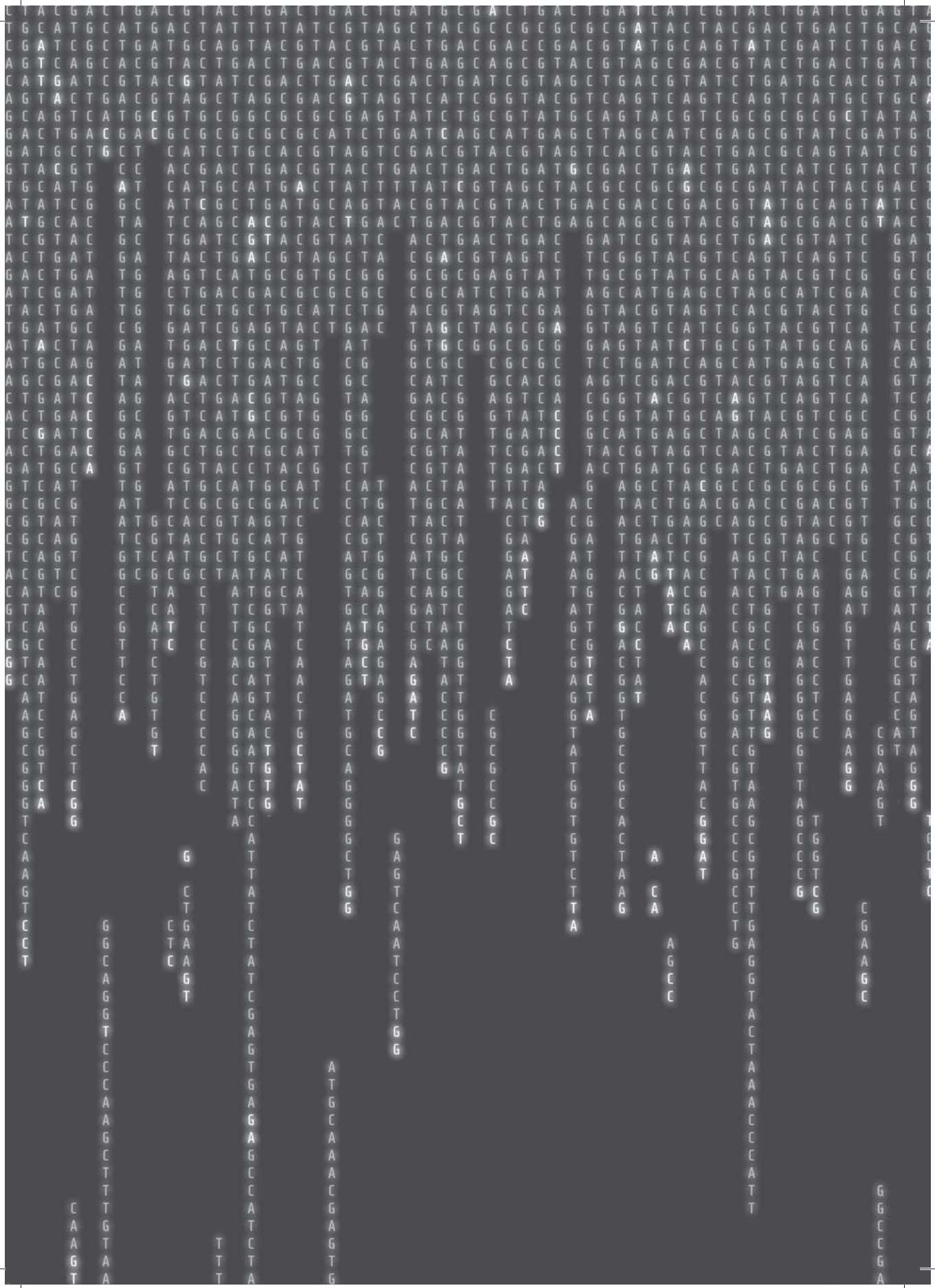
Cracking the code-ing sequence for Parkinson’s disease

Jansen, I.E.

2017

document version
Publisher's PDF, also known as Version of record



CHAPTER 5

BIOLOGICAL PATHWAYS AS CAUSAL GENETIC RISK FACTORS FOR PARKINSON'S DISEASE

This chapter is in preparation for publication as: Jansen IE, Nalls M, Gibbs R, Dijkstra AA, Ingrassia A, Sharma M, van Rooij J, Kraaij R, Uitterlinden AG, van de Berg W and Heutink P. Biological pathways as causal genetic risk factors for Parkinson's disease.

ABSTRACT

Various biological pathways have been linked to Parkinson's disease (PD). The identification of deregulated cellular processes is essential to increase our understanding of PD pathogenesis which enhances the development of improved treatment for PD. We therefore explored the genetic involvement of PD-related mechanisms, either focusing on pathways implied by previous genetic discoveries or on cellular processes that are deregulated on the gene-expression level at the onset of PD pathology. Well-established PD genes were excluded from the pathways prior to the genetic analyses to assure the identification of novel genetic association signals. Using SKAT, we tested for a joint effect of coding variants for a total of 16 biological pathways by analyzing 3 distinct whole exome sequencing (WES) datasets (IPDGC = 1,167 PD cases and 1,685 controls; Merck = 1,138 PD cases and 434 controls, PPMI = 364 PD cases and 159 controls). The *mitochondrial dysfunction, renin-angiotensing signaling* and *caveolar-mediated endocytosis signaling* pathways showed significant associations to PD in the IPDGC discovery dataset, which replicated in the two additional WES datasets. For the *mitochondrial dysfunction* pathway, subsequent gene-based analyses revealed that genes of the NADH:ubiquinone oxidoreductase subunit were mostly affected, thereby implying the specific deregulation of mitochondrial complex I in PD pathogenesis. Future research covering the genetic and functional levels of these 3 pathways will improve our understanding of PD etiology and ultimately facilitate the discovery of adequate targets for optimal therapies to cure PD.

INTRODUCTION

Genetic studies on Parkinson's disease (PD) have identified numerous genetic factors which increased the knowledge on underlying affected biological mechanisms.[1] Mitochondrial dysfunction has been implied to be a cause of PD by the identification of mutations in the Mendelian genes *Parkin, PINK1* and *DJ-1,* discovered through linkage analysis of PD families with autosomal recessive inheritance patterns.[2] Similarly, the establishment of *SNCA* and *LRRK2* by both family and population based studies,[3,4] suggested the synaptic transmission system (endo- and exocytosis) to be affected in PD pathogenesis.[5,6] *SNCA* and *LRRK2* have furthermore functions in the lysosomal-mediated autophagy mechanism,[7] a pathway genetically linked to PD of which the relation was strengthened by the additional discovery of *GBA* and *VPS35.*[7,8]

Establishing the biological mechanisms that contribute to PD pathogenesis is important as it serves as a guide for functional follow-up experiments and ultimately will reveal molecular processes to target for PD therapies. The current understanding of PD-related biological mechanisms is not sufficient for the development of drugs that halt the disease process. PD is a heterogeneous disease, which is reflected in the variety of clinical symptoms, pathology and genetic causes.[9] A treatment that would cure a biological system, instead of a single biological unit, could benefit a larger group of people with distinct genetic causes that converge on similar functional defects. It is therefore crucial to expand our knowledge on PD-related biological pathways, either by obtaining more detailed understanding of established mechanisms or by the discovery of novel cellular processes.

Besides genetics, other levels of biological data have implied the deregulation of specific molecular processes in PD pathogenesis. Multiple transcriptomics and proteomics studies have highlighted defects in pathways that overlap with the genetically based studies,[10-13] and additional pathways, such as retinoid metabolism, iron metabolism and glial activation.[13,14] One caveat concerning study designs including data on gene-expression or protein levels, is the complexity to differentiate between genuine causal deregulations and side-effects. For drug development one would exclusively want to address the biological units that are truly causing the pathology or symptoms at an early stage, rather than a mechanism that is deregulated at a later stage resulting in end-stage symptoms. One study of Dijkstra et al. addressed this issue by comparing genome-wide gene-expression profiles in the substantia nigra brain region of early-stage Parkinson's disease patients (Braak stage 1 and 2), not presenting symptoms, to healthy controls (Braak stage 0).[15] Taking this study one step further, analyzing variation within the genes that are involved in the observed deregulated pathways, could establish putative causal implications of these molecular processes in relation to PD etiology.

To further elucidate the genetic involvement of complete molecular mechanisms in PD susceptibility, we therefore explored multiple PD-related pathways by analyzing

3 independent whole exome sequencing datasets comprising a total of 2,669 PD cases and 2,278 controls. We selected biological mechanisms that were implied by previous genetic discoveries. By excluding the known PD genes, we aimed to identify novel genes that contribute to a more detailed understanding of sub processes of these pathways. A second approach encompassed the search for novel genetically involved cellular processes for PD, by studying differences in genetic variation between PD subjects and healthy controls for cellular processes that have been shown to be deregulated at an early stage of PD pathology (identified by Dijkstra et al.[15]). The identification of novel genes within known PD pathways or the observation of genetic variation within complete novel cellular processes will build a more comprehensive picture of PD etiology, thereby contributing to the knowledge required for the development of fruitful therapies for PD.

METHODS

Subjects

This study encompasses 3 independent cohorts of which the subjects have given written informed consent and participation in genetic studies was approved by relevant local ethical committees for medical research. The first cohort consists of 2,852 samples (post individual QC), including 1,167 PD cases that originate from the International Parkinson's Disease Genomics Consortium (IPDGC), and 1,685 neurologically healthy controls, of which 484 are part of the IPDGC cohort and 1,201 of the Rotterdam Study[16] version 1 (RSX1). We refer to this dataset as IPDGC-RSX1 cohort and it has served as the initial exploration dataset. The other two cohorts were used as replication datasets to confirm initial discoveries. The second cohort consists of 1,138 PD patients originating from Merck Research Laboratories and 434 neurologically healthy controls accessed through the public repository dbGaP. The last cohort, encompassing 364 PD cases and 159 controls, is part of the Progressive Parkinson's Marker Initiative, of which whole exome sequencing data is publically accessible through their database (www.ppmi-info.org/data). Demographic statistics on the distinct cohorts are described in Table 1.

Whole exome sequencing

WES library preparation was performed with different capture kits for the distinct cohorts (Table 1). For all cohorts, paired-end sequencing was accomplished with a HiSeq2000 and the same algorithms were used for read processing. For alignment of the 100-bp sequencing reads to the human reference genome (hg19), Burrows-Wheeler Aligner (BWA)[17] was applied. The Genome Analysis Toolkit (GATK)[18] v3.3-0 was used for calling of single nucleotide variants (SNVs) and small insertions/deletions (indels). The IPDGC and RSX1 samples were integrated by joint variant calling of individual gVCF files. Coverage statistics are described in Table 1.

Table 1. Cohort information

	IPDGC-RSX1	Merck	PPMI
n samples	2.852	1.572	523
% male	55,3%	53,9%	64,4%
cases: *n*	1.167	1.138	364
cases: age of onset	41.2 (10.9)	64.4 (9.0)	60.1 (9.7)
cases: family history	40,4%	NA	25,3%
controls: *n*	1685	434	159
controls: age	63.8 (17.1)	62.1 (16.2)	61.6 (10.5)
Coverage	94,4%	205,4	98,0%
Capture kit	Nimblegen	BGI reduced	Nextera
	Illumina		
Common/rare thr.	0,013	0,031	0,018

n = number of; Nimblegen = Roche Nimblegen exome capture kit; Illumina = Truseq Illumina capture kit; BGI = BGI reduced exome capture kit ; Nextera = Nextera rapid capture expanded exome kit. Coverage statistic for the IPDGC-RSX1 and PPMI datsetes are defined as the % of the exome covered for at least 10x. For the Merck dataset, coverage defines the mean depth.

Quality control procedures

Genotypes with a Phred-based genotype quality score above 20 and a depth above 8 were included for subsequent analyses. Individuals were excluded when showing gender ambiguity, having dubious heterozygosity/genotype calls, evidence present of relatedness or being a population outlier. The latter two QC calculations were based on LD-pruned common variants. We included bi-allelic variants for our genetic analyses and variants with low call rates or deviating from Hardy-Weinberg equilibrium were removed. For the IPDGC-RSX1 cohort, variants were only included if located within regions targeted by both capture kits, resulting in a total of 462,946 variants. The Merck and PPMI datasets consisted of 98,254 and 131,393 post-QC variants, respectively.

Pathways

Gene-sets were based on biological pathways that have been suggested to contribute to PD pathogenesis. Ingenuity Pathway Analysis (IPA) version 01-07 (Ingenuity Systems, www.ingenuity.com) was exploited as a source for canonical pathways that are related to the genetic-literature-based biological PD mechanisms.[1] Four canonical pathways were identified, of which the genes coding for the involved proteins were extracted to generate the gene-sets (Table 2). PD genes that have previously been associated to PD[19] are excluded as we are searching for novel genetic factors.

In addition, to search for the genetic involvement of novel biological pathways we utilized a micro-array gene-expression study[15] as a source for novel biological pathways. As we are specifically interested in cellular mechanisms that are detected to malfunction at the beginning of the PD pathology process, we selected the canonical IPA pathways (Table 2) that are deregulated ($p \leq 0.0001$) in PD patients with Braak stages 1 or 2. Thirteen canonical pathways were selected of which one overlaps with a canonical pathway based on genetic literature.

Variant selection

The aim of this study was to explore the polygenetic burden of common and rare variants in pre-specified gene-sets. As benign variants have the potential to dilute association signals of genuine null alleles, we selected variants based on functionality to enrich for deleterious variants. Predicting the underlying genetic model prior to performing statistical tests is complex, especially considering the distinct types of variants previously associated to PD. We therefore generated distinct groups of functional subsets based on functional annotation and CADD[20] deleteriousness predictions, being (i) amino-acid changing variants, including both nonsynonymous and disruptive variants, (ii) amino-acid changing variants predicted to be part of the 10% most harmful variants of the total genome (scaled C-score > 10), (iii) amino-acid changing variants predicted to belong to the 1% most harmful variants of the genome (scaled C-score > above 20), and (iii) disruptive variants. For the rare variants subset, we additionally tested all coding variants while weighting the variants on the unscaled CADD scores.

Polygenetic burden analysis

Using the R-package SKAT[21,22] we tested for the polygenetic burden of coding variants within the pathways of interest to PD, investigating the genetic involvement of all genes simultaneously and each gene individually. We analyzed the joint effect of exclusively rare variants, exclusively common variants and variants without a minor allele frequency inclusion threshold: the common & rare dataset. The threshold to separate common and rare variants (Table 1) was based on the sample size, as suggested by the SKAT-C which is a sub-package specifically designed to combine the effect of common and rare variants.[23]

For the rare variants, the gene-set and individual gene association tests were performed with SKAT-O[22], which automatically selects the most significant result comparing the one-sided burden model (assuming all variants to have the same direction and magnitude of effect) with the two-sided SKAT model (allowing variants with distinct directions and magnitudes of effects). SKAT-C was used to test for the polygenetic burden of common and common & rare variants. For the latter, SKAT-C assures the equal contribution of the common and rare variants to the test statistic. To correct for confounding factors such as population stratification and technical differences, we included the following

covariates in all analyses: 20 multi-dimensional scaling components, WES coverage (pre-QC missingness) and gender.

For each pathway of interest we compared the nominal *p*-value of the original gene-set to the association signal of 1,000 randomly drawn gene-sets matched on the number of genes. Per gene association significance levels were calculated by implementing the resampling method of SKAT which is accomplished by 10,000 permutations shuffling the affection status. Empirical *p*-values are calculated by (n1+1)/(n+1), where n1 = the number of resampling *p*-values smaller than the original sample *p*-value and n = the number of resampling. If applicable we combined the *p*-values of the replication datasets with the Fisher's method to obtain a meta *p*-value.

Table 2. Gene-sets selected for polygenetic burden analyses in PD.

Pathways	***n* genes**	**PD genes**	***n* tested genes**
Genetic literature			
Autophagy	21	x	21
Caveolar-mediated endocytosis signaling	17	x	17
Clathrin-mediated endocytosis signaling	47	*GAK, SYNJ1*	45
Mitochondrial dysfunction	128	*SNCA, parkin, PINK1, DJ-1, LRRK2*	116
PD differential gene-expression study			
Actin cytoskeleton signaling	34		34
Angiopoeitin signaling	22		22
Axonal guidance signaling	77		77
Clathrin-mediated endocytosis signaling	47	*GAK, SYNJ1*	45
Dopamine-DARPP32 Feedback in cAMP signaling	11		11
FAK signaling	22		22
FLT3 signaling in hematopoieitic progenitor cells	16		16
Gap junction signaling	22		22
Huntinton's Disease signaling	64	*SNCA*	63
mTOR signaling	26		26
Rac signaling	30		30
Renin-angiotensin signaling	26		26
Signaling by Rho Family GTPases	29		29

For Mitochondrial dysfunctioning, 7 additional genes were removed as they consist of mitochondrial DNA which is not covered by standard WES capture kits.

RESULTS

Rare variants

The IPDGC WES dataset is currently the largest WES dataset of unrelated Parkinson's disease cases and was therefore implemented as the discovery dataset. We explored for polygenic influences of rare variants by aggregating rare variants over a group of genes belonging to a particular biological pathway, aiming to obtain optimal power for our study design by increasing the number of variants that are simultaneously addressed. Considering the 4 genetics-based and 13 transcriptomics-based pathways (Table 2), the genetics-based *mitochondrial dysfunction* pathway showed a consistent significant genetic influence of rare variants (empirical $p < 0.05$) on PD for the distinct classifications of functional variants (Table 3; Supplementary Table 1). This finding suggests that, besides the previously published PD genes that were excluded from the gene-set, additional nuclear genes that are involved in mitochondrial processes harbor rare genetic variants that affect PD pathogenesis. None of the transcriptomics-based pathways showed an aggregated effect of rare variants on PD.

To further explore the genetic association of the *mitochondrial dysfunction* pathway, we accessed the two additional independent WES datasets of Merck and PPMI. The rare variant association signal was replicated when combining both datasets for the amino-acid changing variants (empirical $p = 0.025$) and all coding variants weighted for their functional CADD score (empirical $p = 0.0038$), which is displayed in Table 3. Although the other functional classes reached nominal significance, they lacked significance after the competitive test. There is a substantial difference in sample size comparing the two replication datasets with the discovery dataset, implying less power to detect an association in the replication datasets. It is therefore anticipated that the absence of significance for the CADD10, CADD20 and disruptive variant subsets is a result of limited power for the Merck and PPMI dataset.

Common and rare variants

A similar set of gene-set analysis was performed for the 16 pathways exclusively considering common variants or the common & rare variants, excluding the 'weighted' and 'disruptive' variant subsets. The function SKAT-C, specifically designed for the analysis of common variants, lacks the option to weight variants on the functional unscaled CADD scores. Furthermore, the disruptive variant subset was omitted as such variants are in general rare. For the common variants, the *caveolar-mediated endocytosis signaling* and *renin-angiotensin signaling* pathways showed initial significant common variant associations ($p = 0.034$ and $p = 0.047$, respectively) to PD (Supplementary Table 2). Although these pathways were slightly above the significance threshold after the competitive test (for both empirical $p = 0.054$), we next tested these pathways in the replication datasets to

Table 3. Rare variant association of mitochondrial dysfunction pathway.

	IPDGC		Merck		PPMI		Merck & PPMI
variant type	*p*(emp)	n var	*p*(emp)	n var	*p*(emp)	n var	meta *p*
weighted	**0.006 (0.046)**	1298	**0.014 (0.024)**	77	**0.015 (0.018)**	134	**0.0038**
AA changing	**0.00028 (0.029)**	879	**0.030 (0.059)**	38	**0.020 (0.066)**	67	**0.025**
CADD10	**0.00011 (0.022)**	781	0.032 (0.078)	27	0.454	48	NA
CADD20	**0.00032 (0.017)**	639	0.019 (0.056)	19	0.239	35	NA
disrupt	**0.018 (0.049)**	91	NA	NA	NA	NA	NA

weighted = all exonic variants weighted on CADD unscaled C-scores; AA changing = amino acid changing subset; CADD10 = subset of variants predicted to belong to 10% most deleterious variants of the genome; CADD20 = subset of variants predicted to belong to 1% most deleterious variants of genome; disruptive = subset of loss-of-function variants; *p* (emp) = nominal *p*-value with empirical *p*-value in parenthesis of which the latter was calculated by comparison to 1,000 randomly drawn gene-sets. The underlined values are representing the replication. NA = not applicable.

further explore the aggregated common variant effects within these biological processes. The *caveolar-mediated endocytosis signaling* pathway did not show a burden of solely common variants ($p > 0.192$) in the Merck or PPMI WES datasets. However, we observed a common variant association (empirical $p = 0.049$) for the *renin-angiotensin signaling* pathway in the Merck dataset (Supplementary Table 3). The PPMI dataset lacked common CADD20 variants in the genes belonging to this biological process, presumably due to the small sample size, and could therefore not be tested. Meta-analysis of the IPDGC and MERCK dataset results in a significant association (combined p =0.018) of common CADD20 variants to PD, implying that the association could be genuine but the sample size should increase allowing to detect a higher number of common CADD20 variants.

For the joint effect of common & rare variants within the predefined biological pathways to PD, we again detected significant association for *mitochondrial dysfunction* and the *caveolar-mediated endocytosis signaling* (empirical $p = 0.025$), and additionally for the *dopamine-DARPP32 feedback in cAMP signaling* pathway (empirical $p = 0.045$; Supplementary Table 4). Comparing the results of the exclusively rare, exclusively common and joint subset of variants for the *mitochondrial dysfunction* gene-set (Supplementary Table 1, 2 and 4) suggests that the observed influence within common & rare variants for the *mitochondrial dysfunction* pathway is driven by the rare variants. The strongest rare variant association is presumably diluted by the common variant association, resulting in a less strong, though still significant, signal for the joint set of common & rare variants. The significant association for the *caveolar-mediated endocytosis signaling* pathway caused by the aggregated effect of both common & rare variants is replicated in the PPMI dataset, but not in the Merck dataset (Supplementary Table 3). The *dopamine-DARP32 in cAMP signaling* pathway shows no association in the replication datasets, implying a false positive result.

Functional interpretation

The *mitochondrial dysfunction* pathway shows a burden of rare genetic variants in all 3 independent WES datasets. As this gene-set excludes the known PD genes, this finding suggests that additional genetic variants in other genes with a mitochondrial function are contributing to PD susceptibility. To test which genes are contributing most and whether these genes are involved in a specific mitochondrial sub process, we analyzed the effect of aggregated rare variants per individual gene within the *mitochondrial dysfunction* pathway. Of note, the genes targeted by the WES library kits are nuclear, therefore no genes transcribed from mitochondrial DNA were investigated. For the IPDGC dataset, 7 genes were detected with an individual rare variant association of which 6 reached the empirical significance level (Supplementary Table 5). Five of these genes are localized to the mitochondrial complexes, which comprise the electron transport chain. This chain is important for energy production in each cell, as it converts ADP into ATP.[24] Similarly,

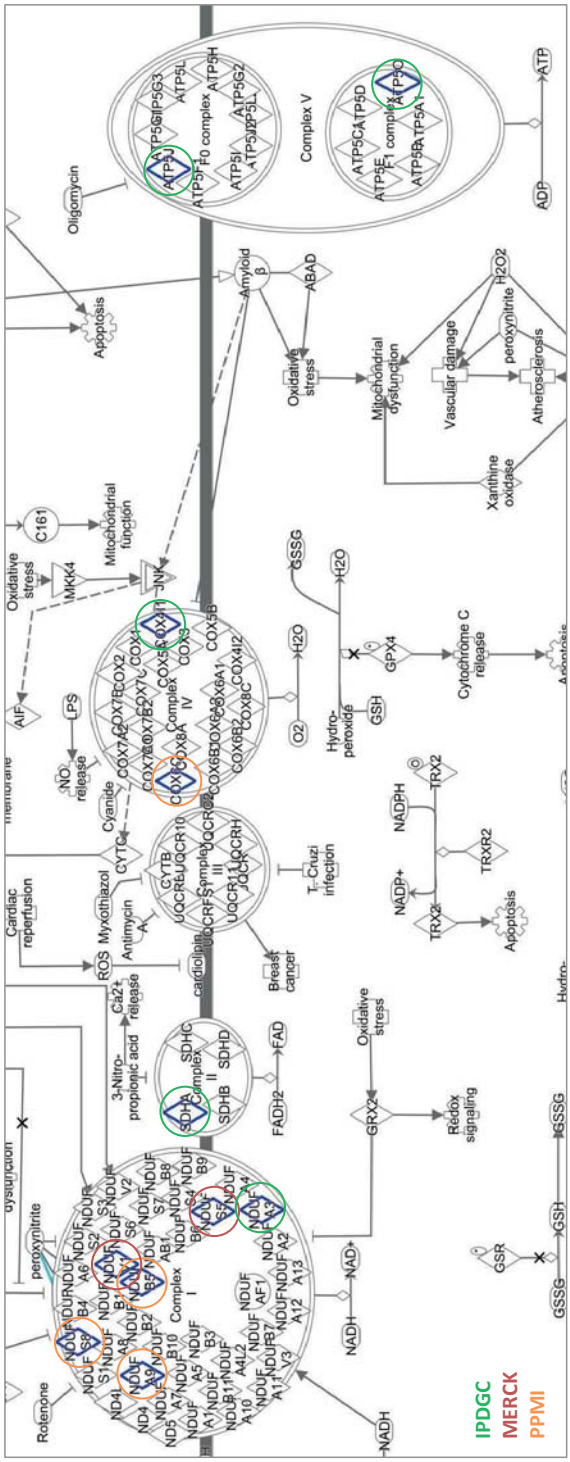


Figure 1. The electron transport chain as part of the mitochondrial system as displayed by the IPA database.

the genes with an individual gene association for the PPMI and Merck datasets are also localized to this respiratory chain. Although no exact replication of individual genes is observed, every WES dataset observed a rare variant association for at least one NADH:ubiquinone oxidoreductase subunit (*NDUF*x), which are involved in mitochondrial complex I. As these genes have unique sequences and are no paralogs of each other, we are confident to exclude the possibility that these significant associations are based on sequencing errors due to inadequate read mapping. Figure 1 displays an enhanced graphical overview of the electron transport chain (Total *mitochondrial dysfunction* pathway displayed in Supplementary Figure 1). Although one gene is contributing to the function of Succinate-quinone oxidoreductase (mitochondrial complex II), most associated genes are involved in mitochondrial complexes I, IV and V.

The *renin-angiotensin signaling* pathway was genetically associated to PD in the IPDGC and MERCK dataset. Both associations were based on only 2 variants, which are distinct between the two cohorts (Supplementary Table 6). All 4 variants have a higher MAF in cases compared to controls, implying a damaging effect of aggregated common variants within this pathway. Visualizing the location of the affected genes in the pathway diagram (Supplementary Figure 2), reveals the functional involvement of processes at the start of the biological pathway, including functions initiating the pathway (*ACE* and *AGT*) and mechanisms influencing the transcription processes in the nucleus (*NOX1* and *MEKK1*). The *caveolar-mediated endocytosis signaling* pathway, which showed a significant joint effect of common and rare variants on PD, constitutes only one individual gene association of *FYN* in the IPDGC WES dataset (p= 0.0006 and 0.009 for the AA-chancing and CADD10 variant subsets, respectively). It therefore seems that the variants distributed over all the genes involved in *caveolar-mediated endocytotsis signaling* pathway (Supplementary Figure 3) are collectively contributing to PD susceptibility.

DISCUSSION

To determine the genetic effects of biological mechanisms on PD pathogenesis, we explored 3 independent whole exome sequencing datasets by establishing differences in overall aggregated allele frequencies between PD patients and healthy controls. Sixteen pathways were selected for investigation, either based on functional involvement of known PD genes[1] or on a transcriptomics study reporting deregulated biological pathways based on gene expression levels observed at the onset (Braak stage 1 or 2) of α-synuclein PD pathology.[15] Focusing on the genetics-based pathways *autophagy, caveolar-* and *clathrin-mediated endocytosis,* and *mitochondrial dysfunction*, we aimed to improve our current understanding of the genetic link with PD by discovering associated sub-processes within the established pathways. The second objective was the discovery of novel genetically involved biological mechanisms which was addressed by the analysis of

the transcriptomics-based cellular processes. By grouping the genes that are encoding for the proteins that are involved in the molecular processes of interest we generated gene-sets, which were tested for their combined association to PD.

For the genetics-based pathways, the *mitochondrial dysfunction* and *caveolar-mediated endocytosis signaling* pathways showed a joint effect of coding variants to PD. Rare variants that were weighted according to their functional effect (predicted by CADD) and located within the genes of the *mitochondrial dysfunction* pathway, showed a difference in grouped allele frequency comparing PD cases versus controls in the IPDGC discovery dataset, and the 2 independent replication datasets, therefore generating a strong convincing evidence. As known PD genes (*parkin, PINK1, DJ-1, LRRK2* and *SNCA*) were excluded from the tested gene-set, it suggests that additional genes implicated in mitochondrial health are deregulated in PD etiology. To our knowledge, this study is the first comprehensive effort to investigate the contribution of variants in nuclear genes with mitochondrial functions to PD etiology. Although the mitochondrial maintenance system is comprised of 1,000 to 1,500 genes,[25] so far only 9 nuclear genes have been studied for their genetic involvement in PD.[26] Small-scale association studies of these genes suggest involvement in mitochondrial biogenesis functions, however no GWAS loci have yet been linked to mitochondrial deregulation, potentially due to the limitation of GWAS studies to exclusively study common variants. In contrast, our sequencing study systematically investigates the association of rare variants enabling the identification of associations that were previously inaccessible. The findings of our population-based study highlights aggregated genetic effects of nuclear mitochondrial genes, other than the known PD genes, thereby strengthening the known relation between mitochondrial health an PD. Furthermore, it emphasizes the need to shift our scientific focus beyond the previously identified mitochondrial PD genes, such as *DJ-1, parkin* and *PINK1*.

The latter genes are involved in mitophagy, the mitochondrial clearance system, and this cellular process has therefore been the focus of the majority of PD-related mitochondrial functional follow-up experiments.[27] However, research has also explored the relation of deregulation of the electron transport chain, more specifically complex I which is also called NADH:ubiquinone oxidoreductase.[28] Animal models frequently use complex I inhibitors as a standard method to induce PD pathology.[29,30] Furthermore, a small study has shown reduced complex I activity in patients with idiopathic PD.[31] Although environmental causes have been extensively described for complex I dysfunctioning in PD patients, such as the usage of specific drugs or exposure to pesticides,[32,33] the implication of genetic causes is limited covered.

Complex I of the respiratory chain consists of 45 subunits, of which 38 and 7 are encoded by nuclear and mitochondrial genes, respectively.[34] Currently, only one gene (*NDUFV2*) has been genetically associated to PD pathogenesis.[35] Our association analysis of the distinct genes within the *mitochondrial dysfunction* pathway show an enrichment

of rare variants for the sub-process of complex I, implying variation in multiple complex I genes to be causal for PD. Respiratory chain complex I might therefore be an adequate target to treat PD. As NADH-quinone oxireductase has already been implied to be dysfunctional through environmental causes, the application of drugs targeting complex I function has been discussed.[36] The drug coenzyme Q10, influencing mitochondrial function including complex I activity, has even been tested in clinical trials, but symptomatic improvements in PD patients have not been convincingly shown.[37] Besides the putative situation that complex I activity is not genuinely affected in PD, an alternative interpretation could be that we still have poor understanding of exact involvement of NADH-quinone oxireductase and more detailed studies are required for the development of adequate complex I related therapies. Furthermore, as PD is a complex multifactorial disease with many heterogeneous features, an alternative option could be that specific subtype(s) are sensitive for complex I treatment. Testing genetic predisposition for NADH-quinone oxireductase malfunctioning might therefore be crucial in such clinical trials.

The second pathway with a significant difference in the aggregated effect of coding variants comparing PD subjects with controls, was the *caveolar-mediated endocytosis signaling*. Testing for the joint effect of both rare and common variants resulted in significant associations in the IPDGC discovery dataset and the PPMI dataset. The degree of replication is not as strong as for the *mitochondrial dysfunction* pathways because an association was lacking for the Merck dataset. The exomes of the Merck dataset has been targeted by a library including a relatively low number of exonic regions, which is potentially causing limited coverage for the genes in the *caveolar-mediated endocytosis signaling* pathway, putatively resulting in decreased power for this specific gene-set. Although the replication in the PPMI dataset suggests a substantial probability of a true association to PD, an alternative interpretation inspired by the invalidation by the Merck dataset encompasses the absence of a genuine genetic association for this endocytosis mechanism. The latter interpretation is reinforced by the study of Dijkstra et al.[15], reporting a deregulation of the clathrin-mediated endocytosis signaling pathway, rather than the mediation through caveolin. We did not observe an association for the clathrin mediated pathway for PD on the genetic variant level. However, additional literature on caveolin-1 (encoded by *CAV1*) shows a relation to PD, thereby in favor of a genuine association of the *caveolar-mediated endocytosis signaling* pathway to PD pathogenesis. Caveolin-1 is the main component of the caveolae plasma membrane and has been shown to be a substrate of *parkin,* the PD gene mainly functional in mitochondrial health, and suggested as a potential therapeutic target.[38,39] Caveolin-1 is encoded by one of the 17 genes that encompassed the tested gene-set. Only one gene within the *caveolar-mediated endocytosis signaling* pathway reached significant individual association to PD, which was *FYN,* a gene that encodes for a protein that interacts with caveolin-1.[40] Besides its association to PD on protein level, the gene *CAV1* harbors homozygous haplotypes

that were reported to be specific for PD patients.[41] However, additional large-scale resequencing studies and functional experiments focusing on the genes involved in the *caveolar-mediated endocytosis* pathway are required to convincingly relate this cellular process to PD etiology.

For the transcriptomics-based pathway, only the *renin-angiotensin* pathway encompassed variants that influence PD susceptibility. This association signal was exclusively driven by a small number of common variants in the IPDGC and Merck dataset, but lacked a replication in the PPMI cohort (364 PD subjects and 159 controls). The lack of signal in the PPMI dataset is anticipated to be a result of insufficient power as this small dataset harbored not a single common variant that was predicted to be pathogenic for this specific pathway. Although traditionally the renin-angiotensin system (RAS) was known for its regulation of blood pressure and water/salt homeostasis, more recently the existence of local RAS systems are acknowledged, including the central RAS for the brain.[42] More specific for PD, an intracellular RAS is observed for dopaminergic neurons, which becomes hyperactive in response to age-related decrease in dopamine activity.[43-47] This over-activation enhances neuronal death, implying the local dopaminergic RAS to be a putative target for PD treatment.

The association signals within the IPDGC and Merck dataset are both composed of an aggregated effect of 2 common variants that are predicted to belong to the 1% most damaging variants of the genome. The 4 variants are present in the 4 distinct genes *ACE, AGT, NOX1* and *MEKK1*, which all have previously been described in relation to PD. High levels of the angiotensin-converting enzyme, encoded by *ACE,* are present within the nigria-striatal pathway and basal ganglia.[48,49] These protein levels are reported to be decreased in the cerebrospinal fluid of PD patients.[50] Furthermore, administration of an ACE-inhibitor in multiple rodent models seem to protect against a loss of dopaminergic neurons.[51-54] Similarly in humans, a small clinical trial showed an improvement in motor symptoms for PD patients that received the ACE-inhibitor perindopril.[55] These studies targeting and investigating ACE on protein levels are implicating a role of ACE-related mechanisms in PD pathogenesis. However, the influence of genetic variants on PD susceptibility remains inconclusive, even though multiple studies have been performed. Genetic research has focused on a 287-bp deletion that was suggested to be enriched in PD.[56,57] Yet recent meta-analyses contradict these results as they are unable to detect a significant association.[58,59] Our findings might therefore be valuable to continue the genetic discussion on *ACE* for its role in PD. Although we were unable to investigate the 287-bp deletion, we did observe a nonsynonymous variant in exon 5 of *ACE* with a double minor allele frequency in the PD patients (MAF cases = 0.026, MAF controls = 0.013) compared to controls. A more comprehensive sequencing study with a larger sample size is required to establish the true genetic involvement of *ACE* in PD etiology.

NOX1 and *MEKK1* are related to PD through oxidative stress signaling. The central nervous system requires controlled production of reactive oxygen species (ROS), a byproduct of mitochondrial metabolism.[60] Overproduction of ROS causes oxidative stress that potentially results in neurodegeneration.[61,62] Specific for PD, mitochondrial dysfunction is suggested to generate oxidative stress, which was emphasized by the discovery of *DJ-1, parkin* and *PINK1*.[1] In relation to the findings in this study, *MEKK1* is a direct target of *DJ-1* and contributes to oxidative-stress induced cell death.[63] Besides dysfunctional mitochondria,[64] an alternative mechanism has been proposed to be a cause for oxidative stress in neurodegeneration, which involves the membrane transporter NADPH oxidase enzyme subunit, encoded by 6 genes including *NOX1*.[65] *NOX1* is expressed in dopaminergic neurons and stress-induced activation of NADPH oxidase 1 results in oxidative stress and neurodegeneration,[66] potentially through modulation of α-synuclein expression.[67] In summary, literature on the renin-angiotensin mechanism and our identified genes implicate a contribution to PD pathogenesis, likely through an interaction with oxidative stress. Of note, this study suggests deregulation of the RAS as a primary pathogenic source rather than a side effect of PD neuropathology. Additional functional experiments on the interaction of renin-angiotensin and oxidative stress processes should elucidate the general pathogenic mechanism contributing to PD susceptibility.

Besides the *renin-angiotensin signaling* pathway, no other transcriptomics-based pathway harbored variants with an aggregated effect on PD. Although no significant associations were detected, it is impossible to draw definitive conclusions on the genetic contribution of these pathways to PD susceptibility. Three scenarios might be causing the absence of an association. One option would be that the deregulation observed on gene-expression level is genuine, but this is not caused by genetic defects. Alternatively, although Dijkstra et al.[1] have performed careful experiments and analyses, their implied pathways might be false positives. The dataset consisted of post-mortem substantia nigra brain tissue of only 5 individuals with early-stage α-synuclein pathology (Braak stage 1 and 2) and 8 controls (Braak stage 0).[15] This small number of samples may have caused spurious outcomes and, perhaps of even more impact, have missed truly associated pathways due to insufficient power. A last option would be that the defined pathways are genuinely associated to PD on both transcriptomics and genomics level, but the tested WES datasets are of insufficient size to reach adequate power to detect these genetic associations. A positive side-effect of gene-set association analysis rather than gene-based analysis, is the increase in statistical power to detect an association signal, specifically in the case of smaller datasets.[15] One important condition that has to be met is that a large part of the included genes are genuinely involved in PD pathogenesis as non-associated genes will dilute the overall association signal.

Future studies focusing on the contribution of complete biological mechanisms to PD susceptibility should be based on larger scale sequencing datasets. Transcriptomics

studies that serve as sources for pathway selection would be more reliable when including a larger number of individuals per α-synuclein Braak stage group. Furthermore, sequencing-based rather than array-based transcriptomics studies would generate a more comprehensive representation of the sample under investigation as it allows for identification of multiple transcripts per gene and poor-established RNA-molecules, such as non-coding RNA. Similarly, larger scale DNA sequencing studies will improve power to detect associations of gene-sets and finally also on individual gene level. Although our current study design included a limited sample size, especially for the replication datasets, we were able to determine grouped genetic effects for 3 biological mechanisms. The *mitochondrial dysfunction* and *caveolar-mediated endocytosis signaling* pathways are implicated to harbor additional genetic factors, besides the known PD genes. Furthermore, our result suggest that the deregulated expression levels of genes part of the renin-angiotensin system, observed in post-mortem brain tissue of early stage PD pathology, might be a consequence of genetic variation in the corresponding genes. Further investigation on genetic and functional level of these 3 pathways will improve our understanding and ultimately facilitate the discovery of adequate targets for therapeutic treatments of PD.

REFERENCES

1. Trinh J, Farrer M. Advances in the genetics of Parkinson disease. *Nature reviews Neurology* 2013; 9(8): 445-54.
2. Bonifati V. Autosomal recessive parkinsonism. *Parkinsonism & related disorders* 2012; 18 Suppl 1: S4-6.
3. Bonifati V. Genetics of Parkinson's disease--state of the art, 2013. *Parkinsonism & related disorders* 2014; 20 Suppl 1: S23-8.
4. Simon-Sanchez J, Schulte C, Bras JM, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* 2009; 41(12): 1308-12.
5. Lashuel HA, Overk CR, Oueslati A, Masliah E. The many faces of alpha-synuclein: from structure and toxicity to therapeutic target. *Nature reviews Neuroscience* 2013; 14(1): 38-48.
6. Shin N, Jeong H, Kwon J, et al. LRRK2 regulates synaptic vesicle endocytosis. *Experimental cell research* 2008; 314(10): 2055-65.
7. Manzoni C, Lewis PA. Dysfunction of the autophagy/lysosomal degradation pathway is a shared feature of the genetic synucleinopathies. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2013; 27(9): 3424-9.
8. Miura E, Hasegawa T, Konno M, et al. VPS35 dysfunction impairs lysosomal degradation of alpha-synuclein and exacerbates neurotoxicity in a Drosophila model of Parkinson's disease. *Neurobiology of disease* 2014; 71: 1-13.
9. Kalia LV, Lang AE. Parkinson's disease. *Lancet* 2015; 386(9996): 896-912.
10. Duke DC, Moran LB, Kalaitzakis ME, et al. Transcriptome analysis reveals link between proteasomal and mitochondrial pathways in Parkinson's disease. *Neurogenetics* 2006; 7(3): 139-48.
11. Monti C, Bondi H, Urbani A, Fasano M, Alberio T. Systems biology analysis of the proteomic alterations induced by MPP(+), a Parkinson's disease-related mitochondrial toxin. *Frontiers in cellular neuroscience* 2015; 9: 14.
12. Srivastava G, Singh K, Tiwari MN, Singh MP. Proteomics in Parkinson's disease: current trends, translational snags and future possibilities. *Expert review of proteomics* 2010; 7(1): 127-39.
13. Dumitriu A, Golji J, Labadorf AT, et al. Integrative analyses of proteomics and RNA transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci in Parkinson disease. *BMC medical genomics* 2016; 9: 5.
14. Basso M, Giraudo S, Corpillo D, Bergamasco B, Lopiano L, Fasano M. Proteome analysis of human substantia nigra in Parkinson's disease. *Proteomics* 2004; 4(12): 3943-52.
15. Dijkstra AA, Ingrassia A, de Menezes RX, et al. Evidence for Immune Response, Axonal Dysfunction and Reduced Endocytosis in the Substantia Nigra in Early Stage Parkinson's Disease. *PloS one* 2015; 10(6): e0128651.
16. Hofman A, Brusselle GG, Darwish Murad S, et al. The Rotterdam Study: 2016 objectives and design update. *European journal of epidemiology* 2015; 30(8): 661-708.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009; 25(14): 1754-60.
18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation

DNA sequencing data. *Genome research* 2010; 20(9): 1297-303.
19. Bras J, Guerreiro R, Hardy J. SnapShot: Genetics of Parkinson's disease. *Cell* 2015; 160(3): 570-.e1.
20. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants. 2014; 46(3): 310-5.
21. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *American journal of human genetics* 2011; 89(1): 82-93.
22. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics* 2012; 91(2): 224-37.
23. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *American journal of human genetics* 2013; 92(6): 841-53.
24. Nunnari J, Suomalainen A. Mitochondria: in sickness and in health. *Cell* 2012; 148(6): 1145-59.
25. Meisinger C, Sickmann A, Pfanner N. The mitochondrial proteome: from inventory to function. *Cell* 2008; 134(1): 22-4.
26. Gaweda-Walerych K, Zekanowski C. The impact of mitochondrial DNA and nuclear genes related to mitochondrial functioning on the risk of Parkinson's disease. *Current genomics* 2013; 14(8): 543-59.
27. Pickrell AM, Youle RJ. The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease. *Neuron* 2015; 85(2): 257-73.
28. Lenaz G, Fato R, Genova ML, Bergamini C, Bianchi C, Biondi A. Mitochondrial Complex I: structural and functional aspects. *Biochimica et biophysica acta* 2006; 1757(9-10): 1406-20.
29. Bove J, Prou D, Perier C, Przedborski S. Toxin-induced models of Parkinson's disease. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics* 2005; 2(3): 484-94.
30. Betarbet R, Sherer TB, MacKenzie G, Garcia-Osuna M, Panov AV, Greenamyre JT. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature neuroscience* 2000; 3(12): 1301-6.
31. Parker WD, Jr., Boyson SJ, Parks JK. Abnormalities of the electron transport chain in idiopathic Parkinson's disease. *Annals of neurology* 1989; 26(6): 719-23.
32. Langston JW, Forno LS, Tetrud J, Reeves AG, Kaplan JA, Karluk D. Evidence of active nerve cell degeneration in the substantia nigra of humans years after 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine exposure. *Annals of neurology* 1999; 46(4): 598-605.
33. Freire C, Koifman S. Pesticide exposure and Parkinson's disease: epidemiological evidence of association. *Neurotoxicology* 2012; 33(5): 947-71.
34. Mimaki M, Wang X, McKenzie M, Thorburn DR, Ryan MT. Understanding mitochondrial complex I assembly in health and disease. *Biochimica et biophysica acta* 2012; 1817(6): 851-62.
35. Hattori N, Yoshino H, Tanaka M, Suzuki H, Mizuno Y. Genotype in the 24-kDa subunit gene (NDUFV2) of mitochondrial complex I and susceptibility to Parkinson disease. *Genomics* 1998; 49(1): 52-8.
36. Marella M, Seo BB, Yagi T, Matsuno-Yagi A. Parkinson's disease and mitochondrial complex I: a perspective on the Ndi1 therapy. *Journal of bioenergetics and*

biomembranes 2009; 41(6): 493-7.

37. Negida A, Menshawy A, El Ashal G, et al. Coenzyme Q10 for Patients with Parkinson's Disease: A Systematic Review and Meta-Analysis. *CNS & neurological disorders drug targets* 2016; 15(1): 45-53.
38. Cha SH, Choi YR, Heo CH, et al. Loss of parkin promotes lipid rafts-dependent endocytosis through accumulating caveolin-1: implications for Parkinson's disease. *Molecular neurodegeneration* 2015; 10: 63.
39. Hashimoto M, Takenouchi T, Rockenstein E, Masliah E. Alpha-synuclein up-regulates expression of caveolin-1 and down-regulates extracellular signal-regulated kinase activity in B103 neuroblastoma cells: role in the pathogenesis of Parkinson's disease. *Journal of neurochemistry* 2003; 85(6): 1468-79.
40. Wary KK, Mariotti A, Zurzolo C, Giancotti FG. A requirement for caveolin-1 and associated kinase Fyn in integrin signaling and anchorage-dependent cell growth. *Cell* 1998; 94(5): 625-34.
41. Darvish H, Heidari A, Hosseinkhani S, et al. Biased homozygous haplotypes across the human caveolin 1 upstream purine complex in Parkinson's disease. *Journal of molecular neuroscience : MN* 2013; 51(2): 389-93.
42. Saavedra JM. Brain angiotensin II: new developments, unanswered questions and therapeutic opportunities. *Cellular and molecular neurobiology* 2005; 25(3-4): 485-512.
43. Rey P, Lopez-Real A, Sanchez-Iglesias S, Munoz A, Soto-Otero R, Labandeira-Garcia JL. Angiotensin type-1-receptor antagonists reduce 6-hydroxydopamine toxicity for dopaminergic neurons. *Neurobiology of aging* 2007; 28(4): 555-67.
44. Rodriguez-Pallares J, Rey P, Parga JA, Munoz A, Guerra MJ, Labandeira-Garcia JL. Brain angiotensin enhances dopaminergic cell death via microglial activation and NADPH-derived ROS. *Neurobiology of disease* 2008; 31(1): 58-73.
45. Joglar B, Rodriguez-Pallares J, Rodriguez-Perez AI, Rey P, Guerra MJ, Labandeira-Garcia JL. The inflammatory response in the MPTP model of Parkinson's disease is mediated by brain angiotensin: relevance to progression of the disease. *Journal of neurochemistry* 2009; 109(2): 656-69.
46. Zawada WM, Banninger GP, Thornton J, et al. Generation of reactive oxygen species in 1-methyl-4-phenylpyridinium (MPP+) treated dopaminergic neurons occurs as an NADPH oxidase-dependent two-wave cascade. *Journal of neuroinflammation* 2011; 8: 129.
47. Sonsalla PK, Coleman C, Wong LY, et al. The angiotensin converting enzyme inhibitor captopril protects nigrostriatal dopamine neurons in animal models of parkinsonism. *Experimental neurology* 2013; 250: 376-83.
48. Strittmatter SM, Thiele EA, Kapiloff MS, Snyder SH. A rat brain isozyme of angiotensin-converting enzyme. Unique specificity for amidated peptide substrates. *The Journal of biological chemistry* 1985; 260(17): 9825-32.
49. Chai SY, Mendelsohn FA, Paxinos G. Angiotensin converting enzyme in rat brain visualized by quantitative in vitro autoradiography. *Neuroscience* 1987; 20(2): 615-27.
50. Zubenko GS, Volicer L, Direnfeld LK, Freeman M, Langlais PJ, Nixon RA. Cerebrospinal fluid levels of angiotensin-converting enzyme in Alzheimer's disease, Parkinson's disease and progressive supranuclear palsy. *Brain research* 1985;

328(2): 215-21.
51. Jenkins TA, Mendelsohn FA, Chai SY. Angiotensin-converting enzyme modulates dopamine turnover in the striatum. *Journal of neurochemistry* 1997; 68(3): 1304-11.
52. Jenkins TA, Wong JY, Howells DW, Mendelsohn FA, Chai SY. Effect of chronic angiotensin-converting enzyme inhibition on striatal dopamine content in the MPTP-treated mouse. *Journal of neurochemistry* 1999; 73(1): 214-9.
53. Lopez-Real A, Rey P, Soto-Otero R, Mendez-Alvarez E, Labandeira-Garcia JL. Angiotensin-converting enzyme inhibition reduces oxidative stress and protects dopaminergic neurons in a 6-hydroxydopamine rat model of Parkinsonism. *Journal of neuroscience research* 2005; 81(6): 865-73.
54. Munoz A, Rey P, Guerra MJ, Mendez-Alvarez E, Soto-Otero R, Labandeira-Garcia JL. Reduction of dopaminergic degeneration and oxidative stress by inhibition of angiotensin converting enzyme in a MPTP model of parkinsonism. *Neuropharmacology* 2006; 51(1): 112-20.
55. Reardon KA, Mendelsohn FA, Chai SY, Horne MK. The angiotensin converting enzyme (ACE) inhibitor, perindopril, modifies the clinical features of Parkinson's disease. *Australian and New Zealand journal of medicine* 2000; 30(1): 48-53.
56. Lin JJ, Yueh KC, Chang DC, Lin SZ. Association between genetic polymorphism of angiotensin-converting enzyme gene and Parkinson's disease. *Journal of the neurological sciences* 2002; 199(1-2): 25-9.
57. Lin JJ, Yueh KC, Lin SZ, Harn HJ, Liu JT. Genetic polymorphism of the angiotensin converting enzyme and L-dopa-induced adverse effects in Parkinson's disease. *Journal of the neurological sciences* 2007; 252(2): 130-4.
58. Song GG, Lee YH. The insertion/deletion polymorphism in the angiotensin-converting enzyme and susceptibility to schizophrenia or Parkinson's disease: A meta-analysis. *Journal of the renin-angiotensin-aldosterone system : JRAAS* 2015; 16(2): 434-42.
59. Su G, Dou H, Zhao L, et al. The angiotensin-converting enzyme (ACE) I/D polymorphism in Parkinson's disease. *Journal of the renin-angiotensin-aldosterone system : JRAAS* 2015; 16(2): 428-33.
60. Ray PD, Huang BW, Tsuji Y. Reactive oxygen species (ROS) homeostasis and redox regulation in cellular signaling. *Cellular signalling* 2012; 24(5): 981-90.
61. Hernandes MS, Britto LR. NADPH oxidase and neurodegeneration. *Current neuropharmacology* 2012; 10(4): 321-7.
62. Kim GH, Kim JE, Rhie SJ, Yoon S. The Role of Oxidative Stress in Neurodegenerative Diseases. *Experimental neurobiology* 2015; 24(4): 325-40.
63. Mo JS, Kim MY, Ann EJ, Hong JA, Park HS. DJ-1 modulates UV-induced oxidative stress signaling through the suppression of MEKK1 and cell death. *Cell death and differentiation* 2008; 15(6): 1030-41.
64. Mudo G, Makela J, Di Liberto V, et al. Transgenic expression and activation of PGC-1alpha protect dopaminergic neurons in the MPTP mouse model of Parkinson's disease. *Cellular and molecular life sciences : CMLS* 2012; 69(7): 1153-65.
65. Nayernia Z, Jaquet V, Krause KH. New insights on NOX enzymes in the central nervous system. *Antioxidants & redox signaling* 2014; 20(17): 2815-37.
66. Choi DH, Cristovao AC, Guhathakurta S, et al. NADPH oxidase 1-mediated oxidative stress leads to dopamine neuron death in Parkinson's disease. *Antioxidants & redox*

signaling 2012; 16(10): 1033-45.

67. Cristovao AC, Guhathakurta S, Bok E, et al. NADPH oxidase 1 mediates alpha-synucleinopathy in Parkinson's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2012; 32(42): 14465-77.

SUPPLEMENTAL DATA

Table 1. Rare variant association results of gene-set analysis in IPDGC WES dataset

	weighted		AA changing		CADD10		CADD20		disrupt	
Pathway	*p* (emp.)	*n* var	*p* (emp.)	*n* var	*p* (emp.)	*n* var	*p* (emp.)	*n* var	*p* (emp.)	*n* var
Autophagy	0.758	526	0.794	346	0.917	311	0.860	237	1	45
Caveolar-mediated endocytosis signaling	0.397	357	0.125	211	0.252	167	0.206	134	1	20
Clathrin-mediated endocytosis signaling	0.28	1182	0.056	762	0.111	654	0.068	510	1	94
Mitochondrial dysfunction	**0.006 (0.046)**	1298	**0.00028 (0.029)**	879	**0.00011 (0.022)**	781	**0.00032 (0.017)**	639	**0.018 (0.049)**	91
Gap junction signaling	0.485	439	0.051	259	0.133	210	0.247	157	0.245	30
Angiopoietin signaling	0.190	540	0.169	374	0.088	330	0.064	249	0.167	34
Dopamine-DARPP32 feedback in cAMP signaling	0.217	252	0.13	140	0.062	124	0.285	96	0.995	15
Signaling by rho family GTPases	0.519	770	1	501	0.922	447	0.693	356	0.325	34
mTOR signaling	1	637	0.863	326	0.885	326	0.672	244	0.348	42
Huntington's Disease signaling	0.337	1492	0.193	898	0.201	785	0.122	583	0.183	85
Axonal guidance signaling	0.177	2285	0.519	1398	0.424	1232	0.686	943	0.433	107
Renin-angiotensin signaling	0.797	510	0.546	327	0.379	278	0.729	207	0.579	46
Rac signaling	1	765	1	478	0.912	422	0.911	334	0.213	51
Actin cytoskeleton signaling	0.699	983	0.889	617	0.868	529	0.754	422	0.089	43
FAK signaling	0.552	589	0.590	358	0.645	313	0.647	237	0.162	30
FLT3 signaling in hematopoietic progenitor cells	0.785	329	0.605	183	0.62	163	0.601	156	0.651	9
Clathrin-mediated endocytosis signaling	0.28	1182	0.056	762	0.111	654	0.068	510	1	94

weighted = subset of all exonic variants weighted on CADD unscaled C-scores; AA changing = amino acid changing subset; CADD10 = subset of variants predicted to belong to 10% most deleterious variants of the genome; CADD20 = subset of variants predicted to belong to 1% most deleterious variants of genome; disruptive = subset of loss-of-function variants. *p* (emp) = nominal *p*-value with empirical *p*-value in parenthesis of which the latter was calculated by comparison to 1,000 randomly drawn gene-sets. The *p*-values in bold reached nominal significance and were selected for competitive testing. Significant empirical *p*-values are underlined.

Table 2. Common variant association results of gene-set analysis in IPDGC WES dataset.

	AA changing		CADD10		CADD20	
Pathway	*p* (emp.)	*n* var	*p* (emp.)	*n* var	*p* (emp.)	*n* var
Autophagy	0.374	10	0.405	4	0.355	1
Caveolar-mediated endocytosis signaling	**0.034 (0.054)**	8	**0.047 (0.066)**	5	0.250	2
Clathrin-mediated endocytosis signaling	0.348	21	0.154	14	0.758	6
Mitochondrial dysfunction	0.322	32	0.635	24	0.314	12
Gap junction signaling	0.430	8	0.456	4	0.794	2
Angiopoietin signaling	0.782	12	0.496	7	0.310	4
Dopamine-DARPP32 feedback in cAMP signaling	0.125	1	0.125	1	0.125	1
Signaling by rho family GTPases	0.347	18	0.377	14	0.785	11
mTOR signaling	0.126	6	0.135	3	0.191	1
Huntington's Disease signaling	0.579	32	0.452	16	0.359	9
Axonal Guidance signaling	0.222	52	0.152	43	0.189	26
Renin-Angiotensin signaling	0.086	6	0.074	4	**0.033 (0.054)**	2
Rac signaling	0.308	13	0.544	9	0.453	5
Actin cytoskeleton signaling	0.068	16	0.186	9	0.239	5
FAK signaling	0.758	17	0.320	10	0.360	5
FLT3 signaling in hematopoietic progenitor cells	0.825	3	0.615	2	0.739	1
Clathrin-mediated endocytosis signaling	0.348	21	0.154	14	0.758	6

weighted = subset of all exonic variants weighted on CADD unscaled C-scores; AA changing = amino acid changing subset; CADD10 = subset of variants predicted to belong to 10% most deleterious variants of the genome; CADD20 = subset of variants predicted to belong to 1% most deleterious variants of genome; disruptive = subset of loss-of-function variants. *p* (emp) = nominal *p*-value with empirical *p*-value in parenthesis of which the latter was calculated by comparison to 1,000 randomly drawn gene-sets. The *p*-values in bold reached nominal significance and were selected for competitive testing. Significant empirical p-values are underlined.

Table 3. Common variant association of *renin-angiotensin* and *caveolar-mediated endocytosis signaling* pathway.

		IPDGC		MERCK		PPMI	
Pathway	variant type	*p*(emp)	n var	*p*(emp)	n var	*p*(emp)	n var
Renin angiotensin	AA changing	0.086	6	0.198	5	0.756	3
(common variants)	CADD10	0.074	4	0.267	4	0.666	1
	CADD20	**0.033 (0.054)**	2	**0.040 (0.049)**	2	NA	NA
Caveolar-mediated endocytosis	AA changing	**0.014 (0.025)**	219	0.345	12	0.087	21
(common and rare variants)	CADD10	**0.040 (0.070)**	172	0.299	9	**0.024 (0.024)**	14
	CADD20	0.13	136	0.351	4	0.163	8

AA changing = amino acid changing subset; CADD10 = subset of variants predicted to belong to 10% most deleterious variants of the genome; CADD20 = subset of variants predicted to belong to 1% most deleterious variants of genome; *p* (emp) = nominal *p*-value with empirical *p*-value in parenthesis of which the latter was calculated by comparison to 1,000 randomly drawn gene-sets. The underlined values are representing the replication. NA = not applicable.

Table 4. Common and rare variant association results of gene-set analysis in IPDGC WES dataset.

	AA changing		CADD10		CADD20	
Pathway	*p* (emp.)	*n* var	*p* (emp.)	*n* var	*p* (emp.)	*n* var
Autophagy	0.623	356	0.707	315	0.637	238
Caveolar-mediated endocytosis signaling	**0.014 (0.025)**	219	**0.040 (0.070)**	172	0.130	136
Clathrin-mediated endocytosis signaling	0.083	783	0.086	668	0.115	516
Mitochondrial dysfunction	**0.0081 (0.169)**	911	**0.0025 (0.085)**	805	**0.0014 (0.047)**	651
Gap junction signaling	**0.043 (0.079)**	267	0.114	214	0.345	159
Angiopoietin signaling	0.352	386	0.158	337	0.058	253
Dopamine-DARPP32 feedback in cAMP signaling	0.067	141	**0.034 (0.045)**	125	0.125	97
Signaling by rho family GTPases	0.633	519	0.584	461	0.693	367
mTOR signaling	0.282	371	0.301	329	0.321	245
Huntington's Disease signaling	0.407	930	0.169	801	0.079	592
Axonal guidance signaling	0.241	1450	0.136	1275	0.237	969
Renin-angiotensin signaling	0.187	333	0.071	282	0.100	209
Rac signaling	0.588	491	0.869	431	0.709	339
Actin cytoskeleton signaling	0.185	633	0.417	538	0.350	427
FAK signaling	0.613	375	0.352	323	0.392	242
FLT3 signaling in hematopoietic progenitor cells	0.635	186	0.612	165	0.863	114
Clathrin-mediated endocytosis signaling	0.083	783	0.086	668	0.115	516

weighted = subset of all exonic variants weighted on CADD unscaled C-scores; AA changing = amino acid changing subset; CADD10 = subset of variants predicted to belong to 10% most deleterious variants of the genome; CADD20 = subset of variants predicted to belong to 1% most deleterious variants of genome; disruptive = subset of loss-of-function variants. *p* (emp) = nominal *p*-value with empirical *p*-value in parenthesis of which the latter was calculated by comparison to 1,000 randomly drawn gene-sets. The *p*-values in bold reached nominal significance and were selected for competitive testing. Significant empirical *p*-values are underlined.

Table 5. Individual gene associations for *mitochondrial dysfunction.*

	Gene	*p* (emp.)	n var	Location
IPDGC	*CASP9*	0.029 (0.021)	17	cytoplasm
	FURIN	0.00045 (0.0002)	19	cytoplasm
	COX4I1	0.006 (0.0035)	7	complex 4
	NDUFA3	0.041 (0.071)	3	complex 1
	ATP5J	0.0023 (0.0003)	4	complex 5
	ATP5O	0.035 (0.030)	17	complex 5
	SDHA	0.0059 (0.0078)	42	complex 2
MERCK	*NDUFS5*	0.033 (0.036)	1	complex 1
	NDUFV1	0.00001 (0.0001)	2	complex 1
PPMI	*NDUFS8*	0.045 (0.059)	2	complex 1
	NDUFA9	0.0002 (0.0008)	1	complex 1
	NDUFB5	0.011 (0.018)	1	complex 1
	COX6C	0.012 (0.015)	2	complex 4

p (emp) = nominal *p*-value with empirical *p*-value in parenthesis of which the latter was calculated by comparison to 10,000 permutations of affection status.

Table 6. Single variant statistics of renin-angiotensin signaling pathway.

Dataset	Gene	Position	Var type	MAF cases	MAF controls
IPDGC	*ACE*	17:61557773	missense	0.026	0.013
	NOX1	X:100105195	missense	0.029	0.026
MERCK	*AGT*	1:230845977	missense	0.161	0.142
	MEKK1	5:56177843	missense	0.023	0.016

Position = genomic position; var type = variant type; MAF cases = minor allele frequency cases; MAF controls = minor allele frequency controls.

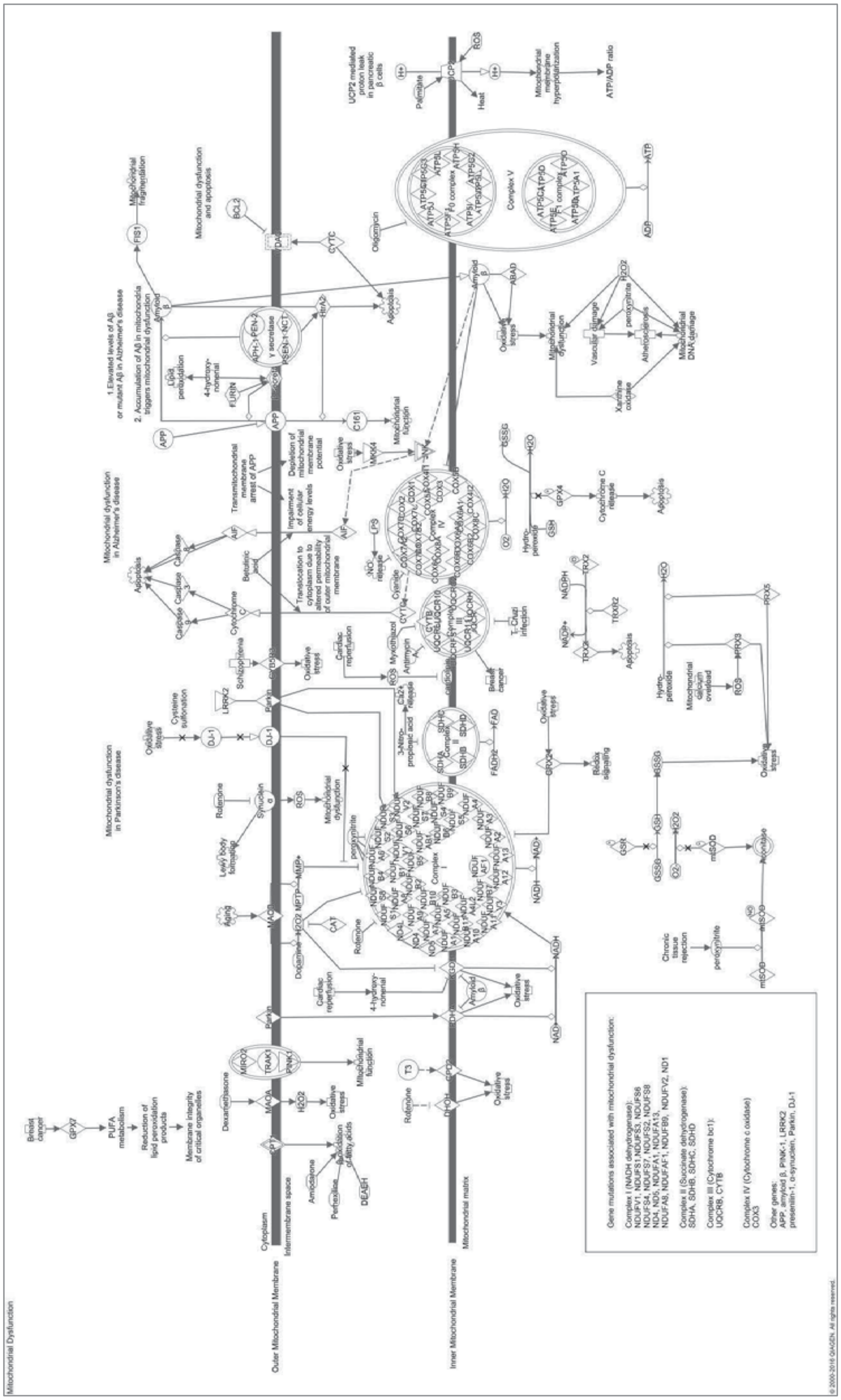


Figure 1. Complete overview of *Mitochondrial dysfunction signaling* pathway of the IPA database.

5

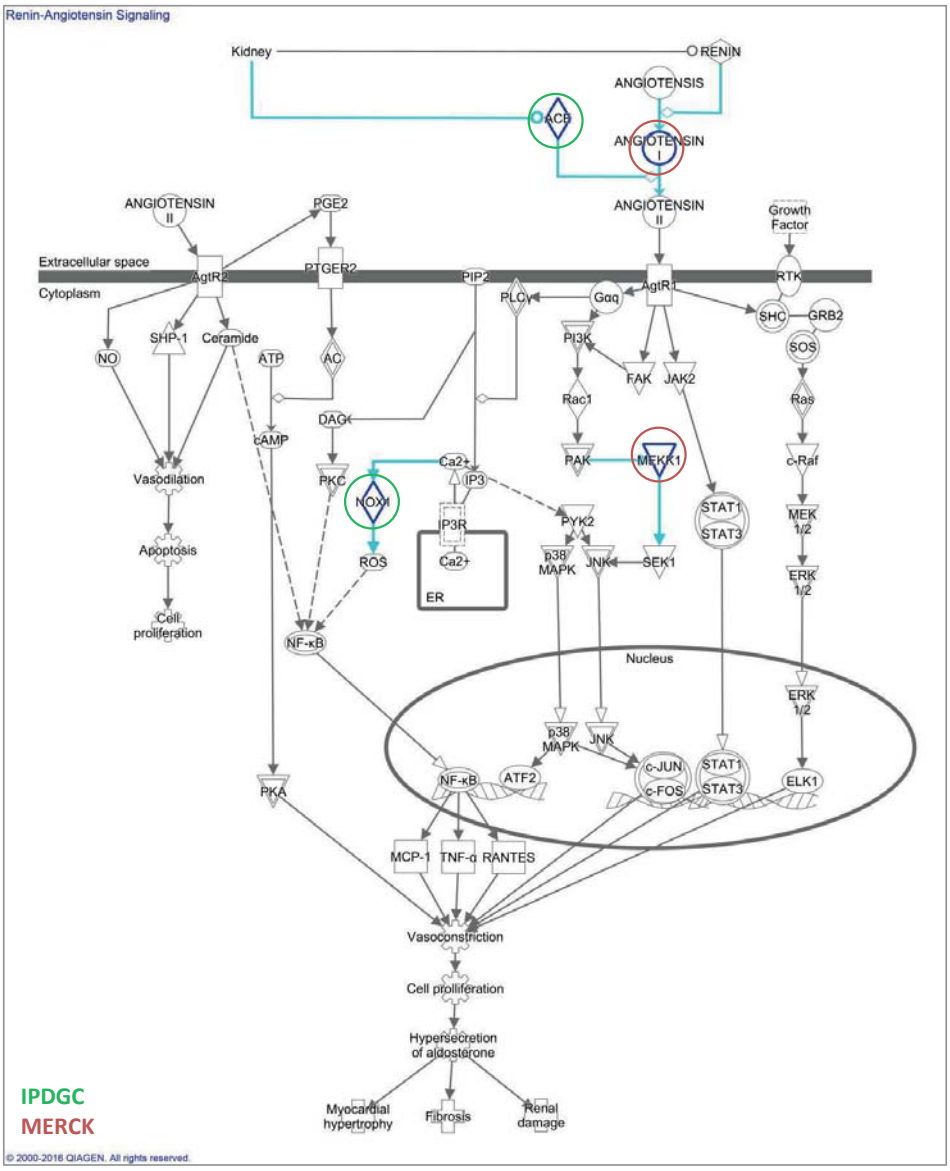


Figure 2. Complete overview of *renin-angiotensin signaling* pathway of the IPA database.

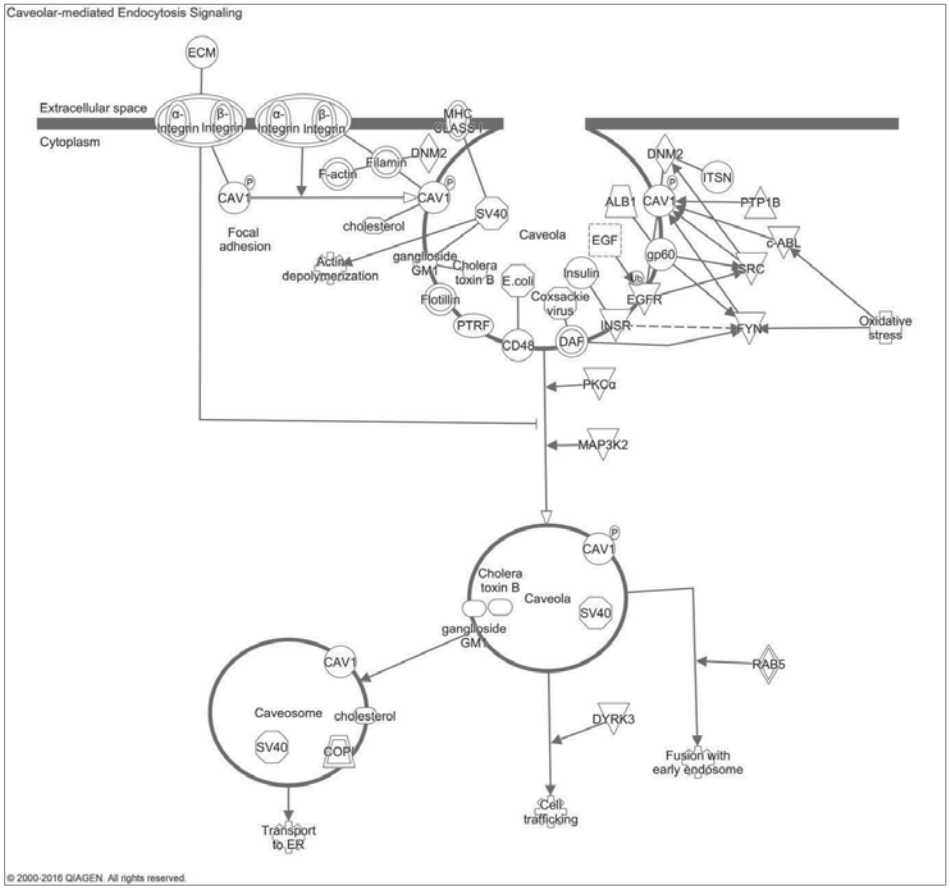


Figure 3. Complete overview of *caveolar-mediated endocytosis signaling* pathway of IPA database.