Functional Variants in *LRRK2* Confer Pleiotropic Effects on Crohn's Disease and Parkinson's Disease Risk

A coding Crohn's disease (CD)-associated risk variant in the *LRRK2* gene, N2081D, affects age of onset, disease location, and stress response in human macrophages and, together with the coding CD-protective *LRRK2* N551K variant, confers pleiotropic effects on both CD and Parkinson's disease.

Ken Y. Hui,[1,2] Heriberto Fernandez-Hernandez,[3] Jianzhong Hu,[3] Adam Schaffner,[4,5] Nathan Pankratz,[6] Nai-Yun Hsu,[3] Ling-Shiang Chuang,[3] Shai Carmi,[7] Nicole Villaverde,[3] Xianting Li,[4] Manual Rivas,[8,9] Adam P. Levine,[10] Xiuliang Bao,[3] Philippe R. Labrias,[3] Talin Haritunians,[11] Darren Ruane,[12] Kyle Gettler,[1,13] Ernie Chen,[3] Dalin Li,[11] Elena R. Schiff,[10] Nikolas Pontikos,[10] Nir Barzilai,[14] Steven R. Brant,[15,16] Susan Bressman,[17] Adam S. Cheifetz,[18] Lorraine N. Clark,[19,20] Mark J. Daly,[8,9,21,22] Robert J. Desnick,[3] Richard H. Duerr,[23,24] Seymour Katz,[25,26,27] Todd Lencz,[28] Richard H. Myers,[29] Harry Ostrer,[30] Laurie Ozelius,[3,31] Haydeh Payami,[32,33] Yakov Peter,[34,35] John D. Rioux,[36,37] Anthony W. Segal,[10] William K. Scott,[38] Mark S. Silverberg,[39,40] Jeffery M. Vance,[38] Iban Ubarretxena-Belandia,[5] Tatiana Foroud,[41] Gil Atzmon,[12,42] Itsik Pe'er,[43] Yiannis Ioannou,[3] Dermot P.B. McGovern,[11] Zhenyu Yue,[4] Eric E. Schadt,[3,44] Judy H. Cho,[1,3,13,45,*] Inga Peter[3,44,46,*]

[1]Section of Digestive Diseases, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA 06520

[2]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA 06520

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[4]Departments of Neurology and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[5]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[6]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA 55455

[7]Braun School of Public Health and Community Medicine, The Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel, 9112102

[8]Department of Medical and Population Genetics, Broad Institute, Cambridge, MA, USA 02142

[9]Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA 02114

[10]Centre for Molecular Medicine, Division of Medicine, University College, London, UK WC1E 6JF

[11]Translational Genomics Group, F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA 90048

[12]Department of Immunology and Inflammation, Regeneron Pharmaceuticals, Tarrytown, NY 10591

[13]Department of Genetics, Yale University, New Haven, CT, USA 06520

[14]Departments of Genetics and Medicine, Albert Einstein College of Medicine, Bronx, NY, USA 10461

[15]Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA 21231

[16]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA 21231

[17]Alan and Barbara Mirken Department of Neurology, Beth Israel Medical Center, New York, NY USA 10003

[18]Division of Gastroenterology, Beth Israel Deaconess Medical Center, Boston, MA, USA 02215

[20]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY, USA 10032

[21]Taub Institute for Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA 10032

[21]Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital, Boston, MA USA 02114

[22]Department of Genetics, Harvard Medical School, Boston, MA, USA 02115

[23]Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA 15261

[2]Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA 15261

[25]New York University School of Medicine, New York City, NY, USA

[26]North Shore University-Long Island Jewish Medical Center, Manhasset, NY, USA

[27]St. Francis Hospital, Roslyn, NY, USA

[28]Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY, USA 11030

[29]Department of Neurology, Boston University School of Medicine, Boston, MA, USA 02114

[30]Departments of Pathology and Pediatrics, Albert Einstein College of Medicine, Bronx, NY 10461

[31]Deparment of Neurology, Massachusetts General Hospital, Boston, MA, USA 02114

[32]Departments of Neurology and Genetics, University of Alabama at Birmingham, Birmingham, AL, USA 35294

[33]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA 35805

[34]Department of Biology, Touro College, Queens, NY, USA 10033

[35]Department of Pulmonary Medicine, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY, USA 10033

[36]Research Center, Montreal Heart Institute, Montreal, Quebec, Canada H1T1C8

[37]Faculté de Médecine, Université de Montréal, Montreal, Quebec, Canada H1T1C8

[38]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA 33136

[39]Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, Ontario, Canada M5T3L9

[40]Department of Medicine, University of Toronto, Toronto, Ontario, Canada M5G1X5

[41]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA 46202

[42]Faculty of Natural Sciences, University of Haifa, Haifa, Israel 3498838

[43]Center for Computational Biology and Bioinformatics, Columbia University, New York, NY, USA 10032

[44]Institute for Genetics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029

[45]Section of Gastroenterology and Hepatology, Department of Pediatrics, Yale University School of Medicine, New Haven, CT, USA 06520

[46]Corresponding author: inga.peter@mssm.edu

[*]Equal contribution

ABSTRACT

Crohn's disease (CD), a form of inflammatory bowel disease, has higher prevalence in Ashkenazi Jewish (AJ) than in non-Jewish (NJ) European populations. To define the role of non-synonymous mutations, we performed exome sequencing of AJ CD patients, followed by array-based genotyping and association analysis in 2,066 AJ CD cases and 3,633 controls. We detected association signals at *LRRK2* conferring CD risk (N2081D, $P=9.5x10^{-10}$) or protection (N551K, tagging R1398H-associated haplotype, $P=3.3x10^{-8}$) that affected CD age of onset, disease location, LRRK2 activity, and autophagy flux. Bayesian network analysis of CD intestinal tissue further implicated *LRRK2* in disease pathogenesis. Analysis of the extended *LRRK2* locus in 24,570 CD cases, PD cases, and controls revealed extensive pleiotropy between CD and PD in both AJ and NJ cohorts. The *LRRK2* N2081D CD risk allele is located in the same kinase domain as G2019S, the major genetic cause of familial and sporadic Parkinson's disease (PD). Like G2019S, N2081D is associated with increased kinase activity, whereas neither N551K nor R1398H on the protective haplotype alter kinase activity. Rather, the histamine allele at R1398H, but not N551K, increases GTPase activity, thereby deactivating LRRK2. We confirm the opposing functions of risk and protective alleles on cytoskeletal function and autophagy in primary human macrophages. The presence of shared *LRRK2* alleles in CD and PD provides refined insight into disease mechanisms and may have major implications for the treatment of these two seemingly unrelated diseases.

INTRODUCTION

The inflammatory bowel diseases (IBD) are comprised of two major subtypes, Crohn's disease (CD) and ulcerative colitis (UC), which are distinguished by the distribution of the chronic inflammatory changes. In UC, the inflammation is relatively superficial and is confined to the colon. CD most commonly affects the terminal ileum (last part of the small intestine) and colon, and is frequently associated with deep, transmural inflammation, often resulting in obstruction and abscess formation requiring resectional surgery.

Approved medical therapies for moderate to severe IBD are the same for CD and UC, and include monoclonal antibodies against the pro-inflammatory TNF cytokine and more recently, antibodies against the $\alpha_4\beta_7$ integrin, which blocks leukocyte trafficking to the intestine. However, present therapies provide prolonged deep remission in only a minority of IBD patients; there is a substantial unmet need for more effective medical therapies, especially for CD patients. Genome-wide association studies (GWAS) have identified over 200 loci associated to IBD (*1, 2*), providing many new potential therapeutic targets. The large majority of these loci are common to CD and UC, implicating numerous pathways, notably the pro-inflammatory interleukin 23 pathway. In particular, R381Q within *IL23R* (interleukin 23 receptor) is a loss-of-function allele that confers protection against developing IBD (*3*). Importantly, monoclonal antibodies blocking the IL-23 pathway have demonstrated efficacy in IBD, as well as a favorable safety profile (*4*). CD-predominant loci include *NOD2* and a number of autophagy genes (e.g. *ATG16L1*, *IRGM*). *NOD2* is an intracellular receptor for bacterial peptidoglycan and is expressed in a wide variety of cells including plasma cells, innate immune leukocytes (e.g. monocytes, macrophages, dendritic cells) and Paneth cells located at the base of small intestinal (but not typically colonic) crypts and extruding potent antimicrobial peptides. Loss-of-function *NOD2* risk alleles are associated with ileal, as opposed to colonic location, earlier age of onset and earlier need for resectional surgery. Among the autophagy-associated signals are the *ATG16L1* T300A allele that results in *ATG16L1* degradation through caspase-3 activation (*5*) and multiple polymorphisms in the 5q33.1 region that cause tissue-specific variation in *IRGM* expression (*6, 7*).

However, a fundamental limitation of common variant-predominant GWAS is the imprecise definition of genes, specific alleles and mechanisms driving most association signals identified thus far, with the large

majority of independent GWAS signals driven by common variants of modest statistical and functional effects. Furthermore, common variation in composite is predicted to contribute only a modest fraction of expected heritability for many diseases. For these reasons, major sequencing efforts to identify rare variants of potentially higher statistical and functional effects are of importance for refining the pathways associated with disease pathogenesis and designing novel therapies.

We hypothesized that uncommon CD susceptibility alleles with higher effects (i.e. odds ratios), which had eluded analysis in common variant-predominant GWAS, play an important role in genetic predisposition to CD and can elucidate new insights into CD pathogenesis. In this study, we sought to identify the strongest functionally relevant associations and to characterize their biological implications. Given that a major epidemiologic feature of IBD is its several-fold higher prevalence in Ashkenazi Jewish (AJ) cohorts (*8, 9*) compared to non-Jewish Europeans (NJs), we performed exome sequencing of AJ CD cases followed by custom array-based genotyping in a large case-control cohort. We identified independent coding CD risk and protective alleles in *LRRK2*, a large multifunctional gene that confers the greatest genetic effects reported thus far in Parkinson's disease (PD), a neurodegenerative movement disorder affecting the basal ganglia and characterized by resting tremor, bradykinesia, rigidity and postural instability (*10*). The presence of shared alleles in CD and PD provides refined insight into disease mechanisms and may have major implications for the treatment of these two seemingly unrelated diseases.

RESULTS

Exome sequencing and HumanExome chip study design

We first performed exome sequencing of 50 AJ CD individuals randomly selected from high quality DNA samples and confirmed by prior chip data (*11*) to cluster as 100% Ashkenazi Jewish in order to ensure optimal utility of cataloguing novel variation (**Supplementary Fig. S1, Supplementary Table S1**). From these results, we selected 4,277 putatively high-yield novel mutations, adding these to the HumanExome beadchip (**Supplementary Fig. S2**, **Supplementary Table S2**). We next performed discovery-phase genotyping and

association analyses in individuals with full genetic AJ ancestry (*11*) (**Supplementary Fig. S3, Supplementary Table S3**).

Top coding region associations in CD

In the discovery-phase cohort of 1,477 unrelated CD cases and 2,614 independent healthy controls, non-synonymous variants at three loci on chromosomes 1, 12, and 16 demonstrated associations that reached a chip-wide significance (**Table 1**). Importantly, in addition to the previously reported *NOD2* and *IL23R* alleles, non-synonymous variants, N2081D in *LRRK2* and S6N in *SLC2A13*, in strong linkage disequilibrium (LD) with each other (r^2=0.91), were identified to be associated with CD risk (minor allele frequency, MAF_{CD}=8.1%, OR=1.73, P=2.56x10^{-9} and MAF_{CD}=8.1%, OR=1.73, P=2.68x10^{-9}, respectively). *LRRK2* N551K was also associated with CD protection (MAF_{CD}=6.6%, OR=0.65, P=7.06x10^{-7}; **Table 1, Fig. 1A, Supplementary Fig. S4**). We then evaluated the evidence for CD association in an independent AJ cohort of 589 CD and 1019 controls (**Supplementary Table S3**); this replicated the association signals at *LRRK2* N2081D (MAF_{CD}=7.4%, OR=1.34, P=4.40x10^{-2}), at *SLC2A13* S6N (MAF_{CD}=7.7%, OR=1.46, P=9.58x10^{-3}), and at *LRRK2* N551K (MAF_{CD}=7.0%, OR=0.72, P=1.27x10^{-2}). Meta-analysis revealed genome-wide significant CD risk at *LRRK2* N2081D (P=9.51x10^{-10}) and at *SLC2A13* S6N (P=1.39x10^{-10}), and protection at *LRRK2* N551K (P=3.28x10^{-8}). A list of all coding variants with discovery-phase association P-values<2x10^{-5} is provided in **Supplementary Table S4**. Notably, R1398H (MAF_{CD}=6.6%, OR=0.71, P=7.33x10^{-5}) and K1423K (MAF_{CD}=5.9%, OR=0.66, P=4.4x10^{-6}) in the *LRRK2* gene, which have previously been reported as combining with N551K to form a protective haplotype in PD (*12-15*), were found to show weaker associations with CD (**Supplementary Table S4**).

Prior studies have implicated distinct common alleles in the *LRRK2* region as being associated to CD (*1, 16, 17*). To further elucidate the genetic structure of the *LRRK2* signal, we conducted a conditional analysis using the discovery cohort, which demonstrated that this broad association peak was entirely dependent on the coding mutation at N2081D in *LRRK2* (**Fig. 1B**): *SLC2A13* S6N, as well as the association signal from the previously reported GWAS hits, including non-synonymous variant rs3761863 (M2397T) (*16, 18*), were

substantially attenuated. Conditioning on N2081D genotypes verified the independence of the protective association signal at *LRRK2* N551K linked to lower CD risk (OR=0.67, P=1.4x10^{-6}; **Fig. 1B**). Conditioning on N551K or R1398H from the protective haplotype as a covariate had minimal effect on the association signal. Interestingly, in phased haplotype association analysis (**Supplementary Table S5**), the 2081D risk variant occurs completely on the background of the protein-destabilizing allele M2397(*18*) (MAF_{CD}=45%; pairwise D'=1.0, r^2=0.09), while the 551K protective variant co-resides with the stabilizing 2397T(*18*) allele (pairwise D'=0.94, r^2=0.06). Conditioning on both N551K and N2081D together effectively eliminated the association signal at M2397T (conditioned P=0.015; unconditioned P=5.9x10^{-7}).

The multi-function kinase, *LRRK2*, has attracted considerable attention since variants in this gene were recognized as major risk factors for PD (*19*). Of note, the G2019S mutation in *LRRK2*, the best known genetic cause of familial and sporadic PD worldwide and located in the same kinase domain as N2081D, showed suggestive, but not genome-wide significant, association with CD (unconditioned OR=1.9, P=4.8x10^{-3}) and no LD with N2081D (r^2=0.0) in the AJ cohort.

Further replication and validation of the shared CD and PD risk within the LRRK2 locus

To replicate our findings in the NJ cohorts and explore the pleiotropic effect of *LRRK2* variation on CD and PD risk, we expanded our analysis to include a total of 8,314 independent AJ and 16,401 independent NJ participants comprising 6,538 CD cases, 5,570 PD cases, and 12,607 healthy controls genotyped in previous studies (**Supplementary Table S3**). After performing imputation and quality control measures, we conducted association testing on the set of *LRRK2* variants in these datasets (see **Supplementary Material and Methods**). Like in the discovery cohort, in both AJs and NJs, we observed a multi-marker CD-associated signal within *LRRK2* (**Supplementary Table S6**) that was fully conditioned on N2081D (**Supplementary Fig. S5A-B**). Also, conditioning on N551K or R1398H as a covariate had minimal effect on the broad association peak. Importantly, in the NJ dataset, association results showed similar marginal effects for N2081D (OR_{AJ}=1.7 [1.4-2.0] vs. OR_{NJ}=1.6 [1.3-2.0]) and N551K (OR_{AJ}=0.67 [0.57-0.79] vs. OR_{NJ}=0.89 [0.79-1.0]) or R1398H

(OR_{AJ}=0.71 [0.60-0.84] vs. OR_{NJ}=0.88 [0.78-0.99]) but with substantially lower MAF's, especially for N2081D (MAF_{AJ_CD}=8.0% vs. MAF_{NJ_CD}=2.9%; **Table 2).** Of note, G2019S did not have nominally significant CD association (P = 0.12), likely due to subtle stochastic fluctuation in allele frequencies during imputation.

To examine the genetic link between CD and PD, we then assessed PD association with *LRRK2* N2081D and N551K/R1398H in AJ and NJ cohorts, observing association signals for all polymorphisms (**Table 2**). Specifically, the OR estimates of the protective variants, 551K and R1398H, were similar between CD and PD with slight differences between AJ and NJ cohorts (N551K: OR_{AJ_CD}=0.67 [0.57–0.79] and OR_{AJ_PD}=0.77 [0.67-0.90]; OR_{NJ_CD}=0.89 [0.79-1.0] and OR_{NJ_PD}=0.87 [0.77-1.0], and R1398H: OR_{AJ_CD}=0.71 [0.60–0.84] and OR_{AJ_PD}=0.84 [0.72-0.98]; OR_{NJ_CD}=0.88 [0.78-0.99] and OR_{NJ_PD}=0.88 [0.77-1.0]). However, in both populations, the risk allele, N2081D, showed higher ORs in association with CD (OR_{AJ_CD}=1.7 [1.4–2.0], OR_{NJ_CD}=1.6 [1.3–2.0]) than with PD (OR_{AJ_PD}=1.1 [1.0–1.4], OR_{NJ_PD}=1.3 [CI 1.0–1.6]). Conditioning on N2081D or N551K demonstrated no difference, with G2019S remaining by far the dominant PD signal (**Supplementary Fig. S5C-D**).

To determine the degree of pleiotropy in the region, we selected variants at least nominally (P < 0.05) associated with both CD and PD and assessed their direction and magnitude of effect across diseases. Following LD pruning (i.e. removal of correlated mutations with pairwise r^2>0.8, thus ensuring statistical independence among the remaining mutations), we detected a consistent pattern of correlated effect sizes, with 23 of 26 independent variants (88%) exhibiting effects in the same direction for both diseases in the AJ dataset (binomial P=5.2x10^{-6}) and, similarly, 25 of 29 variants (86%) in the NJ data (P=7.6x10^{-6}; **Fig. 2**). Taken together, our findings suggest extended pleiotropy between CD and PD throughout the *LRRK2* locus.

Network analysis of IBD tissues further implicates LRRK2 in CD

Given strong LD within the *LRRK2* locus containing several plausible candidate genes, including *SLC2A13* and *MUC19* (**Supplementary Table S6**), we conducted network analysis to explore which of these genes participate in biological pathways involved in CD pathogenesis. We constructed an IBD Bayesian network

using previously described methodology (*20*), from gene expression data for 8,382 genes. The expression data were collected in 203 intestinal biopsies that included ileum, ascending colon, descending colon and transverse colon, and inflamed and non-inflamed sigmoid and rectum, all collected at baseline from 54 anti-TNFα resistant CD patients enrolled in the Ustekinumab (anti-IL12/IL23) clinical trial (*21, 22*). Among the full set of genes, we defined a specific subset, located within IBD-associated loci previously defined in an Immunochip-based large-scale genetic analysis (*1*) with the goal to project these genes onto the intestinal network and identify co-expressed genes that act together. We then excluded genes previously associated with PD (*23*), including *LRRK2*, as well as genes within 1 Mb of *LRRK2* to see whether either *LRRK2* or other genes will be "recovered" by the network as being co-expressed with the IBD-associated genes. We found that the largest connected sub-network of genes, which represents a set of co-expressed IBD-associated genes, contained *LRRK2*, but no other genes in the genomic neighborhood of *LRRK2* (**Fig. 3**), thus implicating *LRRK2* in particular in IBD pathogenesis. Of note, of the 622 genes in this sub-network, there were 102 (16.4%) IBD-associated genes, a 2.5-fold enrichment compared to the full intestinal network (hypergeometric $P=7.6x10^{-8}$). Importantly, *LRRK2* was closely connected to *GPR65*, a proton-sensing G-protein coupled receptor associated to IBD and altered lysosomal function (*24*) and to *HLA-DPA1*, an α-subunit of the major histocompatibility complex protein/peptide-antigen receptor and graft-versus-host disease antigen complex linked to both IBD (*25*) and PD (*26*).

Effect of LRRK2 mutations on protein kinase and GTPase activity

Prior studies in PD suggest a central role for increased LRRK2 kinase activity in disease risk resulting from gain-of-function mutations in the *LRRK2* kinase domain. Given that both PD-risk G2019S and CD-risk N2081D are located in the kinase domain (**Fig. 4A**), we investigated the effect of CD-associated *LRRK2* mutations on kinase activity. Specifically, we quantified phosphorylation of a newly identified LRRK2 substrate, Rab10 (*27*) by LRRK2 WT, and LRRK2 bearing G2019S, R1398H, N551K, N551K+R1398H and N2081D mutations that were expressed and purified from HEK293T cells (**Fig. 4B**). We demonstrated a ~30% increase in phosphorylated Rab10 (pRab10) in the presence of *LRRK2* N2081D mutation compared to WT (**Fig. 4B**) and

also confirmed a previous report that G2019S increased pRab10 (*27*). In contrast, no change was observed in the R1398H, N551K, or N551K+R1398H carrier cells. Roc, a Ras/GTPase domain in complex proteins, is also a common site of PD-linked *LRRK2* mutations, which presumably retain a higher fraction of LRRK2 in a GTP-bound 'on'-state, thereby promoting neurodegeneration (*28, 29*). Importantly, the PD-protective R1398H mutation, which is in strong LD with the CD-protective N551K, is located in the Roc domain (**Fig. 4A**). To determine the effects of *LRRK2* variants on LRRK2 GTPase activity, we compared the ratio of GDP/GTP-bound LRRK2 *in vitro* across the variants (**Fig. 4C**). We found that the GTPase activity was increased in both R1398H and N551K+R1398H-transfected cells, but not in G2019S, N2081D, or N551K mutants (**Fig. 4C**).

Role of LRRK2 mutations in cytoskeletal and autophagy function in human-patient macrophages

To further investigate the properties of the *LRRK2* mutants (**Fig. 4A**), we characterized human monocyte-derived M1 macrophages collected from CD patients who carried *LRRK2* N2081D (n=4), N551K+R1398H (all samples selected for their 551K carrier status also carried 1398H)(n=5), or neither mutation (n=4) in response to cellular serum-nutrient starvation (**Fig. 5**). No differences were detected in total LRRK2 expression by mutation status. As LRRK2 has been reported to influence acetylation of α-tubulin, thus regulating cellular protein trafficking via the microtubule cytoskeleton, we determined the effect of the *LRRK2* mutations on α-tubulin protein acetylation (**Fig. 5A**). Lower acetylation was detected in N2081D carrier cells under normal and PBS-stressed conditions, suggesting impaired resting acetylation activity and a lack of response to cellular stress. In contrast, the highest basal acetylation was detected in non-carriers and carriers of the protective 551K+1398H mutations, which proportionally decreased following cell starvation. As α-tubulin acetylation is associated with autophagy (*30*), one of the major pathophysiological processes involved in CD (and in PD) development, we next investigated the effect of the mutations on autophagy markers, LC3-II, an autophagsome-bound form of the microtubule-associated protein 1 light chain 3β (LC3B), and sequestosome-SQSTM1/p62 (p62), a ubiquitin-associated protein facilitating cargo recognition. Following nutrient starvation, we observed a smaller reduction in p62 in N2081D cells compared to N551K+R1398H cells, while all cells displayed a similar LC3-II ratio (stress/control) regardless of *LRRK2* genotype (**Fig. 5A**). Despite little change

in LC3-II, which is sometimes insensitive to autophagy alteration, a low response of p62 to stress suggested an impairment of cargo clearance. Finally, using a lysosome permeable fluorescent pH indicator (lysosensor), we compared lysosomal acidity, a key factor in autophagy, in response to stress, between the *LRRK2* N2081D and N551K mutant macrophages (**Fig. 5B**). We found that the relative change in mean fluorescent intensity following starvation, although varying among individuals, was decreased (alkaline) in risk N2081D carriers and increased (acidic) in carriers of the protective 551K+R1398H mutation (**Fig. 5B**). These data suggest that N2081D and N551K+R1398H mutations in CD patient macrophages have opposing effects on LRRK2 protein function that, in turn, can alter the autophagy-lysosome response to cellular stress.

Additive effects and phenotypic impact of LRRK2 variants

In contrast with the dominant effect of the G2019S variant in PD risk, we observed an additive effect of N2081D mutations on CD risk, as testing for dominant and recessive disease models did not show any increase in association statistical significance (**Supplementary Table S4**). To assess the strength of the combined effect across the *LRRK2* variants, we calculated additive burden scores (defined as the log sum of number of risk-conferring alleles carried by each individual, weighted by CD odds ratio – which is highly correlated with PD odds ratio as shown in **Fig. 2**) based upon their genotypes. The additive effects of the *LRRK2* risk alleles strongly correlated with both CD and PD risk (**Supplementary Fig. S6**), indicating an overall similar genetic architecture throughout the *LRRK2* locus underlying both diseases. There was no evidence of interaction effects between any of the nominally associated variants.

Moreover, because of a recent study implicating essential roles for both *NOD2* and *LRRK2* in proper lysozymal sorting in Paneth cells (*31*), a group of secretory cells in the ileum with a vital role in maintaining the function of the epithelial barrier, we next examined the effect of *LRRK2* N2081D risk alleles on CD disease location. While 80.5% of CD patients homozygous for the wild-type allele had ileal involvement, heterozygous and homozygous carriers of 2018D demonstrated ileal involvement in 86.1% and 90.9% of individuals, respectively (P=0.01, chi-square test, **Table 3**). Also, carriage of the 2081D allele was significantly associated with a younger age of onset (26.5 years for non-carriers, 24.6 for heterozygote carriers, and 20.8 for

homozygote carriers; P=0.002, linear regression). Neither *LRRK2* N551K nor R1398H showed any meaningful correlation with age of onset or ileal involvement (**Table 3**).

DISCUSSION

In this study, we performed exome sequencing followed by array-based exome chip genotyping in several independent cohorts of AJ CD cases and controls. Among protein-coding variants, in addition to the well-established *NOD2* and *IL23R* associations, we observed genome-wide significant associations to chromosome 12q12 S6N in *SLC2A13* and N2081D in *LRRK2* ($P<5x10^{-8}$), in high LD with each other ($r^2=0.91$), and an independent protective CD-association signal at *LRRK2* N551K. All previous GWAS association signals in or near *LRRK2*, including the common coding variant, M2397T (*16*), reported in one study to lower post-transcriptional LRRK2 protein (*18*), were significantly attenuated after conditioning on N2081D. Given the high LD between S6N in *SLC2A13* and N2081D in *LRRK2,* we applied co-expression approaches to define the likely contributing gene. In our Bayesian network analysis of IBD intestinal tissue, we observed a highly connected subnetwork with *LRRK2*, with no other genes within the chromosome 12q12 region, including *SLC2A13*, demonstrating similar connectivity. *SLC2A13*, solute carrier family 2 member 13, is a glucose transporter that is not expressed in the gut or the immune system and has not been previously linked to IBD, further suggesting that the observed 12q12 signal is driven by the *LRRK2* gene. Intriguingly, *LRRK2* was tightly linked with *GPR65*, where the IBD-associated risk allele, I231L, is associated with impaired lysosomal function (*24*) and *HLA-DPA1*, with variants in this locus linked to both IBD (*25*) and PD (*26*).

Notably, both *LRRK2* N2081D and N551K were also associated with PD in both AJ and NJ cohorts (**Table 2**). While previous reports have documented *LRRK2* N2081D conferring PD risk, and the N551K-R1398H-K1423K haplotype conferring protection (*12-15*), we now demonstrate that these specific non-synonymous variants in *LRRK2* genetically link CD to PD. Importantly, despite the same direction of the effect, the effect size for the risk variant N2081D was substantially higher for CD compared to PD (**Table 2**). Of interest, G2019S, the maximally-associated risk allele in PD (*32, 33*) occurring in the same domain as N2081D

(**Fig. 4A**) - though not in LD with it - showed suggestive association with CD in the AJ discovery cohort only. Further association analysis of independent common variants in >24,500 PD and CD cases and controls suggested additional extensive genetic pleiotropy between CD and PD within the extended *LRRK2* locus with a consistent pattern of correlated effect sizes (**Fig. 2**) in both AJ and NJ datasets. Intriguingly, a recent independent report has suggested that PD is associated with an increased risk of IBD (*34*). Taken together, these results point toward potential shared genetic and epidemiological links between these two diseases and can help identify a subgroup of patients with CD who are at a higher risk for developing PD.

Numerous functional roles for *LRRK2* have been reported, including vesicular trafficking and endocytosis, protein synthesis, immune response regulation, inflammation, and cytoskeleton homoeostasis, among others (*35*). In addition to their association with PD and CD risk, variations in the *LRRK2* locus have been also independently linked to excessive inflammatory responses in patients with leprosy (*36*) and risk of particular types of cancer (*37*). In the gastrointestinal tract of CD patients, *LRRK2* expression is restricted to *lamina propria* macrophages, dendritic cells and B-lymphocytes and is induced by interferon-γ, which is consistent with its role in IBD (*38*). A recent study has found high expression of *LRRK2* in Paneth cells in the ileum demonstrating that both *NOD2* and *LRRK2* are required for proper lysozyme sorting within Paneth cells (*31*). Our correlations of N2081D in *LRRK2* to earlier age of onset and ileal location mirror previously reported *NOD2* risk allele phenotypic correlations. Specifically, we showed that carriers of 2 copies for the risk allele at N2081D had almost a 6-year earlier age of onset compared to non-carriers and predominantly ileal disease involvement, which may be consistent with the recent report of the *LRRK2* effects in Paneth cells (*39*) that are exclusively located in the small intestine. These findings are of substantial clinical significance as a large recent phenotype-genotype analysis of all IBD associated loci identified only a handful of mutations, including in *NOD2*, having considerable effects on age of onset and disease location in CD; in that study, the *LRRK2* N2081D variant was not specifically tested (*40*). Defining altered Paneth cell function stratified on various *LRRK2* and *NOD2* genotype combinations should be a focus of future studies.

The majority of PD-causing mutations fall within the kinase and RocCOR domains, resulting in increased kinase activity or GTP-binding, leading to neurodegeneration. Our findings showed that both kinase

domain disease-associated mutations, G2019S (PD) and N2081D (CD) increased the phosphorylation of the LRRK2 substrate Rab10. Previous studies have reported that G2019S increases phosphorylation of several RAB-family members leading to an abnormal cytosol-membrane Rab protein distribution, which could result in the disruption of the process of autophagy (*27*). Consistent with this report, our studies in human monocyte-derived macrophages of CD patients carrying the N2081D mutation demonstrated faulty stress responses directly related to autophagy, including impaired autophagic cargo clearance, lysosomal acidification as well as defective tubulin acetylation, defects also characteristic of PD models (*41*).

Moreover, we also showed the link between the protective Roc domain R1398H mutation and an increase in GTPase activity (*42*). Importantly, although our statistical analysis prioritized the N551K mutation as significantly associated with a reduced risk of CD, in our biochemical analysis, N551K alone did not yield any detectable effect. Based on a high LD between N551K and R1398H mutations and the fact that all N551K human carriers that were analyzed also carried R1398H, we tested the combined effect of N551K+R1398H on GTPase activity and concluded that the actual physiological protective effect is driven by R1398H and not N551K. Of note, human macrophages from N551K+R1398H carriers also demonstrated an enhanced autophagy response to stress.

However, we speculate that the precise nature of the lysosomal alterations likely differs between these two diseases. That is, autosomal recessive mutations in the *GBA* (glucosylceramidase beta) gene, the most common lysosomal storage defect and the cause of Gaucher's disease, also prevalent in AJ populations, are highly associated with PD (with most cases involving dominant transmission), but in this study not found to be associated with CD. This would suggest that PD and CD pathophysiologies differ by cell-specific properties of the lysosome (neurons or glia versus inflammatory or Paneth cells, respectively), or with respect to distinct hydrolytic targets, namely glycolipids and bacterial peptidoglycan, respectively. Nevertheless, naturally occurring protective alleles, such as *LRRK2* R1398H, are of particular importance, as they define a desired functional effect for therapeutic development. Just as the loss-of-function, protective R381Q in IL23R would predict that blocking the IL-23 pathway would be safe and effective, our present findings suggest that targeting *LRRK2*-mediated signaling may be beneficial in the treatment of both CD and PD.

Among the study limitations is the fact that our CD cohorts were not explicitly screened for PD and vice versa, potentially allowing for the inclusion of individuals with both diseases in one disease category (either CD or PD). However, both CD and PD are relatively rare in the general population (~0.2% and ~1%, respectively) and misclassification of such patients would be expected to have minimal impact on any analyses. Also, we studied the AJ population given its higher CD prevalence, but this focus limited our cohort size and thus the power to identify new, rarer contributing alleles. Because the exome-sequencing phase of our study involved only 50 individuals, there are certainly many rare AJ-specific variants that were not tested in the association phases, and some of these likely play a role in CD pathogenesis. Finally, our Bayesian network analysis, while offering a method to examine gene function in an unbiased manner apart from disease association, did so indirectly and with only gene-expression data from whole tissue used to construct our network.

In summary, we have strongly implicated the contribution of *LRRK2* in CD risk through multiple complementary approaches, including genome-wide screening, Bayesian network analysis, genotype-phenotype correlations, and functional studies. *LRRK2* N2081 risk and N551K/R1398H protective alleles, as well as numerous other variants within the *LRRK2* locus, also revealed extended pleiotropy between CD and PD risk, providing a potential biological basis for clinical co-occurrence. Our findings may lead to new implications of *LRRK2* as a drug target.

MATERIALS AND METHODS

Study design

We first performed exome sequencing of 50 AJ CD individuals (44 independent individuals and 3 full-sibling pairs) (**Supplementary Materials**; **Supplementary Fig. S1**), having sufficient power to detect novel variants with MAF>0.015 (**Supplementary Table S1),** in order to catalog variation in the AJ population that may confer risk for CD (*43*). Because little AJ genetic variation was available from prior public genome sequencing, we sought to extend the coverage of available commercial genotyping platforms by adding novel variants detected in our exome sequencing results. In particular, we favored polymorphic sites that were less likely to be tagged in a previous well-powered genome-wide association study of CD in the AJ population. From these results, we selected 4,277 putatively high-yield novel mutations that were added to the base content of the Illumina HumanExome 1.0 array to create a semi-custom genotyping platform (**Supplementary Fig. S1** and **S2**, **Supplementary Table S2**), with which we performed discovery-phase genotyping and association analyses in 1,477 CD cases and 2,614 controls with full genetic AJ ancestry (*11*) (**Supplementary Fig. S3, Supplementary Table S3**), providing sufficient power to detect associations with modest effect sizes (**Supplementary Table S1**). The top association signals were then replicated in an independent cohort of589 CD cases and 1,019 controls, recruited throughout North America, Europe, and Israel (**Supplementary Table S3**). Disease diagnosis was confirmed using standard criteria as described elsewhere and full AJ ancestry was validated using principal components analysis (*11, 44*). Our second stage genetic association analysis included a total of 8,619 independent AJ and 16,401 independent NJ participants comprising CD cases, PD cases, and healthy controls, genotyped in previous studies (**Supplementary Table S3**) (*45, 46*). PD diagnoses were supported by standard UK Brain Bank criteria (*47*), with a modification to allow the inclusion of cases that had a family history of PD. We performed imputation of genotypes across diseases and within populations in order to allow direct comparison of genetic association at each site between CD and PD. We next conducted experimental validation studies for *LRRK2* N2081D and N551K/R1398H mutations using HEK293 cell lines and whole blood from human subjects enrolled in our prior studies,consented to be contacted for future research, and recalled based on their *LRRK2* genotype status. Four N551K+R1398 carriers and five N2081D

carriers were matched to five non-carriers, all with CD, for age, sex and disease severity. All experiments were performed in at least 3 biological replicates.

Statistical analysis

Genotyping quality control was performed following guidelines produced by the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) consortium (*48*). This procedure included removing samples with low quality metrics (genotype call rate < 0.96 and/or $p10_{GC}$ < 0.4125) and removing markers with overall low probe intensity. A subset of SNPs was subsequently excluded according to clustering criteria based on fluorescent probe intensities and genotype frequencies, as well as visual inspection of markers with uncertain genotyping quality.

Discovery and replication of novel variation associated with CD

We performed chi square-based association testing on all variants genotyped by the Exome chip. We tabulated all non-synonymous variants with P-values suggestive of CD association ($P<2x10^{-5}$), a threshold we estimated using Bonferroni correction with the approximate number of polymorphic variants genotyped using our platform (**Supplementary Fig. S1**), allowing for strong and widespread correlation among exomic variants (i.e. “chip-wide significance”). We collected genotypes at these markers in independent case and control cohorts with full AJ ancestry (**Supplementary Fig. S3**; **Supplementary Table S4**). These replication data were combined with those generated by Exome chip genotyping for a meta-analysis using the METAL program with default parameters (*49*); coding variation with genome-wide significant P-values ($P<5x10^{-8}$) are presented as positive association signals (**Table 1**).

Imputation-based comparative analysis of CD and PD

Additional NJ CD and PD and AJ PD datasets were added to the AJ CD data (imputation cohorts, **Supplementary Table S3**), and reference-free imputation using MACH was performed in order to facilitate

direct comparisons across groups at specific variants(*50*). Both unconditioned and conditional analyses were conducted using logistic regression on pooled empiric (directly genotyped) and probabilistic (imputed) genotypes.

Network analysis

We constructed an adult IBD Bayesian network, using previously described methodology (*20*), from gene expression data generated on 203 intestinal biopsies that included ileum, ascending colon, descending colon and transverse colon, and inflamed and non-inflamed sigmoid and rectum, all collected at baseline from 54 anti-TNFα resistant CD patients enrolled in the Ustekinumab (anti-IL12/IL23) clinical trial (*21*) with the goal to project these genes onto the intestinal network and identify co-expressed genes that act together. This type of probabilistic causal network structure has previously been demonstrated to represent biologically functional pathways across a broad range of diseases including obesity and diabetes (*20, 51-53*), asthma and COPD (*54, 55*), and Alzheimer's disease (*56*). We next excluded genes previously associated with PD (*23*), including *LRRK2*, as well as genes within 1 Mb of *LRRK2* to see whether either *LRRK2* or other genes will be "recovered" by the network as being co-expressed with the IBD-associated genes. We then identified the largest connected sub-graph from the set of IBD-associated genes projected onto the network. To focus on pathways potentially relevant to CD pathogenesis, we removed from our analysis all genes more than two edge lengths away from any of these IBD-associated genes.

Experimental studies

All experimental values represent mean±standard error, and significance was calculated by ANOVA, mixed model ANOVA with a random effect of a biological sample or order-constrained ANOVA (*57*).

RAB10 In Vitro Kinase Assay

LRRK2 was incubated with Rab10 or inhibitor for 30 min incubation on ice in 30uL kinase buffer (20mM Tris pH 7.5, 1mM DTT, 15mM MnCl 2 , 20mM β-glycerophosphate). Reactions were initiated by adding 50μM cold ATP. After 30 minutes at 37˚C, reactions were stopped by addition of Laemmli buffer and boiling at 95˚C for 10 minutes. Samples were resolved on 4-12% SDS-PAGE pre-cast gels (Invitrogen, Madison, WI, USA). Samples were then subjected to Western blot, using anti-Rab 10 (Cell Signaling, #4262) and anti-pT73 Rab10 (University of Dundee, UK). Licor imaging was used to detect phospho- and total Rab10 on the same membrane and Image Studio Lite was used for quantification.

GTP Hydrolysis Assay

GTPase activity of LRRK2 was measured in 30uL GTPase buffer (20mM Tris pH 7.5, 150mM NaCl, 1mM DTT, 5mM $MgCl_2$, 1mM EDTA) at 30˚C for 90 minutes, where the reaction rate is still in a linear phase as previously established , allowing for quantification by densitometry (*29*). Reactions were initiated with the addition of 50μM cold GTP and [α-^{32}P]GTP (3000Ci/mmol; PerkinElmer Life Sciences, Waltham, MA). Reactions were terminated by adding 0.5M EDTA. 2uL of the reaction mixture were dotted onto Thin-Layer Chromatography (TLC) plates (EMD Millipore, Darmstadt, Germany) and GDP and GTP were separated by TLC using 0.5M KH_2PO_4 pH 3.5 for 60 minutes. The TLC plate was dried for 15 minutes and radioactive signal was captured using a phosphor-screen (GE Lifesciences, Pittsburgh, PA, USA) and a Typhoon scanner. ImageQuant densitometry was used to quantify the phosphor-signal.

Autophagy studies in human samples

M1-macrophages from CD patients were derived from whole peripheral blood monocytes according to the manufacturer's instructions (Promocell, Heidelberg, Germany). Monocytes were polarized to mature M1-macrophages in the DXF M1-macrophage generation medium (M1-medium, resting condition, Promocell) for 12 days and then incubated in PBS and M1 medium for 45 minutes. Cells were then lysed and 10 micrograms of total protein were loaded onto 4-12% Bis-Tris Plus precast SDS-polyacrylamide gels, transferred onto a PVDF membrane and probed with primary rabbit anti-LRRK2 antibody (ab133474, abcam), mouse anti-

acetylated alpha-tubulin (T7451, Sigma-Aldrich, St. Louis, MO), rabbit anti-alpha tubulin (ab4074, abcam), mouse anti-SQSTM1 (sc-28359, Santa Cruz Biotechnology), and rabbit anti-LC3B (NB100-2220, Novus Biologicals). The corresponding HRP-conjugated secondary antibody was applied for detection. Total alpha-tubulin was used as a loading control for normalization and protein densitometry was performed using ImageJ software. LRRK2 degradation was assessed as the ratio of degraded LRRK2 to total LRRK2 (full length + degraded) protein. Alpha-tubulin acetylation was assessed as the ratio of acetylated to total alpha-tubulin.

Next, M1 macrophages (1x10^5 cells per experiment), in M1-medium and PBS, were pulsed with lysosensor green DND-189 (L-7535, Life Technologies) for 45 minutes (*58*). Antibodies for cell surface markers were added and cells incubated for 30 minutes at 4°C. After staining, the cells were washed and analyzed on a CANTOII (BD) multi-parameter flow cytometer and data were analyzed using FlowJo software (Tree Star). A fluorescence minus one (FMO) was used for the FITC lysosensor control samples. The fluorescent ratio was calculated between PBS and M1-medium and compared by the *LRRK2* genotype.

SUPPLEMENTARY MATERIALS

Supplementary Materials and Methods

Supplementary Figures:

Supplementary Tables:

REFERENCES

1. L. Jostins, S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E. Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J. P. Achkar, T. Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo, T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. Buning, A. Cohain, S. Cichon, M. D'Amato, D. De Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D. Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C. Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L. Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E. Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y. Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D. Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R. Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. De Vos, C. Wijmenga, D. C. Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, M. S. Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J. D. Rioux, E. E. Schadt, M. J. Daly, A. Franke, M. Parkes, S. Vermeire, J. C. Barrett, J. H. Cho, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).
2. J. Z. Liu, S. van Sommeren, H. Huang, S. C. Ng, R. Alberts, A. Takahashi, S. Ripke, J. C. Lee, L. Jostins, T. Shah, S. Abedian, J. H. Cheon, J. Cho, N. E. Daryani, L. Franke, Y. Fuyuno, A. Hart, R. C. Juyal, G. Juyal, W. H. Kim, A. P. Morris, H. Poustchi, W. G. Newman, V. Midha, T. R. Orchard, H. Vahedi, A. Sood, J. J. Sung, R. Malekzadeh, H. J. Westra, K. Yamazaki, S. K. Yang, C. International Multiple Sclerosis Genetics, I. B. D. G. C. International, J. C. Barrett, A. Franke, B. Z. Alizadeh, M. Parkes, K. T. B, M. J. Daly, M. Kubo, C. A. Anderson, R. K. Weersma, I. B. D. G. Consortium, Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).
3. R. H. Duerr, K. D. Taylor, S. R. Brant, J. D. Rioux, M. S. Silverberg, M. J. Daly, A. H. Steinhart, C. Abraham, M. Regueiro, A. Griffiths, T. Dassopoulos, A. Bitton, H. Yang, S. Targan, L. W. Datta, E. O. Kistner, L. P. Schumm, A. T. Lee, P. K. Gregersen, M. M. Barmada, J. I. Rotter, D. L. Nicolae, J. H. Cho, A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**, 1461-1463 (2006).
4. S. Singh, R. R. Kroe-Barrett, K. A. Canada, X. Zhu, E. Sepulveda, H. Wu, Y. He, E. L. Raymond, J. Ahlberg, L. E. Frego, L. M. Amodeo, K. M. Catron, D. H. Presky, J. H. Hanke, Selective targeting of the IL23 pathway: Generation and characterization of a novel high-affinity humanized anti-IL23A antibody. *mAbs* **7**, 778-791 (2015).
5. A. Murthy, Y. Li, I. Peng, M. Reichelt, A. K. Katakam, R. Noubade, M. Roose-Girma, J. DeVoss, L. Diehl, R. R. Graham, M. van Lookeren Campagne, A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature* **506**, 456-462 (2014).
6. S. A. McCarroll, A. Huett, P. Kuballa, S. D. Chilewski, A. Landry, P. Goyette, M. C. Zody, J. L. Hall, S. R. Brant, J. H. Cho, R. H. Duerr, M. S. Silverberg, K. D. Taylor, J. D. Rioux, D. Altshuler, M. J. Daly, R. J. Xavier, Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* **40**, 1107-1112 (2008).
7. P. Brest, P. Lapaquette, M. Souidi, K. Lebrigand, A. Cesaro, V. Vouret-Craviari, B. Mari, P. Barbry, J. F. Mosnier, X. Hebuterne, A. Harel-Bellan, B. Mograbi, A. Darfeuille-Michaud, P. Hofman, A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet* **43**, 242-245 (2011).
8. C. N. Bernstein, P. Rawsthorne, M. Cheang, J. F. Blanchard, A population-based case control study of potential risk factors for IBD. *Am J Gastroenterol* **101**, 993-1002 (2006).
9. H. Yang, C. McElree, M. P. Roth, F. Shanahan, S. R. Targan, J. I. Rotter, Familial empirical risks for inflammatory bowel disease: differences between Jews and non-Jews. *Gut* **34**, 517-524 (1993).
10. C. Paisan-Ruiz, S. Jain, E. W. Evans, W. P. Gilks, J. Simon, M. van der Brug, A. Lopez de Munain, S. Aparicio, A. M. Gil, N. Khan, J. Johnson, J. R. Martinez, D. Nicholl, I. M. Carrera, A. S. Pena, R. de Silva, A. Lees, J. F. Marti-Masso, J. Perez-Tur, N. W. Wood, A. B. Singleton, Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**, 595-600 (2004).
11. E. E. Kenny, I. Pe'er, A. Karban, L. Ozelius, A. A. Mitchell, S. M. Ng, M. Erazo, H. Ostrer, C. Abraham, M. T. Abreu, G. Atzmon, N. Barzilai, S. R. Brant, S. Bressman, E. R. Burns, Y. Chowers, L. N. Clark, A. Darvasi, D. Doheny, R. H. Duerr, R. Eliakim, N. Giladi, P. K. Gregersen, H. Hakonarson, M. R. Jones, K. Marder, D. P. McGovern, J. Mulle, A. Orr-Urtreger, D. D. Proctor, A. Pulver, J. I. Rotter, M. S. Silverberg, T. Ullman, S. T. Warren, M. Waterman, W. Zhang, A. Bergman, L. Mayer, S. Katz, R. J. Desnick, J. H. Cho, I. Peter, A genome-wide scan of ashkenazi jewish Crohn's disease suggests novel susceptibility Loci. *PLoS Genet* **8**, e1002559 (2012).

12. O. A. Ross, A. I. Soto-Ortolaza, M. G. Heckman, J. O. Aasly, N. Abahuni, G. Annesi, J. A. Bacon, S. Bardien, M. Bozi, A. Brice, L. Brighina, C. Van Broeckhoven, J. Carr, M. C. Chartier-Harlin, E. Dardiotis, D. W. Dickson, N. N. Diehl, A. Elbaz, C. Ferrarese, A. Ferraris, B. Fiske, J. M. Gibson, R. Gibson, G. M. Hadjigeorgiou, N. Hattori, J. P. Ioannidis, B. Jasinska-Myga, B. S. Jeon, Y. J. Kim, C. Klein, R. Kruger, E. Kyratzi, S. Lesage, C. H. Lin, T. Lynch, D. M. Maraganore, G. D. Mellick, E. Mutez, C. Nilsson, G. Opala, S. S. Park, A. Puschmann, A. Quattrone, M. Sharma, P. A. Silburn, Y. H. Sohn, L. Stefanis, V. Tadic, J. Theuns, H. Tomiyama, R. J. Uitti, E. M. Valente, S. van de Loo, D. K. Vassilatis, C. Vilarino-Guell, L. R. White, K. Wirdefeldt, Z. K. Wszolek, R. M. Wu, M. J. Farrer, C. Genetic Epidemiology Of Parkinson's Disease, Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study. *The Lancet. Neurology* **10**, 898-908 (2011).
13. A. Gorostidi, J. F. Marti-Masso, A. Bergareche, M. C. Rodriguez-Oroz, A. Lopez de Munain, J. Ruiz-Martinez, Genetic Mutation Analysis of Parkinson's Disease Patients Using Multigene Next-Generation Sequencing Panels. *Molecular diagnosis & therapy*, (2016).
14. S. Biskup, A. B. West, Zeroing in on LRRK2-linked pathogenic mechanisms in Parkinson's disease. *Biochim Biophys Acta* **1792**, 625-633 (2009).
15. B. A. Benitez, A. A. Davis, S. C. Jin, L. Ibanez, S. Ortega-Cubero, P. Pastor, J. Choi, B. Cooper, J. S. Perlmutter, C. Cruchaga, Resequencing analysis of five Mendelian genes and the top genes from genome-wide association studies in Parkinson's Disease. *Molecular neurodegeneration* **11**, 29 (2016).
16. J. C. Barrett, S. Hansoul, D. L. Nicolae, J. H. Cho, R. H. Duerr, J. D. Rioux, S. R. Brant, M. S. Silverberg, K. D. Taylor, M. M. Barmada, A. Bitton, T. Dassopoulos, L. W. Datta, T. Green, A. M. Griffiths, E. O. Kistner, M. T. Murtha, M. D. Regueiro, J. I. Rotter, L. P. Schumm, A. H. Steinhart, S. R. Targan, R. J. Xavier, C. Libioulle, C. Sandor, M. Lathrop, J. Belaiche, O. Dewit, I. Gut, S. Heath, D. Laukens, M. Mni, P. Rutgeerts, A. Van Gossum, D. Zelenika, D. Franchimont, J. P. Hugot, M. de Vos, S. Vermeire, E. Louis, L. R. Cardon, C. A. Anderson, H. Drummond, E. Nimmo, T. Ahmad, N. J. Prescott, C. M. Onnie, S. A. Fisher, J. Marchini, J. Ghori, S. Bumpstead, R. Gwilliam, M. Tremelling, P. Deloukas, J. Mansfield, D. Jewell, J. Satsangi, C. G. Mathew, M. Parkes, M. Georges, M. J. Daly, Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-962 (2008).
17. A. Franke, D. P. McGovern, J. C. Barrett, K. Wang, G. L. Radford-Smith, T. Ahmad, C. W. Lees, T. Balschun, J. Lee, R. Roberts, C. A. Anderson, J. C. Bis, S. Bumpstead, D. Ellinghaus, E. M. Festen, M. Georges, T. Green, T. Haritunians, L. Jostins, A. Latiano, C. G. Mathew, G. W. Montgomery, N. J. Prescott, S. Raychaudhuri, J. I. Rotter, P. Schumm, Y. Sharma, L. A. Simms, K. D. Taylor, D. Whiteman, C. Wijmenga, R. N. Baldassano, M. Barclay, T. M. Bayless, S. Brand, C. Buning, A. Cohen, J. F. Colombel, M. Cottone, L. Stronati, T. Denson, M. De Vos, R. D'Inca, M. Dubinsky, C. Edwards, T. Florin, D. Franchimont, R. Gearry, J. Glas, A. Van Gossum, S. L. Guthery, J. Halfvarson, H. W. Verspaget, J. P. Hugot, A. Karban, D. Laukens, I. Lawrance, M. Lemann, A. Levine, C. Libioulle, E. Louis, C. Mowat, W. Newman, J. Panes, A. Phillips, D. D. Proctor, M. Regueiro, R. Russell, P. Rutgeerts, J. Sanderson, M. Sans, F. Seibold, A. H. Steinhart, P. C. Stokkers, L. Torkvist, G. Kullak-Ublick, D. Wilson, T. Walters, S. R. Targan, S. R. Brant, J. D. Rioux, M. D'Amato, R. K. Weersma, S. Kugathasan, A. M. Griffiths, J. C. Mansfield, S. Vermeire, R. H. Duerr, M. S. Silverberg, J. Satsangi, S. Schreiber, J. H. Cho, V. Annese, H. Hakonarson, M. J. Daly, M. Parkes, Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* **42**, 1118-1125 (2010).
18. Z. Liu, J. Lee, S. Krummey, W. Lu, H. Cai, M. J. Lenardo, The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. *Nat Immunol* **12**, 1063-1070 (2011).
19. A. Zimprich, S. Biskup, P. Leitner, P. Lichtner, M. Farrer, S. Lincoln, J. Kachergus, M. Hulihan, R. J. Uitti, D. B. Calne, A. J. Stoessl, R. F. Pfeiffer, N. Patenge, I. C. Carbajal, P. Vieregge, F. Asmus, B. Muller-Myhsok, D. W. Dickson, T. Meitinger, T. M. Strom, Z. K. Wszolek, T. Gasser, Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601-607 (2004).
20. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellini, S. Wang, M. F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429-435 (2008).
21. W. J. Sandborn, C. Gasink, L. L. Gao, M. A. Blank, J. Johanns, C. Guzzo, B. E. Sands, S. B. Hanauer, S. Targan, P. Rutgeerts, S. Ghosh, W. J. de Villiers, R. Panaccione, G. Greenberg, S. Schreiber, S. Lichtiger, B. G. Feagan, C. S. Group, Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N Engl J Med* **367**, 1519-1528 (2012).
22. Q. Li, C. H. Lee, L. A. Peters, L. A. Mastropaolo, C. Thoeni, A. Elkadri, T. Schwerd, J. Zhu, B. Zhang, Y. Zhao, K. Hao, A. Dinarzo, G. Hoffman, B. A. Kidd, R. Murchie, Z. Al Adham, C. Guo, D. Kotlarz, E. Cutz, T. D. Walters, D. S. Shouval, M. Curran, R. Dobrin, C. Brodmerkel, S. B. Snapper, C. Klein, J. H. Brumell, M. Hu, R. Nanan, B.

Snanter-Nanan, M. Wong, F. Le Deist, E. Haddad, C. M. Roifman, C. Deslandres, A. M. Griffiths, K. J. Gaskin, H. H. Uhlig, E. E. Schadt, A. M. Muise, Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. *Gastroenterology* **150**, 1196-1207 (2016).

23. J. Farlow, N. D. Pankratz, J. Wojcieszek, T. Foroud, in *SourceGeneReviews® [Internet],* A. M. Pagon RA, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors, Ed. (University of Washington, Seattle, WA, 2004 May 25 [updated 2014 Feb 27]).
24. K. G. Lassen, C. I. McKenzie, M. Mari, T. Murano, J. Begun, L. A. Baxt, G. Goel, E. J. Villablanca, S. Y. Kuo, H. Huang, L. Macia, A. K. Bhan, M. Batten, M. J. Daly, F. Reggiori, C. R. Mackay, R. J. Xavier, Genetic Coding Variant in GPR65 Alters Lysosomal pH and Links Lysosomal Dysfunction with Colitis Risk. *Immunity* **44**, 1392-1405 (2016).
25. P. Goyette, G. Boucher, D. Mallon, E. Ellinghaus, L. Jostins, H. Huang, S. Ripke, E. S. Gusareva, V. Annese, S. L. Hauser, J. R. Oksenberg, I. Thomsen, S. Leslie, C. International Inflammatory Bowel Disease Genetics, Australia, I. New Zealand, I. B. D. G. C. Belgium, I. B. D. G. C. Italian Group for, N. I. B. D. G. Consortium, I. United Kingdom, C. Wellcome Trust Case Control, I. B. D. G. C. Quebec, M. J. Daly, K. Van Steen, R. H. Duerr, J. C. Barrett, D. P. McGovern, L. P. Schumm, J. A. Traherne, M. N. Carrington, V. Kosmoliaptsis, T. H. Karlsen, A. Franke, J. D. Rioux, High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet* **47**, 172-179 (2015).
26. R. Ferrari, Y. Wang, J. Vandrovcova, S. Guelfi, A. Witeolar, C. M. Karch, A. J. Schork, C. C. Fan, J. B. Brewer, F. T. D. G. C. International, C. International Parkinson's Disease Genomics, P. International Genomics of Alzheimer's, P. Momeni, G. D. Schellenberg, W. P. Dillon, L. P. Sugrue, C. P. Hess, J. S. Yokoyama, L. W. Bonham, G. D. Rabinovici, B. L. Miller, O. A. Andreassen, A. M. Dale, J. Hardy, R. S. Desikan, Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *Journal of neurology, neurosurgery, and psychiatry* **88**, 152-164 (2017).
27. M. Steger, F. Tonelli, G. Ito, P. Davies, M. Trost, M. Vetter, S. Wachter, E. Lorentzen, G. Duddy, S. Wilson, M. A. Baptista, B. K. Fiske, M. J. Fell, J. A. Morrow, A. D. Reith, D. R. Alessi, M. Mann, Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *eLife* **5**, (2016).
28. Y. Xiong, V. L. Dawson, T. M. Dawson, LRRK2 GTPase dysfunction in the pathogenesis of Parkinson's disease. *Biochem Soc Trans* **40**, 1074-1079 (2012).
29. X. Li, Y. C. Tan, S. Poulose, C. W. Olanow, X. Y. Huang, Z. Yue, Leucine-rich repeat kinase 2 (LRRK2)/PARK8 possesses GTPase activity that is altered in familial Parkinson's disease R1441C/G mutants. *J Neurochem* **103**, 238-247 (2007).
30. R. Mackeh, S. Lorin, A. Ratier, N. Mejdoubi-Charef, A. Baillet, A. Bruneel, A. Hamai, P. Codogno, C. Pous, D. Perdiz, Reactive oxygen species, AMP-activated protein kinase, and the transcription cofactor p300 regulate alpha-tubulin acetyltransferase-1 (alphaTAT-1/MEC-17)-dependent microtubule hyperacetylation during cell stress. *J Biol Chem* **289**, 11816-11828 (2014).
31. Q. Zhang, Y. Pan, R. Yan, B. Zeng, H. Wang, X. Zhang, W. Li, H. Wei, Z. Liu, Commensal bacteria direct selective cargo sorting to promote symbiosis. *Nat Immunol* **16**, 918-926 (2015).
32. J. Kachergus, I. F. Mata, M. Hulihan, J. P. Taylor, S. Lincoln, J. Aasly, J. M. Gibson, O. A. Ross, T. Lynch, J. Wiley, H. Payami, J. Nutt, D. M. Maraganore, K. Czyzewski, M. Styczynska, Z. K. Wszolek, M. J. Farrer, M. Toft, Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. *Am J Hum Genet* **76**, 672-680 (2005).
33. L. J. Ozelius, G. Senthil, R. Saunders-Pullman, E. Ohmann, A. Deligtisch, M. Tagliati, A. L. Hunt, C. Klein, B. Henick, S. M. Hailpern, R. B. Lipton, J. Soto-Valencia, N. Risch, S. B. Bressman, LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. *N Engl J Med* **354**, 424-425 (2006).
34. J. C. Lin, C. S. Lin, C. W. Hsu, C. L. Lin, C. H. Kao, Association Between Parkinson's Disease and Inflammatory Bowel Disease: a Nationwide Taiwanese Retrospective Cohort Study. *Inflamm Bowel Dis*, (2016).
35. R. Wallings, C. Manzoni, R. Bandopadhyay, Cellular processes associated with LRRK2 function and dysfunction. *The FEBS journal* **282**, 2806-2826 (2015).
36. V. M. Fava, J. Manry, A. Cobat, M. Orlova, N. Van Thuc, N. N. Ba, V. H. Thai, L. Abel, A. Alcais, E. Schurr, T. Canadian Lrrk2 in Inflammation, A Missense LRRK2 Variant Is a Risk Factor for Excessive Inflammatory Responses in Leprosy. *PLoS neglected tropical diseases* **10**, e0004412 (2016).
37. M. J. Devine, H. Plun-Favreau, N. W. Wood, Parkinson's disease and cancer: two wars, one front. *Nat Rev Cancer* **11**, 812-823 (2011).
38. A. Gardet, Y. Benita, C. Li, B. E. Sands, I. Ballester, C. Stevens, J. R. Korzenik, J. D. Rioux, M. J. Daly, R. J. Xavier, D. K. Podolsky, LRRK2 is involved in the IFN-gamma response and host response to pathogens. *J Immunol* **185**, 5577-5585 (2010).

39. T. C. Liu, T. Naito, Z. Liu, K. L. VanDussen, T. Haritunians, D. Li, K. Endo, Y. Kawai, M. Nagasaki, Y. Kinouchi, D. P. B. McGovern, T. Shimosegawa, Y. Kakuta, T. S. Stappenbeck, LRRK2 but not ATG16L1 is associated with Paneth cell defect in Japanese Crohn's disease patients. *JCI Insight, in press*, (2017).
40. I. Cleynen, G. Boucher, L. Jostins, L. P. Schumm, S. Zeissig, T. Ahmad, V. Andersen, J. M. Andrews, V. Annese, S. Brand, S. R. Brant, J. H. Cho, M. J. Daly, M. Dubinsky, R. H. Duerr, L. R. Ferguson, A. Franke, R. B. Gearry, P. Goyette, H. Hakonarson, J. Halfvarson, J. R. Hov, H. Huang, N. A. Kennedy, L. Kupcinskas, I. C. Lawrance, J. C. Lee, J. Satsangi, S. Schreiber, E. Theatre, A. E. van der Meulen-de Jong, R. K. Weersma, D. C. Wilson, C. International Inflammatory Bowel Disease Genetics, M. Parkes, S. Vermeire, J. D. Rioux, J. Mansfield, M. S. Silverberg, G. Radford-Smith, D. P. McGovern, J. C. Barrett, C. W. Lees, Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* **387**, 156-167 (2016).
41. A. R. Esteves, S. M. Cardoso, LRRK2 at the Crossroad Between Autophagy and Microtubule Trafficking: Insights into Parkinson's Disease. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, (2016).
42. J. Nixon-Abell, D. C. Berwick, S. Granno, V. A. Spain, C. Blackstone, K. Harvey, Protective LRRK2 R1398H Variant Enhances GTPase and Wnt Signaling Activity. *Frontiers in molecular neuroscience* **9**, 18 (2016).
43. W. Zhang, K. Y. Hui, A. Gusev, N. Warner, S. M. Ng, J. Ferguson, M. Choi, A. Burberry, C. Abraham, L. Mayer, R. J. Desnick, C. J. Cardinale, H. Hakonarson, M. Waterman, Y. Chowers, A. Karban, S. R. Brant, M. S. Silverberg, P. K. Gregersen, S. Katz, R. P. Lifton, H. Zhao, G. Nunez, I. Pe'er, I. Peter, J. H. Cho, Extended haplotype association study in Crohn's disease identifies a novel, Ashkenazi Jewish-specific missense mutation in the NF-kappaB pathway gene, HEATR3. *Genes Immun*, (2013).
44. I. Peter, A. A. Mitchell, L. Ozelius, M. Erazo, J. Hu, D. Doheny, M. T. Abreu, D. H. Present, T. Ullman, K. Benkov, B. I. Korelitz, L. Mayer, R. J. Desnick, Evaluation of 22 genetic variants with Crohn's Disease risk in the Ashkenazi Jewish population: a case-control study. *BMC Med Genet* **12**, 63 (2011).
45. M. A. Nalls, N. Pankratz, C. M. Lill, C. B. Do, D. G. Hernandez, M. Saad, A. L. DeStefano, E. Kara, J. Bras, M. Sharma, C. Schulte, M. F. Keller, S. Arepalli, C. Letson, C. Edsall, H. Stefansson, X. Liu, H. Pliner, J. H. Lee, R. Cheng, C. International Parkinson's Disease Genomics, G. I. Parkinson's Study Group Parkinson's Research: The Organized, andMe, GenePd, C. NeuroGenetics Research, G. Hussman Institute of Human, I. Ashkenazi Jewish Dataset, H. Cohorts for, E. Aging Research in Genetic, C. North American Brain Expression, C. United Kingdom Brain Expression, C. Greek Parkinson's Disease, G. Alzheimer Genetic Analysis, M. A. Ikram, J. P. Ioannidis, G. M. Hadjigeorgiou, J. C. Bis, M. Martinez, J. S. Perlmutter, A. Goate, K. Marder, B. Fiske, M. Sutherland, G. Xiromerisiou, R. H. Myers, L. N. Clark, K. Stefansson, J. A. Hardy, P. Heutink, H. Chen, N. W. Wood, H. Houlden, H. Payami, A. Brice, W. K. Scott, T. Gasser, L. Bertram, N. Eriksson, T. Foroud, A. B. Singleton, Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**, 989-993 (2014).
46. V. Vacic, L. J. Ozelius, L. N. Clark, A. Bar-Shira, M. Gana-Weisz, T. Gurevich, A. Gusev, M. Kedmi, E. E. Kenny, X. Liu, H. Mejia-Santana, A. Mirelman, D. Raymond, R. Saunders-Pullman, R. J. Desnick, G. Atzmon, E. R. Burns, H. Ostrer, H. Hakonarson, A. Bergman, N. Barzilai, A. Darvasi, I. Peter, S. Guha, T. Lencz, N. Giladi, K. Marder, I. Pe'er, S. B. Bressman, A. Orr-Urtreger, Genome-wide mapping of IBD segments in an Ashkenazi PD cohort identifies associated haplotypes. *Hum Mol Genet* **23**, 4693-4702 (2014).
47. A. J. Hughes, S. E. Daniel, L. Kilford, A. J. Lees, Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of neurology, neurosurgery, and psychiatry* **55**, 181-184 (1992).
48. M. L. Grove, B. Yu, B. J. Cochran, T. Haritunians, J. C. Bis, K. D. Taylor, M. Hansen, I. B. Borecki, L. A. Cupples, M. Fornage, V. Gudnason, T. B. Harris, S. Kathiresan, R. Kraaij, L. J. Launer, D. Levy, Y. Liu, T. Mosley, G. M. Peloso, B. M. Psaty, S. S. Rich, F. Rivadeneira, D. S. Siscovick, A. V. Smith, A. Uitterlinden, C. M. van Duijn, J. G. Wilson, C. J. O'Donnell, J. I. Rotter, E. Boerwinkle, Best Practices and Joint Calling of the HumanExome BeadChip: The CHARGE Consortium. *PLoS ONE* **8**, e68095 (2013).
49. C. J. Willer, Y. Li, G. R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
50. Y. Li, C. J. Willer, J. Ding, P. Scheet, G. R. Abecasis, MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-834 (2010).
51. V. Emilsson, G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink, J. Zhu, S. Carlson, A. Helgason, G. B. Walters, S. Gunnarsdottir, M. Mouy, V. Steinthorsdottir, G. H. Eiriksdottir, G. Bjornsdottir, I. Reynisdottir, D. Gudbjartsson, A. Helgadottir, A. Jonasdottir, A. Jonasdottir, U. Styrkarsdottir, S. Gretarsdottir, K. P. Magnusson, H. Stefansson, R. Fossdal, K. Kristjansson, H. G. Gislason, T. Stefansson, B. G. Leifsson, U. Thorsteinsdottir, J. R. Lamb, J. R. Gulcher, M. L. Reitman, A. Kong, E. E. Schadt, K. Stefansson, Genetics of gene expression and its effect on disease. *Nature* **452**, 423-428 (2008).

52. H. Zhong, J. Beaulaurier, P. Y. Lum, C. Molony, X. Yang, D. J. Macneil, D. T. Weingarth, B. Zhang, D. Greenawalt, R. Dobrin, K. Hao, S. Woo, C. Fabre-Suver, S. Qian, M. R. Tota, M. P. Keller, C. M. Kendziorski, B. S. Yandell, V. Castro, A. D. Attie, L. M. Kaplan, E. E. Schadt, Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* **6**, e1000932 (2010).
53. H. Zhong, X. Yang, L. M. Kaplan, C. Molony, E. E. Schadt, Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet* **86**, 581-591 (2010).
54. S. Bunyavanich, E. E. Schadt, B. E. Himes, J. Lasky-Su, W. Qiu, R. Lazarus, J. P. Ziniti, A. Cohain, M. Linderman, D. G. Torgerson, C. S. Eng, M. Pino-Yanes, B. Padhukasahasram, J. J. Yang, R. A. Mathias, T. H. Beaty, X. Li, P. Graves, I. Romieu, R. Navarro Bdel, M. T. Salam, H. Vora, D. L. Nicolae, C. Ober, F. D. Martinez, E. R. Bleecker, D. A. Meyers, W. J. Gauderman, F. Gilliland, E. G. Burchard, K. C. Barnes, L. K. Williams, S. J. London, B. Zhang, B. A. Raby, S. T. Weiss, Integrated genome-wide association, coexpression network, and expression single nucleotide polymorphism analysis identifies novel pathway in allergic rhinitis. *BMC Med Genomics* **7**, 48 (2014).
55. S. Yoo, S. Takikawa, P. Geraghty, C. Argmann, J. Campbell, L. Lin, T. Huang, Z. Tu, R. Feronjy, A. Spira, E. E. Schadt, C. A. Powell, J. Zhu, Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. *PLoS Genet* **11**, e1004898 (2015).
56. B. Zhang, C. Gaiteri, L. G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezhnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin, E. Fluder, B. Clurman, S. Melquist, M. Narayanan, C. Suver, H. Shah, M. Mahajan, T. Gillis, J. Mysore, M. E. MacDonald, J. R. Lamb, D. A. Bennett, C. Molony, D. J. Stone, V. Gudnason, A. J. Myers, E. E. Schadt, H. Neumann, J. Zhu, V. Emilsson, Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707-720 (2013).
57. R. Kuiper, I. Klugkist, H. Hoijtink, A Fortran 90 program for confirmatory analysis of variance. *J. Stat. Softw* **34**, 1-31 (2010).
58. M. H. Sahani, E. Itakura, N. Mizushima, Expression of the autophagy substrate SQSTM1/p62 is restored during prolonged starvation depending on transcriptional upregulation and autophagy-derived amino acids. *Autophagy* **10**, 431-441 (2014).
59. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
60. A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, W. J. Kent, The UCSC Genome Browser Database: update 2006. *Nucleic acids research* **34**, D590-598 (2006).
61. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).
62. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* **14**, 178-192 (2013).
63. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-1097 (2007).
64. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
65. L. S. Chuang, N. Villaverde, K. Y. Hui, A. Mortha, A. Rahman, A. P. Levine, T. Haritunians, S. M. Ng, W. Zhang, N. Y. Hsu, J. A. Facey, T. Luong, H. Fernandez-Hernandez, D. Li, M. Rivas, E. R. Schiff, A. Gusev, L. P. Schumm, B. M. Bowen, Y. Sharma, K. Ning, R. Remark, S. Gnjatic, P. Legnani, J. George, B. E. Sands, J. M. Stempak, L. W. Datta, S. Lipka, S. Katz, A. S. Cheifetz, N. Barzilai, N. Pontikos, C. Abraham, M. J. Dubinsky, S. Targan, K. Taylor, J. I. Rotter, E. J. Scherl, R. J. Desnick, M. T. Abreu, H. Zhao, G. Atzmon, I. Pe'er, S. Kugathasan, H. Hakonarson, J. L. McCauley, T. Lencz, A. Darvasi, V. Plagnol, M. S. Silverberg, A. M. Muise, S. R. Brant, M. J. Daly, A. W. Segal, R. H. Duerr, M. Merad, D. P. McGovern, I. Peter, J. H. Cho, A Frameshift in CSF2RB Predominant Among Ashkenazi Jews Increases Risk for Crohn's Disease and Reduces Monocyte Signaling via GMCSF. *Gastroenterology*, (2016).
66. C. Genomes Project, G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, G. A. McVean, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

67. S. Carmi, K. Y. Hui, E. Kochav, X. Liu, J. Xue, F. Grady, S. Guha, K. Upadhyay, D. Ben-Avraham, S. Mukherjee, B. M. Bowen, T. Thomas, J. Vijai, M. Cruts, G. Froyen, D. Lambrechts, S. Plaisance, C. Van Broeckhoven, P. Van Damme, H. Van Marck, N. Barzilai, A. Darvasi, K. Offit, S. Bressman, L. J. Ozelius, I. Peter, J. H. Cho, H. Ostrer, G. Atzmon, L. N. Clark, T. Lencz, I. Pe'er, Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nature communications* **5**, 4835 (2014).
68. S. Guha, J. A. Rosenfeld, A. K. Malhotra, A. T. Lee, P. K. Gregersen, J. M. Kane, I. Pe'er, A. Darvasi, T. Lencz, Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome biology* **13**, R2 (2012).

FIGURE LEGENDS

Figure 1: **Crohn's disease association within the *LRRK2* locus**. **A:** Single-point association without covariates, using Exome chip-genotyped variants only. **B:** Association conditioned on N2081D genotypes, using Exome chip-genotyped variants only.

Figure 2: **Odds ratios for Crohn's disease (CD) and Parkinson's disease (PD) analysis**. Red indicates variants for which both diseases have the same direction of effect; blue indicates opposite-direction effects. Only the variants with at least nominal significance ($P<0.05$) in both CD and PD analysis after linkage disequilibrium pruning are shown. Circle sizes correspond inversely to the significance (P-value) of CD association at each variant. **A:** Ashkenazi Jewish odds ratios: 23 of 26 independent variants (88%) exhibited effects in the same direction for both diseases (binomial test $P=5.2\times10^{-6}$). **B:** Non-Jewish odds ratios: 25 of 29 variants (86%) exhibited effects in the same direction for both diseases ($P=7.6\times10^{-6}$).

Figure 3: **A *LRRK2*-focused sub-network within the inflammatory bowel disease-associated gene network.** The full intestinal Bayesian network was comprised of 8,382 genes, 551 (6.6%) of them were IBD-associated. From the intestinal network, the largest connected sub-network of genes in network that were within a path length of two of IBD-associated genes was identified, and its portion that includes *LRRK2* is depicted.

Figure 4: **Effect of LRRK2 mutations on protein kinase and GTPase activity**. **A.** Schematic representation of LRRK2 domain structure (boxes along the full length protein) and the respective locations of N551K, R1398H, and N2081D amino acid substitutions relative to the previously reported Parkinson's disease-associated G2019S and Crohn's disease-associated M2397T. Abbreviations: ARM, armadillo; ANK, ankyrin repeat region; LRR, leucine-rich repeat; Roc, Ras in complex protein; COR, C terminal of Roc; MAPKKK, MAP kinase kinase kinase, and WD40, WD40 protein-protein interaction domain. **B.** Representative immunoblot

(left panel) and quantification (right panel) of Rab10 phosphorylation by wild-type (WT) and LRRK2 variants in vitro. **C.** GTPase activity of WT and LRRK2 variants. Representative GTP hydrolysis assay (left) and the fraction of hydrolyzed GTP (GDP) over bound GTP (right panel). All values represent the mean of 3 independent experiments±standard error, and significance was calculated by ANOVA. *P≤0.05, **P≤0.01.

Figure 5: **Effects of CD-associated *LRRK2* mutations on human monocyte-derived macrophages.** **A.** Representative immunoblot of LRRK2, acetylated α-tubulin, p62, and LC3B (forms I-II) protein expression under control (medium) or starved (PBS) conditions by *LRRK2* genotype (left panel). Bar graphs depicting normalized protein expression ratios of acetylated over total alpha-tubulin, and response to autophagy-inducing starvation, PBS over medium, for p62 and LC3-II for non-carriers (n=4), N551K (n=4) and N2081D carriers (n=2) with three independent technical repeats for each sample (right panel). **B.** Representative flow cytometric histograms illustrating lysosensor florescence following starvation treatment (PBS, top), medium (middle) and isotype control (bottom) for non-carriers (n=4), N551K (n=5) and N2081D carriers (n=4) (left panel), and quantified mean lysosensor fluorescent ratios of PBS over resting control (medium) cells (right panel). All values represent mean±standard error, and significance was calculated by mixed model ANOVA with a random effect of a biological sample (panel A) or order-constrained ANOVA (*57*) (panel B). *P≤0.05, **P≤0.01.

Acknowledgements

We thank Dr. Alain Diaz of the University of Miami for technical assistance.

Funding

We acknowledge financial support from NIH research grants GM007205, DK098927 (to K.Y.H.), DK62429, DK062422, DK092235, DK106593 (to J.H.C.), DK062413, DK046763-19, AI067068, HS021747 (to D.P.B.M.), DK062431 (to S.R.B), AG042188 (to G.A.), NS050487, NS060113 (to L.N.C.), MH089964, MH095458, MH084098 (to T.L.), AG618381, AG021654, AG038072 (to N.B.), NS071674 (to J.M.V.), NS37167, NS036711 (to T.F.), NS076843 (to R.H.M.), NS036960 (to H.P.), DK062420 (to R.H.D.), CA141743 (to R.H.D.), CA121852 (to I.Pe'er), and NS060809 (to Z.Y.); NSF research grants 08929882 and 0845677 (to I.Pe'er); Human Frontier Science Program (to S.C.); Lewis and Rachel Rudin Foundation (to H.O.); North Shore–LIJ Health System Foundation (to T.L.); Brain & Behaviour Foundation (to T.L.); US-Israel Binational Science Foundation (to T.L.); Nathan Shock Center of Excellence for the Biology of Aging (to N.B.); the Glenn Center for the Biology of Human Aging (to N.B.); New York Crohn's Foundation (to R.J.D; I.Peter; Y.P.); Edwin and Caroline Levy and Joseph and Carol Reich (to S.B.); SUCCESS grant (to J.H.C.; I Peter); the Sanford J. Grossman Charitable Trust (to J.H.C.); the Cedars-Sinai F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute Research Funds, the European Union, the Crohn's and Colitis Foundation of America (CCFA), the Joshua L. and Lisa Z. Greer Chair in IBD Genetics (to D.P.B.M.); The Leona M. and Harry B. Helmsley Charitable Trust (to D.P.B.M; I.Peter), the Parkinson's Disease Foundation (to L.N.C.), Meyerhoff Inflammatory Bowel Disease Center and the Atran Foundation (to S.R.B), University of Pittsburgh Inflammatory Bowel Disease Genetic Research Chair (to R.H.D.), The Robert P. & Judith N. Goldberg Foundation, the Bumpus Foundation and the Harvard NeuroDiscovery Center (to T.F.), The Charles Wolfson Charitable Trust (to A.W.S.;A.P.L.; E.R.S.; N.P.).

Genotyping services for selected Parkinson's disease cohorts were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C.

Author Contributions

Primary analysis and manuscript-writing (K.Y.H.); project conception and design (I.Peter; J.H.C.; R.J.D.; D.P.B.M.); recruitment, sample acquisition, phenotype data collection for Crohn's disease, Parkinson's disease, and control datasets for a total of >24,000 study subjects (N.B.; S.R.B.; A.S.C.; R.H.D.; S.K.; D.P.B.M.; J.D.R.; A.S.; N.P.; A.L.; E.R.S.; M.S.S.; S.B.; R.H.M.; L.O.; H.P.; W.K.S.; J.M.V.; T.F.,G.A.; L.N.C.; T.L.; P.R.L.); data processing, preparation, and analysis (K.Y.H.; N.P.; M.R.; K.G.; S.C.; H.O.; T.H.; D.L.; M.J.D.; I. Pe'er; E.E.S.); functional study design (J.H.; Z.Y.; Y.P.;Y.I.; I.U.B.); performing experiments (H.F.H.; A.S.;J.H.; X.B.; X.L.; D.R.; N.V.; N.Y.H; L.S.C.; E.C.); manuscript-writing (K.Y.H.; I.Peter; J.H.C.; J.H.; S.C.; H.F.H.; E.E.S.).

Competing Interests

The authors declare no competing financial interests.

Unrelated conflicts for Robert J. Desnick:
I have no COI related to the article. However, I do have other, UNRELATED CONFLICTS during the current and past 3 years as listed below in the areas of lysosomal diseases and the porphyrias:

Amicus Therapeutics: Founding Stock, Consultant
Alexion Pharmaceuticals: Stock, Consultant
Genzyme-Sanofi: Consultant, Patent Licensed, Royalties, Research Grant
Kiniksa Pharmaceuticals: Stock, Chair SAB, Consultant
Mitsubishi-Tanabe: Consultant
Synageva Pharmaceuticals: Stock, Chair SAB, Consultant
Recordati Rare Diseases: Consultant, Research Grant
Sangamo Therapeutics: Stock, Consultant, Research Grant
Shire: Royalties

Unrelated conflicts for Yiannis Ioannou:
1. Neurotrope, Inc.
2. Amathus Therapeutics, Inc.

Unrelated conflict for Shai Carmi:
1. MyHeritage

Unrelated conflicts for Adam S. Cheifetz:
1. AbbVie Pharmaceuticals
2. Janssen Pharmaceuticals
3. Takeda Pharmaceuticals
4. Pfeizer Pharmaceuticals
5. Ferring Pharmaceuticals
6. Miraca Life Sciences

Table 1: List of the top signals that reached genome-wide significance in meta-analysis

RefSNP ID	Chr	Coordinate	Gene	Substitution	Discovery (N=1477 case, 2614 ctrl)				Replication (N=589 case, 1019 ctrl)				Meta-analysis
					MAF_{CD} (%)	MAF_{CTRL} (%)	P-value	OR	MAF_{CD} (%)	MAF_{CTRL} (%)	P-value	OR	P-value
rs11209026	1	67705958	*IL23R*	R381Q	3.22	8.03	6.79×10^{-18}	0.38	3.15	8.05	3.36×10^{-8}	0.37	1.38×10^{-24}
rs139518863	12	40499594	*SLC2A13*	S6N	8.10	4.84	2.68×10^{-9}	1.73	7.65	5.36	9.58×10^{-3}	1.46	1.39×10^{-10}
rs7308720	12	40657700	*LRRK2*	N551K	6.64	9.85	7.06×10^{-7}	0.65	7.78	10.45	1.27×10^{-2}	0.72	3.28×10^{-8}
rs33995883	12	40740686	*LRRK2*	N2081D	8.13	4.86	2.56×10^{-9}	1.73	7.40	5.61	4.40×10^{-2}	1.34	9.51×10^{-10}
rs141326733	16	50138853	*HEATR3*	R642S	2.78	1.03	3.16×10^{-9}	2.74	1.87	0.93	2.29×10^{-2}	2.02	4.76×10^{-10}
rs2066842	16	50744624	*NOD2*	P268S	32.42	23.03	2.25×10^{-20}	1.60	32.44	20.07	4.21×10^{-15}	1.91	3.31×10^{-33}
rs2066844	16	50745926	*NOD2*	R702W	3.63	1.88	1.19×10^{-6}	1.97	3.82	2.11	4.25×10^{-3}	1.84	1.76×10^{-8}
rs104895447	16	50750842	*NOD2*	M863V	4.06	1.05	1.57×10^{-19}	3.98	3.57	1.08	1.15×10^{-6}	3.39	1.28×10^{-24}
rs2066845	16	50756540	*NOD2*	G908R	8.73	4.21	5.14×10^{-17}	2.18	7.99	3.29	4.12×10^{-9}	2.55	1.52×10^{-24}
rs2066847	16	50763781	*NOD2*	L1007fs	8.33	2.75	6.27×10^{-30}	3.21	7.47	2.40	7.09×10^{-12}	3.28	3.43×10^{-40}

MAF_{CD}, minor allele frequency in Crohn's disease cases; MAF_{CTRL}, minor allele frequency in controls; OR, odds ratio. P-values for discovery and replication cohorts calculated using X^2 testing. Meta-analysis performed using METAL default method.

Table 2: Allele frequencies and association statistics for *LRRK2* non-synonymous variants in imputed datasets

				N551K			
				CD vs. control association		PD vs. control association	
	MAF_{CD}(%)	MAF_{PD}(%)	MAF_{ctrl}(%)[a]	Odds ratio (95% CI)	P-value	Odds ratio (95% CI)	P-value
Ashkenazi Jewish	6.8	7.7	9.8	0.67 (0.57 - 0.79)	$1.4x10^{-6}$	0.77 (0.67 - 0.90)	$3.9x10^{-4}$
Non-Jewish	6.0	6.2	6.9	0.89 (0.79 - 1.0)	$5.1x10^{-2}$	0.87 (0.77 - 1.0)	$4.4x10^{-2}$
				R1398H			
				CD vs. control association		PD vs. control association	
	MAF_{CD}(%)	MAF_{PD}(%)	MAF_{ctrl}(%)[a]	Odds ratio (95% CI)	P-value	Odds ratio (95% CI)	P-value
Ashkenazi Jewish	6.8	7.6	9.1	0.71 (0.60 - 0.84)	$5.0x10^{-5}$	0.84 (0.72 - 0.98)	$1.6x10^{-2}$
Non-Jewish	6.1	6.2	6.9	0.88 (0.78 - 0.99)	$4.0x10^{-2}$	0.88 (0.77 - 1.0)	$5.6x10^{-2}$
				N2081D			
				CD vs. control association		PD vs. control association	
	MAF_{CD}(%)	MAF_{PD}(%)	MAF_{ctrl}(%)[a]	Odds ratio (95% CI)	P-value	Odds ratio (95% CI)	P-value
Ashkenazi Jewish	8.0	5.9	5.4	1.7 (1.4 - 2.0)	$4.3x10^{-8}$	1.1 (1.0 - 1.4)	$3.6x10^{-2}$
Non-Jewish	2.9	2.4	1.8	1.6 (1.3 - 2.0)	$2.1x10^{-6}$	1.3 (1.0 - 1.6)	$1.7x10^{-2}$

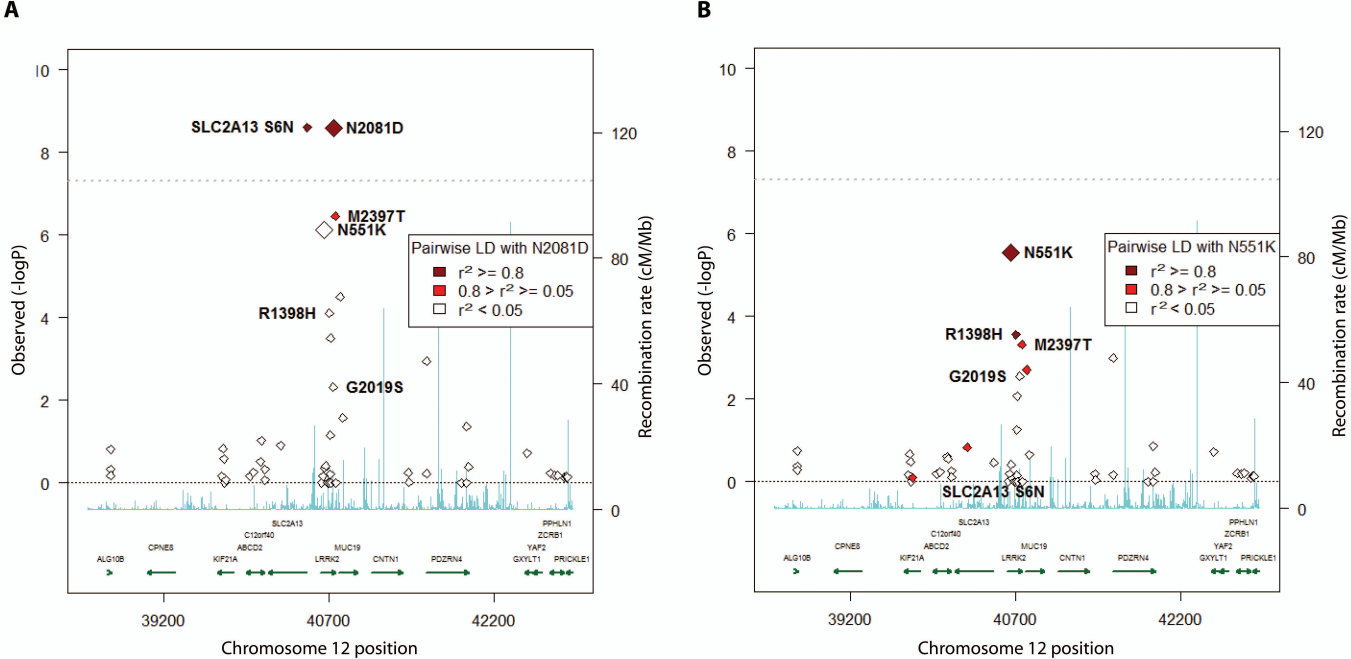
[a] Combined control minor allele frequency (MAF). Each healthy control was randomly assigned to only one disease association analysis to ensure independence.

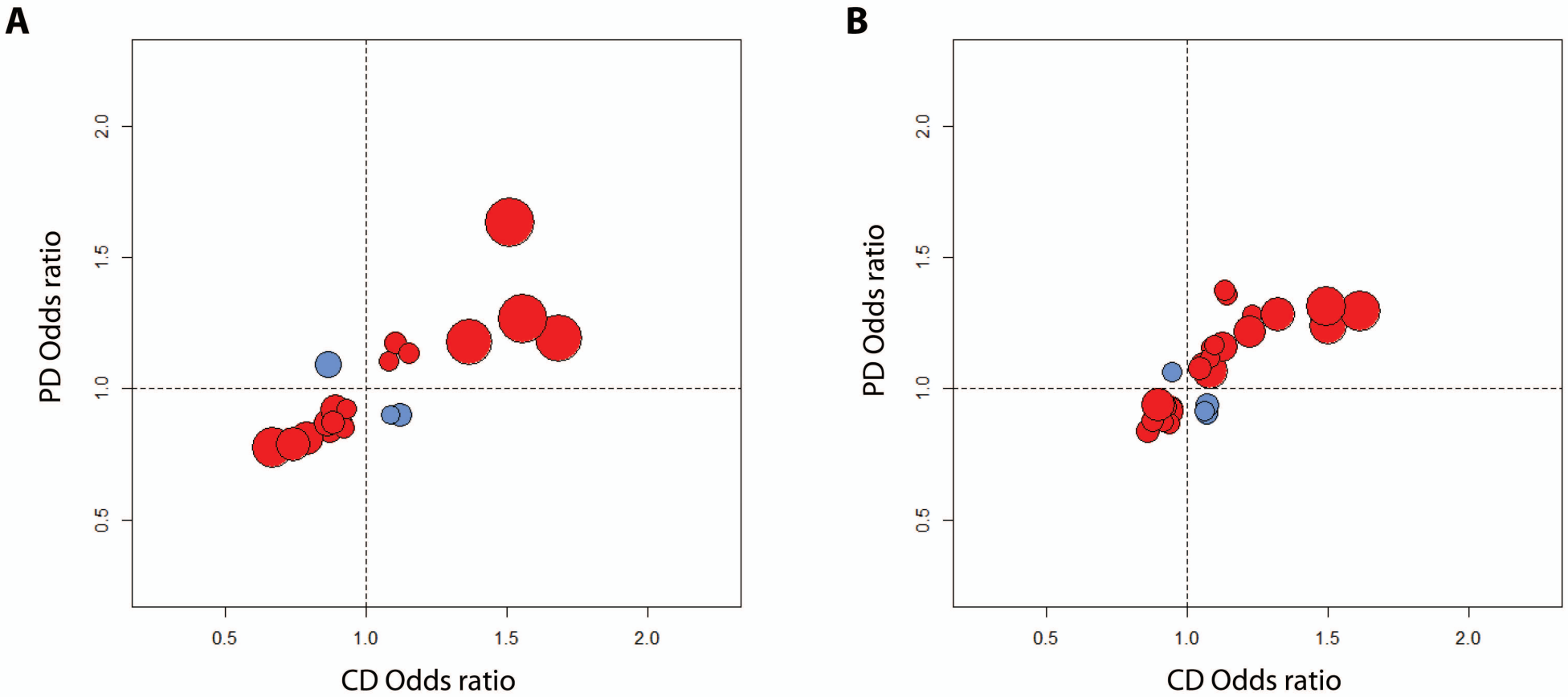
P-values calculated using logistic regression.

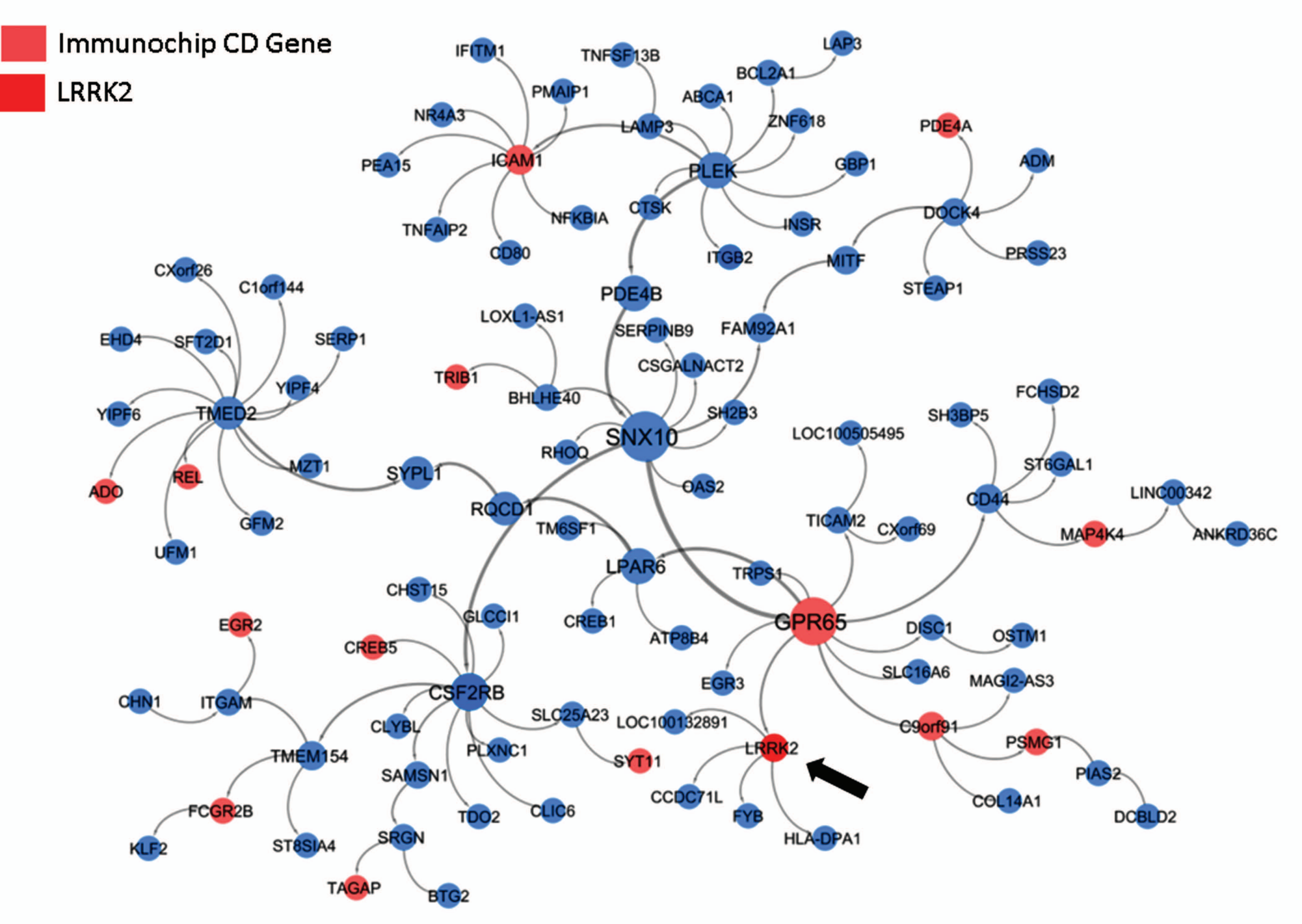
Table 3: Subphenotypic values by *LRRK2* N2081D and R1398H genotype statuses in pooled AJ and NJ CD cohorts

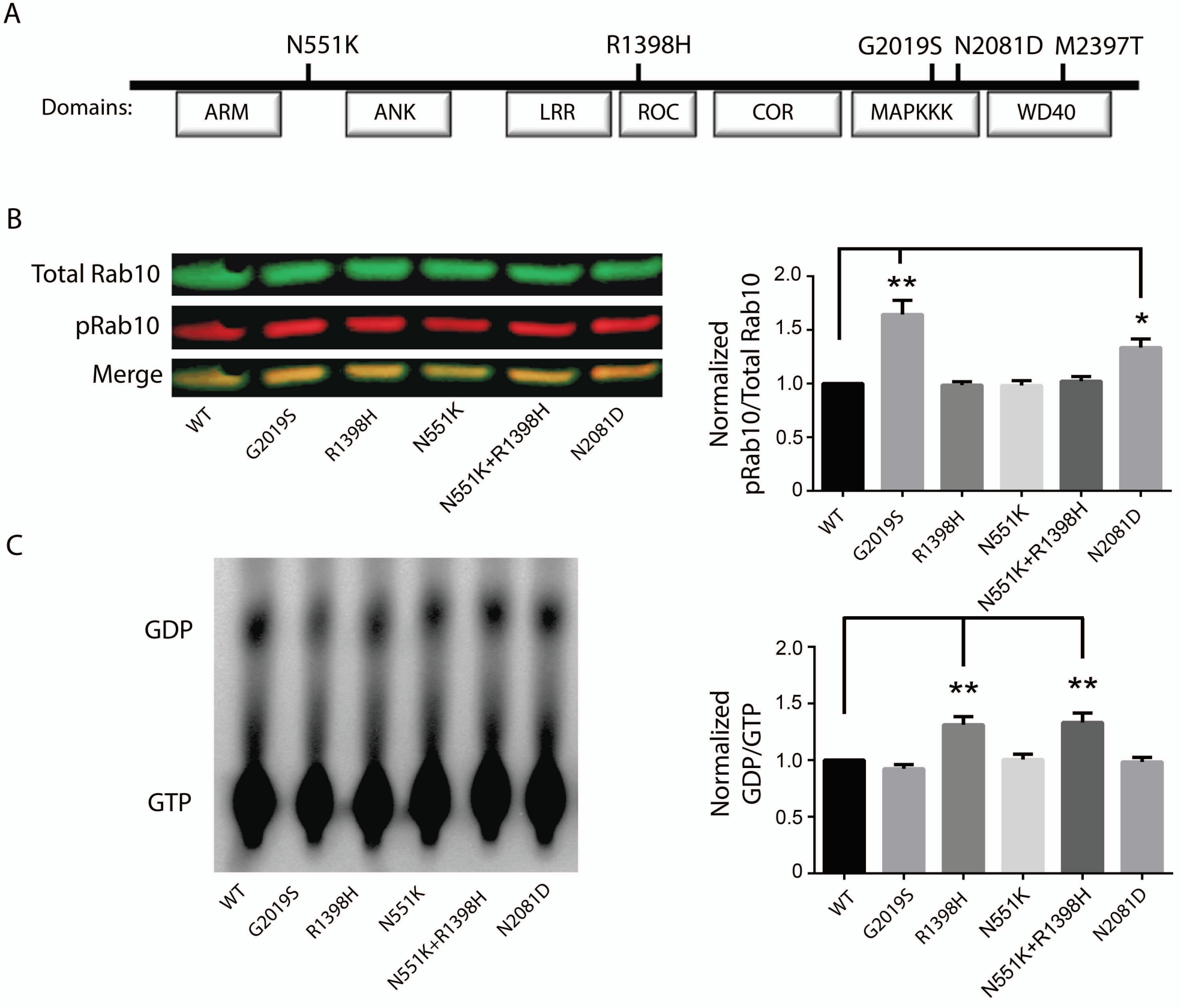
N2081D genotype	Age of CD onset (SD) [N]	Ileal disease location [N]
AA	26.5 (14.0) [5601]	80.5% [5311]
GA	24.6 (13.1) [482]	86.1% [453]
GG	20.8 (9.0) [12]	90.9% [11]
	P=0.002	P=0.01
R1398H genotype		
GG	26.3 (13.9) [5365]	81.1%[5095]
GA	26.4 (14.1) [701]	80.7% [652]
AA	27.2 (19.4) [29]	71.4% [28]
	ns	ns

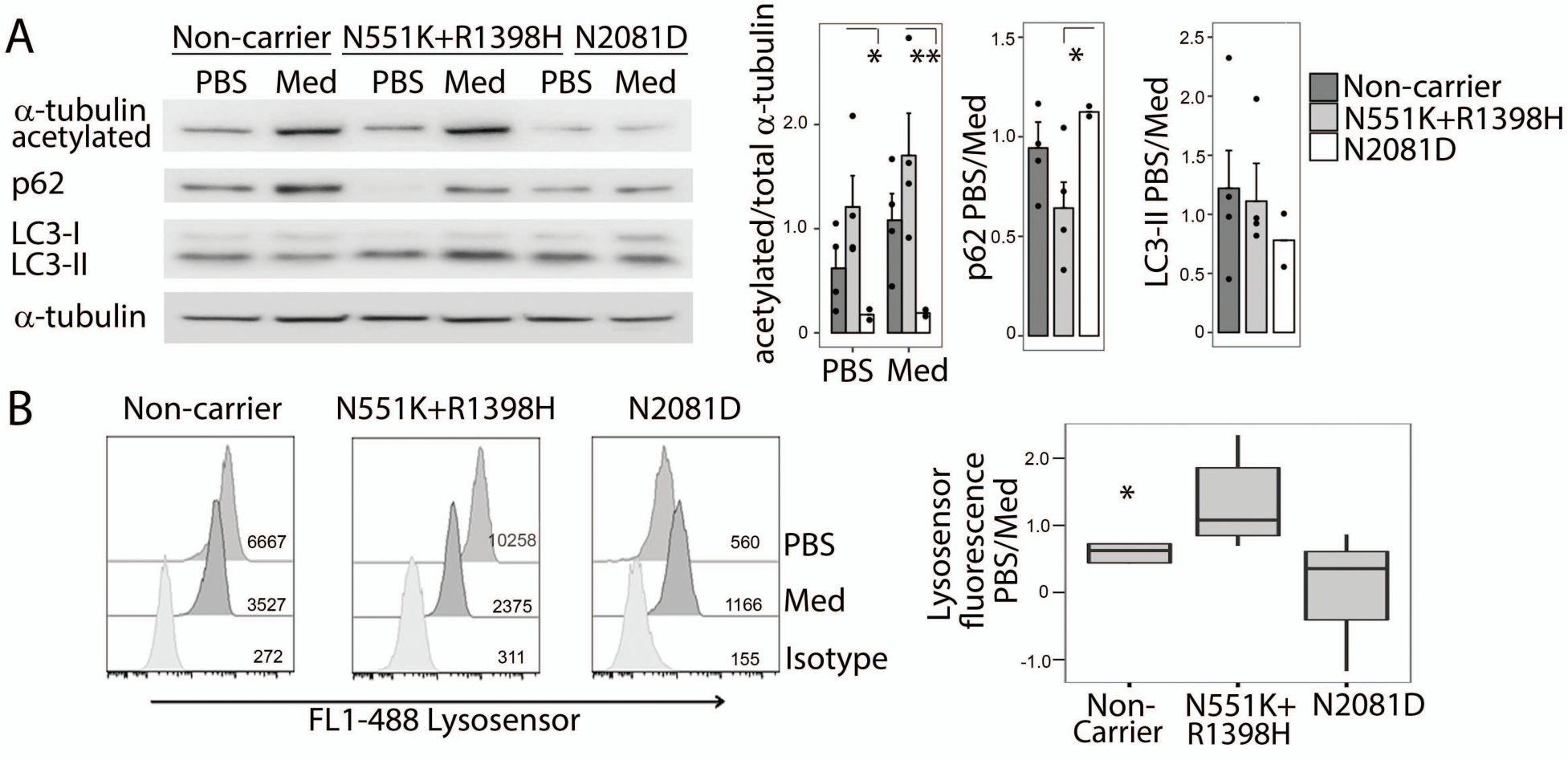
AJ, Ashkenazi Jewish. NJ, non-Jewish European. CD, Crohn's disease. SD, standard deviation. N, group sample size. ns, not significant. Similar results were found for N551K (in strong linkage disequilibrium with R1398H, r^2=0.81). P- values calculated using simple linear regression.


A
B
SLC2A13 S6N
N2081D
M2397T
N551K
R1398H
G2019S
Pairwise LD with N2081D
Pairwise LD with N551K
r² >= 0.8
0.8 > r² >= 0.05
r² < 0.05
Observed (-logP)
Recombination rate (cM/Mb)
Chromosome 12 position
39200
40700
42200
SLC2A13
C12orf40
ABCD2
MUC19
PPHLN1
ZCRB1
YAP2
ALG10B
CPNE8
KIF21A
LRRK2
CNTN1
PDZRN4
GXYLT1
PRICKLE1


A
PD Odds ratio
CD Odds ratio
0.5
1.0
1.5
2.0
B
PD Odds ratio
CD Odds ratio
0.5
1.0
1.5
2.0


Immunochip CD Gene
LRRK2
IFITM1
TNFSF13B
LAP3
BCL2A1
PMAIP1
ABCA1
NR4A3
LAMP3
ZNF618
PDE4A
PEA15
ICAM1
PLEK
GBP1
ADM
CTSK
DOCK4
TNFAIP2
NFKBIA
INSR
CD80
ITGB2
MITF
PRSS23
CXorf26
C1orf144
PDE4B
STEAP1
LOXL1-AS1
SERPINB9
FAM92A1
EHD4
SFT2D1
SERP1
CSGALNACT2
TRIB1
YIPF4
BHLHE40
FCHSD2
YIPF6
TMED2
SH2B3
SH3BP5
SNX10
LOC100505495
RHOQ
ST6GAL1
MZT1
SYPL1
OAS2
CD44
LINC00342
ADO
REL
RQCD1
TICAM2
CXorf69
MAP4K4
ANKRD36C
GFM2
TM6SF1
UFM1
LPAR6
TRPS1
CHST15
GLCCI1
CREB1
GPR65
DISC1
OSTM1
EGR2
ATP8B4
CREB5
SLC16A6
EGR3
MAGI2-AS3
CHN1
ITGAM
CSF2RB
C9orf91
SLC25A23
LOC100132891
CLYBL
PSMG1
TMEM154
PLXNC1
LRRK2
SYT11
PIAS2
SAMSN1
CCDC71L
COL14A1
FCGR2B
TDO2
CLIC6
FYB
DCBLD2
SRGN
HLA-DPA1
KLF2
ST8SIA4
TAGAP
BTG2


A
N551K
R1398H
G2019S
N2081D
M2397T
Domains:
ARM
ANK
LRR
ROC
COR
MAPKKK
WD40
B
Total Rab10
pRab10
Merge
WT
G2019S
R1398H
N551K
N551K+R1398H
N2081D
Normalized pRab10/Total Rab10
2.0
1.5
1.0
0.5
0
**
*
C
GDP
GTP
Normalized GDP/GTP
**
**


A
Non-carrier
N551K+R1398H
N2081D
PBS
Med
PBS
Med
PBS
Med
α-tubulin acetylated
p62
LC3-I
LC3-II
α-tubulin
acetylated/total α-tubulin
PBS
Med
p62 PBS/Med
LC3-II PBS/Med
Non-carrier
N551K+R1398H
N2081D
B
Non-carrier
N551K+R1398H
N2081D
6667
3527
272
10258
2375
311
560
1166
155
PBS
Med
Isotype
FL1-488 Lysosensor
Lysosensor fluorescence PBS/Med
Non-Carrier
N551K+ R1398H
N2081D

Supplementary Materials for: **Functional Variants in *LRRK2* Confer Pleiotropic Effects on Crohn's Disease and Parkinson's Disease**

Table of Contents

Sample collection

A total of 2,066 individuals with Crohn's disease (CD), and 3,633 unaffected controls were enrolled at 14 centers throughout North America, Europe, and Israel (**Table S3**). All participants were unrelated and self-reported as having Ashkenazi Jewish ancestry and provided written consent for genotyping and analysis, under protocols approved by each site's local institutional review board. IBD patients had diagnoses confirmed at each recruiting site by a health care provider, based on standard criteria including clinical presentation, as well as endoscopic, radiologic and/or pathologic confirmation.

Collection and analysis of exome sequencing data

Exome sequencing and variant calling

Genomic DNA was extracted from whole blood; sheared exomic fragments were captured with the NimbleGen 2.1 M human exome array (Roche/NimbleGen, Madison, WI). We sequenced the captured libraries as paired-end 75-bp reads using Illumina GAII (45 samples, comprising 39 unrelated individuals and 3 sibling pairs) and HiSeq (5 unrelated samples) sequencers, with 8 samples bar-coded per sequencing lane. Sequence reads were mapped to the reference genome (hg18) using the BWA program using default parameters (*59*), and coordinates were translated to the hg19 reference genome using LiftOver in the UCSC Genome Browser with default parameters (*60*). The Genome Analysis Toolkit (GATK) v2 was used to call alleles at variant sites (*61*). Sample-level realignment and multi-sample SNP calling were performed using default parameters and the GATKStandard variant filter. Visual confirmation of insertion-deletion polymorphisms, which have a higher error rate in variant calling, was performed in the Integrative Genomics Viewer (*62*).

Among samples sequenced using GAII sequencers, the average number of reads was 71,876,305, with 42,552,977 (58.8%) of reads on-target (within exomic regions), which yielded an average coverage depth of 91.5X. These individuals each had, on average, 31,493 variants detected with at least one non-reference allele. Samples sequenced by HiSeq had a mean read count of 151,352,500, with 58,934,621 (38.9%) on-target, average coverage depth of 128X, and 44,114 variants detected. The variant sets from both machine types were merged, and no platform stratification was carried forward in the analysis.

Imputation of variants detected by exome sequencing

In order to determine whether the variants identified through our exome sequencing were already tagged in our previous AJ GWAS, we performed imputation and assessed the estimated imputation quality. The GWAS cohort consisted of 907 AJ Crohn's independent disease cases and 1,644 matched controls; individuals with only partial AJ ancestry, which were analyzed in the previous GWAS, were excluded here (*11*). Our exome sequencing samples, which were used in the GWAS analysis, served as a reference panel for imputation on the full case-control cohort, using BEAGLE version 3.3.0 with default parameters (*63*). BEAGLE performs an internal estimation of imputation accuracy (r^2, correlation between imputed and true genotypes). We conducted an additional study of empirical imputation accuracy, stratifying the exome

sequencing markers by minor allele count and BEAGLE estimated accuracy, and then masking 25 randomly selected markers within each group. Pearson correlation between observed and inferred allele dosage was calculated after imputation. We determined that for variants with BEAGLE accuracies above 0.7 and minor allele frequencies over 5%, there was high agreement between the estimated and empirical imputation accuracies. We therefore used MAF $\geq$ 5% and BEAGLE $r^2 \geq 0.7$ as the joint criteria for defining variation already successfully assayed by the GWAS platforms; 8,758 exomic coding variants were defined as well-imputed (**Figure S2**), with all other markers eligible for addition to the HumanExome platform as custom content (**Figure S1**). As a fine-mapping exercise, we performed case-control association analysis among the well-imputed variants and found no evidence for novel CD association signals at a Bonferroni-corrected significance threshold of 10^{-5}.

Custom assay design

HumanExome base content

The Illumina HumanExome beadchip v1.0 contains 230,296 non-synonymous and canonical splice-site markers, located in over 20,000 RefSeq-listed genes that were identified in various large exome-sequencing studies. The largest number of samples was contributed by the NHLBI exome sequencing study (4260), and autoimmune disease samples were represented in the GO type II diabetes (T2D), Hispanic T2D genes, and Pfizer-MGH-Broad T2D study cohorts. The vast majority of samples were of European ancestry, with little Ashkenazi Jewish representation reported. The targeting of predominantly rare coding-region variants unsurprisingly results in a low proportion of genomic variation captured (0.088 of variation with MAF > 1.0% tagged with $r^2 > 0.8$).

Selection of variants as custom content

From the list of candidate non-synonymous variants identified in our exome sequencing, we removed those already included in the HumanExome base content. We then evaluated the markers using the Illumina Assay Design Tool. We eliminated the variants that had a Final Score less than 0.7; this value represents the probability that a probe designed to assay a given variant will be successful. Common background markers tagging previously established IBD loci in an Immunochip-based study were also added to the genotyping platform (*1*). This yielded the final set of variants that was used as custom content on the genotyping platform.

Because the goal of the exome sequencing phase of the study was variant detection, rather than association analysis, the sibling-pair data was not stratified or separated from the unrelated samples. Given the small number of sibling pairs sequenced, we did not perform any linkage-based analysis of these data, nor were we able to assess for specific variants that were differentially carried by familial and sporadic CD cases.

Selection of variants and samples for analysis

Genotype and sample quality filtering

Genotyping data were collected at three genotyping centers (Philadelphia, PA, Manhasset, NY, and Los Angeles, CA) using the same custom genotyping array. The data were combined and preliminary genotypes were called jointly using input from all three centers in GenomeStudio. Following guidelines produced by the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) consortium, we enforced quality control using SNP metrics based on fluorescent probe intensities and genotype frequencies, as well as visual inspection of markers with uncertain genotyping quality (*48*). First, samples with low quality metrics (genotype call rate < 0.96 and/or $p10_{GC}$ < 0.4125) were removed, and the markers re-clustered in GenomeStudio. Markers with overall low probe intensity were removed. A subset of SNPs was excluded based on meeting any of these genotype clustering criteria:

- Θ_{AA} mean $\geq$ 0.25
- Θ_{BB} mean < 0.8
- Θ_{AB} mean < 0.19 and $\geq$ 0.83
- Θ_{AA} deviation > 0.0355
- Θ_{BB} deviation > 0.0355
- Θ_{AB} deviation < 0.0107 and $\geq$ 0.08
- Cluster separation < 0.35
- Heterozygote excess < -0.2 and > 0.03
- For chromosome X only: Heterozygote excess < -0.6
- R_{mean}(AA), R_{mean}(AB), or R_{mean}(BB) < 0.2
- Call rate < 0.99

Another set of SNPs was flagged for manual review by the following criteria. During manual review, genotype cluster boundaries were adjusted to optimize cluster separation and genotype call rate.

- Θ_{AA} mean 0.15 to 0.25
- Θ_{BB} mean 0.8 to 0.9
- Θ_{AB} mean 0.19 to 0.28 and 0.78 to 0.83
- Θ_{AA} deviation 0.027 to 0.0355
- Θ_{BB} deviation 0.027 to 0.0355
- Θ_{AB} deviation 0.046 to 0.08
- Cluster separation 0.35 to 0.45
- Heterozygote excess -0.055 to -0.2 and 0.017 to 0.03
- For chromosome X only: Heterozygote excess > 0
- $Frequency_{AB} \geq 0.508$
- R_{mean}(AA), R_{mean}(AB), or R_{mean}(BB) 0.2 to 0.25
- Call frequency 0.99 to 0.999

- All chromosome Y
- All chromosome MT
- Rep errors > 0

An additional set of variants were flagged as potentially having a mis-called heterozygote cluster. These markers, defined by the criteria below, were also manually reviewed.

- $\text{Frequency}_{AB} = 0$ and $\text{frequency}_{AA} > 0$ and $\text{frequency}_{BB} > 0$
- $\text{Frequency}_{AA} = 1$ and call frequency < 1
- $\text{Frequency}_{BB} = 1$ and call frequency < 1
- $\text{Frequency}_{AB} = 0$ and MAF > 0

After all manual review was completed, additional quality thresholds for sample exclusion (genotype call rate < 0.968, $p50_{GC} < 0.758$, or average heterozygosity > 0.31) were applied. We observed no differences between genotyping centers in the quality statistics. Related samples were identified using pairwise identity-by-descent detection in PLINK (*64*) and removed. Samples with discrepancy between self-reported gender and genotypic gender were excluded. Following association analysis, cluster plots of the 198 significant coding markers were re-examined to ensure high-quality genotype calling.

Among all genotyped non-synonymous markers, 19,361 markers were removed by genotype quality filtering, and 153,978 were monomorphic in the AJ cohort, which yielded the final set of 61,234 markers for analysis (**Figure S1**).

Principal components analysis

We created a set of 10,312 null independent autosomal polymorphisms by removing markers with pairwise linkage disequilibrium (LD) of $r^2 > 0.05$, those with minor allele frequency (MAF) below 0.05, custom content, and variants within established IBD loci. Principal components analysis (PCA) was performed using the princomp() function in R version 2.15.1 (**Figure S3**). The PCA was conducted on our dataset in conjunction with a reference cohort comprising currently unpublished, non-Jewish European-ancestry samples on which exome chip data was available. Boundaries defined along the top three principal components were used to define outlier samples for removal from the AJ cohort. We determined that many of the excluded samples had self-reported less than 100% AJ ancestry. No significant correlations between the top principal components and case-control status, geographic location, or genotyping center were observed. Using membership in the PCA cluster as genetic validation of AJ ancestry, we conducted all further analyses using only samples with 100% AJ ancestry: 1,477 CD cases, and 2,614 independent healthy controls (**Table S3**). Because we excluded recent ancestry-admixed subjects and the AJ population is a homogenous isolate, we did not include population substructure as an additional covariate in downstream analyses, in order to maximize our power to detect rare-variant association signals.

Association testing

Power calculations

For the exome sequencing phase of our study, we performed power calculations using binomial distribution probability estimates (**Table S1**). Because the cohort consisted of 44 independent CD cases and 3 pairs of siblings, the calculation was performed assuming 97 independent chromosomes. For the association testing power estimates, we used cumulative probability densities under a normal distribution.

Quantile-quantile (Q-Q) plots

We created a Q-Q plot of chi-square association statistics (**Figure S4**) using a reduced set of non-synonymous variants in order to assess the validity of the null distribution assumption throughout our exomic dataset. We removed markers with high pairwise linkage disequilibrium ($r^2 \geq 0.5$) or with a minor allele count of 1, since, given our study's sample size, such low-frequency variants could not achieve significance ($P < 0.05$) even without any correction for multiple testing. Genomic inflation was calculated using this dataset and found to be within the standard range of previous GWA studies ($\lambda = 1.095$).

Evaluation of conditional independence

Logistic regression, in which coding variants and background associated markers served as covariates, was used to define independent association signals. Conditionally dependent pairs of variants were defined as those whose conditioned P-values were at least an order of magnitude less significant than the individual single-point P-values (in logistic regression with no covariates).

Non-additive association and epistasis

In single-point analysis, none of the associated markers showed deviation from standard allelic (additive) association (**Table S4**, "Alternate models" tab), and there was no evidence of interaction effects between any of the nominally significant variants.

Association with age of onset and Crohn's disease location

Crohn's disease age of diagnosis was available for 6,095 CD cases, and disease location was available for 5,775 CD cases, using data from both AJ and NJ individuals in the NIDDK IBDGC repository database. Association testing for these two phenotypes was performed using linear regression and chi-square testing, respectively.

Selection of markers for trends among independent highly associated markers

To evaluate for significant patterns among the markers highly associated with CD, we performed LD pruning (pairwise $r^2 < 0.5$) to create a set of independent polymorphisms. From these, we used the set of 100 most associated markers ($P < 1.3 \times 10^{-3}$) to perform several analyses described in the following three sections.

Imputation of *LRRK2* locus variants in CD and PD cohorts

Sample selection

We expanded our analysis to evaluate all polymorphisms, including noncoding variation, throughout the 5 Mb-region on chromosome 12 symmetrically flanking *LRRK2*, in analysis of individuals with CD or PD, as well as in comparable non-Jewish European ancestry samples whose ancestry was validated using principal component analysis as previously described (*65*). Additional healthy control genotypes from the 1,000 Genomes Project representing independent individuals from the CEU and TSI populations were extracted from the Phase 1 data release (11/23/2010) and included single-nucleotide polymorphisms (SNPs) and short indels (*66*). Genotypes from the Ashkenazi Genome Consortium (TAGC) representing healthy controls of AJ ancestry were extracted from the public Phase 1 release (9/9/2014, http://browser.1000genomes.org) and also included both SNPs and short indels (*67*).

Imputation of genotypes

In the process of combining data from the various sources described above, we ensured that the datasets had consistent strand alignment and variant positions. Markers that were exclusive to the whole-genome sequencing datasets (TAGC and 1,000 Genomes for the AJ and NJ analyses, respectively) were omitted to increase overall imputation accuracy, as we did not treat these as reference datasets for imputation, given their relatively small number of individuals included. A total of 4,124 variants were used as imputation input in the AJ data, compared to 2,256 markers used for imputation in NJs; this discrepancy was due to a greater variety of genotyping platforms (and therefore a larger set of genotyped markers) used in the AJ datasets (**Table S2**). Reference-free imputation was performed using MACH, with 300 haplotype states and 50 Markov chain rounds (*50*). Integrated imputation of CD, PD, and control samples was performed in a single process, with no phenotype data used as input for the algorithm. The AJ and NJ datasets were imputed separately to reduce runtime, and only variants with high imputation quality ($R^2 > 0.7$) were retained, which yielded 1,436 Ashkenazi and 643 non-Jewish polymorphisms, including 486 overlapping variants present in both populations' datasets (**Figure S1**). We noted that imputation quality was generally better in the Jewish data than in NJ, indicating a greater extent of haplotype sharing in the AJ population, which is consistent with previous population genetic studies (*67*).

Association analysis

Within each population, we conducted separate association analyses for CD and PD. P-values were calculated using logistic regression in order to facilitate direct comparison between unconditioned analysis and those which included certain markers as covariates. As in the first stage analysis focused on novel non-synonymous variation, we again did not use population stratification as a covariate in AJ analyses, noting that previous work has shown that there exists little intra-Ashkenazi population structure using PCA. This prior study demonstrated that the first PC in AJs is already a local signal (human leukocyte antigen, or HLA), not a genome-wide one, implying that there is no room for stratification correction with PCs (*68*). As genome- or exome-wide data were not obtained on many NJ samples, PCA was unable to be performed for that analysis.

Healthy controls were randomly assigned to either CD or PD analysis, to ensure independence of the analyses; using only a subset of controls to estimate each set of association statistics accounted for minor discrepancies in the ORs and P-values for the *LRRK2* variants reported in **Tables 1 and 2**.

Comparisons of overall genetic architecture between CD and PD (as in **Figure 3**) were made using an LD-pruned set of markers (26 markers in the AJ data; 29 in NJ), in which variant pairs with high LD ($r^2 > 0.8$) had been eliminated.

Experimental Studies

Autophagy experiments in human macrophages

Peripheral Blood Mononuclear Cells (PBMC) were isolated from approximately 50mL of forearm vein blood using Cell Preparation Tube (BD Vacutainer® CPT™) with Sodium Heparin (362753) per manufacturer's protocol. PBMCs were incubated in a nunclon delta 6-well plate for 90 minutes in Monocyte Attachment Medium (MAM) (C-28051, promocell, Heidelberg, Germany) at 37 degrees Celsius and 5% of CO2. Then cells were washed twice using MAM and monocytes remained attached. Two milliliters of complete M1 macrophage generation medium purchased from Promocell were added per well. Complete M1 medium included the basal M1 medium (C-28055, Promocell, Heidelberg, Germany) plus supplement mix M1-macrophage generation medium DXF (c-39855, promocell, Heidelberg, Germany) plus M1 cytokine mix (c-39894, Promocell, Heidelberg, Germany). At day 6, one milliliter of complete M1 macrophage medium was added to each well. At day 10, the medium was changed and replaced with fresh complete M1 macrophage medium. At day 12, matured M1-macrophages were incubated in PBS and in complete M1 macrophage medium for 45 minutes. M1-macrophages from these patients were derived from whole peripheral blood monocytes according to the manufacturer's instructions (Promocell, Heidelberg, Germany). Monocytes were polarized to mature M1-macrophages in the DXF M1-macrophage generation medium (M1-medium, resting condition, Promocell) for 12 days and then incubated in PBS and M1 medium for 45 minutes. Cells were then lysed in RIPA buffer (PI89900, Thermo Scientific, Waltham, MA USA) with Halt Protease and Phosphatase Inhibitor Cocktail (PI78440, Thermo Scientific, Waltham, MA USA). Ten micrograms of total protein were loaded onto 4-12% Bis-Tris Plus precast SDS-polyacrylamide

gels, transferred onto a PVDF membrane and probed with primary rabbit anti-LRRK2 antibody (ab133474, abcam), mouse anti-acetylated alpha-tubulin (T7451, Sigma-Aldrich, St. Louis, MO), rabbit anti-alpha tubulin (ab4074, abcam), mouse anti-SQSTM1 (sc-28359, Santa Cruz Biotechnology), and rabbit anti-LC3B (NB100-2220, Novus Biologicals). The corresponding HRP-conjugated secondary antibody was applied for detection. Total alpha-tubulin was used as a loading control for normalization and protein densitometry was performed using ImageJ software. LRRK2 degradation was assessed as the ratio of degraded LRRK2 to total LRRK2 (full length + degraded) protein. Alpha-tubulin acetylation was assessed as the ratio of acetylated to total alpha-tubulin.

Supplementary figures

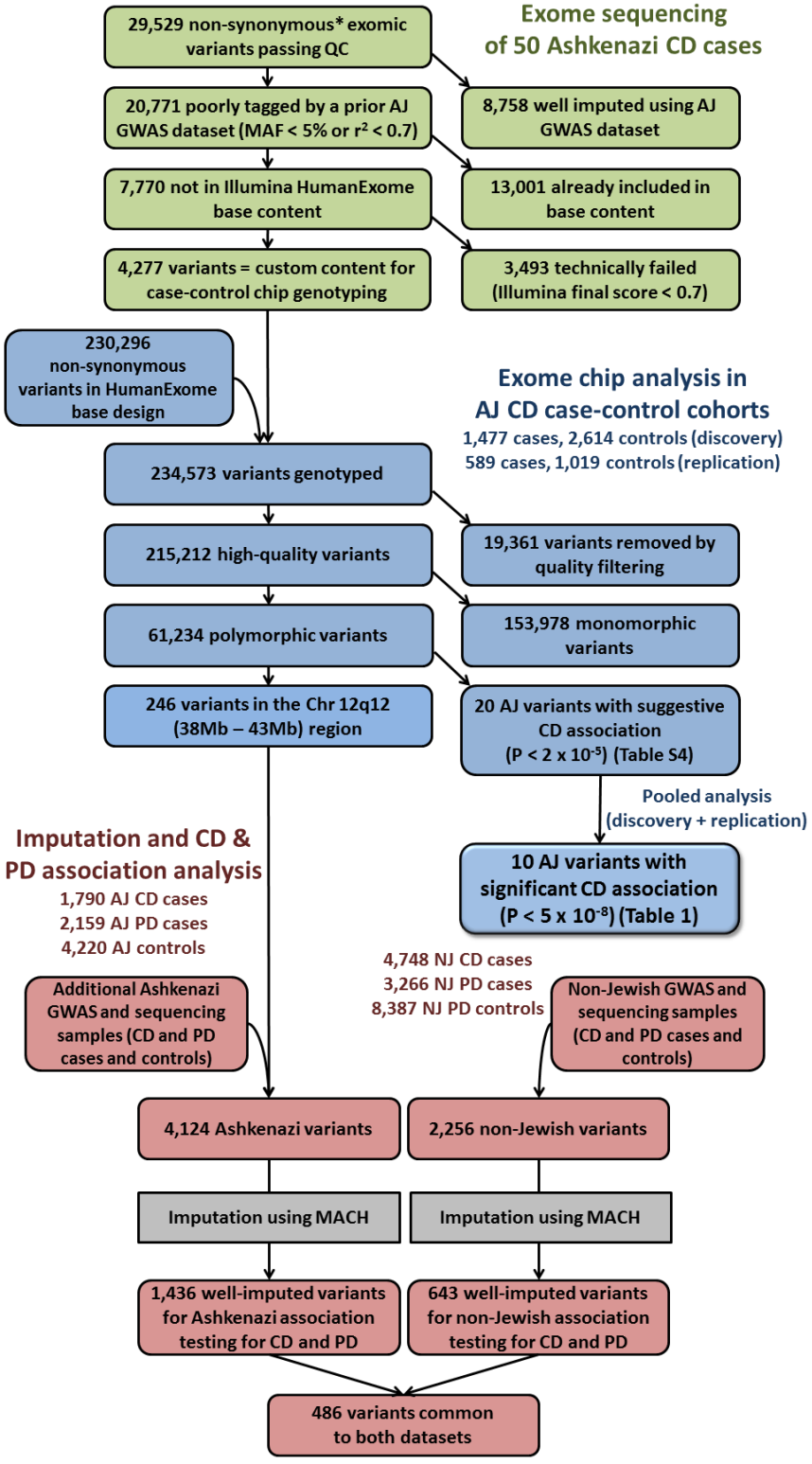


Figure S1: Schematic workflow of genetic analysis, by analytic stages

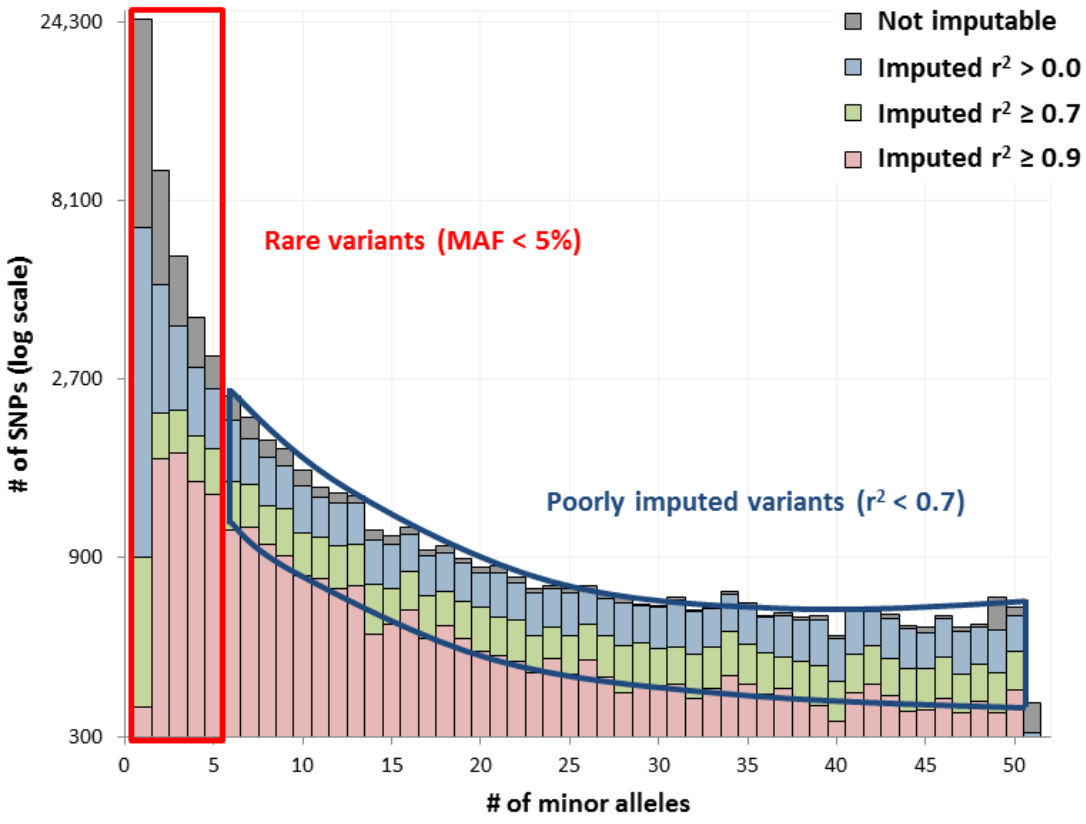


Figure S2: Variants identified through exome sequencing, by MAF and imputation quality.

Colored boxes represent two classes of candidate variants considered for direct genotyping in a case-control cohort: rare variants (red), and common variants with low imputation quality (blue).

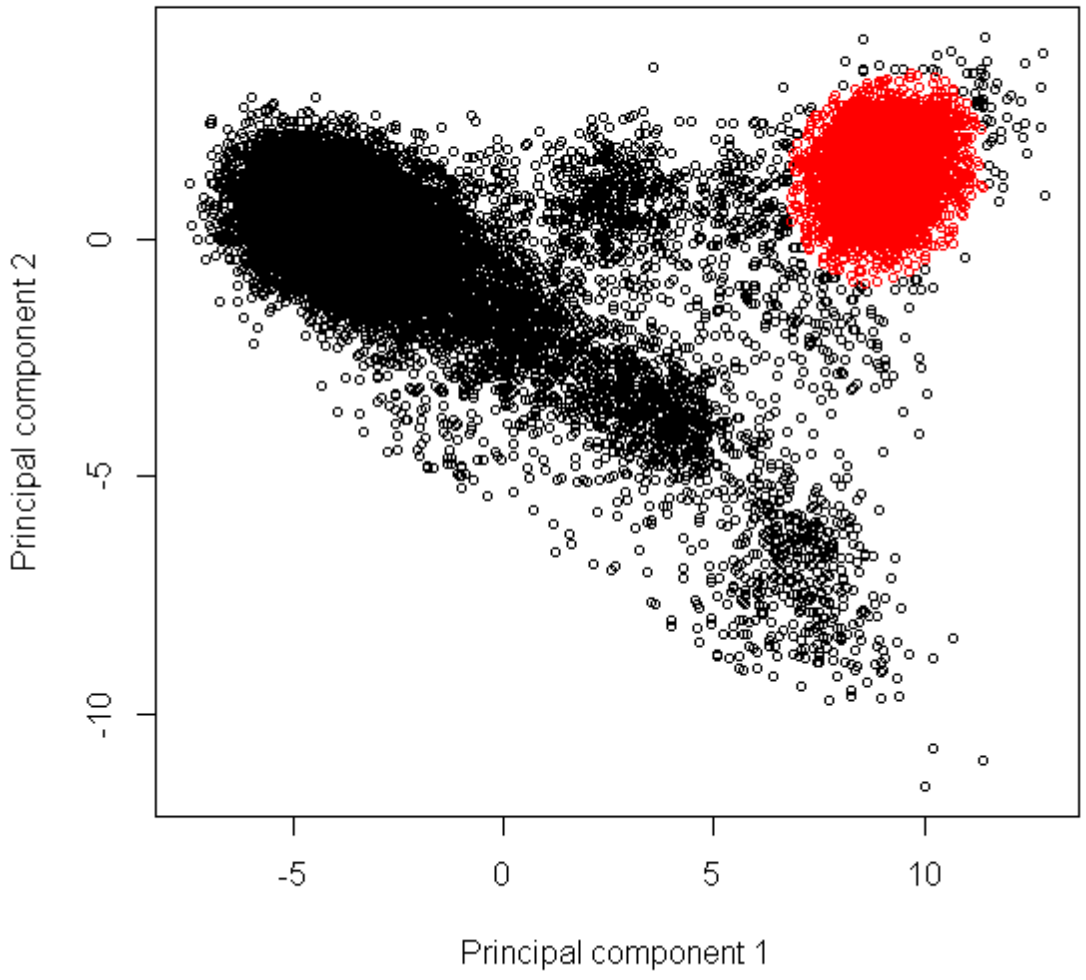


Figure S3: Principal components analysis

Samples defined as having full AJ ancestry are denoted in red.

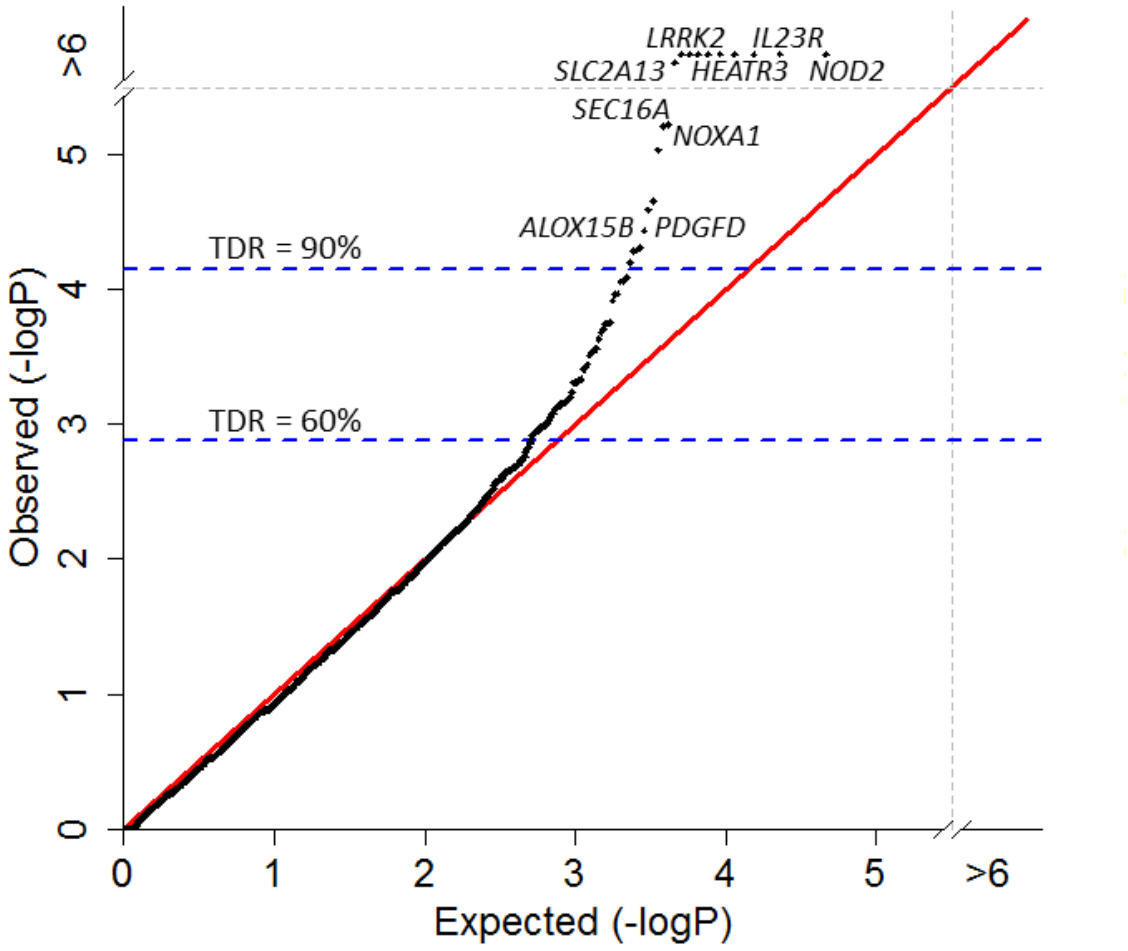


Figure S4: Q-Q plot of CD association results show enrichment of true positive signals below 10^{-3}.

True discovery rate as a function of P-value threshold; the expected numbers of significant markers under the null distribution were calculated through permutation. Blue dotted lines in panels A and B indicate P-value cutoffs corresponding to true discovery rates of 60% and 90%.

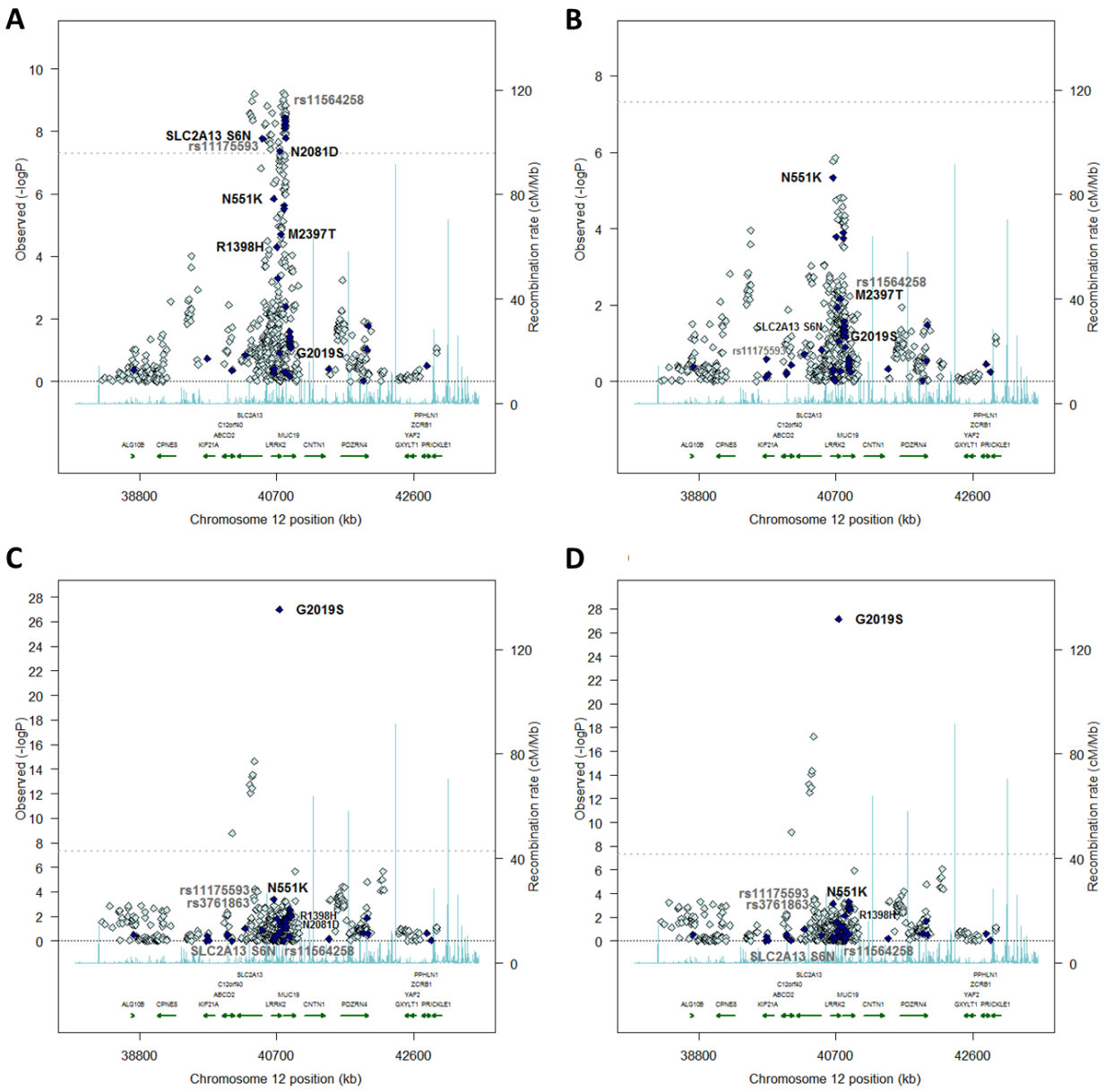


Figure S5: Single-point association with CD and PD in the AJ cohort using imputed genotypes within the *LRRK2* locus, conditioned and unconditioned on the CD-associated coding variant N2081D.

A. Unconditioned analysis of AJ CD. **B.** AJ CD analysis conditioned on N2081D genotypes. **C.** Unconditioned analysis of AJ PD. **D.** AJ PD analysis conditioned on N2081D genotypes. Dark points indicate non-synonymous variants.

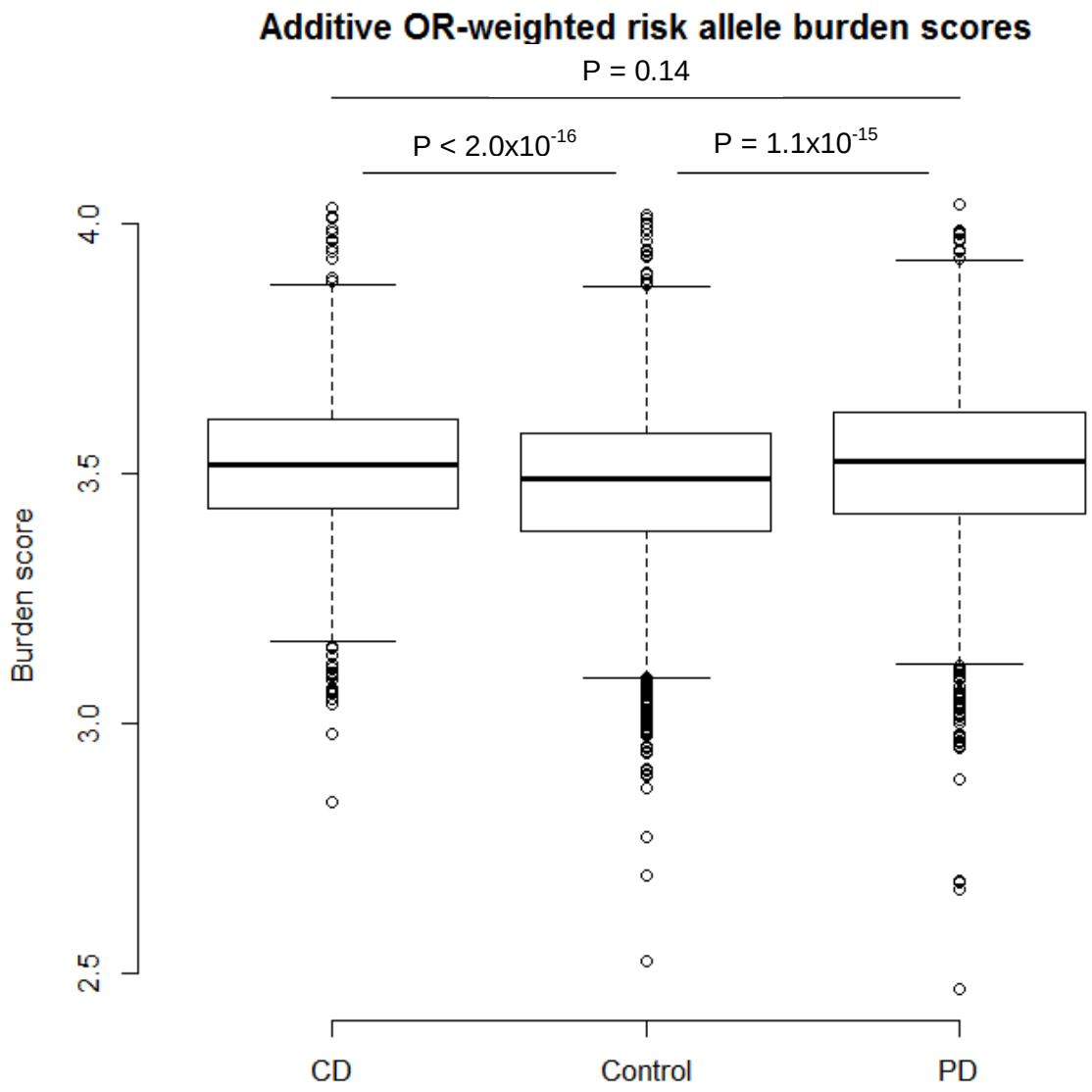


Figure S6: Log odds ratio-weighted additive risk allele burden scores.

The burden scores across the *LRRK2* risk alleles are significantly higher both CD and PD individuals compared to controls, indicating an overall similar genetic architecture throughout the *LRRK2* locus underlying both diseases. Pairwise two-sided *t*-test P-values are shown above the boxplots.

Supplementary tables

	Minor allele frequency (MAF) in case samples										
	0.0001	**0.001**	**0.005**	**0.01**	**0.015**	**0.02**	**0.025**	**0.03**	**0.04**	**0.05**	**0.1**
	Estimated power to detect a variant in sequencing of 97 independent chromosomes, by MAF										
	0.01	0.09	0.39	0.62	0.77	0.86	0.91	0.95	0.98	0.99	1.00

Odds ratio	Estimated Power to Detect Significance (α=0.001) of Single Markers in 1477 Cases and 2614 Controls, by MAF										
0.33	0.00	0.12	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.4	0.00	0.05	0.66	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	0.00	0.02	0.25	0.65	0.89	0.97	0.99	1.00	1.00	1.00	1.00
0.66	0.00	0.00	0.03	0.11	0.21	0.33	0.46	0.57	0.76	0.87	1.00
0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.03
1.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02
1.5	0.00	0.00	0.01	0.04	0.08	0.13	0.19	0.26	0.41	0.55	0.93
2	0.00	0.01	0.06	0.19	0.36	0.54	0.69	0.80	0.93	0.98	1.00
2.5	0.00	0.01	0.11	0.36	0.62	0.81	0.91	0.96	1.00	1.00	1.00
3	0.00	0.01	0.17	0.51	0.78	0.92	0.98	0.99	1.00	1.00	1.00

0% power to observe variant	100% power to observe variant
0% power to detect significant association	100% power to detect significant association

Table S1: Power Calculations

4,277	**Total exomic coding variants**
3,702	Missense substitutions
84	Nonsense substitutions
63	Splice variants
3,849	**Total single-nucleotide polymorphisms (SNPs)**
224	Frameshift mutations
204	In-frame indels
428	**Total indels**

Table S2: Ashkenazi Jewish-enriched exomic variants genotyped as custom content

AJ CD exome chip cohorts

Discovery

Cohort	CD	Control	Genotyping/sequencing platform
NIDDK IBDGC	323	89	Illumina HumanExome
ISSMS	352	51	
Yale University	268	-	
Cedars Sinai Medical Center	384	407	
Hebrew University of Jerusalem	-	1579	
Other	150	488	
Total	1477	2614	

Replication

Cohort	CD	Control	Genotyping/sequencing platform
NIDDK IBD GC	74	187	Illumina HiSeq
ISSMS	272	704	Illumina HiSeq
University College London	243	-	Illumina HiSeq
The Ashkenazi Genomic Consortium	-	128	Complete Genomics
Total	589	1019	

Imputation cohorts

Ashkenazi Jewish

Cohort	CD	PD	Control	Genotyping/sequencing platform
NIDDK IBDGC, ISSMS, CSMC, et al.	1477	-	2614	Illumina HumanExome
	313	-	206	Illumina HiSeq
Ashkenazi Jewish PD GWAS	-	1012	669	Affymetrix 550K, Affymetrix 6.0
	-	1095	805	Illumina 660k, Illumina Omni 1M
Hussman Institute of Human Genomics	-	22	15	Illumina Human610
NeuroGenetics Research Consortium	-	88	41	Illumina Omni1 Quad
PROGENI/GenePD	-	87	30	Illumina HumanCNV370
The Ashkenazi Genome Consortium	-	-	128	Complete Genomics
Total	1790	2304	4220[a]	

Non-Jewish European Ancestry

Cohort	CD	PD	Control	Genotyping/sequencing platform
NIDDK IBDGC	4748	-	4829	Illumina HumanExome
Hussman Institute of Human Genomics	-	563	603	Illumina Human610
NeuroGenetics Research Consortium	-	1893	1943	Illumina Omni1 Quad
PROGENI/GenePD	-	810	832	Illumina HumanCNV370
1000 Genomes CEU+TSI	-	-	180	Illumina GAII, Illumina HiSeq
Total	4748	3266	8387	

[a] Total is less than the sum of component study sample numbers due to sample overlap between studies

NIDDK IBDGC = National Institute of Diabetes and Digestive and Kidney Diseases Inflammatory Bowel Disease Genetics Consortium;

ISMMS = Icahn School of Medicine at Mount Sinai; CSMC = Cedars-Sinai Medical Center

CD, Crohn's disease; PD, Parkinson's disease, Control, unaffected individuals.

Table S3: Sample cohorts description

Table S4: All variants with AJ CD discovery P-values < 2 x 10^{-5}

Provided as a supplementary Excel file. The "Alternate models" tab shows results for χ^2-based models of association (Cochran-Armitage trend test, dominant allele, recessive allele, genotypic effect).

Haplotype	$Freq_{CD}$ (%)	$Freq_{control}$ (%)	P-value
N551/**2081D**/**M2397**	8.20	4.87	1.72×10^{-9}
N551/N2081/**M2397**	37.0	34.6	3.17×10^{-2}
551K/N2081/2397T	6.31	9.70	1.64×10^{-7}
N551/N2081/2397T	48.5	50.9	4.46×10^{-2}

Discovery cohort only; bold indicates minor alleles

Table S5: *LRRK2* phased haplotype association

Table S6: All imputed variants with nominal CD or PD association (P < 0.05) within the *LRRK2* region.

Provided as a supplementary Excel file.