

# JOURNAL OF BIOMEDICINE AND TRANSLATIONAL RESEARCH

Copyright©2021 by Faculty of Medicine Universitas Diponegoro and Indonesian Society of Human Genetics

Original Research Articles

## Patient-Driven Findings of Genetic Associations for PANS and PANDAS

Robert S. Horvath<sup>1\*</sup>, Samuel T. Keating<sup>2</sup>

<sup>1</sup>Father of a PANS child and Citizen Scientist, Canada

<sup>2</sup>Patient Advocate and Citizen Scientist, USA

### Article Info

History

Received: 19 Aug 2021

Accepted: 09 Nov 2021

Available: 31 Dec 2021

### Abstract

**Background:** There are presently very few genetic studies for PANS (Pediatric Acute-Onset Neuropsychiatric Syndrome) or PANDAS (Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcal Infections). More work in genetic associations for PANS and PANDAS (P/P) is needed to increase understanding of these debilitating childhood disorders that have a range of presentations.

**Objective:** This work represents a novel approach that aims to determine genetic associations between P/P and other diseases, disorders and traits (hereafter referred to as phenotypes).

**Methods:** Consumer genetic data (23andMe, AncestryDNA) for 155 patients with P/P were obtained from consenting parents over a period from 2018 to 2020. An analysis plan for this work was registered at Open Science Framework, additional genotypes imputed using Impute.me, and polygenic risk scores for 1,702 phenotypes calculated for each of the 155 P/P patients.

**Results:** One-sample t-tests performed across the 155 individual risk scores revealed that P/P is statistically significantly associated with 21 different groups of Single Nucleotide Polymorphisms (SNPs) that are in turn associated with 21 phenotypes. Some of the 21 phenotypes are previously known to be related to or associated with P/P: a group of SNPs associated with Tourette's Syndrome, and another group associated with Autism Spectrum Disorder or Schizophrenia, and a third associated with "feeling nervous" yielded t-tests with p values of  $1.2 \times 10^{-5}$ ,  $1.2 \times 10^{-11}$  and  $1.0 \times 10^{-5}$  respectively for association with the P/P data. This validated our analysis methodology. Our analysis also revealed novel genetic associations such as between P/P and plasma anti-thyroglobulin levels ( $p=1.3 \times 10^{-7}$ ), between P/P and triglycerides ( $p=5.6 \times 10^{-6}$ ), and between P/P and Lewy body disease ( $p=7.8 \times 10^{-6}$ ), inviting further investigation into the underlying etiology of P/P.

**Conclusion:** P/P is associated with many phenotypes not previously recognized as being connected to P/P. Further work on these connections can lead to better understanding of P/P.

**Keywords:** PANS Genetic Associations; PANS Polygenic Risk Score; Consumer Genetics Tests; Imputation

**Permalink/DOI:** <https://doi.org/10.14710/jbtr.v7i3.12082>

## INTRODUCTION

### Background

PANS and PANDAS present as a sudden onset of Obsessive-Compulsive Disorder (OCD) and/or tics and/or anorexia (food restriction) and other symptoms such as anxiety, emotional lability, irritability, regression in behavior and in school performance, motor and sensory abnormalities, sleep disturbances, and enuresis<sup>1,2</sup>.

PANDAS is associated with a streptococcal infection, whereas PANS may be associated with a number of other infections or environmental triggers.

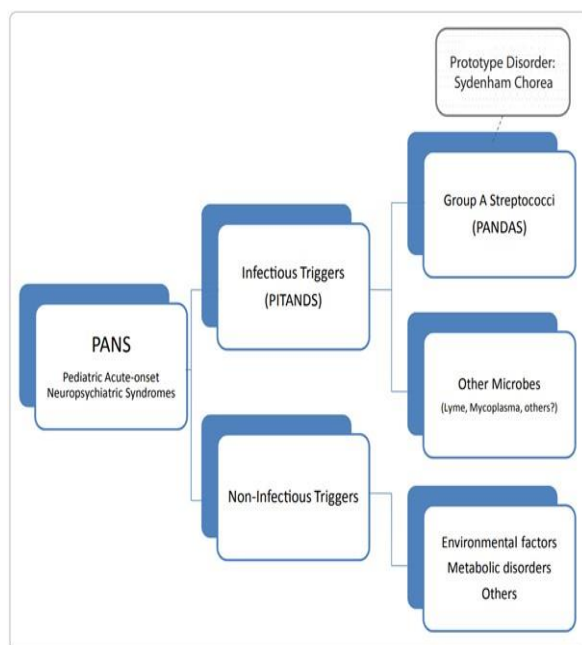
\* Corresponding author:

E-mail: [bobhorvath@alumni.uwaterloo.ca](mailto:bobhorvath@alumni.uwaterloo.ca)

(Robert S. Horvath)

Though some researchers take care in their studies to separate PANDAS from PANS and some also carefully evaluate cases being considered for acceptance<sup>3</sup>, it is the broad experience of the parents of P/P children and treating clinicians, that PANS and PANDAS seem closely related. The diagnoses of both disorders are clinical; there is no definitive biomarker for either. Some patients (36, or 27% of those in this study who have reported which diagnosis they have) have received both PANDAS and PANS diagnoses at different times, and indeed in some parents' experience, PANDAS (streptococcal-triggered) tends to evolve to have other triggers (as is the case for PANS) over a long course of the disorder. Because some patients do not present with typical streptococcal infection symptoms (sore throat), uninformed parents and clinicians, also unaware of the possibility of occult streptococcal infection, obtain neither traditional throat swabs for streptococcus bacteria, nor other tests such as tests of stool, when symptoms first appear. For this reason, many cases that could be PANDAS may have been diagnosed as PANS.

Despite subtle differences in the definitions of the two disorders, PANDAS is illustrated by some of the defining experts to be a subgroup of PANS (see Figure 1, from Swedo et al, 2012<sup>4</sup>).



**Figure 1:** Hierarchy of the Pediatric Acute onset Neuropsychiatric Syndrome © 2012 Swedo SE et al, used under a Creative Commons License<sup>4</sup>.

### Rationale

P/P has no biomarker for diagnosis, typically involves long courses of treatment, and has no established cure. The etiology of P/P is not yet well understood, but genetic studies could help improve understanding and ultimately advance treatment for these disorders. There are few genetic studies on PANS and PANDAS to date<sup>3,5-9</sup>.

### Objectives

This study aims to establish any genetic associations between P/P and other phenotypes, in the hope that

further studies of the associations can establish better understanding of P/P. The phenotypes evaluated are limited to the over 2000 listed at impute.me, but they are very broad in that some are diseases, some are disorders, and many are traits of all kinds. Any association has the potential to shed light on the etiology of P/P. This study is a first report of this kind of genetic association for P/P.

### Impute.me and Associations

The website www.impute.me allows consumer genetic data to be checked for errors, expanded (by imputation), and then analyzed to compute polygenic risk scores for various phenotypes<sup>10</sup>. It matches our analysis needs, given the data we have. Impute.me calculates, for an individual, z-scores for over 2,000 phenotypes. This score allows a user to see what fraction of the general population of the same race has more (or fewer) risk alleles among a group of SNPs associated with the phenotype, than the user's data. Additionally, the algorithms are available to run on local computers as a Docker container, which was a requirement for this study due to data contributor consent form wording.

The P/P cohort's polygenic risk z-scores for a phenotype, when averaged and compared via a t-test to the de facto population mean of zero, can determine if the SNPs associated with that phenotype are statistically significantly relevant in P/P. If they are, this link between P/P and a phenotype is a bona fide association, even though it is indirect. As is possible with any statistical association, the phenotype is linked to a third characteristic (the list of SNPs) which is in turn linked to P/P. There may or may not be a more direct connection between the phenotype and P/P. As with any association study, causes cannot be ascertained in this analysis, only links, with potentially one or more unknown links in the chain of connection.

### MATERIAL AND METHODS

#### Study Design: PANS and PANDAS Together

There has been a wide variation in willingness to diagnose PANS and particularly PANDAS among clinicians. Parents who are nevertheless convinced that their child has P/P (parents repeatedly witness dramatic symptoms that cannot all be video-recorded for clinicians' benefit) will seek other clinicians that could ultimately give them the diagnosis that was refused by one or more other health care providers. In this way, the data collected here could be described as having a low diagnostic bar, because the process of rejection and repeated searching for other clinicians for these diagnoses has been very common in the P/P community (33% see five or more clinicians prior to receiving a P/P diagnosis<sup>11</sup>).

In an effort to obtain results that could potentially be applied to the widest group of patients with the dramatic and very similar symptoms of PANS and PANDAS, this study not only groups PANS and PANDAS patients together, but offers no further scrutiny for the purpose of eliminating potentially borderline candidates. This is a strength of this study in that:

- 1) Results can be applicable to the widest array of patients

- 2) If any of the results actually pertain to a sub-group only, then they are particularly strong if they are detected as statistically significant when diluted by the rest of the participants outside the subgroup.

### Data Collection

The 155 consumer genetic data sets (23andMe, AncestryDNA) used in this analysis were collected directly by the lead author (not from any third party) in 3 different phases. According to the consent form agreed to by participants (who are all members of online self-help groups of parents of P/P children), this data cannot be made available to others. The first 71 data sets were obtained in late 2018 and early 2019, with the end of data collection defined by prior specification of a cutoff date in a plan registered prior to data collection<sup>12</sup>. One of those 71 was eliminated because it was determined to be from a younger sibling of one of the other participants. A second phase of data collection was completed in December 2019, when 5 consumer genetic data sets that were late to the phase 1 cutoff were combined with 65 newly obtained data sets, for a total of 70 in phase 2 (data collection completion for Phase 2 was defined in advance to be 70 data sets). A third phase of data collection was conducted in early 2020, with a pre-defined cutoff time on Feb. 1, 2020. All 3 phases involved uploading the consumer genetic data sets to the website GEDMatch, which was used to exclude closely related individuals for the subsequent analysis. The 3 phases of collection yielded a total of 152 data sets. Analyses performed after each of the phases were to investigate single SNP associations (not the subject of this report, nor formally published in a journal). Finally, for the present study, 3 additional data sets that were late to the phase 3 deadline were included, for a total of 155 consumer genetic data sets.

For the phase 1 data (70 data sets), it was not stipulated that a formal diagnosis by a licensed medical practitioner was required, as it was for the subsequent phase 2 and 3 data collection. However, by later contacting those parents who did not contribute anonymously, formal diagnoses were confirmed for 49 Phase 1 data sets, such that of the final 155 data sets, 134 are known to have formal diagnoses of PANS, PANDAS, or both (given at different times by different medical practitioners). None of the Phase 1 participants that responded to the question of diagnosis reported who they did not have a formal diagnosis from a licensed medical practitioner. Medical records were not obtained. Table 1 shows the distribution of diagnoses among the 155, as well as distributions of sex and race, the latter based on an interpretation of the GEDmatch Admixture tool “World-22”, which was used for 152 of the 155 individuals. The sex distribution is very similar to a survey<sup>13</sup> that included 698 PANS patients that has shown the incidence of P/P as higher in males (65%) than in females.

The data was collected from throughout the United States and Canada, with at least one data set from Europe, primarily through 10 Facebook (FB) support groups. Three of the FB sources were smaller regional groups, but the vast majority of the data were contributed from larger national (indeed, international) groups, all English-speaking. More information is available at [osf.io](http://osf.io)<sup>12</sup>.

**Table 1:** Data Distributions

Diagnosis	Number	percent
PANS	50	32%
PANDAS	48	31%
PANS & PANDAS	36	23%
unknown	21	14%
<b>Y Chromosome</b>		
Yes	94	61%
No	61	39%
<b>Population Mix</b>		
African American	1*	0.7%
Asian	3	2.0%
White	143	94.7%
Native American	5	3.3%

\*represents 2 individuals that are each ~50% African American

### Pre-registered Analysis Plan

A plan for analysis was created for this study and registered<sup>14</sup>. The data was collected long before the plan was conceived, but the plan was created prior to data analysis. Included in the plan was a stepwise approach, where a group of 52 phenotypes related to psychiatric conditions were pre-selected for a first analysis, and then a further group of 301 phenotypes that has known or suspected autoimmune aspects was pre-defined for a second analysis. The psychiatric and autoimmune lists are defined in the registration plan<sup>14</sup>, and the full list of the phenotypes investigated is shown in column A of Supplementary Table S1. Using these pre-defined list of hypotheses, a less stringent multiple comparison correction (using 52 for the number of comparisons) was done for the list of psychiatric phenotypes, and a multiple correction based on  $52+301 = 353$  comparisons was done for a list that also included the autoimmune-related phenotypes. Thirdly, all phenotypes were analyzed, using a correspondingly more stringent multiple comparison correction. The Holm-Šidák criteria for significance for each of these cases are calculated and shown in Table 2, according to the following formula:

$$1-(1-\alpha)^{1/(n+i)}$$

Where:

$\alpha$  = significance level, 0.025 for this two-tailed consideration  
 $n$  = number of multiple comparisons (e.g. 52 for psychiatric list)  
 $i$  = the ranked order of p-values ( $i=1$  for the lowest p-value)

Per the Holm-Šidák method, when the  $i^{\text{th}}$  lowest p-value is higher than the calculated criteria, no further p-values can be considered significant. As can be seen from Table 2, the Holm modification of the Šidák correction (that is, the slightly higher criteria for the 2<sup>nd</sup>-lowest p, 3<sup>rd</sup>-lowest p, etc.) yields a small change in the threshold for significance.

**Table 2:** Holm-Šidák Significance Criteria

No. of Comparisons	lowest p	2 <sup>nd</sup> lowest p	3 <sup>rd</sup> lowest p	4 <sup>th</sup> lowest p
52 (psychiatric list)	0.000487	0.000496	0.000506	0.000517
353 (Psych + Autoimmune)	7.172E-05	7.192E-05	7.213E-05	7.233E-05
1702 (all minus Ukbiobank)	1.488E-05	1.488E-05	1.489E-05	1.490E-05
2042 (all phenotypes)	1.240E-05	1.240E-05	1.241E-05	1.242E-05

**Table 3:** List of Phenotypes Found by This Study to be Associated with P/P

	p-value	direction
1 adverse response to drug (PMID 30420678)	9.42E-06	negative
2 alcohol consumption, drinks per week (PMID 28485404)	7.01E-07	negative
3 alcohol consumption drinks per week (PMID 30643258)	2.47E-07	negative
4 Alzheimer's disease and age of onset (PMID 26830138)	4.88E-12	negative
5 autism spectrum disorder or schizophrenia (PMID 28540026)	1.18E-11	positive
6 basal cell carcinoma (PMID 27539887)	9.15E-08	positive
7 basal cell carcinoma (PMID 31174203)	3.65E-06	positive
8 bulimia nervosa (PMID 23568457)	9.65E-05	negative
9 major depressive disorder (PMID 27479909)	2.46E-06	negative
10 eye color (PMID 23548203)	2.28E-09	negative
11 feeling nervous (PMID 29500382)	1.02E-05	positive
12 interleukin-1-beta levels (PMID 27989323)	1.88E-06	negative
13 Lewy body disease (PMID 25188341)	7.85E-06	positive
14 plasma anti-thyroglobulin levels (PMID 29678681)	1.30E-07	positive
15 post bronchodilator fev1 in COPD (PMID 26634245)	2.05E-27	negative
16 post bronchodilator fev1fvc ratio in COPD (PMID 26634245)	1.19E-06	negative
17 squamous cell carcinoma (PMID 31174203)	1.70E-05	positive
18 systolic blood pressure (PMID 27618448)	1.35E-05	negative
19 Tourette's syndrome (PMID 30818990)	1.17E-04	positive
20 triglycerides (PMID 23726366)	5.61E-06	negative
21 height	2.75E-14	positive

It should be noted that more than half of the phenotype's reported are not independent because they are repeated or related to others (see supplementary Table S1 for a list of all phenotypes investigated). These multiple studies therefore do not all have completely independent SNPs. This renders the significance criteria used as conservative.

### Impute.me Analysis

Details of the imputation and polygenic risk calculation is described elsewhere<sup>10</sup>. The execution of the algorithms was performed between January and early March, 2021 (the algorithms or number of diseases checked by impute.me may change in time). The original consent forms used during data collection allowed uploading of patient data only to one particular website (www.GEDmatch.com) and not to any others. Critical for the present work therefore was the fact that the impute.me algorithms could be implemented on a local computer, using a Docker container<sup>15</sup>. Each data set took approximately 6 hours to run the algorithms

locally; the majority of the computation involved the imputation steps.

### t-test vs z-test

The one-sample t-test mandated by the registration plan was performed after the .json files containing the z-scores (or polygenic risk scores) produced by the impute.me algorithms were parsed into delimited files that could be readily imported into a spreadsheet. This ultimately became the heart of supplementary Table S1 as columns F through FD - one column of these polygenic risk scores for each of the 155 genetic data sets from participants. Select t-test results performed with formulas within the spreadsheet of Table S1 (column B) were also replicated with R and online t-test calculators. It was determined during analysis that a z-test could have been proposed during registration, because the standard deviation of the population of z-scores was known (it is 1 by definition). That is, the t-test approximation of using the sample standard deviation could have been avoided for a stronger test. The z-test p-values derived across the 155 participants are shown in column C of supplementary

Table S1, one value for each phenotype (which are rows). The t-test results (shown in Column B of supplementary Table S1) that are significant are nevertheless reported here per the pre-registered plan; there are fewer true significance scores (21) than produced by the z-test (47). The t-test was clearly problematic for 31 or 9.1% of the UK-biobank phenotypes that produced false significance because of the number of identical z-scores among the 155 data sets. This was due to a preponderance of very low frequency alleles (<0.001) in the very large UK-biobank data sets and studies. These very low frequency alleles tend to not appear in a sample size of only 155, thus creating many identical non-zero (but close to zero) z-score values. The near-zero standard deviation of these repeated values produced a very high t-statistic (for which the standard deviation is in the denominator) and subsequently an artificially low p-value and false significance. When the z-test (with its standard deviation of 1) was instead used on the same data, the 31 UK-biobank p-values that inappropriately appeared significant with the t-test no longer met the criteria for significance. Rather than creating a new criteria outside of the registration plan to delineate which UK-biobank phenotypes were problematic and which were not, it was decided to simply reject all UK-biobank results for the t-test.

One other phenotype, “asparaginase-induced acute pancreatitis in acute lymphoblastic leukemia - onset time” was also excluded because it effectively included only one SNP (because of lack of calls of the remainder of the study’s SNPs among the 155 data sets). Single SNP associations were investigated previously with this data; the present study was established to investigate groups of SNPs associated with phenotypes.

## RESULTS

Associations were investigated in the stepwise fashion described in the registration plan created prior to analysis<sup>14</sup>. There were 4 associations found among the list of 52 psychiatric conditions, then a further 4 from the larger list of 353 that also included autoimmune diseases, and finally an additional 13 associations were found for the strongest significance levels created for all 1702 phenotypes that were analyzed by t-test. Note that the significance levels shown as the last row in Table 2 (for 2042 phenotypes) was reserved for an alternate z-test that incorporated all the UK-biobank studies. This z-test was performed for information only, as it was outside the registration plan.

Table 3 lists the 21 statistically significant associations that were indicated with the t-test. These are the results of this study. A positive direction means that P/P is associated with genetic variations that are in turn associated with an increase in the phenotype or its symptoms.

The full list of all phenotypes with both t-test results and alternate z-test results are shown in supplementary Table S1. The order of the 155 patients in Table S1 (shown as columns) is randomized so as to be different than the order in which their data was received, and also different in order than any of the previous studies that were performed with this data. This was done to help safeguard the anonymity of the patients.

## DISCUSSION

### *Discussion of Selected Associations with PANS/PANDAS*

#### **Alcohol-related Associations**

The two alcohol-related associations in Table 3 (items 2 and 3) may seem to be counter-intuitive in that they are in the negative direction. That is, P/P is inversely associated with genetics that are in turn associated with alcohol drinks per week, as if to suggest that P/P patients might have less genetic susceptibility to alcoholism than the general population. Though it would be expected that the strain of P/P symptoms, if carried into adulthood (which is often not the case) could be expected to create a propensity for escape in alcoholism or other drugs, or that P/P itself may be co-morbid with some mental health disorders that are linked positively to substance abuse, alternate z-tests results that were statistically significant for these two and also 5 additional alcohol-related studies were also consistently in the same negative direction. See supplementary Table S1 for the alternate z-test results, as well as all results for all phenotypes. This consistency is additional evidence (beyond the individual p-values obtained for the two t-tests) that these negative associations are not spurious, and should therefore not be dismissed. Note that the fact that most of of 25 alcohol-related studies in the phenotype list do not score as statistically significantly associated with P/P, does not negate the association established by those few that do. Not all alcohol-related SNPs need to correlate with P/P for there to be some association.

#### **Alzheimer's**

Though Alzheimer's is thought to be related to some autoimmune conditions, item 4 in Table 3 (Alzheimer disease and age of onset) was not included in the autoimmune list because age of onset and interaction studies were deliberately excluded per the registration plan. This phenotype was nonetheless statistically significantly associated with P/P in the negative direction, to the more stringent criteria applied to the set of all phenotypes.

#### **Autism Spectrum Disorder**

Autism Spectrum Disorder (ASD) or Schizophrenia (item 5 in Table 3) has a positive association with P/P through the group of SNPs determined in the study with PMID 26830138 to be associated with ASD or Schizophrenia. ASD is known to be associated with P/P<sup>16</sup>, so this result, along with the Tourette's study association (item 19) is evidence that this analysis can predict appropriately, as tics are also a symptom of P/P. There are 3 cancer-related diseases in Table 3 (items 6, 7 and 17) that are all positively associated with P/P through their SNP association. These were pre-chosen in the registration plan as suspected autoimmune-related cancers. Two other cancer-related phenotypes (one also among the pre-defined autoimmune list) were also found to be significantly associated with P/P among the z-test results, both also with a positive direction of association.

#### **Bulimia Nervosa**

The negative association of P/P with bulimia nervosa (item 8 in Table 3) is for the only study among the phenotype list that includes bulimia nervosa. Although

PANS have a listed symptom of food restriction (at a superficial level, contrary to the binge eating behavior of bulimia nervosa), it is also the case that bulimia nervosa has some association to OCD<sup>17</sup>, which is a major symptom of PANS.

### Depression

Major depressive disorder (item 9 in Table 3) is negatively associated with P/P. The former is related to low dopamine levels<sup>18</sup>, while excess dopamine transmission is theorized to be a mechanism in P/P<sup>2</sup>.

### Interleukin-1 $\beta$

Interleukin-1 $\beta$  levels (item 12) have had conflicting studies on association with OCD. A 2019 study indicates that higher levels are associated with OCD<sup>19</sup>, and a 1997 study indicates the opposite<sup>20</sup>. The association found in this study was in the negative direction.

### Post Bronchodilator Biometrics

Two post bronchodilator biometric traits (items 15 and 16 in Table 3) that indicate a strong association may be an association with chronic obstructive pulmonary disease itself, as these results are only applicable to those with that disease. These two associations are both in the negative direction. The first (item 15) is suspect because of the dramatically different (and not statistically significant) result with the alternate z-test method (whose p-value is shown in the supplementary Table S1 as 0.14). The 155 scores are fairly similar (generally, they are small negative scores), rendering the standard deviation of the sample of 155 (used in the t-test, but not in the z-test) as small, thus giving a large t-statistic and consequently a very low t-test p-value. Item 16 in Table 1 however does not suffer from this issue (the t-test and z-test scores are similar).

### Thyroglobulin, Lewy Body Disease and Triglyceride

Finally, SNPs associated with plasma thyroglobulin levels (item 14), Lewy Body disease (item 13) and triglyceride (item 20) are interesting new associations with P/P.

### Limitations

The results of this work are limited to the 2,042 phenotypes for which polygenic risk scores could be calculated at impute.me in early 2021. There could be disorders and traits that are suspected to be connected to P/P that are not evaluated at impute.me. The phenotypes incorporated at impute.me at the time of this analysis are shown in the first column of supplementary Table S1. Lack of statistically significant connections to impute.me's phenotypes that are not listed in Table 3 does not prove that there is no connection between P/P and any of these. Repeating the present analysis with more P/P participants, including whole genome data, and also analysis of other groups of SNPs found in other/future studies not incorporated at impute.me in early 2021, could lead to associations not indicated in the present work. Individual P/P patients may not themselves have an unusual polygenic risk score for many of the phenotypes in Table 3; rather, this group of 155 on the whole has an unusual average tally for each of the entries in Table 3. For example, for the first

phenotype listed in Table 3 (Adverse Response to Drug, PMID 30420678), though the average of the 155 z-scores is -0.37, there are some individuals with high positive scores (e.g. one with a very high z-score of +3.55) that clearly contradict the overall trend for this trait by this P/P group. Results cannot necessarily be applied to a particular individual, and none of the trends reported (i.e. the positive or negative direction of association) are so general as to apply to all of the 155 candidates.

Most of the 2,130 rows in Table S1 that report phenotypes have an aspect that appears multiple times in the list, either because there have been multiple studies of the same phenotype over time, or because the phenotype is combined with others, or includes an interaction. For example, alcohol appeared in 25 studies for which t-test scores have been obtained in this study, giving alcohol-related traits a better chance of being found to be significant than, for example, Sjogren's syndrome, which was evaluated in only one study. This limitation will decrease slightly as additional traits are added to impute.me's list.

Finally, because the participants are disproportionately white, it is unknown whether these results hold for other races.

### CONCLUSION

Calculation of polygenic risk scores for a broad range of diseases, disorders or traits for a group of patients with a common diagnosis (in this case, PANS/PANDAS) is a novel way of finding new associations for that diagnosis. This approach is related to standard calculations of genetic correlation, but has the advantage of feasibility with the consumer genetics data that we have access to as citizen scientists. Evidence for the validity of this technique has been demonstrated in that the known common feature of Tourette's and of P/P, namely tics (there are also other connections<sup>21</sup>), suggests that the association found in this work between P/P and Tourette's is appropriate. An anecdotally known connection between P/P and autism has been recognized by researchers<sup>16</sup>, and therefore the connection between P/P and autism found in this study (albeit with schizophrenia also included) further contributes evidence for the validity of this kind of association study.

All associations determined in this study are with P/P alone, and not among or between other traits listed in this study. New associations have been found between P/P and some phenotypes, including Plasma anti-thyroglobulin levels, Lewy body disease, Basal cell carcinoma and squamous cell carcinoma and Triglyceride levels in blood, among others. The associations of the above traits to P/P are not direct, and should not be misinterpreted as such. For example, this study does not provide direct evidence of plasma anti-thyroglobulin levels, only the indirect linkage of some genetics commonality. The maximum polygenic risk score among the 155 participants for the trait of plasma anti-thyroglobulin levels is 3.1, and the minimum is -2.5, so this result should not be taken to indicate candidacy for a biomarker for P/P. The implication of this study is that this novel association, and that of P/P to other traits shown in Table 3, should be investigated both clinically, and by analysis of biological pathways that could be common to the trait and P/P. It is hoped that further study

and investigations of these new associations might provide new information that could contribute to further understanding of these debilitating childhood disorders of P/P.

#### ACKNOWLEDGEMENTS

There were no grants or funding support of any kind for this work. There are 154 parents that trusted a fellow P/P parent with the genetic information of their children. This work is dedicated to those parents who contributed data, and to all those who struggle with P/P. The authors would like to thank Dr. Lasse Folkerson, the genetics lead contact at impute.me for contributing to our analysis plan, for his help in establishing a working docker container for impute.me on a Windows computer, for troubleshooting the issue with UKbiobank results, and for his helpful book Understand Your DNA – a Guide. Michaela Holden, whose question on a Facebook group became the impetus for collecting data, is also appreciated for her significant involvement in the early years of data collection and individual SNP assessment. Dr. Bettina Ryll has been an inspiration in patient advocacy through her enthusiasm and advice.

#### REFERENCES

- PANDAS physicians network [Internet]. Pandasppn.org. 2013 [cited 2021 Oct 2]. Available from: <https://www.pandasppn.org/>
- Chang K, Frankovich J, Cooperstock M, Cunningham MW, Latimer ME, Murphy TK, et al. Clinical evaluation of youth with Pediatric Acute-onset Neuropsychiatric Syndrome (PANS): recommendations from the 2013 PANS Consensus Conference. *J Child Adolesc Psychopharmacol*. 2015;25(1):3–13.
- Fernell E, Sundin M, Fasth A, Dinkler L, Galazka M, Gillberg C, et al. Paediatric Acute onset Neuropsychiatric Syndrome: Exploratory study finds no evidence of HLA class II association but high rate of autoimmunity in first-degree relatives. *Acta Paediatr* [Internet]. 2021;(apa.15805). Available from: <http://dx.doi.org/10.1111/apa.15805>
- E. Swedo S. From research subgroup to clinical syndrome: Modifying the PANDAS criteria to describe PANS (pediatric acute-onset neuropsychiatric syndrome). *Pediatr Ther*. 2012;02(02):2.
- Karacao?lan G, Lüleyap Ü, Tahiro?lu AY, Pazarc? P, Yaman A, Çetiner S, et al. Streptokok enfeksiyonu ile ilişkili pediatrik otoimmün nöropsikiyatrik hastal?k (PANDAS) ile insan lökösit antijen (HLA) ilişkisi. *Cukurova Med J*. 2019;44(1):52–8.
- Luleyap HU, Onatoglu D, Yilmaz MB, Alptekin D, Tahiroglu AY, Cetiner S, et al. Association between pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections disease and tumor necrosis factor- $\alpha$  gene-308 g/a, -850 c/t polymorphisms in 4-12-year-old children in Adana/Turkey. *Indian J Hum Genet*. 2013;19(2):196–201.
- Çelik GG, Taş DA, Tahiroglu AY, Erken E, Seydao?lu G, Ray PÇ, et al. Mannose-binding lectin 2 gene polymorphism in PANDAS patients. *Noro Psikiyatr Ars*. 2019;56(2):99–105.
- Trifiletti R, Lachman HM, Manusama O, Zheng D, Spalice A, Chiurazzi P, et al. Identification of ultra-rare genetic variants in Pediatric Acute Onset Neuropsychiatric Syndrome (PANS) by exome and whole genome sequencing [Internet]. *bioRxiv*. 2021 [cited 2021 Jun 25]. p. 2021.05.25.21257256. Available from: <https://www.medrxiv.org/content/10.1101/2021.05.25.21257256v1>
- Genetic association study for PANDAS and PANS: Phase 3. 2020 [cited 2021 Jun 25]; Available from: <https://osf.io/53yve/>
- Folkersen L, Pain O, Ingason A, Werge T, Lewis CM, Austin J. Impute.Me: An open-source, non-profit tool for using data from direct-to-consumer genetic testing to calculate and interpret polygenic risk scores. *Front Genet*. 2020;11:578.
- NEPANS – PANS PANDAS Facts [Internet] Nepans.org [cited 2021 Oct 3]. Available from <http://www.nepans.org/panspandas-info.html>
- OSF [Internet]. *osf.io*. [cited 2021 Oct 3]. Available from: <https://osf.io/m2k6v/>. See also <https://osf.io/mbxvy/>, <https://osf.io/6rcj7/> and <https://osf.io/53yve/> for in-depth descriptions of data collection.
- Calaprice D, Tona J, Parker-Athill EC, Murphy TK. A survey of pediatric acute-onset neuropsychiatric syndrome characteristics and course. *J Child Adolesc Psychopharmacol*. 2017;27(7):607–18.
- OSF [Internet]. *osf.io*. [cited 2021 Oct 3]. Available from: <https://osf.io/ck5un/>
- Docker Hub [Internet]. *hub.docker.com*. [cited 2021 Jun 2]. Available from: <https://hub.docker.com/lassefolkersen/impute-me/>
- Sala R, Amet L, Blagojevic-Stokic N, Shattock P, Whiteley P. Bridging the gap between physical health and autism spectrum disorder. *Neuropsychiatr Dis Treat*. 2020;16:1605–18.
- Thiel A, Broocks A, Ohlmeier M, Jacoby GE, Schüssler G. Obsessive-compulsive disorder among patients with anorexia nervosa and bulimia nervosa. *Am J Psychiatry*. 1995;152(1):72–5.
- Dunlop BW, Nemeroff CB. The role of dopamine in the pathophysiology of depression. *Arch Gen Psychiatry*. 2007;64(3):327–37.
- Karagüzel EÖ, Arslan FC, Uysal EK, Demir S, Aykut DS, Tat M, et al. Blood levels of interleukin-1 beta, interleukin-6 and tumor necrosis factor-alpha and cognitive functions in patients with obsessive compulsive disorder. *Compr Psychiatry*. 2019;89:61–6.
- Brambilla F, Perna G, Bellodi L, Arancio C, Bertani A, Perini G, et al. Plasma interleukin-1 beta and tumor necrosis factor concentrations in obsessive-compulsive disorders. *Biol Psychiatry*. 1997;42(11):976–81.
- Church AJ, Dale RC, Lees AJ, Giovannoni G, Robertson MM. Tourette's syndrome: a cross sectional study to examine the PANDAS hypothesis. *J Neurol Neurosurg Psychiatry*. 2003;74(5):602–7.