mitochondrion organization, cell death, development, cell organization and immune response

200   (complete list of terms, divided in groups, with p-values and sample frequencies is reported in

201   Supplementary Table 4). The terms classified within the general terms group were excluded from

202   following analyses because of their lack of specificity; the terms present in the 12 remaining GO term

203   groups were ranked based on their p-values in significance levels from 10^{-16} to 10^{-7}. The contribution

204   of each functional group toward the enrichment in a particular significant level was calculated as

205   follows (function 1):

$$\% \, functional \, group \, G \, in \, significance \, level \, N = \frac{\sum_{i=1}^{g}(i) * 100}{\sum_{j=1}^{n}(j)}$$

206

207   *Function 1*

208   Where g is the number of terms enriched in the functional group G within the significance level N; n is

209   the total number of enriched terms across all the functional groups within the significance level N.

210   Results are shown in Figure 7. The significant levels 10^{-16}, 10^{-15}, 10^{-14} and 10^{-12} were only composed

211   of terms belonging to the functional groups of transport/localization, cell organization and membrane

212   processes. Once these highly significant terms were taken aside, all the other GO term groups started to

213   present their contributions toward the enrichment (in the significant levels from 10^{-11} to 10^{-7}).

214

215   **LRRK2 interactome: IntaAct and BioGRID merged dataset for LRRK2 self-interactions**

216   The IntAct database contained 29 annotations for LRRK2 self-interactions (homologous interactions);

217   BioGRID, a total of 36 homologous interactions. Only 7 publications were shared between the two

218   databases therefore, in the total of 31 publications reporting LRRK2 self-interaction, 77% of the

219   information was captured independently by IntAct or BioGRID. Following the same procedure used

220   for the heterologous interactions, a merged dataset was prepared, with a final count of 31 publications

221   and 53 annotations for LRRK2 self-interaction, based on different experimental methodologies (Fig.

222   8).

223   The IntAct database (not BioGRID) reports details (when available) on the regions of the LRRK2

224   protein participating in protein interactions. Based on this information, a profile of the LRRK2

225   fragments/residues associated with LRRK2 self-interaction was prepared (Fig. 9). As expected, the

226   catalytic core of LRRK2 was most frequently reported to be associated with LRRK2 dimerization

227   (with a top score of 17 publications), and all the residues tested directly for self-interaction lay in the

228   catalytic core of LRRK2.

229

230   **LRRK2 interactome and disease**

231   The complete LRRK2 interactome comprised of 269 proteins (the filtered interactome combined with

232   the interactors identified by 1 annotation only) was compared with data from PD and IBD GWAs.

233   Since the associated SNP is only a marker for a genomic locus, not an indication of a specific gene, all

234   the coding open reading frames (ORF), for a total of 156 ORFs in the interval +/- 200 Kbp from the

235   associated SNP, have been taken into account. The positive matches between GWAs hits and LRRK2

236   interactors were recorded (Table 1). Four proteins in the LRRK2 interactome (SNCA, RAB7L1, GAK

237   and MAPT) were found to be candidate proteins in the PD GWA. These proteins were annotated

238   multiple times as LRRK2 interactors from different publications. Among these proteins, SNCA is not

239   only associated with the risk of sporadic PD, it is also a gene whose mutations are responsible for

240   inheritance of familial PD. 1004 ORFs were identified within a 200 Kbp region either side of

241   associated SNPs for IBD and 17 proteins in the LRRK2 interactome were found to be positive matches,

242   4 out of 17 were indicated as candidate proteins in the GWA study, the others were just around the

243   associated SNP; the LRRK2 interactors matching the IBD GWA, with the exception CDC37, were

244   annotated in just 1 publication.

245

246   **Discussion**

247   The implication of LRRK2 in different human diseases has made this protein the centre of interest for

248   many research groups, leading to a large number of functional studies. The breadth of the LRRK2

249   literature is complemented by its depth, with a wide range of different approaches and techniques used

250   in investigations. A PubMed search for LRRK2 retrieved 1257 publications on the 30th October 2014.

251   As a consequence of this, LRRK2 is an excellent candidate for *in silico* analysis to critically appraise

252   the literature creating a synthesis of our understanding of the LRRK2 interactome. LRRK2 literature

253   can be divided into two different blocks, the first one comprises functional publications that study

254   LRRK2 without reporting PPI data, the second contains a smaller number of publications with details

255   about physical interactions between LRRK2 and partners. This second set of publications is what is

256   annotated in PPI databases. The critical collection of the LRRK2 interacting proteins can be used to

257   generate the state of the art LRRK2 interactome. The interactome can be then analysed to rationalize

258   the understanding of the existing functional literature, which describe the role of LRRK2 in a multitude

259   of biological processes (not annotated in PPI databases). The result of such analysis can also suggest

260   possible future wet-lab investigations.

261

262   When searching for LRRK2 interactions, we found that the information stored in the IntAct and

263   BioGRID databases overlapped only for around 50%; this is probably a consequence of the curator's

264   choice in paper selection and the database specialization. Moreover IntAct and BioGRID are both

265   members of the IMEX consortium and thus committed to minimize redundant annotations (Orchard et

266   al., 2012b). Therefore, to have a more detailed view of the LRRK2 interactome the two databases were

267   merged. After filtering out interactions with non-human partners and repeated annotations, 422 entries,

268   corresponding to 269 different interactors, were found to describe the LRRK2 interactome. Only 23%

269   of the interactors (62 interactors) were confirmed with 2 or more annotations coming from either

270   different publications or different experimental approaches; the percentage was reduced to 9.3% (25

271   interactors) when the selection criteria were strengthened allowing confirmation in at least 2 different

272   publications and by using at least 2 different techniques. It is noteworthy that LRRK2 is annotated

273   frequently as self-interaction partner. This has been reported in 31 publications and classified with 54

274   entries. After analysis of the LRRK2 interactors that have been confirmed with at least 2 annotations,

275   53% of the interactors fall into 10 protein families; 8 of those were identified independently by

276   different research groups. This first result has at least 2 consequences. First, it suggests that, based on

277   the features of the annotations stored in databases, not all entries are equal. Future functional research

278   should therefore be oriented toward the interactors with stronger experimental evidence, while more

279   biochemical investigation is needed to test the others, to confirm or discard them. Secondly, by looking

280   at the methods used to retrieve the interactions, it has to be taken into account that many annotations

281   come from hypothesis driven approaches: if a protein was previously mentioned, it is more likely that

282   other researchers will plan to investigate it. This results in a degree of ascertainment bias: when a

283   protein is identified as probable interactor, then it is more frequently examined and therefore annotated.

284   This may influence the gathering of annotations around a specific family. Frequently annotated

285   proteins are probably reliable interaction partners. Many interactors are annotated just once; we have

286   considered them less reliable since they have not yet been confirmed. Some of them may represent

287   false positives; others may have just been ignored by replication studies due to lack of biological

288   appeal with regard to further analysis.

289

290   The filtered list of 62 LRRK2 interactors (i.e. interactors that have been confirmed with at least 2

291   annotations) was used for enrichment analysis to identify specific biological processes (indicated by

292   GO terms) that LRRK2 is likely to be associated with. These data can be used to infer the function of

293   the LRRK2 interactome. Moreover, based on the assumption that any other protein annotated to these

294   biological processes may interact (directly or indirectly) with LRRK2, these data guide the deduction

295   of additional potential LRRK2 interacting proteins. Enrichment is performed by algorithms developed

296   to recognize if a descriptor (e.g. a GO term associated with biological process or cellular component) is

297   particularly enriched in the list of LRRK2 interactors in comparison with what is expected due to the

298   recurrence of that term in a standard reference list of proteins or genes. This analysis has been

299   performed using Panther (through the GO Consortium portal) and WebGestalt, two freely available

300   online platforms. Since WebGestalt and Panther apply different statistical approaches and use different

301   versions of the GO annotation and ontology datasets they were used to provide 2 independent analyses.

302   The results using both tools were, however, very similar, pointing towards a comparable functional

303   interpretation of LRRK2 interactome; specifically, all the top 10 WebGestalt hits were also present in

304   the Panther enriched list. It has to be noted that a number of 150 genes/proteins is generally suggested

305 as minimum to obtain a significant enrichment (http://geneontology.org/faq/what-minimum-

306 information-include-functional-analysis-paper). Our analysis included just 62 LRRK2 interactors,

307 nevertheless we found statistically significant enrichment with both of the two algorithms we used.

308 This may suggest that the group of proteins we used as input (the LRRK2 interactome) had a low level

309 of background noise which therefore didn't mask enrichment, and this was probably achieved by using

310 a screened interactome composed only by interactors that have been confirmed in at least one

311 replication study.

312

313 In the WebGestalt analysis, 50% of the LRRK2 interactors contributed toward the enrichment of

314 functional terms related with transport/localization, 68% with cell organization and 29% with

315 regulation of kinase activity. 40% of the LRRK2 interactors contributed toward the enrichment of

316 terms associated with the cytoskeleton, 37% with cell projections and 31% with vesicles. The

317 enrichment of terms related to the regulation of kinase activity was expected, as LRRK2 is an active

318 kinase and likely to regulate other kinases in cascade. However the LRRK2 interactors contributing to

319 this specific enrichment were also present in the GO term groups of vesicles (50%) and

320 transport/localization (72%) thus suggesting LRRK2's interactome may regulate kinase functions in

321 vesicle transport. 60% of the LRRK2 interactors involved in the GO term group of cytoskeleton also

322 contributed toward the enrichment of transport. Similarly, 79% of the interactors involved in the GO

323 term group of vesicles were also annotated as involved in transport, thus suggesting that the LRRK2

324 interactome influences transport processes related with cytoskeleton and vesicles dynamics. These

325 results are in accordance with recent functional literature that sees a role of LRRK2 in vesicle mediated

326 processes like endocytosis (Alegre-Abarrategui et al., 2009), synaptic function (Cirnaru et al., 2014;

327 Parisiadou et al., 2014) and autophagy (Schapansky et al., 2014; Manzoni et al., 2013; Gómez-Suaga et

328 al., 2012). It is of note that the papers referenced in this discussion section are not annotated in the

329 BioGRID neither in the IntAct databases since they are either too recent, or they report only a

330 functional evidence for LRRK2 activity and do not contain PPI records. For this reason they have not

331 contributed toward the enrichment, so they independently support the results.

332

333 The Panther analysis identified more descriptive GO terms as enriched, compared with the WebGestalt

334 analysis (supplementary Table 4). For this reason, we then analysed details of each single, enriched GO

335 term group. For the most significantly enriched group in Panther, i.e transport/localization, the 3 most

336   specific terms were "cytoskeleton-dependent intracellular transport", "cellular protein localization" and

337   "vesicle-mediated transport" thus reiterating the importance of LRRK2 interactome for the process of

338   transport and vesicular trafficking. GTPases and regulators of GTPase activity are a protein group

339   represented in the LRRK2 interactome, it is therefore unsurprising that the GO term GTP catabolic

340   process was enriched in the analysis. Similarly, the enrichment of many terms related to nucleotide

341   catabolism and metabolism was expected due to the presence of ATP/GTPases in the LRRK2

342   interactome. The function of many kinases is to regulate other kinases therefore the enrichment of

343   terms associated with kinase regulation was expected. It was however impossible to predict from this

344   analysis whether the 11 LRRK2 associated kinases (Supplementary Table 5) are being regulated by

345   LRRK2 or whether the LRRK2 interactome is regulating LRRK2's activity. The child terms in the GO

346   term group of cell death specifically indicated the regulation of this process and the special cell death

347   defined as apoptosis. Very interestingly, 3 out of the 5 terms enriched in the GO domain of regulation

348   of mitochondrion organization were related to the regulation of mitochondrial processes in apoptosis

349   (supplementary Table 4). The combination of this information may suggest that the key enriched

350   process is apoptosis rather than the general process of cell death, and that it is the apoptotic processes

351   taking place at the mitochondria, rather than mitochondria generic processes, which are relevant to the

352   function of the LRRK2 interactome. When the signalling processes were looked at in detail, all the

353   enriched terms appeared to be very general, with the exception of "Fc receptor signalling pathway" (p-

354   value = $1.32e^{-11}$, 21% of LRRK2 interactors contributing to the enrichment of this term). This is a very

355   specific GO term that in the ancestor GO chart is directly connected with cell surface signaling and

356   immune response, and supports the recent findings that microglial Fc receptors have been associated

357   with alpha-synuclein-induced pro-inflammatory signaling in PD (Cao, Standaert & Harms, 2012).

358   There were 8 terms enriched within the immune response domain; 4 of them were parents of the

359   specific term "immune response-regulating cell surface receptor signaling pathway" and thus also have

360   a parental relationship with the enriched GO term "Fc receptor signaling pathway". It is known that

361   LRRK2 is expressed in the immune system (Thévenet et al., 2011) and that its total level and

362   phosphorylation state are modulated by immune cell stimulation (Moehle et al., 2012; Kubo et al.,

363   2010). Finally, the developmental process domain presented with 15 enriched terms; 5 of them were

364   general terms; the remaining 10 terms were associated with development of neurons and the nervous

365   system. Very little is known about LRRK2 and development; however it has been shown that total

366   LRRK2 levels vary according to different developmental stages (Giesert et al., 2013; Zechel et al.,

367   2010).

368

369   In the light of these observations we propose the function of LRRK2 interactome to be associated with

370   transport and trafficking, possibly regulating enzymatic events associated with cytoskeleton and

371   vesicles. However, we recognize that LRRK2 probably has a role as a hub protein, thus controlling

372   many different functions at a time, and this is likely to be one of the reasons why it has proved so

373   challenging to reach a consensus on LRRK2 physiological function. Based on our results we suggest

374   the role of LRRK2 can also be modulated depending on the cell type analysed (i.e. immune *versus*

375   nervous system) and possibly the developmental stage, so the outcome of wet lab research may depend

376   on the cellular system used to infer interaction. Not only does this give a possible explanation for the

377   number of different functions that are described for LRRK2 in different experimental models; it also

378   suggests a possible reason for the plethora of different diseases LRRK2 is associated with and opens to

379   discussions of a LRRK2-oriented therapy. If LRRK2 regulates different biological processes in

380   different tissues, a hypothetical targeting of LRRK2 during disease may generate unexpected, and

381   unwanted, side effects. It would be beneficial to analyse LRRK2 interactome according on the cell type

382   in which the interaction was reported. Unfortunately at the moment the majority of the annotations

383   comes from *in vitro* studies and cancer cell lines; we don't have enough annotations from primary

384   cultures and physiological model systems to perform cell type specific analysis yet.

385   The analysis of the 200 Kbp around the associated SNPs for PD and IBD revealed that some of the

386   proteins encoded in these regions are included in the LRRK2 interactome. The influence of associated

387   SNPs on the control of the regions around them is unknown. They are markers for loci involved in the

388   risk for disease, they may modify the gene in which they actually reside or alternatively regulate

389   multiple genes around them. The presence of 21 LRRK2 interactors around associated SNPs for PD

390   and IBD reinforces the role of LRRK2 in these diseases and provides a system to prioritise candidates

391   to be evaluated in the context of the molecular mechanism of sporadic PD and IBD.

392

393   **Conclusions**

394   The analysis of the LRRK2 interactome we described above provides an overview of the published

395   research on this protein, and a system in which to identify the interacting proteins with the most

396   reliable supporting experimental evidence. The enrichment analysis of this interactome provides an

397   indication of the pathways and processes LRRK2 can influence within its cellular environment. This

398   analysis indicates that LRRK2 may be involved in more than one function, perhaps depending on the

399   cellular type and developmental stage. Intriguingly, some of the proteins in the LRRK2 interactome

400   have also been suggested to be genetic risk factors for PD and IBD. Our approach can therefore guide

401   future research priorities as it suggests focusing on these interactors and their relationship to LRRK2

402   and disease. Moreover the multiple interactions and cellular functions associated with LRRK2 suggest

403   caution in using it as a drug target given the potential for off-target problems. Importantly, *in silico*

404   analyses such as those described above can provide a starting point for further hypothesis driven wet

405   laboratory based investigations – opening up new avenues that may reveal important insights into the

406   biology of LRRK2.

407

408   **References**

409   Alegre-Abarrategui J, Christian H, Lufino MM, Mutihac R, Venda LL, Ansorge O, Wade-Martins R.

410   2009. LRRK2 regulates autophagic activity and localizes to specific membrane microdomains in a

411   novel human genomic reporter cellular model. *Human Molecular Genetics* 18:4022-4034.

412

413   Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS,

414   Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE,

415   Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The

416   Gene Ontology Consortium. *Nature Genetics* 25:25-29

417

418   Breitkreutz BJ, Stark C, Tyers M. 2003. The GRID: the General Repository for Interaction Datasets.

419   *Genome Biology* 4: R23.

420

421   Cao S, Standaert DG, Harms AS. 2012. The gamma chain subunit of Fc receptors is required for alpha-

422   synuclein-induced pro-inflammatory signaling in microglia. *Journal of Neuroinflammation* 9:259.

423

424   Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L,

425   Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M,

426   Oughtred R, Dolinski K, Tyers M. 2013. The BioGRID interaction database: 2013 update. *Nucleic*

427   *Acids Research* 41: D816-23.

428

429 Cirnaru MD, Marte A, Belluzzi E, Russo I, Gabrielli M, Longo F, Arcuri L, Murru L, Bubacco L,

430 Matteoli M, Fedele E, Sala C, Passafaro M, Morari M, Greggio E, Onofri F, Piccoli G. 2014. LRRK2

431 kinase activity regulates synaptic vesicle trafficking and neurotransmitter release through modulation

432 of LRRK2 macro-molecular complex. *Frontiers in Molecular Neuroscience* 7:49.

433

434 Dzamko N, Deak M, Hentati F, Reith AD, Prescott AR, Alessi DR, Nichols RJ. 2010. Inhibition of

435 LRRK2 kinase activity leads to dephosphorylation of Ser(910)/Ser(935), disruption of 14-3-3 binding

436 and altered cytoplasmic localization. *Biochemical Journal* 430: 405-413.

437

438 Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM,

439 Tanner CM, Langston JW, Wojcicki A, Eriksson N. 2011. Web-based genome-wide association study

440 identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genetics*

441 7:e1002141.

442

443 Giesert F, Hofmann A, Bürger A, Zerle J, Kloos K, Hafen U, Ernst L, Zhang J, Vogt-Weisenhorn DM,

444 Wurst W. 2013. Expression analysis of Lrrk1, Lrrk2 and Lrrk2 splice variants in mice. *PLoS One*

445 8:e63778.

446

447 Gómez-Suaga P, Luzón-Toro B, Churamani D, Zhang L, Bloor-Young D, Patel S, Woodman PG,

448 Churchill GC, Hilfiker S. 2012. Leucine-rich repeat kinase 2 regulates autophagy through a calcium-

449 dependent pathway involving NAADP. *Human Molecular Genetics* 21:511-525.

450

451 Greggio E, Zambrano I, Kaganovich A, Beilina A, Taymans JM, Daniëls V, Lewis P, Jain S, Ding J,

452 Syed A, Thomas KJ, Baekelandt V, Cookson MR. 2008. The Parkinson disease-associated leucine-rich

453 repeat kinase 2 (LRRK2) is a dimer that undergoes intramolecular autophosphorylation. *Journal*

454 *Biological Chemistry* 283:16906-16914.

455

456 Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M,

457 Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G, Sherman D,

458   Apweiler R. 2004. IntAct: an open source molecular interaction database. *Nucleic Acids Research* 32:
459   D452-5.

460

461   International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Trust Case Control
462   Consortium 2 (WTCCC2). 2011. A two-stage meta-analysis identifies several new loci for Parkinson's
463   disease. *PLoS Genetics* 7: e1002142.

464

465   Inzelberg R, Cohen OS, Aharon-Peretz J, Schlesinger I, Gershoni-Baruch R, Djaldetti R, Nitsan Z,
466   Ephraty L, Tunkel O, Kozlova E, Inzelberg L, Kaplan N, Fixler Mehr T, Mory A, Dagan E,
467   Schechtman E, Friedman E, Hassin-Baer S. 2012. The LRRK2 G2019S mutation is associated with
468   Parkinson disease and concomitant non-skin cancers. *Neurology* 78:781-786.

469

470   Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y,
471   Anderson CA, Essers J, Mitrovic M, Ning K, Cleynen I, Theatre E, Spain SL, Raychaudhuri S, Goyette
472   P, Wei Z, Abraham C, Achkar JP, Ahmad T, Amininejad L, Ananthakrishnan AN, Andersen V,
473   Andrews JM, Baidoo L, Balschun T, Bampton PA, Bitton A, Boucher G, Brand S, Büning C, Cohain
474   A, Cichon S, D'Amato M, De Jong D, Devaney KL, Dubinsky M, Edwards C, Ellinghaus D, Ferguson
475   LR, Franchimont D, Fransen K, Gearry R, Georges M, Gieger C, Glas J, Haritunians T, Hart A,
476   Hawkey C, Hedl M, Hu X, Karlsen TH, Kupcinskas L, Kugathasan S, Latiano A, Laukens D,
477   Lawrance IC, Lees CW, Louis E, Mahy G, Mansfield J, Morgan AR, Mowat C, Newman W, Palmieri
478   O, Ponsioen CY, Potocnik U, Prescott NJ, Regueiro M, Rotter JI, Russell RK, Sanderson JD, Sans M,
479   Satsangi J, Schreiber S, Simms LA, Sventoraityte J, Targan SR, Taylor KD, Tremelling M, Verspaget
480   HW, De Vos M, Wijmenga C, Wilson DC, Winkelmann J, Xavier RJ, Zeissig S, Zhang B, Zhang CK,
481   Zhao H; International IBD Genetics Consortium (IIBDGC), Silverberg MS, Annese V. Hakonarson H,
482   Brant SR, Radford-Smith G, Mathew CG, Rioux JD, Schadt EE, Daly MJ, Franke A, Parkes M,
483   Vermeire S, Barrett JC, Cho JH. 2012. Host-microbe interactions have shaped the genetic architecture
484   of inflammatory bowel disease. *Nature* 491:119-124.

485

486   Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M,
487   Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L,

488 Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H. 2007.

489 IntAct - an open source molecular interaction database. *Nucleic Acids Research* 35: D561-5.

490

491 Kubo M, Kamiya Y, Nagashima R, Maekawa T, Eshima K, Azuma S, Ohta E, Obata F. 2010. LRRK2

492 is expressed in B-2 but not in B-1 B cells, and downregulated by cellular activation. *Journal*

493 *Neuroimmunology* 229: 123-128.

494

495 Lehne B, Schlitt T. 2009. Protein-protein interaction databases: keeping up with growing interactomes.

496 *Human Genomics* 3:291-297.

497

498 Lewis PA, Manzoni C. 2012. LRRK2 and human disease: a complicated question or a question of

499 complexes? *Science Signaling* 5:pe2.

500

501 Manzoni C, Mamais A, Dihanich S, Abeti R, Soutar MP, Plun-Favreau H, Giunti P, Tooze SA,

502 Bandopadhyay R, Lewis PA. 2013. Inhibition of LRRK2 kinase activity stimulates macroautophagy.

503 *Biochimica Biophysica Acta* 1833:2900-2910.

504

505 Marcinek P, Jha AN, Shinde V, Sundaramoorthy A, Rajkumar R, Suryadevara NC, Neela SK, van

506 Tong H, Balachander V, Valluri VL, Thangaraj K, Velavan TP. 2013. LRRK2 and RIPK2 variants in

507 the NOD 2-mediated signaling pathway are associated with susceptibility to Mycobacterium leprae in

508 Indian populations. *PLoS One* 8:e73103.

509

510 Mi H, Muruganujan A, Thomas PD. 2012. PANTHER in 2013: modeling the evolution of gene

511 function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research* 41:

512 D377-86.

513

514 Moehle MS, Webber PJ, Tse T, Sukar N, Standaert DG, DeSilva TM, Cowell RM, West AB. 2012.

515 LRRK2 inhibition attenuates microglial inflammatory responses. *Journal Neuroscience* 32:1602-1611.

516

517 Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL, Kara E, Bras J,

518 Sharma M, Schulte C, Keller MF, Arepalli S, Letson C, Edsall C, Stefansson H, Liu X, Pliner H, Lee

519   JH, Cheng R, International Parkinson's Disease Genomics Consortium (IPDGC), Parkinson's Study

520   Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI), 23andMe,

521   GenePD, NeuroGenetics Research Consortium (NGRC), Hussman Institute of Human Genomics

522   (HIHG), Ashkenazi Jewish Dataset Investigator, Cohorts for Health and Aging Research in Genetic

523   Epidemiology (CHARGE), North American Brain Expression Consortium (NABEC), United Kingdom

524   Brain Expression Consortium (UKBEC), Greek Parkinson's Disease Consortium, Alzheimer Genetic

525   Analysis Group, Ikram MA, Ioannidis JP, Hadjigeorgiou GM, Bis JC, Martinez M, Perlmutter JS,

526   Goate A, Marder K, Fiske B, Sutherland M, Xiromerisiou G, Myers RH, Clark LN, Stefansson K,

527   Hardy JA, Heutink P, Chen H, Wood NW, Houlden H, Payami H, Brice A, Scott WK, Gasser T,

528   Bertram L, Eriksson N, Foroud T, Singleton AB. 2014. Large-scale meta-analysis of genome-wide

529   association data identifies six new risk loci for Parkinson's disease. *Nature Genetics* 46:989-993.

530

531   Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J, Bidwell S, Bridge A, Briganti L, Brinkman FSL,

532   Cesareni G, Chatr-aryamontri A, Chautard E, Chen C, Dumousseau M, Goll J, Hancock REW,

533   Hannick LI, Jurisica I, Khadake J, Lynn DJ, Mahadevan U, Perfetto L, Raghunath A, Ricard-Blum S,

534   Roechert B, Salwinski L, Stümpflen V, Tyers M, Uetz P, Xenarios I, Hermjakob H. 2012. Protein

535   interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nature Methods* 9:

536   345-350.

537

538   Paisán-Ruíz C, Jain S, Evans EW, Gilks WP, Simón J, van der Brug M, López de Munain A, Aparicio

539   S, Gil AM, Khan N, Johnson J, Martinez JR, Nicholl D, Carrera IM, Pena AS, de Silva R, Lees A,

540   Martí-Massó JF, Pérez-Tur J, Wood NW, Singleton AB. 2004. Cloning of the gene containing

541   mutations that cause PARK8-linked Parkinson's disease. *Neuron* 44: 595-600.

542

543   Paisán-Ruiz C, Lewis PA, Singleton AB. 2013. LRRK2: cause, risk, and mechanism. *Journal of*

544   *Parkinsons Disease* 3:85-103.

545

546   Parisiadou L, Yu J, Sgobio C, Xie C, Liu G, Sun L, Gu XL, Lin X, Crowley NA, Lovinger DM, Cai H.

547   2014. LRRK2 regulates synaptogenesis and dopamine receptor activation through modulation of PKA

548   activity. *Nature Neuroscience* 17:367-376.

549

550 Ruiz-Martínez J, de la Riva P, Rodríguez-Oroz MC, Mondragón Rezola E, Bergareche A, Gorostidi A,

551 Gago B, Estanga A, Larrañaga N, Sarasqueta C, López de Munain A, Martí Massó JF. 2014.

552 Prevalence of cancer in Parkinson's disease related to R1441G and G2019S mutations in LRRK2.

553 *Movement Disorders* 29:750-755.

554

555 Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T. 2012. A

556 travel guide to Cytoscape plugins. *Nature Methods* 9:1069-1076.

557

558 Saunders-Pullman R, Barrett MJ, Stanley KM, Luciano MS, Shanker V, Severt L, Hunt A, Raymond

559 D, Ozelius LJ, Bressman SB. 2010. LRRK2 G2019S mutations are associated with an increased cancer

560 risk in Parkinson disease. *Movement Disorders* 25: 2536-2541.

561

562 Schapansky J, Nardozzi JD, Felizia F, LaVoie MJ. 2014. Membrane recruitment of endogenous

563 LRRK2 precedes its potent regulation of autophagy. *Human Molecular Genetics* 15: 4201-4214.

564

565 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.

566 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks.

567 *Genome Research* 13:2498-2504.

568

569 Taymans JM, Vancraenenbroeck R, Ollikainen P, Beilina A, Lobbestael E, De Maeyer M, Baekelandt

570 V, Cookson MR. 2011. LRRK2 kinase activity is dependent on LRRK2 GTP binding capacity but

571 independent of LRRK2 GTP binding. *PLoS One* 6:e23207.

572

573 Thévenet J, Pescini Gobert R, Hooft van Huijsduijnen R, Wiessner C, Sagot YJ. 2011. Regulation of

574 LRRK2 expression points to a functional role in human monocyte maturation. *PLoS One* 6:e21519.

575

576 Wang J, Duncan D, Shi Z, Zhang B. 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt):

577 update 2013. *Nucleic Acids Research* 41:W77-83.

578

579 Zhang B, Kirov S, Snoddy J. 2005. WebGestalt: an integrated system for exploring gene sets in various

580 biological contexts. *Nucleic Acids Research* 33:W741-8.

581

582 Zhang FR, Huang W, Chen SM, Sun LD, Liu H, Li Y, Cui Y, Yan XX, Yang HT, Yang RD, Chu TS,

583 Zhang C, Zhang L, Han JW, Yu GQ, Quan C, Yu YX, Zhang Z, Shi BQ, Zhang LH, Cheng H, Wang

584 CY, Lin Y, Zheng HF, Fu XA, Zuo XB, Wang Q, Long H, Sun YP, Cheng YL, Tian HQ, Zhou FS, Liu

585 HX, Lu WS, He SM, Du WL, Shen M, Jin QY, Wang Y, Low HQ, Erwin T, Yang NH, Li JY, Zhao X,

586 Jiao YL, Mao LG, Yin G, Jiang ZX, Wang XD, Yu JP, Hu ZH, Gong CH, Liu YQ, Liu RY, Wang

587 DM, Wei D, Liu JX, Cao WK, Cao HZ, Li YP, Yan WG, Wei SY, Wang KJ, Hibberd ML, Yang S,

588 Zhang XJ, Liu JJ. 2009. Genomewide association study of leprosy. *New England Journal of Medicine*

589 361:2609-2618.

590

591 Zhang F, Liu H, Chen S, Low H, Sun L, Cui Y, Chu T, Li Y, Fu X, Yu Y, Yu G, Shi B, Tian H, Liu D,

592 Yu X, Li J, Lu N, Bao F, Yuan C, Liu J, Liu H, Zhang L, Sun Y, Chen M, Yang Q, Yang H, Yang R,

593 Zhang L, Wang Q, Liu H, Zuo F, Zhang H, Khor CC, Hibberd ML, Yang S, Liu J, Zhang X. 2011.

594 Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nature*

595 *Genetics* 43:1247-1251.

596

597 Zechel S, Meinhardt A, Unsicker K, von Bohlen Und Halbach O. 2010. Expression of leucine-rich-

598 repeat-kinase 2 (LRRK2) during embryonic development. *International Journal of Developmental*

599 *Neuroscience* 28:391-399.

600

601 Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M, Uitti RJ,

602 Calne DB, Stoessl AJ, Pfeiffer RF, Patenge N, Carbajal IC, Vieregge P, Asmus F, Müller-Myhsok B,

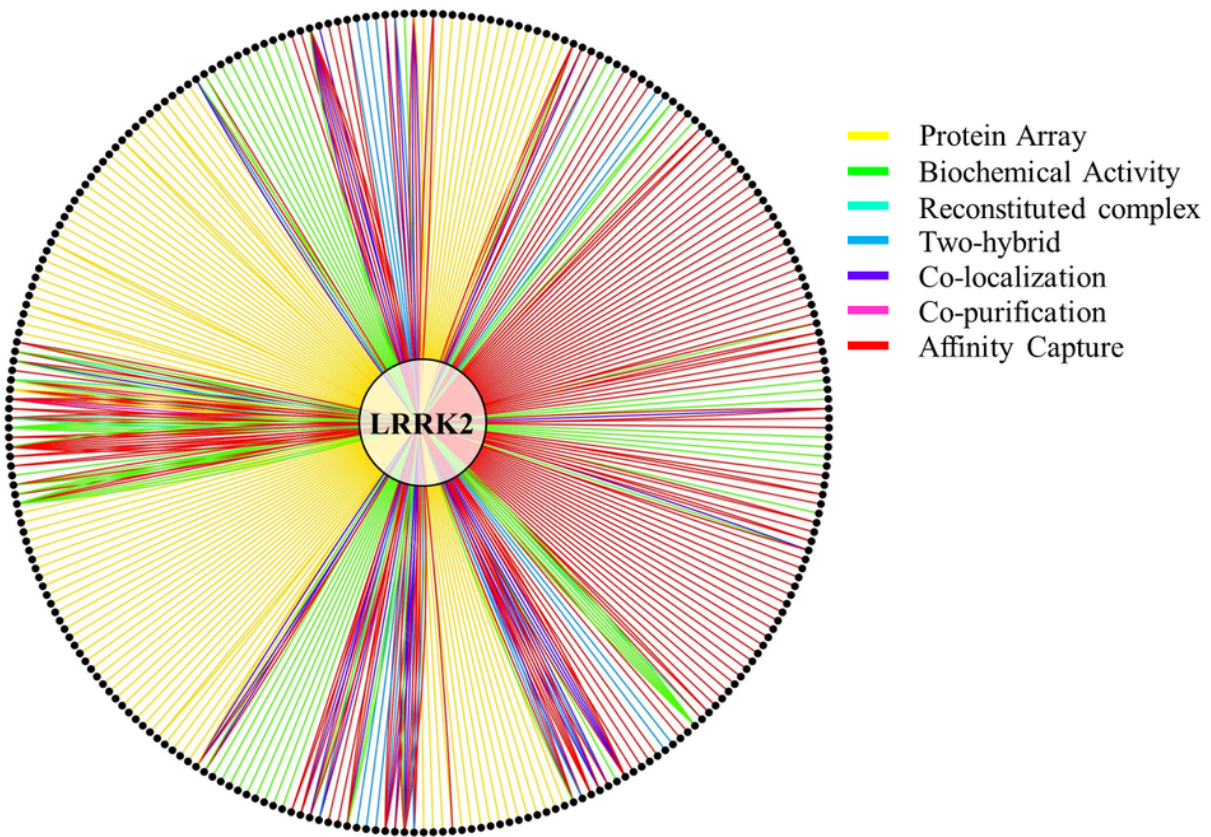
603 Dickson DW, Meitinger T, Strom TM, Wszolek ZK, Gasser T. 2004. Mutations in LRRK2 cause

604 autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* 44:601-607.

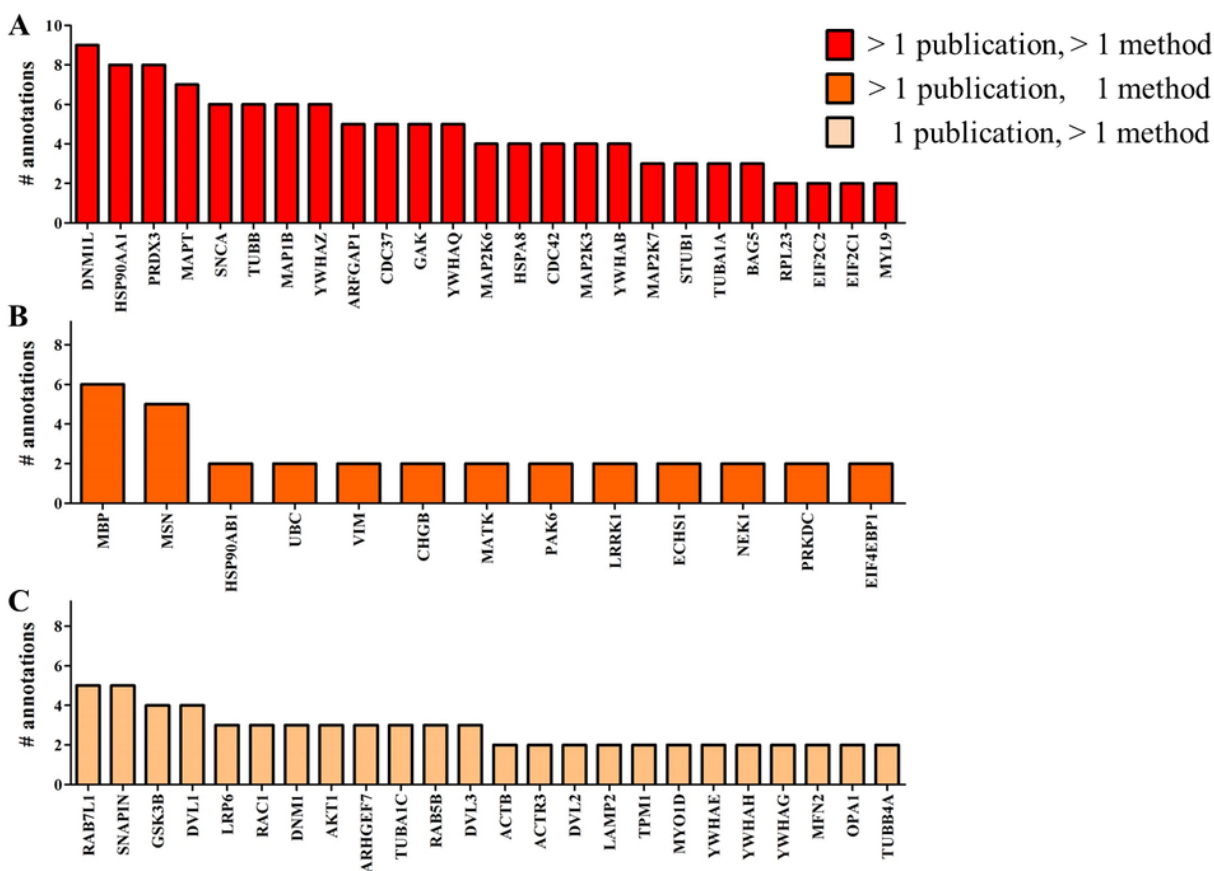
605

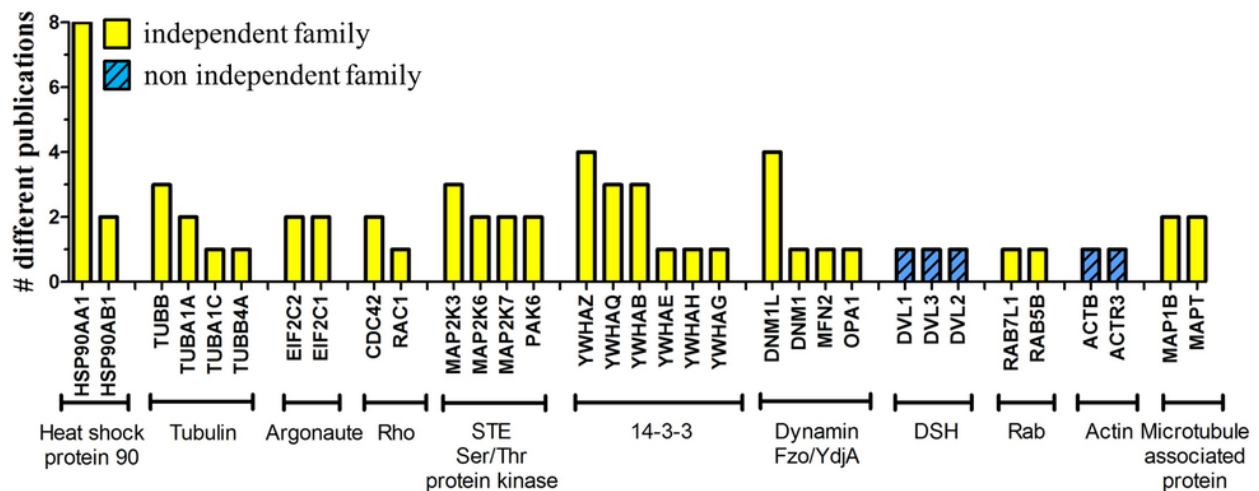
1

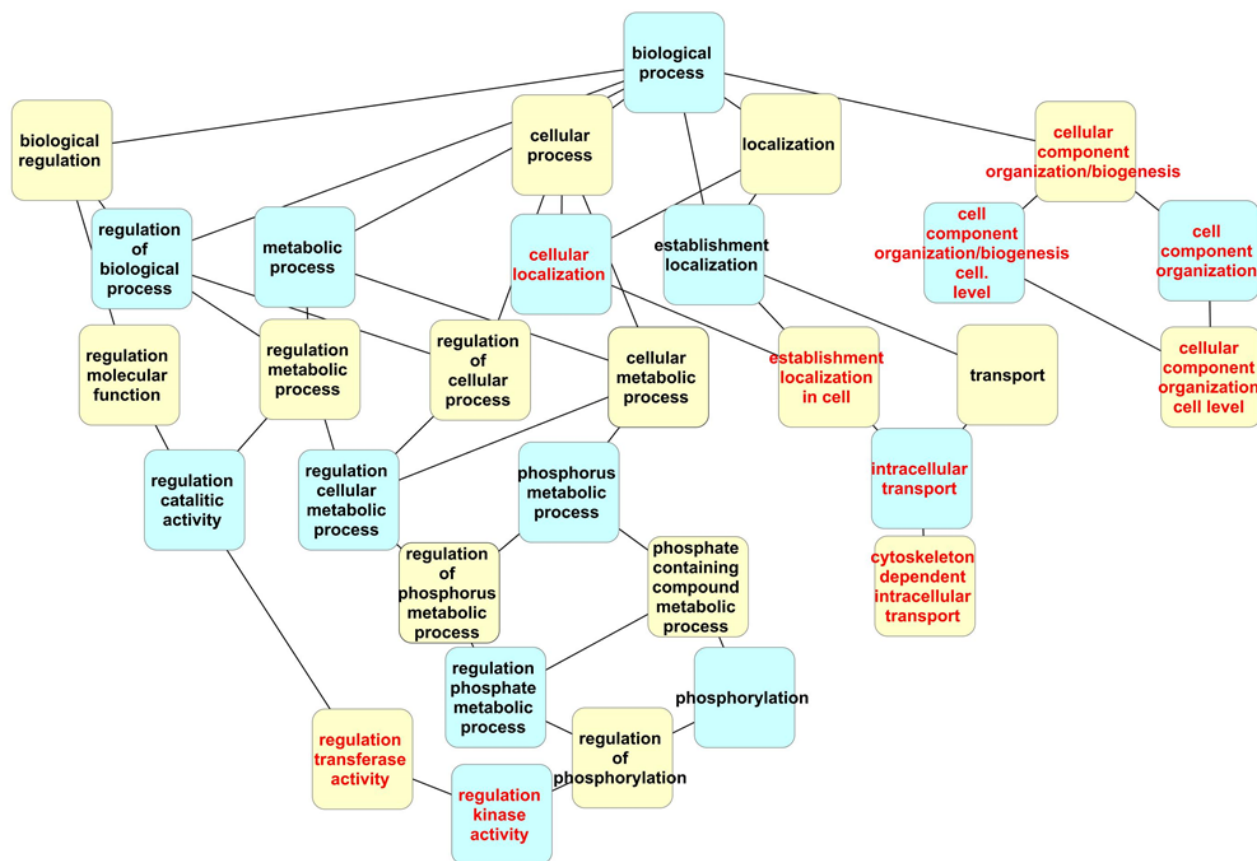
LRRK2 heterologous interactions as reported in the merged dataset.

LRRK2 in the middle of the graph is linked to candidate partners (black dots) through 422 annotations as described in the merged dataset. Some partners have one connection only; others have multiple connections based on the number of annotations. Different colours represent different methods used to infer the interaction; note that since IntAct and BioGRID use different classifications for the interaction detection method, a simplified and harmonized version has been applied to this figure to help the reader. In particular: i) Affinity Capture stands for - Affinity Capture-Western and Affinity Capture-MS (in BioGRID) - Anti Bait Coimmunoprecipitation, Anti Tag Coimmunoprecipitation, Coimmunoprecipitation, Pull Down and Affinity Chromatography Technology (in IntAct); ii) Biochemical Activity stands for - Biochemical Activity (in BioGRID) - Protein Kinase Assay (in IntAct); iii) Co-localization stands for - Co-localization (in BioGRID) - Confocal microscopy and Fluorescence Microscopy (in IntAct); iv) Reconstituted Complex stands for - Fluorescence Polarization Spectroscopy and Surface Plasmon Resonance (in IntAct).

2

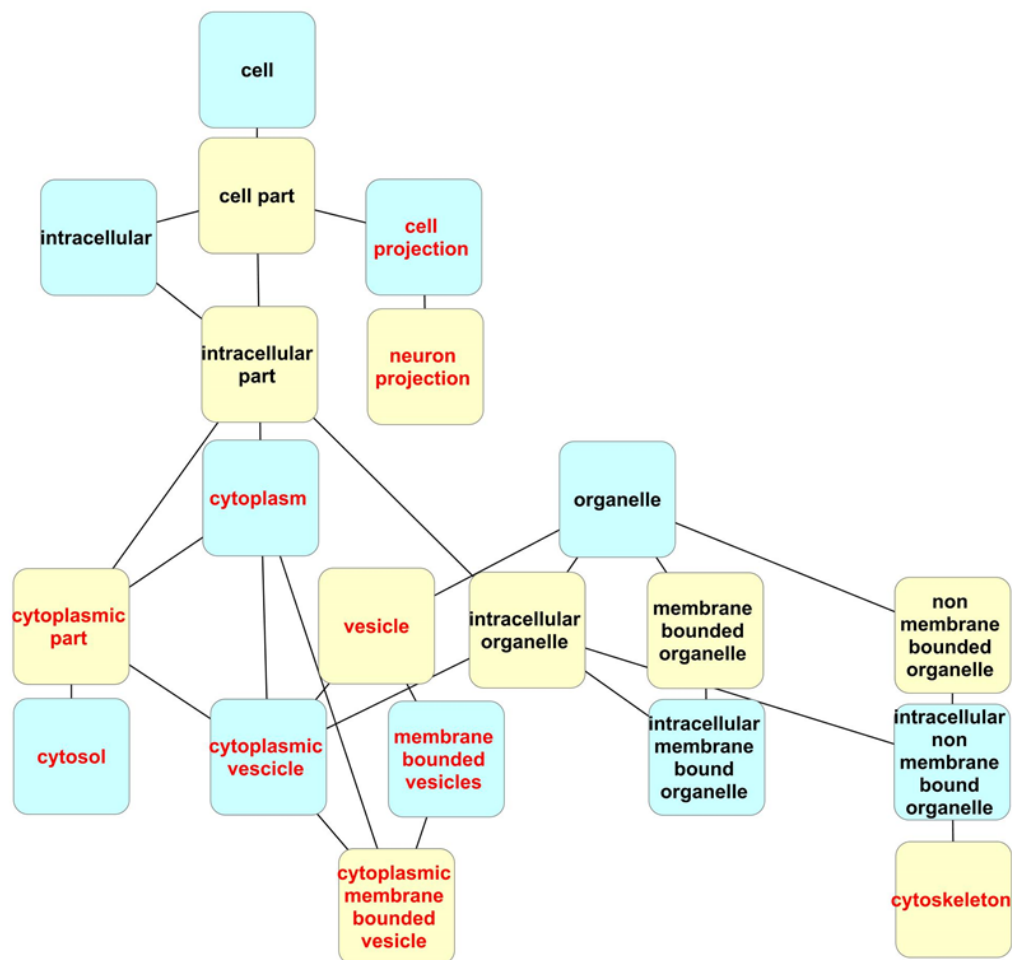
Number of annotations capturing each of the LRRK2 interactors.

Only the interactors reported in 2 or more annotations, and used in the enrichment analysis, are included in the figure, all other interactors were identified in just a single experiment. A) High occurrence LRRK2 interactors, >1 publication, >1 method; B) Medium-high occurrence LRRK2 interactors, >1 publication, 1 method; and C) Medium-low occurrence LRRK2 interactors, 1 publication, >1 method.

Families of LRRK2 interactors.

Each of the 62 proteins in the LRRK2 filtered interactome was associated to a family (if available) based on the UniProtKB "family&domains" classification. Proteins belonging to the same family were plotted accordingly to the total number of publications (y-axis); each family was then classified based on the authorships as non-independent (different family members described by the same research group) or independent (different family members described by at least two different research groups).

Dendrogram of "GO biological process" terms enriched for LRRK2 interactors in WebGestalt.

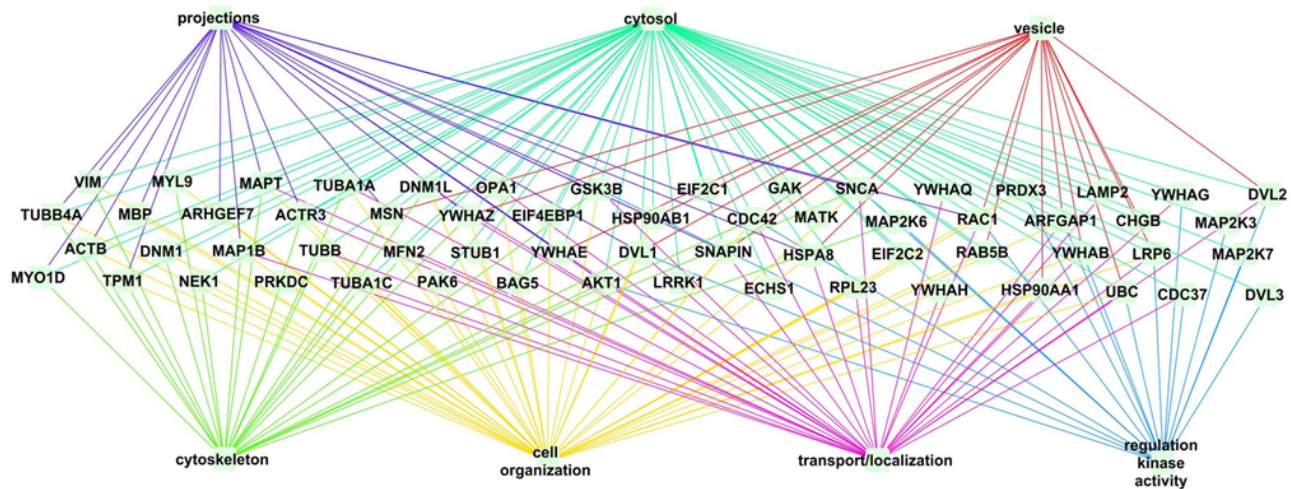
Hierarchical levels of the dendrogram are alternatively represented in blue and yellow; red text indicates the top 10 GO terms. The table lists details of the enriched terms: GO term and ID, number of proteins in the GO term category (Ref.), number of LRRK2 interactors associated with the GO term (Genes), p-value adjusted for multiple testing.

Biological Process GO term	GO ID.	Ref.	Genes	Adjusted p-value
Cellular localization	GO:0051641	1977	31	$3.80e^{-09}$
Intracellular transport	GO:0046907	1131	24	$6.72e^{-09}$
Cellular component organization at cellular level	GO:0071842	3317	38	$3.93e^{-08}$
Cellular component organization	GO:0016043	4105	42	$4.55e^{-08}$
Establishment localization in cell	GO:0051649	1756	28	$5.38e^{-08}$
Regulation of kinase activity	GO:0043549	632	18	$5.55e^{-08}$
Regulation of transferase activity	GO:0051338	653	18	$9.48e^{-08}$
Cellular component organization or biogenesis at cellular level	GO:0071841	3420	38	$1.03e^{-07}$
Cellular component organization or biogenesis	GO:0071840	4222	42	$1.21e^{-07}$
Cytoskeleton-dependent intracellular transport	GO:0030705	66	8	$3.01e^{-07}$

Dendrogram of "GO cell component" terms enriched for LRRK2 interactors in WebGestalt.

Hierarchical levels of the dendrogram are alternatively represented in blue and yellow; red text indicates the top 10 GO terms. The table lists details of the enriched terms: GO term and ID, number of proteins in the GO term category (Ref.), number of LRRK2 interactors associated with the GO term (Genes), p-value adjusted for multiple testing.

Cellular Component GO term	GO ID	Ref.	Genes	Adjusted p-value
cytosol	GO:0005829	2372	40	$8.60e^{-18}$
cytoplasmic part	GO:0044444	6772	56	$1.41e^{-14}$
cytoplasm	GO:0005737	9130	59	$5.29e^{-11}$
cell projection	GO:0042995	1230	23	$2.51e^{-09}$
neuron projection	GO:0043005	651	17	$1.36e^{-08}$
cytoplasmic vesicle	GO:0031410	928	19	$5.32e^{-08}$
vesicle	GO:0031982	970	19	$1.12e^{-07}$
cytoplasmic membrane-bounded vesicle	GO:0016023	862	18	$1.27e^{-07}$
cytoskeleton	GO:0005856	1790	25	$1.28e^{-07}$
membrane-bounded vesicle	GO:0031988	883	18	$1.87e^{-07}$

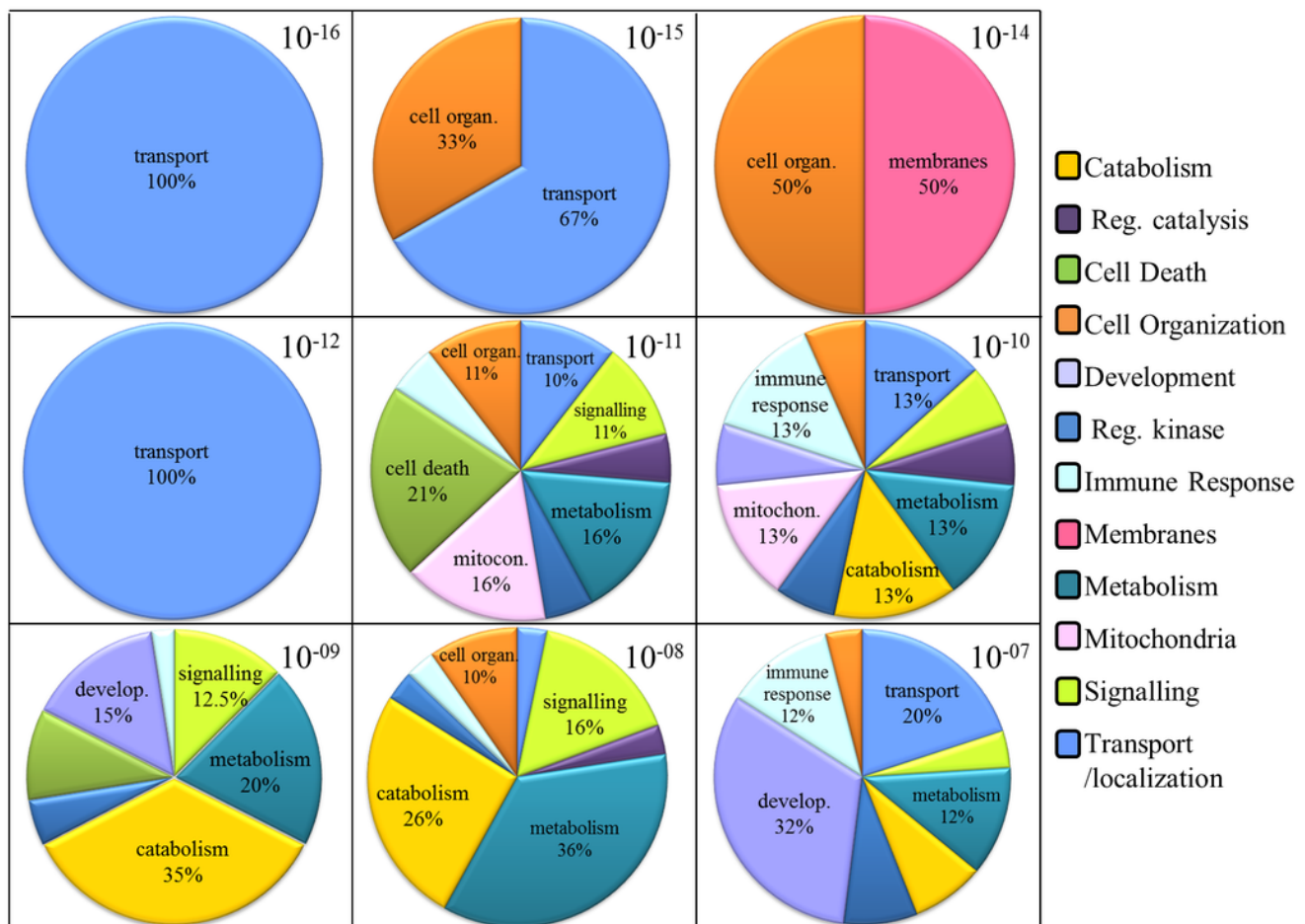
6

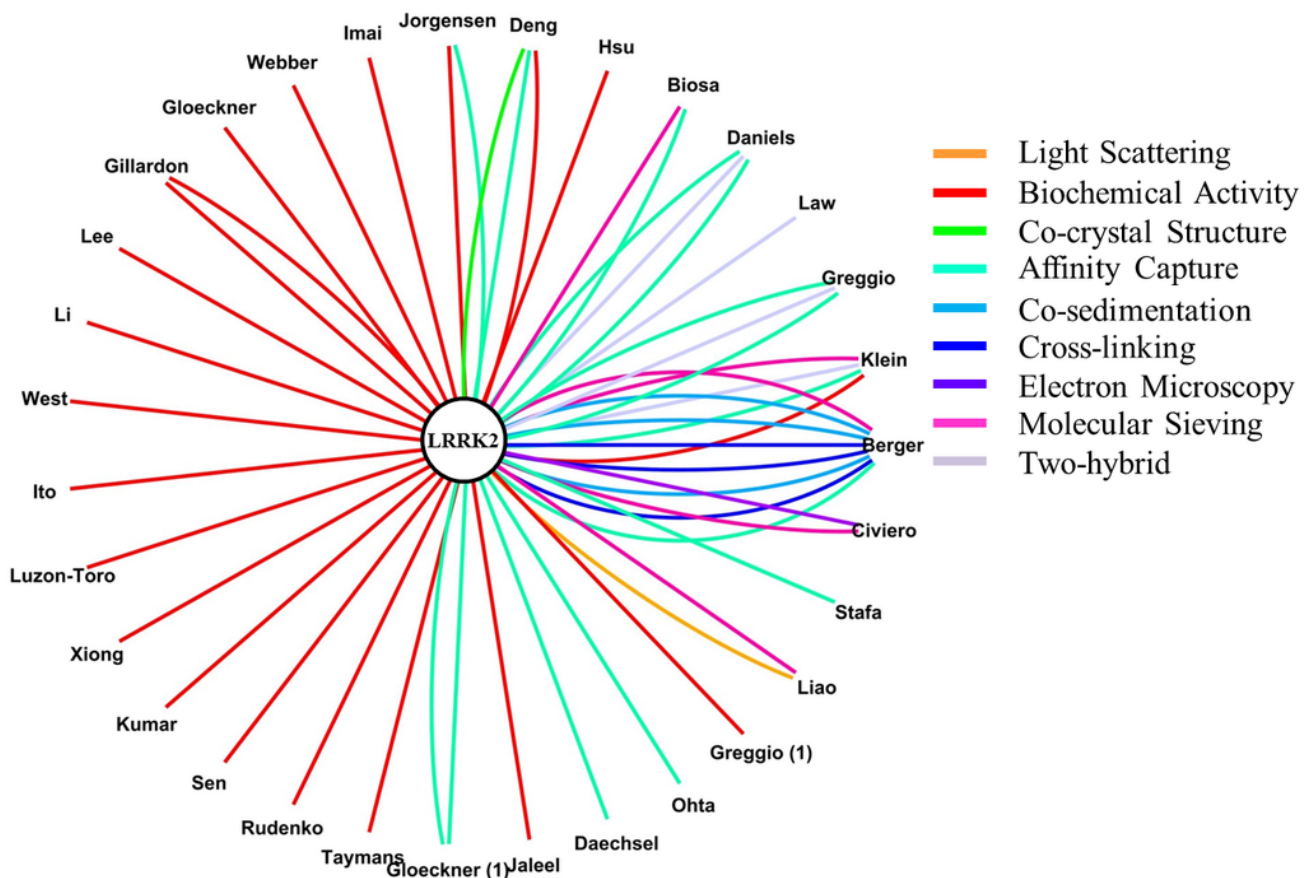
LRRK2 interactors enrichment for "GO cell component" and "GO biological process" terms.

The LRRK2 filtered interactome is shown after WebGestalt enrichment analysis. The enriched GO terms were grouped in: transport/localization, cell organization, regulation of kinase activity, cytosol, vesicles, cytoskeleton and cell projections. Every protein in the filtered interactome was connected to the GO term groups it has participated toward the enrichment. In the table the 7 groups are listed in columns (A-G) and rows (1-7). The percentage of proteins from the LRRK2 interactome that contributed toward the enrichment of the GO term group listed in the row and that were also reported in the GO term group reported in the column was calculated (intersection between the group in the row and the group in the column). In the last row of the table the percentage of the proteins that contributed toward the enrichment of each GO term group was calculated against the total number of LRRK2 interactors (i.e. 62 proteins). Cells discussed in the text are highlighted.

		1	2	3	4	5	6	7
		vesicles	projections	cytoskeleton	cytosol	transport	cell organization	reg. kinase
A	vesicles	100%	53%	32%	100%	79%	74%	47%
B	projections	43%	100%	61%	96%	61%	87%	26%
C	cytoskeleton	24%	56%	100%	100%	60%	84%	24%
D	cytosol	32%	37%	42%	100%	53%	68%	31%
E	transport	48%	45%	48%	100%	100%	81%	42%
F	cell organization	33%	48%	50%	95%	60%	100%	26%
G	reg. kinase	50%	33%	33%	100%	72%	61%	100%

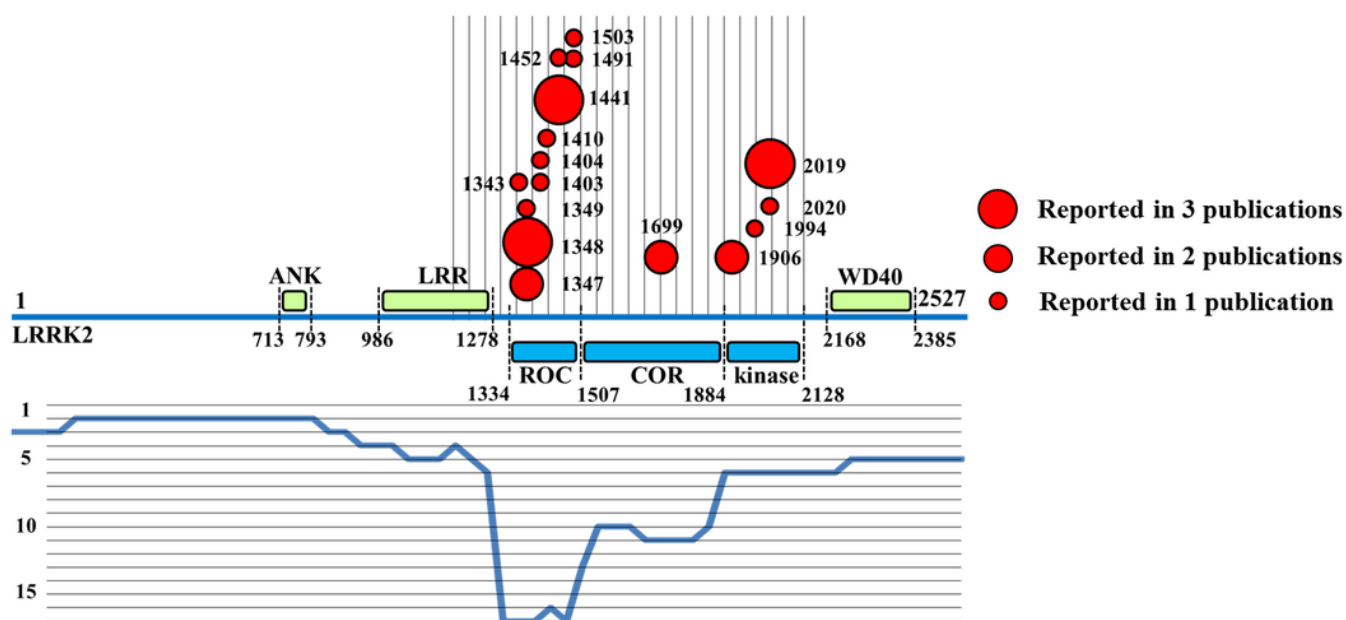
total in LRRK2 interactome	31%	37%	40%	95%	50%	68%	29%

"GO biological process" terms enriched for LRRK2 interactors in GO supported by Panther.

Pie charts showing the composition of every significance level of enrichment. The legend shows the 12 GO term groups, although in the charts the group name was reported only for those that reached the 10% contribution toward the enrichment.

Annotations for LRRK2 self-interaction.

LRRK2 in the middle of the graph is linked to the first author of publications describing self-interaction. Different colours represent different methods used to infer the interaction; note that since IntAct and BioGRID use different classifications for the interaction detection method, a simplified and harmonized version has been applied to this figure to help the reader. In particular: i) Affinity Capture stands for - Affinity Capture-Western (in BioGRID) - Anti Tag Coimmunoprecipitation, and Pull Down (in IntAct); ii) Biochemical Activity stands for - Biochemical Activity (in BioGRID) - Protein Kinase Assay (in IntAct).

9

Profile of LRRK2 self-interaction.

Each residue described as involved in LRRK2 self-interaction is reported with a dot, the dimension of the dot is proportional to the number of publications annotating that specific residue. A profile covering the entire length of the LRRK2 protein is shown in blue, in the bottom half of the image; the y value of the profile represents the number of publications in which the fragment in the profile was reported as associated with LRRK2 self-interaction.

LRRK2 interactors associated to the genetics of PD or IBD.

A, number of annotations; P, number of publications; HTP, one of the detection methods is high throughput. The approximate distance in base pair between the risk SNP and the protein coding sequence was calculated and the protein is indicated as candidate protein according to the original GWAs study. Methods are as follows: IP, immunoprecipitation; CM, confocal microscopy; FM, fluorescent microscopy; PA, protein array; BA, biochemical activity; MS, affinity capture mass spectrometry and 2H, two hybrid.

2   **Table 1**

Protein	A	P	Methods	HTP	Familial PD	GWA PD	GWA IBD	Associated SNP	Distance (Kbp)	Candidate
SNCA	6	2	IP, CM	-	X	X		rs356182	~20	Yes
RAB7L1	5	1	PA, IP, FM, CM	+		X		rs823118	~14	Yes
GAK	5	2	PA, IP, FM, CM	+		X		rs34311866	~25	Yes
MAPT	7	2	IP, BA	-		X		rs17649553	0	Yes
CDC37	5	4	MS, IP	+			X	rs11879191	0	No
DVL1	4	1	IP, CM, 2H	+			X	rs12103	~23	No
GNA12	1	1	MS	+			X	rs798502	0	Yes
CALM3	1	1	MS	+			X	rs1126510	~9.7	Yes
GLTPD1	1	1	PA	+			X	rs12103	~13	No
CEP72	1	1	PA	+			X	rs11739663	~18	No
ARPC2	1	1	MS	+			X	rs2382817	~32	Yes
RIPK2	1	1	BA	-			X	rs7015630	~73	Yes
PSMG1	1	1	PA	+			X	rs2836878	~82	No
DBN1	1	1	MS	+			X	rs12654812	~89	No
RPAP3	1	1	PA	+			X	rs11168249	~108	No
CD2BP2	1	1	PA	+			X	rs11150589	~116	No
PLK1	1	1	PA	+			X	rs7404095	~163	No
ACTR2	1	1	MS	+			X	rs6740462	~169	No
STIP1	1	1	PA	+			X	rs559928	~178	No
MYO1B	1	1	MS	+			X	rs1517352	~179	No
CLTC	1	1	MS	+			X	rs1292053	~189	No