Toxicology Reports 8 (2021) 1762–1768

Contents lists available at ScienceDirect

Toxicology Reports

journal homepage: www.elsevier.com/locate/toxrep

Replication of chromosomal loci involved in Parkinson's disease: A quantitative synthesis of GWAS

Dimitrios Rikos [a,b], Vasileios Siokas [a], Tatyana I. Burykina [c], Nikolaos Drakoulis [d], Efthimios Dardiotis [a], Elias Zintzaras [b,e,*]

[a] Department of Neurology, University Hospital of Larissa, Faculty of Medicine, University of Thessaly, Larissa, Greece
[b] Department of Biomathematics, Faculty of Medicine, University of Thessaly Larissa, Greece
[c] Department of Analytical and Forensic Medical Toxicology, Sechenov University, 119048 Moscow, Russian Federation
[d] Research Group of Clinical Pharmacology and Pharmacogenomics, Faculty of Pharmacy, School of Health Sciences, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, 17551 Athens, Greece
[e] The Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Tufts University School of Medicine, Boston, MA, United States

ARTICLE INFO

Handling Editor: Dr. Aristidis Tsatsakis

Keywords:
GWAS
Meta-analysis
Heterogeneity
Parkinson's disease
METRADISC-XL

ABSTRACT

Introduction: Parkinson's disease is a neurodegenerative disorder with a complex etiology coming from interactions between genetic and environmental factors. Research on Parkinson's disease genetics has been an effortful struggle, while new technologies and novel study designs served as indispensable boosters. Until now, 90 loci and 20 disease-causing gene mutations have been identified. In this study we describe a novel non-parametric approach to GWAS meta-analysis and its application in PD genetics.

Methods: A literature search was conducted to identify Genome-Wide Association Studies (GWAS) regarding Parkinson's disease. We applied predefined inclusion criteria and extracted the reported SNPs and their respective position and statistical significance. We divided all chromosomes in approximately equal genetic distance segments called bins and recorded the most significant SNP from each bin and each study and ranked them in terms of their *p*-value. Ranks from each bin were summed, averaged and added in a heterogeneity-based analysis using the METRADISC-XL software. Weighted and unweighted analysis was performed.

Results: Five-hundred and forty-three SNPs and their respective *p*-values from 15 studies were matched in their corresponding bins. The METRADISC-XL analysis resulted in 7 bins with a significant *p*-value. A bin on chromosome 4 where the SNCA gene is located found with genome-wide significant association with Parkinson's Disease.

Conclusion: This is the first time a non-parametric method is applied in GWAS meta-analysis. The results add some insight on the overall understanding of Parkinson's disease genetics and serve as a first step of further convergent analysis with Genome-wide linkage studies.

1. Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting 1% of the individuals over the age of 60 and 4% of the population older than 85 [1]. The disease has three core clinical characteristics, tremor, rigidity and bradykinesia, and numerous non-motor features that are now recognized to be present years before the manifestation of the typical parkinsonian syndrome, in a so-called prodromal phase. The disease's neuropathological hallmark is neurodegeneration in specific brain areas, mainly the substantia nigra, due to the accumulation of a-synuclein and other proteins [2].

The pathogenesis of the disease is still not fully understood, and it is considered a multifactorial disease, with both a genetic and an environmental component hence most PD cases are sporadic and only 5–10 % of PD patients suffer from a monogenic form. To date, at least 90 loci and 20 disease-causing genes for parkinsonism have been identified [3].

Genetic epidemiology is a relatively new scientific approach to investigating the role of genetic factors in determining disease in families and populations. Genetic linkage and association studies were followed by Genome Wide Linkage (GWLS) and Association Studies (GWAS) as new genotyping methods emerged, resulting in a large amount of data. Meta-analysis of available data has a major contribution

* Corresponding author at: Department of Biomathematics, Faculty of Medicine, University of Thessaly, Papakyriazi 22, 41222 Larissa, Greece.
E-mail addresses: zintza@med.uth.gr, zintza@uth.gr (E. Zintzaras).

https://doi.org/10.1016/j.toxrep.2021.10.008
Received 2 June 2021; Received in revised form 15 September 2021; Accepted 9 October 2021
Available online 12 October 2021

in revealing the true genetic component of sporadic disease under the common variant – common disease context. Meta-analyzing GWASs demands great computational effort and purpose-build software. Furthermore, since every GWAS may use different marker sets and genotyping platforms a classical meta-analysis approach should use genotype imputation with unclear effects on its performance [4]. In many cases the needed datasets to perform a GWAS meta-analysis are either incomplete or require the collaboration of many research teams globally to have access to the full spectrum of information.

In this study, we sought to perform a GWAS meta-analysis by implementing a comprehensive software (the METa-analysis of Ranked DISCovery datasets- METRADISC-XL), which can overcome the aforementioned limitations and to produce data that can be combined with relevant information from other study designs (GWLS and GWAS) as part of our team's effort to pursue a genomic convergence approach regarding Parkinson's disease. Previous similar approaches were made using the HEGESMA (Heterogeneity and Genome Search Meta Analysis) software and applied on Genome-wide scan meta-analyses [5–9]. In the case of GWAS meta-analysis though a larger number of markers (SNPs) and more missing values were anticipated, and the METRADISC-XL software was chosen to overcome these barriers. The METRADISC-XL (available online at http://biomath.med.uth.gr/metradisc/) is a software for non-parametric meta-analysis of ranked discovery datasets [10, 11] which is here used for the first time for this purpose.

2. Material and methods

2.1. Search strategy

A thorough literature search was conducted in online databases PubMed and EMBASE for GWAS concerning PD from its inception to the 30th of June 2020. Combinations of key words such as "Parkinson's disease"; "Genome-wide association study"; "GWAS"; "genome-wide"; "linkage disequilibrium"; "whole genome association" were used. To strengthen the depth and validity of our search, findings were compared and cross-validated with the HuGE navigator/GWAS integrator [12] and GWAS catalog [13] entries, where "Parkinson's disease" was selected as the trait of interest.

2.2. Inclusion criteria

Eligible for inclusion were English language studies which followed a classical GWAS approach with well-characterized sporadic Parkinson's disease cases and available association/statistical significance, in a genotype or most-significant level. Studies which examined other forms of PD (juvenile or early onset PD) or described associations with clinical characteristics (*e.g.* age at onset, motor and cognitive outcomes) or interactions (*e.g.* gene-environment interaction, coffee consumption) were considered ineligible. In case of overlapping samples, the study with the larger sample was included.

2.3. Data extraction

From each eligible study the following data were extracted: publication details (first author, year of study, title); number of cases and controls genotyped; all available SNPs either in the article, the supplemental files or the publicly available databases with their respective *p*-value and position. Only originally genotyped SNPs where included. Any replication results were discarded as with any overlapping samples.

2.4. Bins

All chromosomes were divided in approximately equal genetic distance segments called bins. Bin length was set to approximately 30 cM as usually used in the Genome Scan Meta-Analysis (GSMA) approach [8, 14]. The bins were coded by the number of the respective chromosome and the order of the bin in the form "chromosome.bin order". For example, bin 1.1 is the first bin of the first chromosome (Supplementary Table S1). The physical location of every bin (starting and ending base pair) was pinpointed by intergrading a Marshfield map and its respective DS markers, and the UCSC Genome Browser on Human (GRch38/hg38 Assembly).

2.5. SNPs matching

From each GWAS study, the most significant, in terms of reported *p*-value, SNP obtained within each bin was recorded. To facilitate this procedure due to the large number of entries, we matched the SNPs of each study to the corresponding chromosome and respective bin and finally recorded only the most significant, in a step-by-step algorithmic approach using original code in Python language through a Jupiter notebook. Original code is publicly available at https://dataverse.harvard.edu/ and https://github.com/ [15].

2.6. Heterogeneity based meta-analysis

For each study, the bins were ranked (1–120) according to their *p*-value significance. The smallest *p*-values were accredited the higher rank (120). Bins with no corresponding *p*-value were considered as missing values and attributed the code number "-99" to be recognized as such by the software [10,11]. When equal *p*-values were noted, we considered them as tied ranks and performed the mid-rank method *i.e.* they ranked by their median rank. The resulted ranks of each bin were summed and averaged across studies. The average rank of each bin (R) would serve as an indication of association or not of this bin with the trait, in this case Parkinson's Disease. To further strengthen this indication, we investigated the consistency of the results for the same bin across studies, namely the between studies heterogeneity. This was assessed using the Q statistic which is defined as the sum of the squared deviations from the mean of the ranks of each study [8,16].

To implement the above-mentioned methodology, we used the METa-analysis of Ranked DISCovery datasets (METRADISC-XL) software. The METRADISC-XL software is a generalization of the METRADISC software based on the same methodology as described previously and implemented in microarray meta-analysis [10,11]. In this case the biological variable of interest are the chromosomal bins. As described previously each bin from each study is ranked based on the most significant *p*-value. Since, due to missing values, different number of bins may be ranked at each study (which may be common amongst all studies or in some of them) these raw ranks are adjusted by the maximum number of tested bins (nmax) in any of the combined studies. Therefore, the ranks of each study are multiplied by the nmax divided by the number of ranked bins in this study.

The significance of the metrics (R and Q) is assessed using a Monte Carlo method. The ranks of each study are randomly permuted for several times (in this case 100.000 times) and the software calculates the simulated metrics to create null distributions for them. Since there are missing values (not all bins have available ranking in all the studies) each bin is tested against the null distribution corresponding to the group of bins having available information (rank) from the same studies. These groups are called information classes and they are defined by the missing data. The significance of the metrics is defined as the percentage of simulated metrics that exceed or are equal to the observed metric.

The METRADISC-XL software allows for both unweighted and weighted analysis. We performed both and in the case of weighted analysis we used the weight function $(n1i*n2i)/(n1i + n2i)$ where n1i is the number of cases and n2i the number of controls in study i.

3. Results

The database search resulted in 1.412 entries, 55 studies of which were initially selected as relevant. GWAS catalog under the trait

"Parkinson's disease" resulted in 22 studies which were cross-referenced with the GWAS integrator entries. After duplication removal and application of the selection criteria as described in the methods section, 19 GWASs with a total of 191.397 available-for-extraction SNPs, were selected for further analysis (Table 1) [17–35]. Taking into consideration the significant amount of missing values (less than 5 available SNPs) from the studies of Beecham [18], Davis [20], Satake [31] and Vacic [35] they were also removed from the final analysis resulting in a total of 15 included GWASs.

Application of the original code matched 543 SNPs and their respective *p*-values in their corresponding bins (the most significant SNP in terms of *p*-value in each bin and in each study) (Table 1) while 1.257 bins had missing values. Based on data availability from various studies these bins belonged to 92 information classes.

Application of the METRADISC-XL software for 15 studies, 120 bins, 100.000 permutations and 92 information classes revealed 7 statistically significant for association with-PD trait bins Fig. 1. Their corresponding right sided *p*-value for the adjusted R for both weighted and unweighted analysis and corresponding *p*-value of the Q statistic are shown on Table 2.

The bins with the most significant *p*-values in both weighted and unweighted analysis were bin 1.1 (chr1: 1-11404933), 3.2 (chr3: 30697536-88674208), 4.4 (chr4: 70530434-98736813), 12.2 (chr12:12486720-43878111), 17.2 (chr17: 12661253-42207671) and 17.3 (chr17:42207672-71465399). Bins 12.4 (chr12: 73700113-103294741) and 19.4 (chr19: 50558304-58617616) where significant in unweighted analysis only (Table 2). Heterogeneity metric was marginally low for bins 4.4 and 3.2 (right-sided *p*-value = 0,11) and rather large for the rest bins.

Table 1
Demographic characteristics of included studies.

No.	Author	Year	Initial sample size (cases/ controls)	Ethnicity	Extracted SNPs (n)	Matched bins (n)
1	Bandre-Ciga	2016	240/192	Caucasian	28	21
2	Beecham*	2013	484/ 1.145	Caucasian	1	1
3	Davis*	2013	31/767	Amish	3	3
4	Do	2011	3.426/ 29.624	Caucasian	390	65
5	Edwards	2010	604/619	Caucasian	72	33
6	Foo	2016	779/ 13.227	East Asian (Han Chinese)	96	32
7	Fung	2006	267/270	Caucasian	26	17
8	Hamza	2010	2.000/ 1.986	Caucasian	89	16
9	Hu Y	2015	250/250	Chinese	22	21
10	Liu	2011	268/178	Ashkenazi	55	32
11	Pickrell	2016	9.619/ 324.522	Caucasian	25	20
12	Saad	2010	1.039/ 1.984	Caucasian	50	21
13	Satake*	2009	988/ 2.521	Japanese	20	4
14	Simon-Sanchez	2009	1.713/ 3.978	Caucasian	345	87
15	Simon-Sanchez	2011	772/ 2.024	Caucasian	30	8
16	Spencer	2010	1.705/ 5.175	Caucasian	55	24
17	Vacic*	2014	1.130/ 2.611	Ashkenazi	4	4
18	Maraganore	2005	381/363	Caucasian	190059	120
19	Chang	2017	6.476/ 302.042	Caucasian	27	26
	Total		32208/ 693847		191397 ** (191369)	555 ** (543)

* Studies removed due to large number of missing values.
** The included data set.

4. Discussion

Exploration of the heritability in Parkinson's disease has been a long and fascinating journey with numerous successes and drawbacks. Technological advantages were a booster in this effort, while the complexity of the matter was, and still is, a holdback. Until now, 20 disease-causing genes and 90 SNPs have been identified to be associated with the risk of developing Parkinson's disease [36–40]. In this effort, GWASs and their meta-analyses have so far added insight of great value, consuming however great effort.

These approaches revealed a small, yet significant portion of the heritability of the disease. in GWASs or by implementing GWASs with other clinical phenotypes [36].

In this novel approach, we sought to investigate whether a quantitative synthesis is capable of effectively pooling available data from GWASs. Our goal was to identify genomic regions in a genome-wide-hypothesis free fashion, with significant pooled value serving to indicate candidate regions for further investigation.

This method is easy to be understood by clinicians and is not restricted by a distribution assumption nor by the different effect size measures or different techniques used in the initial GWASs. Nonparametric approaches have successfully been used in Genome Wide Linkage Scans and microarray meta-analysis [8,41]. Furthermore, in our effort to apply convergent genomics in PD, this is the first step to be followed by a similar meta-analysis of Genome-Wide linkage scans (Genome Scan Meta-Analysis, GSMA), and combine our findings based on the notion that "true" hits on both study designs have a better positive predictive value and serve as better candidate regions.

In this study we combined the initial, originally genotyped SNPs from each study. The combination of 534 SNPs and their ranking between 15 studies and 120 bins in 92 information classes using this methodology managed to result in one significant in the genome level bin (*p*-value<0,000042, threshold adjusted for 120 bins) and six bins with less significant association (*p*-value<0,05) with the trait in question. Forty-three of the top significant as initially genotyped and reported by the studies SNPs (n = 138) are located on a significant bin (Table 3).

The most significant recognized bin is 4.4 (chr4: 70530434-98736813). At least 22 SNPs were reported as top-ranking SNPs in their initial genotyping, from 10 different studies within this region (Table 3). In seven studies, this bin had SNPs with the most significant *p*-value thus assigned the maximum ranking (120) in our analysis. This resulted in a right-sided *p*-value for Q of 0,11. Furthermore, 37 out of the 67 SNPs reported as having an association with the PD trait in GWAS catalog (data downloaded on July 27, 2020) are located into bin 4.4. In this region rests the SNCA gene (Chr4: 89700345-89838315) which is a well-recognized risk gene for PD with very high confidence to represent an actual PD gene [36,42]. Bin 3.2 also showed some consistency among studies with ranking at the top quartile in 6 out of seven studies where data existed, but with an average rank *p*-value of 0,03.

Bins 17.2 and 17.3 where significant at the 0,01 level but with substantial heterogeneity. Bin 17.3 contains the MAPT gene, which also had been nominated association with increased PD risk [43,44]. This bin along with bin 4.4 may represent polymorphic risk loci were multiple common and rare risk alleles co-exist as described earlier [45].

This is the first time, to the best of our knowledge, of such an approach to GWAS meta-analysis being tested. Despite our enthusiasm, we should mention that this effort has some limitations. This method relies on the most significant statistical value in each bin from each study, and the consequent summation and averaging of their ranks. A great number of bins, though, remained without a designation due to missing values. GWAS datasets are reported to be publicly available but

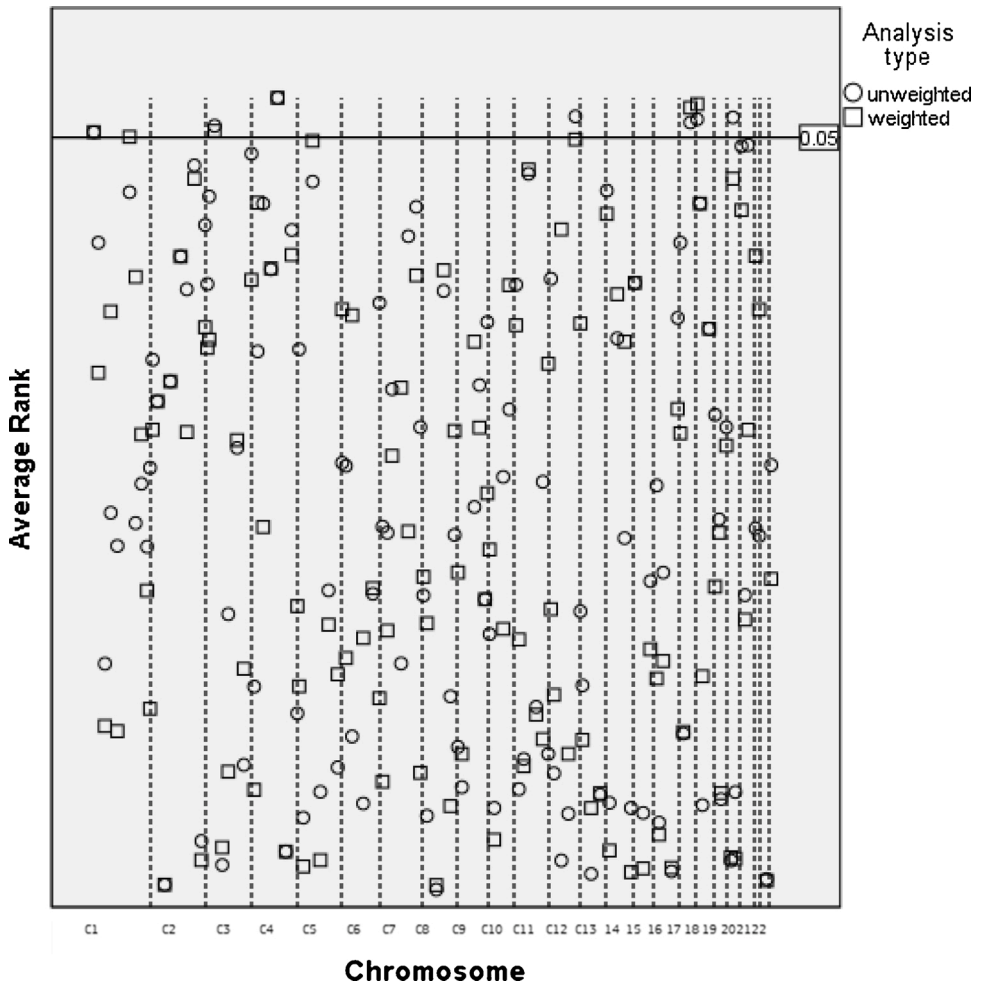

Fig. 1. Weighted (square) and unweighted (circle) significance level of the average ranks of 120 bins in size-adjusted chromosomes. Bins with significant p-value<0,05 are shown above the 0.05 (solid) reference line.

Table 2
Bins with high unweighted and/or weighted adjusted average ranks (Rmean, Rw/mean) and the corresponding significance and heterogeneity metrics.

	Bin	4.4	12.4	19.4	17.3	17.2	3.2	1.1
	study 1	657,1	−99,0	−99,0	680,0	662,9	−99,0	−99,0
	study 4	219,7	−99,0	−99,0	216,0	217,8	186,5	−99,0
	study 5	436,4	−99,0	−99,0	367,3	−99,0	378,2	−99,0
	study 6	450,0	−99,0	−99,0	−99,0	−99,0	435,0	−99,0
	study 7	−99,0	−99,0	−99,0	−99,0	840,0	−99,0	−99,0
	study 8	892,5	−99,0	−99,0	847,5	870,0	−99,0	−99,0
	study 9	628,6	−99,0	−99,0	−99,0	−99,0	−99,0	628,6
adjusted rank	study 10	−99,0	−99,0	−99,0	345,0	356,3	431,3	−99,0
	study 11	720,0	−99,0	−99,0	708,0	−99,0	−99,0	−99,0
	study 12	685,7	680,0	−99,0	−99,0	662,9	−99,0	−99,0
	study 14	158,2	−99,0	139,8	143,7	156,9	127,9	−99,0
	study 15	1800,0	−99,0	−99,0	−99,0	1755,0	−99,0	−99,0
	study 16	590,0	−99,0	−99,0	575,0	585,0	580,0	−99,0
	study 18	10,0	108,0	115,0	66,0	73,0	11,0	116,0
	study 19	553,8	−99,0	−99,0	549,2	−99,0	−99,0	−99,0
	Rmean	45141167	19968572	1496758	23183384	50969116	11496949	18539999
	right sided p-value for Rmean	0,00	0,02	0,02	0,03	0,03	0,04	0,04
	right sided p - value Rw/mean	0,00	0,05	0,10	0,01	0,01	0,04	0,04
	right sided p value for Q-mean	0,12	0,14	0,50	0,26	0,50	0,12	0,97

accessible only through consortia, specific organizations, and authorized users. In our study, the METRADISC-XL software can deal with missing values by creating null distributions from the same information class, yet such large amount of missing information drives to incomplete results. This discrepancy may have also contributed to the substantial observed heterogeneity. However, this approach is unlikely to generate false positive results.

Another issue is the well-known problem of matching a genetic map with a sequence-based physical map. Problems with assembly and incorrect identification of marker positions may lead to errors in the

Table 3
Top significant, initially genotyped, SNPs as reported from each study and their corresponding BINs. Study number corresponds to the number on Table 1. BINs in bold are the significant ones.

SNP	Position (bp)	Study No	BIN
rs12063142	18813023	5	1.2
rs1543467	86548977	5	1.5
rs17344386	83254011	18	
rs35749011	155135036	19	1.6
rs2986574	182173237	7	1.7
rs823118	205723572	1	1.8
rs823156	204031263	4	
rs823118	205754444	11	
rs823118	205723572	19	
rs10797576	232664611	1	1.9
rs849898	228153917	9	
rs870575	45356764	8	2.3
rs12613026	42867793	10	
rs10197606	41790447	18	
rs11887431	42179113	18	
rs11674789	41822751	18	
rs6430538	135539967	1	2.6
rs1955337	169129145	1	2.7
rs1474055	169110394	19	
rs11186	189897394	9	2.8
rs1010491	231160521	6	2.9
rs1561374	29458092	6	**3.2**
rs1684524	21936271	10	
rs1352135	21935471	10	
rs6783485	59427797	9	3.3
rs1879553	118615463	9	3.5
rs1879512	113576590	10	
rs7641311	113574386	10	
rs10513789	184242767	4	3.7
rs976683	173767581	5	
rs9290751	182732230	6	
rs12637471	182762437	19	
rs6599389	929113	4	4.1
rs356220	90860363	12	
rs1564282	842313	15	
rs34311866	951947	19	
rs4266290	15735495	11	4.2
rs4698412	15346446	12	
rs12502586	15335662	15	
rs2242330	68129844	7	4.3
rs6826751	68116450	7	
rs3775866	68126775	7	
rs356181	90626139	1	**4.4**
rs356220	90860363	4	
rs6812193	77418010	4	
rs356220	89720189	5	
rs356220	90641340	6	
rs8180209	90644454	6	
rs3775439	90709741	6	
rs6532194	90780902	6	
rs356220	89720189	8	
rs356220	90860363	8	
rs356168	90893454	8	
rs2736990	90897564	8	
rs1350855	91413829	8	
rs6812193	76277833	11	
rs2736990	90897564	12	
rs2736990	90897564	14	
rs3857059	90894261	14	
rs11931074	90858538	14	
rs2736990	90897564	15	
rs3857059	90894261	15	
rs11931074	90858538	15	
rs356182	90626111	19	
rs4862792	188438344	7	4.8
rs13153459	44515935	9	5.2
rs1916642	72488303	10	5.3
rs6879012	72498637	10	
rs26990	112814742	12	5.4
rs3129882	32517508	8	6.2
rs3129882	32441753	8	
rs4713118	27709015	11	
rs276555	137415146	6	6.5
rs6912319	137452537	6	
rs10256359	23084258	11	7.2
rs320682	137038092	6	7.5
rs17068332	3820589	16	8.2
rs16887478	38561200	18	8.3
rs10815285	5804424	18	9.1
rs10746953	76917840	9	9.3
rs2724788	12490835	16	10.2
rs1892302	12486578	16	
rs1480597	44481115	7	10.3
rs7097094	44530696	7	
rs10999501	72171365	10	10.4
rs188789342	119612816	11	10.5
rs117896735	121536327	19	10.6
rs12294719	36684837	12	11.2
rs1533588	36687460	12	
rs12419750	36589978	12	
rs7128419	36613848	12	
rs687432	57926788	11	11.3
rs10501570	84095494	7	11.4
rs329648	133765367	1	11.6
rs34637584	39020469	4	12.2
rs148294058	42655580	11	
rs1472402	40549297	18	
rs7954761	82691472	12	**12.4**
rs11060180	123303586	4	12.5
rs11060180	122819039	11	
rs11060180	123303586	19	
rs9513249	97507450	5	13.3
rs12870589	97572967	5	
rs9323124	47466177	9	14.2
rs11158026	55348869	19	
rs1816879	58318356	5	15.2
rs17463995	46791064	18	
rs1881335	5206420	10	16.1
rs4888984	78066835	7	16.3
rs11868035	17715101	1	**17.2**
rs12185268	41279463	4	
rs11868035	17655826	4	
rs281357	19683106	7	
rs199533	42184098	8	
rs199528	42198305	8	
rs17690703	41281077	12	
rs199533	42184098	14	
rs169201	42145386	14	
rs393152	41074926	14	
rs1981997	41412603	14	
rs2532274	41602941	14	
rs2532269	41605885	14	
rs8070723	41436901	14	
rs17563986	41347100	15	
rs1981997	41412603	15	
rs8070723	41436901	15	
rs2532274	41602941	15	
rs393152	41074926	15	
rs17649553	43994648	1	**17.3**
rs17649553	43994648	19	
rs1362858	32986600	9	18.2
rs12456492	40673380	1	18.3
rs4130047	38932233	4	
rs4130047	43098270	11	
rs1406968	19649880	5	20.1
rs3746736	23372613	18	20.2
rs1984279	23261192	18	
rs151358	57043454	10	20.4
rs2823357	15836776	4	21.1

order of markers on physical maps [46]. Finally, since this is a meta-analysis based on GWAS, it carries all the inherent disadvantages of its type.

5. Conclusions

Overall, this study is the first attempt to handle the GWAS meta-analysis with a non-parametric rank-based approach. Though several drawbacks may have limited the value of our results, this study adds some insight in the overall understanding of Parkinson's disease genetics and serves as a first step of further convergent analysis [47], while possibly introducing a new, useful tool to the scientific community.

Author statement

Dimitrios Rikos: Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original Draf **Vasileios Siokas**: Formal analysis, Investigation, Writing - Original Draf **Tatyana I, Burykina**: Investigation, Data Curation, Writing - Review & Editing **Nikolaos Drakoulis**: Investigation, Writing - Review & Editing, Visualization **Efthimios Dardiotis**: Investigation, Supervision **Elias Zintzaras**: Conceptualization, Methodology, Supervision

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Code availability

Code is publicly available at https://dataverse.harvard.edu/ and https://github.com/.

Availability of data and material

Data and material are available upon request by the corresponding author.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank professors Georgios Hadjigeorgiou and Georgia Xiromerisiou for their overall contribution and support. We highly appreciate Aris Andreou for his essential contribution in writing the code used as described in methods.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.toxrep.2021.10.008.

References

[1] O.B. Tysnes, A. Storstein, Epidemiology of Parkinson's disease, J. Neural Transm. (Vienna, Austria : 1996) 124 (8) (2017) 901–905.
[2] D.W. Dickson, Neuropathology of Parkinson disease, Parkinsonism Relat. Disord. 46 (Suppl 1) (2018). S30-s3.
[3] H. Deng, P. Wang, J. Jankovic, The genetics of Parkinson disease, Ageing Res. Rev. 42 (2018) 72–85.
[4] J. Li, Guo Y-f, Y. Pei, H.-W. Deng, The impact of imputation on meta-analysis of genome-wide association studies, PLoS One 7 (4) (2012).
[5] Y. Cao, M. Liao, X. Huang, Z. Mo, F. Gao, Meta-analysis of genome-wide linkage studies of atopic dermatitis. Dermatitis : contact, atopic, occupational, Drug 20 (4) (2009) 193–199.
[6] T.A. Trikalinos, A. Karvouni, E. Zintzaras, T. Ylisaukko-oja, L. Peltonen, I. Järvelä, et al., A heterogeneity-based genome search meta-analysis for autism-spectrum disorders, Mol. Psychiatry 11 (1) (2006) 29–36.
[7] M. Tziastoudi, I. Stefanidis, K. Stravodimos, E. Zintzaras, Identification of chromosomal regions linked to diabetic nephropathy: a meta-analysis of genome-wide linkage scans, Genet. Test. Mol. Biomarkers 23 (2) (2019) 105–117.
[8] E. Zintzaras, J.P. Ioannidis, HEGESMA: genome search meta-analysis and heterogeneity testing, Bioinformatics (Oxford, England) 21 (18) (2005) 3672–3673.
[9] E. Zintzaras, G. Kitsios, G.A. Harrison, H. Laivuori, K. Kivinen, J. Kere, et al., Heterogeneity-based genome search meta-analysis for preeclampsia, Hum. Genet. 120 (3) (2006) 360–370.
[10] E. Zintzaras, J.P. Ioannidis, Meta-analysis for ranked discovery datasets: theoretical framework and empirical demonstration for microarrays, Comput. Biol. Chem. 32 (1) (2008) 38–46.
[11] E. Zintzaras, J.P. Ioannidis, METRADISC-XL: a program for meta-analysis of multidimensional ranked discovery oriented datasets including microarrays, Comput. Methods Programs Biomed. 108 (3) (2012) 1243–1246.
[12] GWAS Integrator: a Bioinformatics Tool to Explore Human Genetic Associations Reported in Published Genome-wide Association Studies, 2011 [Internet]. Available from: https://phgkb.cdc.gov/PHGKB/gWAHitStartPage.action https://www.nature.com/articles/ejhg201191.pdf.
[13] The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019, 2019 [Internet]. Available from: https://www.ebi.ac.uk/gwas/home.
[14] E. Zintzaras, G. Kitsios, D. Kent, N.J. Camp, L. Atwood, P.N. Hopkins, et al., Genome-wide scans meta-analysis for pulse pressure, Hypertension (Dallas, Tex : 1979) 50 (3) (2007) 557–564.
[15] A. Andreou, A Jupyter Notebook Used to Match a List of Research Papers, Base Pairs and an Associated p-value to Their Corresponding Bin and Filter for the Smallest Value: andreouA./basepair_to_bin, 2019.
[16] E. Zintzaras, J.P. Ioannidis, Heterogeneity testing in meta-analysis of genome searches, Genet. Epidemiol. 28 (2) (2005) 123–137.
[17] S. Bandrés-Ciga, T.R. Price, F.J. Barrero, F. Escamilla-Sevilla, J. Pelegrina, S. Arepalli, et al., Genome-wide assessment of Parkinson's disease in a Southern Spanish population, Neurobiol. Aging 45 (213) (2016) e3–e9.
[18] G.W. Beecham, D.W. Dickson, W.K. Scott, E.R. Martin, G. Schellenberg, K. Nuytemans, et al., PARK10 is a major locus for sporadic neuropathologically confirmed Parkinson disease, Neurology 84 (10) (2015) 972–980.
[19] D. Chang, M.A. Nalls, I.B. Hallgrímsdóttir, J. Hunkapiller, M. van der Brug, F. Cai, et al., A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci, Nature 49 (10) (2017) 1511–1516.
[20] M.F. Davis, A.C. Cummings, L.N. D'Aoust, L. Jiang, D.R. Velez Edwards, R. Laux, et al., Parkinson disease loci in the mid-western Amish, Hum. Genet. 132 (11) (2013) 1213–1221.
[21] Tung J.Y. Do CB, E. Dorfman, A.K. Kiefer, E.M. Drabant, U. Francke, et al., Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease, PLoS Genet. 7 (6) (2011), e1002141.
[22] T.L. Edwards, W.K. Scott, C. Almonte, A. Burt, E.H. Powell, G.W. Beecham, et al., Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease, Ann. Hum. Genet. 74 (2) (2010) 97–109.
[23] J.N. Foo, L.C. Tan, I.D. Irwan, W.L. Au, H.Q. Low, K.M. Prakash, et al., Genome-wide association study of Parkinson's disease in East Asians, Hum. Mol. Genet. 26 (1) (2017) 226–232.
[24] H.C. Fung, S. Scholz, M. Matarin, J. Simón-Sánchez, D. Hernandez, A. Britton, et al., Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data, Lancet Neurol. 5 (11) (2006) 911–916.
[25] T.H. Hamza, C.P. Zabetian, A. Tenesa, A. Laederach, J. Montimurro, D. Yearout, et al., Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease, Nat. Genet. 42 (9) (2010) 781–785.
[26] Y. Hu, L. Deng, J. Zhang, X. Fang, P. Mei, X. Cao, et al., A pooling genome-wide association study combining a pathway analysis for typical sporadic Parkinson's disease in the Han Population of Chinese Mainland, Mol. Neurobiol. 53 (7) (2016) 4302–4318.
[27] X. Liu, R. Cheng, M. Verbitsky, S. Kisselev, A. Browne, H. Mejia-Sanatana, et al., Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population, BMC Med. Genet. 12 (2011) 104.
[28] D.M. Maraganore, M. de Andrade, T.G. Lesnick, K.J. Strain, M.J. Farrer, W. A. Rocca, et al., High-resolution whole-genome association study of Parkinson disease, Am. J. Hum. Genet. 77 (5) (2005) 685–693.
[29] J.K. Pickrell, T. Berisa, J.Z. Liu, Detection and interpretation of shared genetic influences on 42 human traits, nature 48 (7) (2016) 709–717.
[30] M. Saad, S. Lesage, A. Saint-Pierre, J.C. Corvol, D. Zelenika, J.C. Lambert, et al., Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population, Hum. Mol. Genet. 20 (3) (2011) 615–627.
[31] W. Satake, Y. Nakabayashi, I. Mizuta, Y. Hirota, C. Ito, M. Kubo, et al., Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease, Nat. Genet. 41 (12) (2009) 1303–1307.
[32] J. Simón-Sánchez, C. Schulte, J.M. Bras, M. Sharma, J.R. Gibbs, D. Berg, et al., Genome-wide association study reveals genetic risk underlying Parkinson's disease, Nat. Genet. 41 (12) (2009) 1308–1312.
[33] J. Simón-Sánchez, J.J. van Hilten, B. van de Warrenburg, B. Post, H.W. Berendse, S. Arepalli, et al., Genome-wide association study confirms extant PD risk loci among the Dutch, Eur. J. Hum. Genet. 19 (6) (2011) 655–661.
[34] C.C. Spencer, V. Plagnol, A. Strange, M. Gardner, C. Paisan-Ruiz, G. Band, et al., Dissection of the genetics of Parkinson's disease identifies an additional association

5' of SNCA and multiple associated haplotypes at 17q21, Hum. Mol. Genet. 20 (2) (2011) 345–353.
[35] V. Vacic, L.J. Ozelius, L.N. Clark, A. Bar-Shira, M. Gana-Weisz, T. Gurevich, et al., Genome-wide mapping of IBD segments in an Ashkenazi PD cohort identifies associated haplotypes, Hum. Mol. Genet. 23 (17) (2014) 4693–4702.
[36] C. Blauwendraat, M.A. Nalls, A.B. Singleton, The genetic architecture of Parkinson's disease, Lancet Neurol. 19 (2) (2020) 170–178.
[37] E. Dardiotis, D. Rikos, V. Siokas, A.M. Aloizou, Z. Tsouris, E. Sakalakis, et al., Assessment of TREM2 rs75932628 variant's association with Parkinson's disease in a Greek population and Meta-analysis of current data, Int. J. Neurosci. 131 (6) (2021) 544–548.
[38] M.A. Nalls, C. Blauwendraat, C.L. Vallerga, K. Heilbron, S. Bandres-Ciga, D. Chang, et al., Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies, Lancet Neurol. 18 (12) (2019) 1091–1102.
[39] V. Siokas, A.M. Aloizou, ADORA2A rs5760423 and CYP1A2 rs762551 polymorphisms as risk factors for Parkinson's disease, J. Clin. Med. 10 (3) (2021).
[40] V. Siokas, S. Arseniou, A.M. Aloizou, Z. Tsouris, I. Liampas, M. Sgantzos, et al., CD33 rs3865444 as a risk factor for Parkinson's disease, Neurosci. Lett. 748 (2021), 135709.
[41] X. Kong, V. Mas, K.J. Archer, A non-parametric meta-analysis approach for combining independent microarray datasets: application using two microarray datasets pertaining to chronic allograft nephropathy, BMC Genomics 9 (1) (2008) 98.
[42] G. Xiromerisiou, E. Dardiotis, V. Tsimourtou, P.M. Kountra, K.N. Paterakis, E. Z. Kapsalaki, et al., Genetic basis of Parkinson disease, Neurosurg. Focus 28 (1) (2010) E7.
[43] E. Pascale, M.E. Di Battista, A. Rubino, C. Purcaro, M. Valente, F. Fattapposta, et al., Genetic Architecture of MAPT Gene Region in Parkinson Disease Subtypes, Front. Cell. Neurosci. 10 (2016) 96.
[44] N. Seto-Salvia, J. Clarimon, J. Pagonabarraga, B. Pascual-Sedano, A. Campolongo, O. Combarros, et al., Dementia risk in Parkinson disease: disentangling the role of MAPT haplotypes, Arch. Neurol. 68 (3) (2011) 359–364.
[45] A. Singleton, J. Hardy, A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci, Hum. Mol. Genet. 20 (R2) (2011) R158–62.
[46] A.T. DeWan, A.R. Parrado, T.C. Matise, S.M. Leal, The map problem: a comparison of genetic and sequence-based physical maps, Am. J. Hum. Genet. 70 (1) (2002) 101–107.
[47] A.M. Aloizou, V. Siokas, E.M. Sapouni, N. Sita, I. Liampas, A.G. Brotis, et al., Parkinson's disease and pesticides: are microRNAs the missing link? Sci. Total Environ. 744 (2020), 140591.