博士論文

論文題目　Identification of genetic factors of sleep disorders

（睡眠障害の遺伝要因の探索）

氏　　名　　山崎　茉莉亜

Contents

Abstract

Abbreviations

I. General introduction

II. Estimating heritability in sleep disorders

1. Introduction

2. Materials and methods

3. Results

4. Discussion

III. Genome-wide CNV (copy number variation) study of sleep disorders in the Japanese population

1. Introduction

2. Materials and methods

3. Results

4. Discussion

IV. Conclusion

V. Acknowledgements

References

Tables and Figures

Abstract

In humans, narcolepsy is a sleep disorder that is characterized by sleepiness, cataplexy, and rapid eye movement (REM) sleep abnormalities. Essential hypersomnia (EHS) is another type of sleep disorders, characterized by excessive daytime sleepiness without cataplexy. A human leukocyte antigen (HLA) class II allele, specifically the *HLA-DQB1*06:02*, is a major genetic factor for narcolepsy. Almost all narcolepsy patients possess this *HLA* allele, while 30–50% of EHS patients and 12% of healthy individuals in Japan carry this *HLA* allele. The pathogenesis of narcolepsy and EHS is thought to be partially shared.

In order to evaluate contribution of common SNPs on narcolepsy onset and common genetic background among sleep disorders, I conducted a polygenic analysis, using 393 narcoleptic patients, 38 EHS with *HLA-DQB1*06:02* patients, 119 EHS without *HLA-DQB1*06:02* patients, other neuropsychiatric disorders, including 376 panic disorders, 213 autism and 56 schizophrenia, and 1,582 healthy individuals. Narcolepsy heritability was estimated as 58.1% ($P_{HLA\text{-}DQB1*06:02} = 2.30\times10^{-48}$, $P_{others} = 6.73\times10^{-2}$) including *HLA-DQB1*06:02* effects and 1.3% ($P_{others} = 2.43\times10^{-2}$) excluding *HLA-DQB1*06:02* effects. The results also indicated that common small-effect SNPs contributed to the development of narcolepsy. Reported susceptible SNPs for

narcolepsy in the Japanese population, such as *CPT1B*, *TRA@*, *P2RY11*, were found to explain 0.8% of narcolepsy heritability (P_{others} = 9.74×10^{-2}). EHS with *HLA-DQB1*06:02* were estimated to have higher shared genetic background with narcolepsy than EHS without *HLA-DQB1*06:02* even if the effects of *HLA-DQB1*06:02* were excluded (EHS with *HLA-DQB1*06:02*: 40.4%, $P_{HLA\text{-}DQB1*06:02}$ = 7.02×10^{-14}, P_{others} = 1.34×10^{-1}, EHS without *HLA-DQB1*06:02*: 0.4%, P_{others} = 3.06×10^{-1}).

In this study, I have investigated possible associations between rare large-scale copy number variations (CNVs) and sleep disorders, including 327 narcoleptic patients, 33 EHS with *HLA-DQB1*06:02* patients, 104 EHS without *HLA-DQB1*06:02* patients and 459 healthy individuals. Overall, significant enrichments of rare large-scale CNVs (frequency ≤ 1%, size ≥ 100 kb) in the patients were found (narcolepsy: case-control ratio of CNV count = 1.54, P = 5.00×10^{-4}, EHS with *HLA-DQB1*06:02*: case-control ratio of CNV count = 1.46, P = 3.24×10^{-2}, EHS without *HLA-DQB1*06:02*: case-control ratio of CNV count = 1.43, P = 2.17×10^{-3}). In particular, rare large-scale CNVs in *PARK2* showed significant associations with narcolepsy and EHS without *HLA-DQB1*06:02* (narcolepsy: P = 3.07×10^{-2}, EHS without *HLA-DQB1*06:02*: P = 2.91×10^{-2}).

Abbreviations

ADHD attention-deficit hyperactivity disorder

bp base-pair

BAF b allele frequency

CDCV common disease-common variant

CDRV common disease-rare variant

CNVs copy number variations

CPT1B carnitine palmitoyltransferase 1B

CTSH cathepsin H

CSF cerebrospinal fluid

EHS essential hypersomnia

FDR false discovery rate

GWAS genome-wide association study

Hcrt2 hypocretin (orexin) receptor 2

HLA human leukocyte antigen

HMM hidden markov model

IL10RB interleukin 10 receptor, beta

LRR log R ratio

MSLT multiple sleep latency test

PARK2 parkinson protein 2, E3 ubiquitin protein ligase

P2RY11 purinergic receptor P2Y, G-protein coupled, 11

REM rapid eye movement

SLE systemic lupus erythematosus

SOREMP sleep onset rapid eye movement period

TNFSF4 tumor necrosis factor superfamily member 4

TRA@ T cell receptor alpha locus

TRB@ T cell receptor beta locus

ZNF365 zinc finger protein 365

I General introduction

Narcolepsy with cataplexy (narcolepsy) is a sleep disorder with characteristics of excessive daytime sleepiness, cataplexy (sudden loss of muscle tone in response to strong emotion), and pathological manifestations of rapid eye movement (REM) sleep—including hypnagogic hallucinations, sleep paralysis, and sleep-onset REM sleep. Onset of narcolepsy is usually around adolescence, and prevalence rate of the disease in Japan is 0.16%-0.18%, while the counterpart in the United States and Europe is 0.02%-0.06%: men and women are equally affected[1, 2]. Relative risk of narcolepsy in first-degree family members of patients with narcolepsy is 10- to 40- fold higher than that in the general population[1].

Multiple factors are associated with the development of narcolepsy, including genetic variations at several loci. Initially, it was discovered that narcolepsy was tightly associated with a human leukocyte antigen (*HLA*) class II allele, specifically *HLA-DQB1*06:02*[3-7]. Eighty five to one hundred percent of patients carry *HLA-DQB1*06:02*. This *HLA* allele is thought to be necessary, but not sufficient, for the development of narcolepsy because 10-40% of individuals in the general population carry *HLA-DQB1*06:02*. Recent studies, utilizing genome-wide association studies (GWASs), have identified several narcolepsy susceptibility loci. A locus near *CPT1B* (carnitine palmitoyltransferase 1B), which indicated a new pathogenic mechanism

related to fatty acid oxidation, was detected[8]. *TRA@* (T cell receptor alpha), *TRB@* (T cell receptor beta), *P2RY11* (purinergic receptor P2Y, G-protein coupled, 11), *CTSH* (cathepsin H), *TNFSF4* (tumor necrosis factor superfamily member 4), *ZNF365* (zinc finger protein 365) and *IL10RB* (Interleukin 10 receptor, beta) have been also identified[9-12]. These discoveries provided the evidence that narcolepsy is caused by disruption of autoimmunity.

The destruction of hypothalamic neurons that secretes a wake-promoting neuropeptide hypocretin (orexin) causes narcolepsy[13, 14]. Most of the patients with narcolepsy have low or undetectable levels of hypocretin-1 in the cerebrospinal fluid (CSF)[15, 16]. However, it has been reported that mutations and polymorphisms in prepro-hypocretin and hypocretin-receptor genes cannot explain onset of narcolepsy, except but rare cases[17-20].

Diagnoses of narcolepsy, according to International Classification of Sleep Disorders 2nd edition (ICSD-2) (AASM2005), follows: (i) the patient has a complaint of excessive daytime sleepiness occurring almost daily for at least three months; (ii) a definite history of cataplexy, defined as sudden and transient episodes of loss of motor tone triggered by emotions, is present; (iii) the diagnosis of narcolepsy with cataplexy should, whenever possible, be confirmed by nocturnal polysomnography followed by an

multiple sleep latency test (MSLT); (iv) the hypersomnia is not better explained by another sleep disorder or neurological disorder, mental disorder, medication use, or substance use disorder.

Essential hypersomnia (EHS) is another type of sleep disorders, characterized by excessive daytime sleepiness. However, the patients do not exhibit cataplexy. Both genetic and environmental factors are considered to contribute to the development of EHS[15, 21]. EHS was diagnosed based on the following three clinical items in central nervous system hypersomnias: (i) recurrent daytime sleep episodes that occur basically everyday over a period of at least six months; (ii) absence of cataplexy; (iii) the hypersomnia is not better explained by another sleep disorder, medical or neurological disorder, mental disorder, medication use or substance use disorder[22-25]. The diagnostic criteria for EHS correspond to narcolepsy without cataplexy and a part of idiopathic hypersomnia without long sleep time if I apply the criteria according to ICSD-2.

Narcolepsy without cataplexy and idiopathic hypersomnia without long sleep time do not exhibit cataplexy. Excessive daytime sleepiness of narcolepsy without cataplexy is typically associated with naps that are refreshing nature. Meanwhile, that of idiopathic hypersomnia without long sleep time results in unintended naps with or

without refreshing nature. EHS includes patients with short refreshing naps in idiopathic hypersomnia without long sleep time. The nocturnal sleep and total amount of sleep are basically normal in both disorders. MSLT is required for the diagnosis in ICSD-2. Idiopathic hypersomnia without long sleep time is differentiated from narcolepsy without cataplexy by the absence of REM-related features, most notably two or more sleep-onset REM periods (SOREMPs) on MSLT. However, I consider that the nature of naps (short or long, refreshing or non-refreshing) is more appropriate to distinguish a group of patients for study than MSLT, because MSLT is not optimized for the diagnosis of related hypersomnia patients using the number of SOREMPs even though MSLT is useful for the diagnosis of narcolepsy with cataplexy. In addition, MSLT requires a whole day and impose an economic burden to patients. Therefore, I focused on the symptoms themselves and adopted the definition of EHS.

Previous studies have reported that EHS and narcolepsy share some of susceptibility genes[22-26]. Especially, about 30–50% of EHS patients carry *HLA-DQB1*06:02*[22, 23]. In addition, EHS is associated with narcolepsy susceptibility SNPs within *CPT1B* and *TRA@*[22, 26]. The pathogenesis of EHS is considered to be partially common with that of narcolepsy[25].

In the first section, I evaluated the contribution of common variants for

narcolepsy onset. This estimation will be useful for making future strategy, how to uncover reaming genetic factors[27, 28]. Also I assessed the extent of genetic share between narcolepsy and EHS. The assessment of the genetic share between narcolepsy and EHS can provide “phenotypic classification” based on genetic basis.

In the second section of this study, I tried to detect rare susceptible variants, especially copy number variations (CNVs). Majority of human narcolepsy and EHS genetic research targeted to discover common susceptible variants based on common disease-common variant hypothesis (CDCV hypothesis)[8-10, 22, 26, 29]. It was important to evaluate the possibility that genetic architecture of narcolepsy and EHS might be based on common disease-rare variant hypothesis (CDRV hypothesis)[29].

Ⅱ Estimating heritability in sleep disorders

1. Introduction

In this section, I evaluated contribution of common variants for narcolepsy onset[29]. This will be useful for making future strategy, how to uncover reaming genetic factors[27, 28]. Also I assessed common genetic background of narcolepsy and EHS, based on CDCV hypothesis. Previous studies have reported that EHS and narcolepsy share some of susceptibility genes[22-26]. Therefore, the pathogenesis of EHS is thought to be partially common with that of narcolepsy. However, the extent of the shared genetic background has not been investigated, thus this assessment aimed to estimate the common genetic basis among the sleep disorders.

A large number of GWASs were conducted for various complex diseases and phenotypes until now, and they discovered a certain number of susceptible SNPs to many traits. However, discovered SNPs only explained a small proportion of heritability[1], which was estimated from family studies[27, 28]. These gaps were called as “missing heritability”. The potential causes of “missing heritability” were thought to be as follows: the effect sizes of individual SNPs are too small to reach a genome-wide significance ($P < 5\times10^{-8}$) in GWASs: causal SNPs are not in sufficient correlation with

[1] Heritability is the proportion of the phenotypic variance among a group, which is attributable to genetics.

SNPs on the commercial arrays to be detected by the association studies: epistasis effects of SNPs exist. To overcome the problem of small effect size, many researchers have been making a big effort to enlarge sample sizes in order to discover novel susceptibility SNPs[30-33]. At the same time, the polygenic risks based on CDCV hypothesis were assessed, so that the genetic architecture of diseases have been uncovered, such as extent of missing heritability and shared genetic background among diseases[34-37]. For example, a recent study provided not only the existence of substantial proportion of missing heritability in neuropsychiatric diseases, such as schizophrenia, bipolar disorders, and attention deficit hyperactivity disorder (ADHD), but also genetic sharing among these disorders[34]. Furthermore, the assessment of genetic similarity among the diseases could lead to phenotypic re-classification based on genotypes.

Recent developments in statistical analysis make it possible to estimate heritability of diseases using GWAS data. There were mainly two new methods to estimate polygenic effect on diseases. In the first method, an additive polygenic risk score, which was based on SNPs below a certain *P* value threshold in a discovery GWAS, was tested in an independent set of samples, so called “polygenic analysis”[34, 36]. The other was based on pairwise relationship of all samples, which meant

genome-wide similarity between all pairs of individuals determined using all SNPs[35, 37]. The concept of this method was that if genetic factor contribute to a phenotype, case-case pairs and control-control pairs are more similar across the genome than case-control pairs. In this study, I applied the first method, because, once polygenic risks of a certain phenotype were defined, polygenic risks could be applied for different phenotypes. In the other words, I could assess the common genetic background among different diseases.

The pathogenesis of EHS is considered to be partially common to that of narcolepsy[25]. Previous studies have reported that narcolepsy and EHS were associated with the same susceptibility genes[22-26]. Especially, about 30-50% of EHS patients carried *HLA-DQB1*06:02* allele[22, 23]. In addition, EHS was associated with narcolepsy susceptibility SNPs in *CPT1B* and *TRA@*[22, 26]. However, to mention in detail, only EHS with *HLA-DQB1*06:02* showed a significant association with *TRA@* while EHS without *HLA-DQB1*06:02* did not, thus in this study EHS with/without *HLA-DQB1*06:02* were evaluated separately[26].

In this study, in order to evaluate contribution of common variants for narcolepsy, I conducted a polygenic analysis using Affymetric 6.0 data of 393 narcoleptic patients. In addition, to assess overlap of polygenic risks for narcolepsy and

EHS, I also analyzed data of 38 EHS with *HLA-DQB1*06:02* and 119 EHS without *HLA-DQB1*06:02* patients.

2. Materials and methods

Subject

Participants in the study were 426 narcoleptic patients, 46 EHS with *HLA-DQB1*06:02* patients, 125 EHS without *HLA-DQB1*06:02* patients, 432 panic disorders, 246 autism, 66 schizophrenia and 2,072 healthy individuals. Japanese patients with narcolepsy were unrelated individuals living in Tokyo or neighboring areas. All patients with narcolepsy carried the allele, *HLA-DQB1*06:02*. Among 171 EHS patients, 46 individuals possessed *HLA-DQB1*06:02* (EHS with *HLA-DQB1*06:02*), while other 125 EHS patients did not (EHS without *HLA-DQB1*06:02*)[22]. Each patient with narcolepsy or EHS had received a diagnosis at the Neuropsychiatric Research Institute and diagnosed by specialists for sleep disorders. Data of neuropsychiatric disorders such as panic disorders, autism and schizophrenia was parts of previous studies and analyzed in this study[38-41].

Among healthy controls, 562 unrelated healthy Japanese individuals living in Tokyo or neighboring areas. The rest of the healthy controls were genotyped in previous studies, and their data was also analyzed in this study[38-42]. The control subjects did not have a history of narcolepsy, central nervous system (CNS) hypersomnia or other neuropsychiatric diseases. Age and gender were not matched between cases and controls. Ethical approval was obtained from the local institutional review boards of all

participating organizations. All individuals gave written informed consent for their participation in the study.

Genotyping and quality control

Genomic DNA from all participants was genotyped for 906,622 SNPs by Affymetrix Genome-Wide SNP Array 6.0 (Affymetrix, Santa Clara, CA) (http://www.affymetrix.com). Genotype calling was done by Birdseed algorithm in Birdsuite (http://www.broadinstitute.org/), except a sample that had less than 90% of genotyping call rate. Quality control procedures were performed, using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) (Figure 1)[43]. Samples with overall call rates lower than 97% were excluded. Samples, who were reported to have family relationship with other participants or PIHAT value greater than 0.125 (PIHAT: mean probability of being identity-by-descent, calculated in PLINK), were excluded, until only one sample remained out of reported or estimated family members. Outliers in principal component analysis were also excluded in order to eliminate population stratification, using EIGENSOFT (http://www.hsph.harvard.edu/alkes-price/software/) as described later[44]. SNPs were excluded based on following criteria: (i) minor allele frequency (MAF) was less than 0.05; (ii) p-value from the Hardy-Weinberg Equilibrium

(HWE) test either for the patient group or the healthy control group was less than 0.001; (iii) SNP call rate was less than 99%; (iv) SNPs resided within *HLA* region (chr6: 27,539,703-35,377,701) or sex chromosomes. After the quality controls, 476,446 SNPs in 393 narcolepsy individuals, 38 EHS with *HLA-DQB1*06:02* individuals, 119 EHS without *HLA-DQB1*06:02* individuals, 376 panic disorders, 213 autism, 56 schizophrenia and 1,582 healthy individuals were analyzed in this study (Figure 1).

Principal component analysis

In order to eliminate population stratification, a principal component analysis was performed using EIGENSOFT. In the evaluation, 91 JPT (Japanese in Tokyo, Japan), 90 CHB (Han Chinese in Beijing, China), 180 CEU (Utah residents with Northern and Western European ancestry) and 180 YRI (Yoruba in Ibadan, Nigeria), deriving from Hapmap project (http://hapmap.ncbi.nlm.nih.gov/) were included[45]. HapMap populations and the present study sample populations were combined after the quality controls procedures as described above.

Statistical analysis

A polygenic analysis principally estimates the capability of common SNPs to

explain phenotypic variance in a group[34, 36]. The study design, shown in Figure 2, was applied. Samples were divided into two groups. One group (discovery-stage) was used for a genome-wide association analysis in order to calculate narcolepsy risk-effect of each common SNP. The other group (test-stage) was used for estimation of the capability of common SNPs to explain phenotypic variance, using collective risk-effects of common SNPs, which is, in another term, "polygenic risk". A permutation analysis, which randomized sample division of narcoleptic patients and healthy controls into two groups, discover-stage and test-stage, was performed, repeating one thousand times. In addition, in the test-stage, the same healthy controls across different diseases within one permutation were utilized.

A genome-wide association analysis for narcolepsy was conducted in the discovery-stage. Risk-effects of individual SNPs were calculated based on chi-square values. For an example, suppose that chi-square value based on allelic model is 30 at SNP-X, the risk-effects of SNP-X, were calculated as 0, 30 and 60, according to the number of risk allele, 0, 1 and 2. In the test-stage, polygenic risks for each participant by summing risk-effects of all SNPs were obtained, named as "polygenic risk score". Several statistical thresholds for calculation of polygenic risks were employed: (i) SNP inclusion criteria based on the discovery-stage association analysis; $P < 0.001, 0.01, 0.1,$

0.2, 0.3, 0.4 and 0.5; (ii) SNP inclusion criteria based on pruning: $r^2 < 0.25, 0.5, 0.75, 1$ within a 200-kb window in order to exclude secondary associations due to linkage disequilibrium. Polygenic risks were calculated for each threshold. A model of logistic regression analysis was followed;

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 ,$$

where x_1 was 0~1 collective polygenic risk score for each individual. Nagelkerkes' pseudo R^2 was utilized to reflect estimated the capability of polygenic risks to explain disease onset. Narcolepsy is tightly associated with *HLA-DQB1*06:02*, thus so as to eliminate the *HLA* effect, the region (chr6: 27,539,703-35,377,701) was excluded from polygenic risk score[3-7]. Instead, SNP rs7744293, which was in the highest correlation with *HLA-DQB1*06:02* (r^2 = ~0.8), was included into the logistic regression model as a surrogate maker, so that estimation of the heritability of narcolepsy reflect the effect of the allele, *HLA-DQB1*06:02*;

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 ,$$

where x_1 was 0~1, collective polygenic risk score for each individual and x_2 was 0 or 1, the positivity of the risk allele of SNP rs7744293, which surrogated the positivity of *HLA-DQB1*06:02*. In addition, to assess the heritability explained by the reported susceptible SNPs in the Japanese population, *CPT1B*, *TRA@* and *P2RY11*, only SNPs

which reside within 500kb up- and down-stream of the reported SNPs were utilized for calculation of polygenic risk score[8-10].

Polygenic risks among healthy controls were assessed. Healthy controls were split into a pseudo-case group and a pseudo-control group. Then, the same analytical procedures as described above were adopted. One thousands-permutation analysis was carried out, randomizing both pseudo case-control status and discovery-test stage.

Logistic regression analysis was conducted by PredictABEL packages (http://www.genabel.org/) provided in R[46]. I also estimated desirable sample size for detecting remaining genetic factors using "power.prop.test3" function provided in R (http://aoki2.si.gunma-u.ac.jp/R/power_prop_test2.html). Analytical and quality control procedures were conducted using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/) and R (http://www.r-project.org/)[43, 47].

3. Results

Samples, consisting of 426 patients with narcolepsy, 46 EHS with *HLA-DQB1*06:02*, 125 EHS without *HLA-DQB1*06:02*, 432 panic disorders, 246 autism, 66 schizophrenia and 2,072 healthy individuals, were genotyped by Affymetrix Genome-Wide SNP Array 6.0 and Birdseed algorithm. A principal component analysis was conducted using EIGENSOFT, and an outlier was excluded (Figure 3). Populations suffered form different diseases did not clustered separately (Figure 4). After the quality control procedures, 476,446 SNPs in 393 narcoleptic patients, 38 EHS with *HLA-DQB1*06:02* patients, 119 EHS without *HLA-DQB1*06:02* patients, 376 panic disorder patients, 213 autism, 56 schizophrenia and 1,582 healthy controls, remained for a subsequent analysis (Figure 1).

A genome-wide association analysis on narcolepsy in the discovery-stage was performed. For each permutation, narcolepsy risk-effects of SNPs were calculated. The average lamda value of one thousand genome-wide association analyses was 1.004, when I applied the pruning criteria of $r^2 < 0.25$ for SNP inclusion (Figure 5).

Polygenic risk was computed for each narcoleptic patient and healthy control in the test-stage, using seven different *P* value thresholds ($P < 0.001, 0.01, 0.1, 0.2, 0.3,$ 0.4 and 0.5). As a result of one thousand-permutation analysis, narcolepsy heritability

was estimated as 58.1%, when a SNP pruning criterion was $r^2 = 0.25$ ($P_{HLA\text{-}DQB1*06:02} = 2.30 \times 10^{-48}$, $P_{others} = 6.73 \times 10^{-2}$) (Table 1). Heritability of narcolepsy other than the *HLA* region was estimated as 1.3% ($P_{others} = 2.43 \times 10^{-2}$) (Table 2 and Figure 6). It was observed that the estimates of heritability became large until SNPs with *P* value < 0.5 in the discovery-stage were included in the analysis (Figure 7). This suggested that common small-effect SNPs contributed to the onset of narcolepsy.

Next, SNPs pruning criteria were evaluated whether they influenced polygenic analyses. At the first calculation of polygenic risks, SNPs with $r^2 < 0.25$ were used. Here, the analyses were extended by applying different pruning criteria 0.25, 0.5, 0.75, 1 of r^2. Narcolepsy polygenic risk was significantly associated with the onset of narcolepsy, being concordant across the criteria, 1.3% ~ 2.1% ($P_{others} = 3.54 \times 10^{-3} \sim 2.43 \times 10^{-2}$) (Table 3). In addition, narcolepsy heritability explained by the reported susceptibility SNPs, was calculated. Using only SNPs residing within 500kb up- and down-stream of the narcolepsy susceptible SNPs in *CPT1B*, *TRA@* and *P2RY11*, the results were 0.7%~1.1% according to the pruning criteria ($P_{others} = 3.82 \times 10^{-2} \sim 1.19 \times 10^{-1}$) (Table 3).

Next, I aimed to assess the capability of polygenic risk of "narcolepsy" to explain the onset of EHS with/without *HLA-DQB1*06:02*. The risk-effects of SNPs,

which were calculated in narcolepsy vs healthy control GWASs in the discovery-stage, were applied to obtain polygenic risk score of EHS patients. Common genetic background between narcolepsy and EHS with *HLA-DQB1*06:02* was estimated including *HLA-DQB1*06:02* effects (narcolepsy: 58.1%, $P_{HLA\text{-}DQB1*06:02} = 2.30 \times 10^{-48}$, $P_{others} = 6.73 \times 10^{-2}$, EHS with *HLA-DQB1*06:02*: 40.4%, $P_{HLA\text{-}DQB1*06:02} = 7.02 \times 10^{-14}$, $P_{others} = 1.34 \times 10^{-1}$) (Table 1). This suggested that substantial proportion of heritability was shared between narcolepsy and EHS with *HLA-DQB1*06:02*. Narcolepsy polygenic risk, excluding *HLA-DQB1*06:02* effects, did not explain the onset of EHS either with or without *HLA-DQB1*06:02* (Table 2 and Figure 6). However, EHS with *HLA-DQB1*06:02* patients tended to have larger polygenic risk of "narcolepsy" than EHS without *HLA-DQB1*06:02* patients (EHS with *HLA-DQB1*06:02*: 1.4%, $P_{others} = 1.56 \times 10^{-1}$, EHS without *HLA-DQB1*06:02*: 0.4%, $P_{others} = 3.06 \times 10^{-1}$). (Table 2 and Figure 6).

Polygenic analyses were extended to other neuropsychiatric diseases, including 376 panic disorder patients, 213 autisms, and 56 schizophrenias. Polygenic risk of "narcolepsy" did not explain phenotypic variance of these neuropsychiatric diseases (panic disorder: 0.1%, $P_{others} = 5.20 \times 10^{-1}$, autism: 0.2%, $P_{others} = 4.75 \times 10^{-1}$, schizophrenia: 0.5%, $P_{others} = 3.79 \times 10^{-1}$) (Table 2 and Figure 6).

I also assessed the possibility that healthy controls in the present study led to spurious estimations. All healthy controls were divided into pseudo-case and pseudo-control. Based on pseudo-case and pseudo-control, polygenic risks among healthy controls were estimated using the same methodology. As a result, there was no identical heritability among healthy controls (0.2%, $P_{others} = 5.10 \times 10^{-1}$).

The sample sizes were estimated to detect remaining common small-effect susceptible SNPs in narcolepsy. It was found that the estimates of heritability became large until SNPs with $P < 0.5$ in the discovery-stage were included in the analysis (Figure 7). In the current GWAS, applying all 393 narcolepsy vs all 1,582 healthy controls, the ORs of SNPs with $P < 0.5$ ranged from 1.1 to 1.6 (Table 4). The sample size was calculated for detecting common small-effect susceptible SNPs with their ORs ranging from 1.1 to 1.6 (Figure 8). More than 42,000 narcoleptic patients were required to detect with the genome-wide significant level, when ORs of targeted SNPs was 1.1; risk allele frequency in controls was 0.2; proportion of case-controls (case/controls) was 0.25; α-level was $P = 5 \times 10^{-8}$.

4. Discussion

In this study, a polygenic analysis of sleep disorders was performed. After stringent quality controls, 476,446 SNPs in 393 narcoleptic patients, 38 EHS with *HLA-DQB1*06:02* patients, 119 EHS without *HLA-DQB1*06:02* patients, other neuropsychiatric disorders, including 376 panic disorders, 213 autism and 56 schizophrenia, and 1,582 healthy individuals were analyzed.

In this study, I evaluated the extent of narcolepsy genetic factors, both including and excluding the *HLA* region. Heritability of narcolepsy was accounted as around 58.1% (Tables 1) and 1.3% excluding the *HLA* effects (Table 2 and Figure 6). *HLA-DQB1*06:02* is thought to be essential in narcolepsy because the all patients possess this allele. However, this allele itself could not fully explain the onset of narcolepsy, because 15-25% of individuals in the general populations carry *HLA-DQB1*06:02* whereas the prevalence of the disease is around 0.1%[1]. First-degree relatives of individuals with narcolepsy possess this allele with a more than 50% of possibility, however the risk of the first-degree relatives was 1-2%[48]. Two thirds of monozygotic twins were discordant for development of narcolepsy[1]. These reports showed that the onset of narcolepsy was associated with other genetic or environmental factors. In this study, the effects of genetic factors were evaluated, and the extent of genetic factors residing outside the *HLA* region was also estimated.

Narcolepsy heritability accounted as 58.1% ($P_{HLA\text{-}DQB1*06:02} = 2.30 \times 10^{-48}$, $P_{others} = 6.73 \times 10^{-2}$) (Table 1), in which heritability of narcolepsy explained by the region other than *HLA* was estimated as 1.3% ($P_{others} = 2.43 \times 10^{-2}$) (Table 2). Previous reports about twin studies in narcolepsy-like symptoms in the Finish twin cohort, including 1,322 monozygotic and 2,463 dizygotic adult twin-pairs, showed that genetic factors contributed for 35-39% of narcolepsy-like symptom[49]. However, the previous study diagnosed narcoleptic patients according to questionnaire assessment. The positivity of *HLA-DQB1*06:02* in each patient was not confirmed. In the heritability assessment, phenotypic classification influenced the accuracy of estimation critically. Thus, the previous study possibly underestimated the heritability. In this study, all narcoleptic patients were confirmed to possess *HLA-DQB1*06:02* and diagnosed by hypersomnia specialists.

Pruning criteria for SNP inclusion were assessed whether it might influence the estimation of heritability (Table 3). If SNPs were pruned too strictly, including only SNPs with less than 0.25 of r^2 in the analysis, it is possible to miss true susceptibility SNPs, so that the estimation would decrease. If SNPs were not pruned at all, using all SNPs regardless of their linkage disequilibrium, secondary SNPs, which tended to have smaller effects than primary susceptible SNPs, would be included in the analysis. In

such a case, the differences of polygenic risk scores between patients and healthy controls would decrease because polygenic risk scores were the average of chi-square-weighed additive risks at each locus. In other words, the effects of the primary susceptible SNPs became to be weakened by the effects of secondary SNPs. Thus, appropriate pruning criteria are considered to be important for the accuracy of heritability assessments. Previous polygenic analysis reports used different pruning criteria, which ranged from 0.1 to 0.5 of r^2[34, 36, 50]. The estimation would be, to some extent, depending on pruning criteria. Then, I applied the different pruning criteria, so that the estimation in this thesis could be confirmed. Narcolepsy polygenic risk was significantly associated with onset of narcolepsy at every criterion.

Narcolepsy heritability explained by the region other than *HLA* accounted as 2.1%, at the highest value, with the criteria, 0.75 of r^2 (Table 3). This meant that, with the criteria, 0.75 of r^2, the common small-effect SNPs could be efficiently detected without missing true signals or including redundant signals. This suggested that, when I targeted to discover common small-effect SNPs in GWAS, I could reduce the number of tested SNPs by using this pruning criterion and possibly could loosen the stringent genome-wide significant level. For an example, with the criteria of 0.75 of r^2, 228,555 SNPs remained and tested in this narcolepsy GWAS, thus the significant threshold

adjusted by Bonferroni correction is 2.19×10^{-7}. It is much higher than the current genome-wide association threshold, 5.00×10^{-8}.

Until now, several genetic factors other than the *HLA* region were reported in the Japanese population, including *CPT1B*, *TRA@* and *P2RY11*. Here, the extent of these factors to explain the heritability was calculated (Table 3). SNPs, which reside within 500kb up- and down-stream of the reported SNPs, were included in the analysis. As a result, heritability was estimated to be 0.7%~1.1%, depending on pruning criteria ($P_{others} = 3.82 \times 10^{-2} \sim 1.19 \times 10^{-1}$). The differences between heritability explained by the region other than *HLA* and heritability explained by *CPT1B*, *TRA@* and *P2RY11*, were about 0.5%~1.0%, suggesting that a certain proportion of narcolepsy susceptible genetic factors was not yet discovered.

EHS with/without *HLA-DQB1*06:02* were assessed whether polygenic risks of "narcolepsy" contribute to onset of the diseases. Narcolepsy polygenic risks did not explain phenotypic variance of either EHS with/without *HLA-DQB1*06:02*. However, EHS with *HLA-DQB1*06:02* tended to have higher polygenic risks than EHS without *HLA-DQB1*06:02* (Tables 1 and 2, and Figure 6). If *HLA-DQB1*06:02* was included in the analysis, in total, about 40.4 % of EHS with *HLA-DQB1*06:02* disease onset were explained by narcolepsy genetic background (Table 1). This result suggested that

EHS with *HLA-DQB1*06:02* be more similar to narcolepsy than EHS without *HLA-DQB1*06:02*. This result was concordant with previous studies[8, 9, 22, 26]. In the reported association between EHS and *TRA@*, only EHS with *HLA-DQB1*06:02* showed a significant association, while EHS without *HLA-DQB1*06:02* did not[26]. The definition of EHS with/without *HLA-DQB1*06:02* did not follow ICSD-2, indeed. However, classification of EHS with/without *HLA-DQB1*06:02* seem to reflect genetic features of the diseases.

Other neuropsychiatric disorders were also evaluated. Narcolepsy polygenic risk did not explain onset of these neuropsychiatric disorders. In other words, narcolepsy and these neuropsychiatric disorders did not share genetic background due to common variants. From the viewpoint of feature of symptoms, it is hardly said that narcolepsy is similar to other neuropsychiatric disorders, such as autism, schizophrenia, and panic disorders. Thus, the results were concordant with the expectation.

In studies of heritability, it is essential to assess whether data of healthy control samples possibly affect estimations of heritability. The genetic feature among control samples was validated. Control samples were divided into pseudo-case and pseudo-control, and the same polygenic analysis was performed. There was no significant identical heritability among healthy controls, indicating that healthy controls

were unbiased and usable for this analysis.

Furthermore, sample sizes to identify remaining narcolepsy genetic factors were estimated (Figure 8). In the future, an effort should be made to collect around 42,000 narcoleptic patients. At the same time, sample size for healthy controls should be 4 times larger than that of the patients in the estimation. Therefore it would be necessary to make consortium probably by means of large-scale international collaboration, both the patients and healthy controls.

Relatively small sample size was used in this study as compared to previous polygenic analysis studies. Previous research analyzed at least 4,000 samples, and the largest studies utilized ~ 30,000 samples[34-37]. To overcome the weakness of the small sample size, I applied a permutation procedure, which randomized discovery GWAS set samples and test set samples. This permutation procedure was reported to be able to estimate heritability as same as larger sample size[34]. In addition, division of healthy controls among discovery- and test-stage was contrived. The same healthy controls in the test-stage were utilized across different diseases within one permutation. By this design, I could avoid from making sample size of each group smaller when several diseases were assessed. Also the same controls in the test-stage across different diseases made it possible to compare heritability directly among them.

Limitation of this study may be the assumption of the method. This polygenic analysis assumed additive effects of each SNP across genome. The dominant effects and epistasis effects were not taken into consideration for the estimation. There were many reported susceptible SNPs with dominant models in mammal[51-53]. Epistasis effects were also reported[54]. Therefore, genetic architecture of the diseases attributable to common variants was not fully described by this method.

In this study, I utilized the data from the biggest collaboration in narcolepsy research across Japan. The extent of heritability in narcolepsy was estimated. Also the nominal shared genetic background due to common variants between narcolepsy and EHS with *HLA-DQB1*06:02* were uncovered. However, many efforts for catching up with the estimated desirable sample sizes are needed to discover remaining genetic factors. Therefore, not only the collaboration across Japan but also large-scale international collaboration would be necessary for the future research.

Ⅲ Genome-wide CNV (copy number variation) study of sleep disorders in the Japanese population[2]

[2] The study in this section has been already published.
Yamasaki, M., Miyagawa, T., Toyoda, H., Khor, SS., Koike, A., Nitta, A., Akiyama, K., Sasaki, T., Honda, Y., Honda, M., Tokunaga, K., Genome-wide analysis of CNV (copy number variation) and their associations with narcolepsy in a Japanese population. J Hum Genet, 2014. 59(5):p.235-240

1. Introduction

In this section, I assessed whether any rare variants contributed to the development of narcolepsy and essential hypersomnia (EHS). Majority of narcolepsy and EHS genetic research has targeted to discover common susceptible variants based on common disease-common variant hypothesis (CVCV hypothesis)[8-10, 22, 26]. It was important to evaluate the possibility that genetic architecture of narcolepsy and EHS might be based on common disease-rare variant hypothesis (CVRV hypothesis).

Recent technological developments made it possible to conduct genome-wide studies on copy number variations (CNVs), especially rare large-scale CNVs. CNVs are insertions, deletions, or duplications of genomic sequences ranging from a kilobase to multiple megabasepairs in length, which occupy 5-12% of the human genome. Recent studies have revealed that rare large-scale CNVs are associated with several neuropsychiatric diseases such as autism, schizophrenia or sever neurodevelopmental disorders[55-58]. These studies reported that the average number of rare large-scale CNVs per person in those patients was higher than that of healthy individuals, indicating that rare large-scale CNVs are involved in the development of neuropsychiatric diseases. The studies also reported that several neuropsychiatric diseases were associated with a certain CNV region, suggesting shared genetic

susceptibility among the different diseases[56, 59-64].

Here, I hypothesized an impact of CNVs on narcolepsy and EHS, because it has been considered that narcolepsy and EHS has a neuropsychiatric etiology[14-16, 19]. I conducted a genome-wide CNV study in order to identify novel genetic susceptibility loci for narcolepsy and EHS.

2. Materials and Methods

Subjects

Japanese patients with narcolepsy (n = 426) were unrelated individuals living in Tokyo or neighboring areas. All patients with narcolepsy carried the allele, *HLA-DQB1*06:02*. Among 171 EHS patients, 46 individuals possessed *HLA-DQB1*06:02* (EHS with *HLA-DQB1*06:02*), while other 125 EHS patients did not (EHS without *HLA-DQB1*06:02*)[22]. Each patient had received a diagnosis at the Neuropsychiatric Research Institute. In this study, all patients were diagnosed by specialists for sleep disorders.

Controls comprised 562 unrelated healthy Japanese individuals living in Tokyo or neighboring areas. The control subjects did not have a history of narcolepsy and central nervous system (CNS) hypersomnia. Age and gender were not matched between cases and controls because narcolepsy and EHS affects men and women equally. Ethical approval was obtained from the local institutional review boards of all participating organizations. All individuals gave written informed consent for their participation in the study.

Preparation of gene chip data

A genome-wide CNV analysis was performed with DNA samples from 426 narcoleptic patients, 46 EHS with *HLA-DQB1*06:02* patients, 125 EHS with *HLA-DQB1*06:02* patients and 562 healthy controls. The Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA) was used according to the manufacturer's protocols (http://www.affymetrix.com). This array comprises about 906,600 SNP probes and about 946,000 CNV probes. Samples with call rates lower than 99% were excluded from the CNV analysis because low-quality genotyping is strongly correlated with inaccurate CNV detection (Figure 9)[65].

CNV detection

PennCNV (http://www.openbioinformatics.org/) was utilized in order to detect CNVs[66]. This software uses called genotypes and normalized intensity to create reference cluster positions to compute relative differences in the signal from each sample in the form of B allele frequency (BAF) and log R ratio (LRR) (Figure 10). BAF is a normalized measure of relative signal intensity ratio of the B and A alleles in the SNP array. LRR is a measure of normalized total signal intensity relative to expected value, (= $\log2(R_{observed}/R_{expected})$, R is a sum of probe intensities). For normalization of intensity data, Affymertix Power Tools software (http://www.affymetrix.com/) was

utilized before the CNV detection.

I also made use of Birdsuite (http://www.broadinstitute.org/) in order to detect CNVs[67]. This software consists of 4 parts: (i) Canary assigns copy number across regions of common copy number polymorphisms (CNPs) using expectation–maximization algorithm and a one-dimensional Gaussian mixture model; (ii) Birdseed calls genotypes of SNPs; (iii) Birdeyes identifies rare CNVs via a hidden Markov model (HMM); (iv) Fawkes generates an integrated sequence and copy number genotype at every locus (for example, including genotypes such as A-null, AAB and BBB in addition to AA, AB and BB calls).

Quality control

Quality controls were done after the CNV detection (Figure 9). In the raw detection data from PennCNV, samples with CNV call counts > 100 were excluded because an unusually large number of CNVs indicates that a sample contains low-quality DNA. Then, CNVs that included fewer than 10 probes were removed in order to increase the reliability of the analysis because CNVs with lower numbers of probes cause false detection of CNVs[55, 57, 65].

Statistical analysis

Statistical analyses of the CNV data were performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/)[43]. Gene information was based on glist-hg18 list provided by PLINK and NCBI Build 36[68]. In this analysis, rare large-scale CNVs, that were ≤ 1% in frequency and that were ≥ 100 kb in size, were targeted. Frequency was defined in each group: (i) narcolepsy patients and healthy individuals; (ii) EHS with *HLA-DQB1*06:02* patients and healthy individuals; (iii) EHS without *HLA-DQB1*06:02* patients and healthy individuals. In a global burden analysis, the average number of all rare large-scale CNVs per person was on comparison between cases and healthy controls. A region-based analysis was performed in order to identify specific CNV-associated regions, targeting CNVs with its size ≥ 30kb. A significant region identified in the region-based analysis was validated using a TaqMan assay as described later (Applied Biosystems) (http://www.appliedbiosystems.jp/).

A pathway analysis was designed to discover disease-related functional pathways, basically according to previous studies[57, 69, 70]. Any gene-sets, which are more frequently affected by rare large-scale CNVs in the patients compared with that of healthy controls, were searched. CNV with its size ≥ 100k were included in this analysis[55, 69, 70]. Gene sets were derived from the Gene Ontology

(http://www.geneontology.org/) and from KEGG (http://www.genome.jp/kegg/)[71, 72]. The analysis was conducted using Perl (http://www.perl.org/) and R (http://www.r-project.org/)[47]. Gene-sets with sizes > 700 genes or with sizes < 5 genes were excluded, because large and small gene-sets are less likely to produce useful biological meaning.

Functional clusters were created to interpret the relation of a large number of significant gene sets. Cytoscape (http://www.cytoscape.org/) was used to create functional clusters of significant gene sets[73]. Edge indicates that two gene sets share "support-genes", which are more frequently affected in the patients. Edge width is the proportion of the shared support-genes versus the total number of genes within two gene sets. Node indicates a gene set. Node size is proportional to the number of genes in a gene set.

The global burden analysis and the region-based analysis were evaluated using one-tailed 100,000-permutation tests. False-discovery rate (FDR) was applied to correct for multiple testing in the pathway analysis. To estimate FDR, 10,000-permutation tests were also used to correct for the dependency of gene sets because an individual gene could fall into more than one gene set. Corrections for multiple testing were not applied to the region-based analysis. All programming and file

conversions were done with Perl or R script.

CNV validation by quantitative PCR

Quantitative PCR (qPCR) was performed on LightCycler 480 System (Roche Applied Science). TaqMan Copy Number Assay (Applied Biosystems) was selected via Life Techonologies website (http://www.appliedbiosystems.jp/). As a reference gene, I used commercially available TaqMan Copy Number Reference Assay RNase P. Each sample was examined in 10 μl reaction mixture (TaqMan® Genotyping Master Mix: 5μl, TaqMan® Copy Number Assay 20× working stock: 0.5μl, TaqMan Copy Number Reference Assay 20×: 0.5μl, Nuclease-free water: 2μl and 5ng/μl genomic DNA: 2μl). The qPCR thermal cycling conditions were as follows: initiation at 95 °C for 10 minutes for hot start, followed by 40 cycles of 95 °C for 15 seconds and 60 °C for 1 minute. The PCR efficiency of each assay was extracted from the calibration curves of mixed DNA from three different samples with normal copy numbers, serially diluted from 40 ng to 2.5 ng of genomic DNA. The data analysis was further performed using the ΔΔCT method. CNV ratio was normalized ratio of ΔΔCT, based on the calibration samples with normal copy numbers ($=\Delta\Delta CT_{sample}/\Delta\Delta CT_{calibration}$). Experiments in triplicate were repeated 4 times. Differences in CNV ratio between each case and a control group were

tested using t-test.

3. Results

SNP Gene chip data were used to detect CNVs within the genomes of 426 Japanese narcoleptic patients, 46 EHS with *HLA-DQB1*06:02* patients, 125 EHS without *HLA-DQB1*06:02* patients and 562 healthy controls. Each sample with an call rate of <99% was excluded from further analysis before detection of CNVs was performed; consequently, the samples including 345 narcoleptic patients, 37 EHS with *HLA-DQB1*06:02* patients, 111 EHS without *HLA-DQB1*06:02* patients and 470 healthy controls were used for CNV detection (Figure 9). PennCNV was applied to detect CNVs; after the detection step, whole samples or individual CNVs were eliminated during the ensuing quality control steps. The CNV call count threshold was set to >100 for sample exclusion; this step left 327 narcoleptic patients, 33 EHS with *HLA-DQB1*06:02* patients, 104 EHS without *HLA-DQB1*06:02* patients and 459 healthy control samples. Individual CNVs that included fewer than 10 probes were removed from subsequent analyses. Then, all small CNVs and all common CNVs were excluded to target only rare large-scale CNVs (frequency: ≤1%, CNV size: ≥100 kb).

The global burden of rare large-scale CNVs, which were stratified by size and CNV type, among narcoleptic patients was compared with that among control subjects (Table 5). Overall, a significant enrichment of rare large-scale CNVs in the patients was observed (case-control ratio of CNV count = 1.54, $P = 5.00\times10^{-4}$). In particular, a higher

occurrence of duplications with sizes between 100 and 200 kb in the patients was detected (case-control ratio of CNV count = 1.93, $P = 2.00\times10^{-5}$) (Table 5). An excess of rare large-scale deletions with sizes between 500 kb and 1Mb was evident in the patients (case-control ratio of CNV count = 3.12, $P = 2.18\times10^{-2}$) (Table 5). The number of genes disrupted by rare large-scale CNVs was significantly higher in the patients than in control subjects (case-control ratio of number of genes disrupted by CNVs = 2.18, $P = 5.75\times10^{-3}$) (Table 6). Especially, gene disruptions by duplications with sizes of 200 to 500 kb were more frequent in the patients in the same manner of Table 5 (case-control ratio of number of genes disrupted by CNVs = 2.62, $P = 9.70\times10^{-4}$) (Table 6).

The global burden of rare large-scale CNVs was assessed between EHS with *HLA-DQB1*06:02* and healthy controls (Table 7). In total, a significant enrichment of rare large-scale CNVs in the patients was found (case-control ratio of CNV count = 1.46, $P = 3.24\times10^{-2}$) (Table 7). Particularly, deletions with sizes between 100 and 200kb were more frequently detected in the patients than control individuals (case-control ratio of CNV count = 2.13, $P = 1.35\times10^{-2}$) (Table 7). In terms of the number of genes disrupted by CNVs, I did not found any significant enrichments (Table 8).

The global burden of rare large-scale CNVs was compared between EHS without *HLA-DQB1*06:02* and healthy controls (Table 9). Overall, a significant

enrichment in the patients was found (case-control ratio of CNV count = 1.43, $P = 2.17\times10^{-3}$) (Table 9). Especially, a significant higher occurrence rate with sizes smaller than 200kb, was seen, regardless of CNV types (Table 9). The number of genes disrupted by rare large-scale CNVs was significantly higher in the patients than in control individuals (case-control ratio of number of genes disrupted by CNVs = 1.80, $P = 3.25\times10^{-2}$) (Table 10). In detail, only gene disruptions by duplications with sizes between 100kb and 500kb were more recurrent in the patients than in controls (100kb-200kb: case-control ratio of number of genes disrupted by CNVs= 2.51, $P = 8.02\times10^{-3}$, 200kb-500kb: case-control ratio of number of genes disrupted by CNVs = 2.98, $P = 4.12\times10^{-2}$) (Table 10).

A region-based analysis was performed to identify regions associated with narcolepsy susceptibility (Figure 11). In this analysis, the definition of rare large-scale CNVs was CNVs with its size ≥ 30kb and with its frequency ≤ 1%, so that the possibility to miss true associations could be reduced. The *PARK2* (parkinson protein 2, E3 ubiquitin protein ligase) region (Chr 6: 162,685,167-162,762,467) was detected as a region of nominal significant association in narcolepsy ($P = 3.07\times10^{-2}$) (Figure 11). A significant association in EHS without *HLA-DQB1*06:02* was also detected, using CNVs deriving from different CNVs detection software, Birdsuite ($P = 2.91\times10^{-2}$)

(Figure 11). Four narcoleptic patients and two EHS without *HLA-DQB1*06:02* carried duplications with copy numbers = 3 within the *PARK2* region, but no controls carried duplications in this region. In order to ensure reliability of the duplications in the *PARK2* region in these four patients, these CNVs were validated using qPCR (Figures 12 and 13).

To assess the overall influence of rare large-scale CNVs in the pathogenic narcolepsy, a pathway analysis was conducted because association studies that involve only individual rare CNVs and a limited number of samples often have insufficient power to detect disease-causing variants. Comprehensive collections of gene-sets were derived from KEGG and the Gene Ontology, and then each gene-set was assessed whether they were more frequently affected by rare large-scale CNV in narcoleptic patients than in healthy controls. A total of 32 gene sets showed significant associations when the FDR was set to 5% (Figure 14 and Table 11). Four functional clusters were created using these 32 gene sets: immune responses, acetyltransferase activity, cell cycle regulation and regulation of cell development.

4. Discussion

In this study, a genome-wide CNV analysis was performed using data comprising 327 narcoleptic patients, 33 EHS with *HLA-DQB1*06:02* patients, 104 EHS without *HLA-DQB1*06:02* patients and 459 healthy control samples. Both narcoleptic patients and EHS with/without *HLA-DQB1*06:02* patients were found to carry a larger global burden of rare large-scale CNVs compared to controls (Tables 5, 7 and 9). The number of genes disrupted by rare large-scale CNVs was also higher in narcoleptic patients than that in controls (Table 6). EHS with *HLA-DQB1*06:02* did not show significantly higher number of gene disrupted by rare large-scale CNVs, however the overall tendency of larger number of gene disruption in the patients could be still observed (Table 8). EHS without *HLA-DQB1*06:02* were found to have more genes disrupted by rare large-scale CNVs (Table 10). However, the manner was different from CNV burden in EHS without *HLA-DQB1*06:02* (Tables 9 and 10). Case-control ratio of CNV count in deletions was larger compared to that of number of genes disrupted, while case-control ratio of CNV count in duplication was smaller than that of number of genes disrupted. This might be because the density of genes is different at each region. These results were confirmed by using another CNV detection software, Birdsuite. At this moment, a further replication study is needed to confirm these results because the

sample size of this study was relatively small.

In this study, two kinds of software were utilized to detect CNVs, PennCNV and Birdsuite. At this moment, accuracy of rare and large-scale CNV detection has some space for improvement. For an example, the recovery rates, that are the ability to call previously identified CNVs, were reported to be 55% for PennCNV and 70% for Birdsuite, when CNVs were detected by more than 20 probes[74]. However, they also reported that, when number of probes used in CNV calling increased, the recovery rate also increased. In this study, I focused on CNVs with larger than 100kb in size, thus the expected number of probes used in CNV calling in this study was 100. Hence I could anticipate that the recovery rate in this study was higher than the reported statistics in the point of probe number. In addition, when CNV frequency was under 20%, the recovery rates were reported to be 85% for PennCNV and 80% for Birdsuite[74]. According to the report, when frequency of CNV decreased, the recovery rate increased. CNVs with less than 1% of frequency were a main target in this study. I could assume that the recovery rate in this study was higher than the reported statistics from the viewpoint of frequency. Thus, in total, the CNV detection in this study was reliable.

To identify the influence of rare large-scale CNVs at specific loci on sleep disorders, a region-based analysis was conducted. Rare large-scale CNVs in the *PARK2*

region were found to be significantly associated with narcolepsy and EHS without *HLA-DQB1*06:02* (Figures 11, 12, and 13). Previous reports have shown that EHS and narcolepsy share some of susceptibility genes. About 30–50% of EHS patients carry *HLA-DQB1*06:02* allele[22-24]. Narcolepsy associated SNPs in *CPT1B* and *TRA@* identified by GWAS were also associated with EHS[22, 26]. The pathogenesis of EHS is thought to be partially identical to that of narcolepsy[22, 23, 26, 75]. Nonetheless the replication study will be necessary to validate the result.

Interestingly, the *PARK2* region is associated with other neuropsychiatric diseases such as Parkinson disease, autism and ADHD (attention-deficit hyperactivity disorder). A large number of *PARK2* mutations have been reported in patients with Parkinson disease: these mutations include point mutations, small insertions or deletions, and single or multiple exon copy number variations[76]. According to reports, eight patients with autism were found to carry *PARK2* deletions, and three patients with autism carry a duplication of *PARK2*[56, 62, 77]. A study reported the same *PARK2* region as the candidate region for ADHD[59]. They found three deletions and nine duplications in ADHD patients. Taken together, these findings indicate that variations in the *PARK2* region may result in shared genetic susceptibility across multiple neuropsychiatric diseases.

PARK2 encodes for parkin, a ubiquitin ligase. Animal studies confirmed that knockout of this gene in mice disrupted dopaminergic transmission[78, 79]. Altered central dopaminergic synaptic transmission has been implicated in several neurological and psychiatric disorders, such as Parkinson's disease, schizophrenia, and ADHD[80-82]. Individuals with these diseases demonstrate sleep abnormalities[83-85]. A study demonstrated that partial dopamine depletion causes disturbances of REM sleep without affecting motor functions. Then dysfunction of *PARK2* might lead to sleep symptoms across neuropsychiatric diseases. However, reported variants in neuropsychiatric diseases other than Parkinson diseases were rare, thus *PARK2* itself might explain only small part of genetic contribution to sleep symptoms in these diseases.

In the previous studies on other neuropsychiatric disorders such as intellectual ability, schizophrenia, autism and bipolar disorders, some of the regional CNVs were found to be associated with these diseases [86]. For an example, 16p11.2 duplication was reported in intellectual ability, schizophrenia, autism and bipolar disorder and one of the most recurrent rare and large-scale CNVs across the disorders. In this study, no rare and large-scale CNVs overlapped with previously reported pathogenic CNVs other than *PARK2*.

A pathway analysis was carried out to assess the influence rare large-scale genic CNVs throughout genome on narcolepsy. Using a false-discovery rate of 5%, 32 gene sets were significantly associated with narcolepsy (Figure 14 and Table 11). Functional clustering of these 32 gene sets revealed four groups of gene sets: immune response, acetyltransferase activity, cell cycle regulation and regulation of cell development.

Narcolepsy is well known to be associated with *HLA* and *TRA@*[3-6, 9], and autoantibodies against Tribbles homolog 2 (Trib2) have been recently detected in some patients with narcolepsy[87-89]. An enrichment of CNVs disrupting functional gene-sets involved in immune response was observed; supporting the hypothesis that narcolepsy is caused by an immune attack on hypocretin-producing neurons. The largest group of gene sets in the functional map contained genes within the acetyltransferase activity pathways. Associations between schizophrenia and histone deacetylase genes were reported in a Chinese and a Korean population[90, 91]. To my knowledge, no published reports document a relationship between narcolepsy and alterations in acetyltransferase activity; nevertheless, modified acetyltransferase activity might be an unrecognized component in the pathogenesis of narcolepsy. Novel observations included gene sets involved in cell cycle arrest, cell development and cell

morphogenesis functional groups. Further studies will be needed to interpret the relation between narcolepsy and these pathways.

The finding of this study provided the first insight of the involvement of multiple rare large-scale CNVs, both genome-wide and at specific loci, in the pathogenesis of sleep disorders, even though the replication study would be necessary. Moreover, these findings revealed new genetic and functional targets in narcolepsy that may lead to an integrated understanding of the pathogenic mechanism of narcolepsy. Although several genetic factors leading or contributing to sleep disorders have been identified, it is necessary to develop a more comprehensive understanding of the genetic underpinnings, especially the role of CNVs, in the diseases.

IV Conclusion

In the first section, I evaluated the contribution of common variants for narcolepsy onset. As a result, Narcolepsy heritability was estimated as 58.1% ($P_{HLA\text{-}DQB1*06:02} = 2.30\times10^{-48}$, $P_{others} = 6.73\times10^{-2}$) including *HLA-DQB1*06:02* effects and 1.3% ($P_{others} = 2.43\times10^{-2}$) excluding *HLA-DQB1*06:02* effects. Reported susceptible SNPs for narcolepsy in the Japanese population, such as *CPT1B*, *TRA@*, *P2RY11*, were found to explain 0.8% of heritability ($P_{others} = 9.74\times10^{-2}$). If we undertake a further research to detect narcolepsy genetic susceptibility factors, over 42,000 patients were required. Therefore, not only the collaboration across Japan but also large-scale international collaboration would be necessary for the future narcolepsy research.

Also I assessed the extent of genetic share between narcolepsy and EHS. EHS with *HLA-DQB1*06:02* patients were estimated to have higher shared genetic background with narcolepsy than EHS without *HLA-DQB1*06:02* patients even if the effects of *HLA-DQB1*06:02* were excluded (EHS with *HLA-DQB1*06:02*: 1.4%, $P_{others} = 1.56 \times 10^{-1}$, EHS without *HLA-DQB1*06:02*: 0.4%, $P_{others} = 3.06 \times 10^{-1}$). The assessment of the genetic share between narcolepsy and EHS can propose “phenotypic classification” based on genetic architecture.

In the second section of this study, I tried to detect rare susceptible variants, especially copy number variations. Overall, significant enrichments of rare large-scale

CNVs (frequency ≤ 1%, size ≥ 100 kb) in the patients were found (narcolepsy: case-control ratio of CNV count = 1.54, $P = 5.00\times10^{-4}$, EHS with *HLA-DQB1*06:02*: case-control ratio of CNV count = 1.46, $P = 3.24\times10^{-2}$, EHS without *HLA-DQB1*06:02*: case-control ratio of CNV count = 1.43, $P = 2.17\times10^{-3}$). In particular, rare large-scale CNVs in *PARK2* region showed significant associations with narcolepsy and EHS without *HLA-DQB1*06:02* (narcolepsy: $P = 3.07\times10^{-2}$, EHS without *HLA-DQB1*06:02*: $P = 2.91\times10^{-2}$). The result indicated that variations in the *PARK2* region might result in genetic susceptibility to multiple neuropsychiatric diseases. In total, the findings threw light on the possibility that common disease-rare variant hypothesis (CDRV hypothesis) were included in the pathogenesis of narcolepsy and EHS. In the future, devoted collection of samples will be essential for a replication study.

IV Acknowledgements

I owe my deepest gratitude to Professor Katsushi Tokunaga for supervising my study and providing me opportunity to complete my doctor degree study in the University of Tokyo. I am heartily thankful to my supervisor, Assistant Professor Taku Miyagawa, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. I am very thankful to Associate Professor Akihiko Mabuchi, who provided me precious and critical advice and comments. I would like to show my gratitude to Dr. Yutaka Honda (Japan Somnology Centre, Neuropsychiatric Research Institute) and Dr. Makoto Honda (Sleep Control Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science) for providing detailed diagnosis and blood samples of patients with narcolepsy and EHS. I would also like to show my gratitude to Prof. Tsukasa Sasaki (Department of Physical and Health Education, Graduate School of Education, The University of Tokyo), Prof. Ryozo Kuwano (Department of Molecular Genetics, Bioresource Science Branch, Center for Bioresources, Brain Research Institute, Niigata University), Prof. Kiyoto Kasai, Dr. Takeshi Otowa (Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo) and Dr. Takafumi Shimada (Division for Counseling and Support, The University of Tokyo) for kind support regarding with samples of healthy controls and

neuropsychiatric patients. I would like to thank Dr. Asako Koike (Central Research Laboratory, Hitachi Ltd.) for instructing me about programming skills. I am also thankful to many of my colleagues to support me during the course of this study. Lastly, I offer my regards and blessings to all of the members of Department of Human Genetics, Graduate School of Medicine, The University of Tokyo.

References

1. Mignot, E., *Genetic and familial aspects of narcolepsy.* Neurology, 1998. **50**(2 Suppl 1): p. S16-22.
2. Overeem, S., Mignot, E., van Dijk, J. G. and Lammers, G. J., *Narcolepsy: clinical features, new pathophysiologic insights, and future perspectives.* J Clin Neurophysiol, 2001. **18**(2): p. 78-105.
3. Juji, T., Satake, M., Honda, Y. and Doi, Y., *HLA antigens in Japanese patients with narcolepsy. All the patients were DR2 positive.* Tissue Antigens, 1984. **24**(5): p. 316-319.
4. Langdon, N., Welsh, K. I., van Dam, M., Vaughan, R. W. and Parkes, D., *Genetic markers in narcolepsy.* Lancet, 1984. **2**(8413): p. 1178-1180.
5. Matsuki, K., Juji, T., Tokunaga, K., Naohara, T., Satake, M. and Honda, Y., *Human histocompatibility leukocyte antigen (HLA) haplotype frequencies estimated from the data on HLA class I, II, and III antigens in 111 Japanese narcoleptics.* J Clin Invest, 1985. **76**(6): p. 2078-2083.
6. Miyagawa, T., Hohjoh, H., Honda, Y., Juji, T. and Tokunaga, K., *Identification of a telomeric boundary of the HLA region with potential for predisposition to human narcolepsy.* Immunogenetics, 2000. **52**(1-2): p. 12-18.
7. Mignot, E., Lin, L., Rogers, W., Honda, Y., Qiu, X., Lin, X., Okun, M., Hohjoh, H., Miki, T., Hsu, S., Leffell, M., Grumet, F., Fernandez-Vina, M., Honda, M. and Risch, N., *Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups.* Am J Hum Genet, 2001. **68**(3): p. 686-699.
8. Miyagawa, T., Kawashima, M., Nishida, N., Ohashi, J., Kimura, R., Fujimoto, A., Shimada, M., Morishita, S., Shigeta, T., Lin, L., Hong, S. C., Faraco, J., Shin, Y. K., Jeong, J. H., Okazaki, Y., Tsuji, S., Honda, M., Honda, Y., Mignot, E. and Tokunaga, K., *Variant between CPT1B and CHKB associated with susceptibility to narcolepsy.* Nat Genet, 2008. **40**(11): p. 1324-1328.
9. Hallmayer, J., Faraco, J., Lin, L., Hesselson, S., Winkelmann, J., Kawashima, M., Mayer, G., Plazzi, G., Nevsimalova, S., Bourgin, P., Hong, S. C., Honda, Y., Honda, M., Hogl, B., Longstreth, W. T., Jr.,

Montplaisir, J., Kemlink, D., Einen, M., Chen, J., Musone, S. L., Akana, M., Miyagawa, T., Duan, J., Desautels, A., Erhardt, C., Hesla, P. E., Poli, F., Frauscher, B., Jeong, J. H., Lee, S. P., Ton, T. G., Kvale, M., Kolesar, L., Dobrovolna, M., Nepom, G. T., Salomon, D., Wichmann, H. E., Rouleau, G. A., Gieger, C., Levinson, D. F., Gejman, P. V., Meitinger, T., Young, T., Peppard, P., Tokunaga, K., Kwok, P. Y., Risch, N. and Mignot, E., *Narcolepsy is strongly associated with the T-cell receptor alpha locus.* Nat Genet, 2009. **41**(6): p. 708-711.

10. Kornum, B. R., Kawashima, M., Faraco, J., Lin, L., Rico, T. J., Hesselson, S., Axtell, R. C., Kuipers, H., Weiner, K., Hamacher, A., Kassack, M. U., Han, F., Knudsen, S., Li, J., Dong, X., Winkelmann, J., Plazzi, G., Nevsimalova, S., Hong, S. C., Honda, Y., Honda, M., Hogl, B., Ton, T. G., Montplaisir, J., Bourgin, P., Kemlink, D., Huang, Y. S., Warby, S., Einen, M., Eshragh, J. L., Miyagawa, T., Desautels, A., Ruppert, E., Hesla, P. E., Poli, F., Pizza, F., Frauscher, B., Jeong, J. H., Lee, S. P., Strohl, K. P., Longstreth, W. T., Jr., Kvale, M., Dobrovolna, M., Ohayon, M. M., Nepom, G. T., Wichmann, H. E., Rouleau, G. A., Gieger, C., Levinson, D. F., Gejman, P. V., Meitinger, T., Peppard, P., Young, T., Jennum, P., Steinman, L., Tokunaga, K., Kwok, P. Y., Risch, N., Hallmayer, J. and Mignot, E., *Common variants in P2RY11 are associated with narcolepsy.* Nat Genet, 2011. **43**(1): p. 66-71.

11. Faraco, J., Lin, L., Kornum, B. R., Kenny, E. E., Trynka, G., Einen, M., Rico, T. J., Lichtner, P., Dauvilliers, Y., Arnulf, I., Lecendreux, M., Javidi, S., Geisler, P., Mayer, G., Pizza, F., Poli, F., Plazzi, G., Overeem, S., Lammers, G. J., Kemlink, D., Sonka, K., Nevsimalova, S., Rouleau, G., Desautels, A., Montplaisir, J., Frauscher, B., Ehrmann, L., Hogl, B., Jennum, P., Bourgin, P., Peraita-Adrados, R., Iranzo, A., Bassetti, C., Chen, W. M., Concannon, P., Thompson, S. D., Damotte, V., Fontaine, B., Breban, M., Gieger, C., Klopp, N., Deloukas, P., Wijmenga, C., Hallmayer, J., Onengut-Gumuscu, S., Rich, S. S., Winkelmann, J. and Mignot, E., *ImmunoChip study implicates antigen presentation to T cells in narcolepsy.* PLoS Genet, 2013. **9**(2): p. e1003270.

12. Han, F., Faraco, J., Dong, X. S., Ollila, H. M., Lin, L., Li, J., An, P., Wang, S., Jiang, K. W., Gao, Z. C., Zhao, L., Yan, H., Liu, Y. N., Li, Q. H., Zhang, X. Z., Hu, Y., Wang, J. Y., Lu, Y. H., Lu, C. J., Zhou, W.,

Hallmayer, J., Huang, Y. S., Strohl, K. P., Pollmacher, T. and Mignot, E., *Genome wide analysis of narcolepsy in China implicates novel immune loci and reveals changes in association prior to versus after the 2009 H1N1 influenza pandemic.* PLoS Genet, 2013. **9**(10): p. e1003880.

13. Chemelli, R. M., Willie, J. T., Sinton, C. M., Elmquist, J. K., Scammell, T., Lee, C., Richardson, J. A., Williams, S. C., Xiong, Y., Kisanuki, Y., Fitch, T. E., Nakazato, M., Hammer, R. E., Saper, C. B. and Yanagisawa, M., *Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation.* Cell, 1999. **98**(4): p. 437-451.
14. Lin, L., Faraco, J., Li, R., Kadotani, H., Rogers, W., Lin, X., Qiu, X., de Jong, P. J., Nishino, S. and Mignot, E., *The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene.* Cell, 1999. **98**(3): p. 365-376.
15. Mignot, E., Lammers, G. J., Ripley, B., Okun, M., Nevsimalova, S., Overeem, S., Vankova, J., Black, J., Harsh, J., Bassetti, C., Schrader, H. and Nishino, S., *The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias.* Arch Neurol, 2002. **59**(10): p. 1553-1562.
16. Nishino, S., Ripley, B., Overeem, S., Lammers, G. J. and Mignot, E., *Hypocretin (orexin) deficiency in human narcolepsy.* Lancet, 2000. **355**(9197): p. 39-40.
17. Gencik, M., Dahmen, N., Wieczorek, S., Kasten, M., Bierbrauer, J., Anghelescu, I., Szegedi, A., Menezes Saecker, A. M. and Epplen, J. T., *A prepro-orexin gene polymorphism is associated with narcolepsy.* Neurology, 2001. **56**(1): p. 115-117.
18. Olafsdottir, B. R., Rye, D. B., Scammell, T. E., Matheson, J. K., Stefansson, K. and Gulcher, J. R., *Polymorphisms in hypocretin/orexin pathway genes and narcolepsy.* Neurology, 2001. **57**(10): p. 1896-1899.
19. Peyron, C., Faraco, J., Rogers, W., Ripley, B., Overeem, S., Charnay, Y., Nevsimalova, S., Aldrich, M., Reynolds, D., Albin, R., Li, R., Hungs, M., Pedrazzoli, M., Padigaru, M., Kucherlapati, M., Fan, J., Maki, R., Lammers, G. J., Bouras, C., Kucherlapati, R., Nishino, S. and Mignot, E., *A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains.* Nat Med, 2000. **6**(9): p. 991-997.

20. Hungs, M., Lin, L., Okun, M. and Mignot, E., *Polymorphisms in the vicinity of the hypocretin/orexin are not associated with human narcolepsy.* Neurology, 2001. **57**(10): p. 1893-1895.
21. Anderson, K. N., Pilsworth, S., Sharples, L. D., Smith, I. E. and Shneerson, J. M., *Idiopathic hypersomnia: a study of 77 cases.* Sleep, 2007. **30**(10): p. 1274-1281.
22. Miyagawa, T., Honda, M., Kawashima, M., Shimada, M., Tanaka, S., Honda, Y. and Tokunaga, K., *Polymorphism located between CPT1B and CHKB, and HLA-DRB1*1501-DQB1*0602 haplotype confer susceptibility to CNS hypersomnias (essential hypersomnia).* PLoS One, 2009. **4**(4): p. e5394.
23. Honda, Y., Juji, T., Matsuki, K., Naohara, T., Satake, M., Inoko, H., Someya, T., Harada, S. and Doi, Y., *HLA-DR2 and Dw2 in narcolepsy and in other disorders of excessive somnolence without cataplexy.* Sleep, 1986. **9**: p. 1-2.
24. Honda, Y., Takahashi, Y., Honda, M., Watanabe, Y., Sato, T., Miki, T. , *Genetic Aspects of Narcolepsy. Sleep and Sleep Disorders: From Molecule to Behavior.* Academic Press, New York, , 1998.
25. Komada, Y., Inoue, Y., Mukai, J., Shirakawa, S., Takahashi, K. and Honda, Y., *Difference in the characteristics of subjective and objective sleepiness between narcolepsy and essential hypersomnia.* Psychiatry and Clinical Neurosciences, 2005. **59**(2): p. 194-199.
26. Miyagawa, T., Honda, M., Kawashima, M., Shimada, M., Tanaka, S., Honda, Y. and Tokunaga, K., *Polymorphism located in TCRA locus confers susceptibility to essential hypersomnia with HLA-DRB1*1501-DQB1*0602 haplotype.* J Hum Genet, 2010. **55**(1): p. 63-65.
27. Maher, B., *Personal genomes: The case of the missing heritability.* Nature, 2008. **456**(7218): p. 18-21.
28. Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. and Visscher, P. M., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p.

747-753.

29. Wright, S., *Evolution and the Genetics of Populations.* Chicago: University of Chicago Press., 1968. **1**(Genetic and Biometric Foundations.).

30. Allen, H. L., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segre, A. V., Speliotes, E. K., Wheeler, E., Soranzo, N., Park, J. H., Yang, J., Gudbjartsson, D., Heard-Costa, N. L., Randall, J. C., Qi, L., Smith, A. V., Magi, R., Pastinen, T., Liang, L., Heid, I. M., Luan, J., Thorleifsson, G., Winkler, T. W., Goddard, M. E., Lo, K. S., Palmer, C., Workalemahu, T., Aulchenko, Y. S., Johansson, A., Zillikens, M. C., Feitosa, M. F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N. L., Hayward, C., Hottenga, J. J., Jacobs, K. B., Knowles, J. W., Kutalik, Z., Monda, K. L., Polasek, O., Preuss, M., Rayner, N. W., Robertson, N. R., Steinthorsdottir, V., Tyrer, J. P., Voight, B. F., Wiklund, F., Xu, J. F., Zhao, J. H., Nyholt, D. R., Pellikka, N., Perola, M., Perry, J. R. B., Surakka, I., Tammesoo, M. L., Altmaier, E. L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D. I., Chen, C., Coin, L., Cooper, M. N., Dixon, A. L., Gibson, Q., Grundberg, E., Hao, K., Junttila, M. J., Kaplan, L. M., Kettunen, J., Konig, I. R., Kwan, T., Lawrence, R. W., Levinson, D. F., Lorentzon, M., McKnight, B., Morris, A. P., Muller, M., Ngwa, J. S., Purcell, S., Rafelt, S., Salem, R. M., Salvi, E., Sanna, S., Shi, J. X., Sovio, U., Thompson, J. R., Turchin, M. C., Vandenput, L., Verlaan, D. J., Vitart, V., White, C. C., Ziegler, A., Almgren, P., Balmforth, A. J., Campbell, H., Citterio, L., De Grandi, A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J. G., Freimer, N. B., Geus, E. J. C., Glorioso, N., Haiqing, S., Hartikainen, A. L., Havulinna, A. S., Hicks, A. A., Hui, J. N., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelaeinen, T. O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J. J., Lokki, M. L., Marusic, A., Maschio, A., Meitinger, T., Mulas, A., Pare, G., Parker, A. N., Peden, J. F., Petersmann, A., Pichler, I., Pietilainen, K. H., Pouta, A., Riddertrale, M., Rotter, J. I., Sambrook, J. G., Sanders, A. R., Schmidt, C. O., Sinisalo, J., Smit, J. H., Stringham, H. M., Walters, G. B., Widen, E.,

Wild, S. H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W. L., Nelis, M., Peters, M. J., Ripatti, S., Meurs, J. B. J., Aben, K. K., Ardlie, K. G., Beckmann, J. S., Beilby, J. P., Bergman, R. N., Bergmann, S., Collins, F. S., Cusi, D., den Heijer, M., Eiriksdottir, G., Gejman, P. V., Hall, A. S., Hamsten, A., Huikuri, H. V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L. J., Lehtimaki, T., Melander, O., Mosley, T. H., Musk, A. W., Nieminen, M. S., O'Donnell, C. J., Ohlsson, C., Oostra, B., Palmer, L. J., Raitakari, O., Ridker, P. M., Rioux, J. D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A. R., Siscovick, D. S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G. J., Viikari, J., Heath, A. C., Martin, N. G., Montgomery, G. W., Province, M. A., Kayser, M., Arnold, A. M., Atwood, L. D., Boerwinkle, E., Chanock, S. J., Deloukas, P., Gieger, C., Gronberg, H., Hall, P., Hattersley, A. T., Hengstenberg, C., Hoffman, W., Lathrop, G. M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A. F., Assimes, T. L., Barroso, I., Hofman, A., Mohlke, K. L., Boomsma, D. I., Caulfield, M. J., Cupples, L. A., Erdmann, J., Fox, C. S., Gudnason, V., Gyllensten, U., Harris, T. B., Hayes, R. B., Jarvelin, M. R., Mooser, V., Munroe, P. B., Ouwehand, W. H., Penninx, B. W., Pramstaller, P. P., Quertermous, T., Rudan, I., Samani, N. J., Spector, T. D., Volzke, H., Watkins, H., Wilson, J. F., Groop, L. C., Haritunians, T., Hu, F. B., Kaplan, R. C., Metspalu, A., North, K. E., Schlessinger, D., Wareham, N. J., Hunter, D. J., O'Connell, J. R., Strachan, D. P., Schadt, H. E., Thorsteinsdottir, U., Peltonen, L., Uitterlinden, A. G., Visscher, P. M., Chatterjee, N., Loos, R. J. F., Boehnke, M., McCarthy, M. I., Ingelsson, E., Lindgren, C. M., Abecasis, G. R., Stefansson, K., Frayling, T. M., Hirschhorn, J. N. and Consortium, P., *Hundreds of variants clustered in genomic loci and biological pathways affect human height.* Nature, 2010. **467**(7317): p. 832-838.

31. Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Magi, R., Randall, J. C., Vedantam, S., Winkler, T. W., Qi, L., Workalemahu, T., Heid, I. M., Steinthorsdottir, V., Stringham, H. M., Weedon, M. N., Wheeler, E., Wood, A. R., Ferreira, T., Weyant, R. J., Segre, A. V., Estrada, K., Liang, L. M., Nemesh, J., Park, J. H., Gustafsson, S.,

Kilpelanen, T. O., Yang, J. A., Bouatia-Naji, N., Esko, T., Feitosa, M. F., Kutalik, Z., Mangino, M., Raychaudhuri, S., Scherag, A., Smith, A. V., Welch, R., Zhao, J. H., Aben, K. K., Absher, D. M., Amin, N., Dixon, A. L., Fisher, E., Glazer, N. L., Goddard, M. E., Heard-Costa, N. L., Hoesel, V., Hottenga, J. J., Johansson, A., Johnson, T., Ketkar, S., Lamina, C., Li, S. X., Moffatt, M. F., Myers, R. H., Narisu, N., Perry, J. R. B., Peters, M. J., Preuss, M., Ripatti, S., Rivadeneira, F., Sandholt, C., Scott, L. J., Timpson, N. J., Tyrer, J. P., van Wingerden, S., Watanabe, R. M., White, C. C., Wiklund, F., Barlassina, C., Chasman, D. I., Cooper, M. N., Jansson, J. O., Lawrence, R. W., Pellikka, N., Prokopenko, I., Shi, J. X., Thiering, E., Alavere, H., Alibrandi, M. T. S., Almgren, P., Arnold, A. M., Aspelund, T., Atwood, L. D., Balkau, B., Balmforth, A. J., Bennett, A. J., Ben-Shlomo, Y., Bergman, R. N., Bergmann, S., Biebermann, H., Blakemore, A. I. F., Boes, T., Bonnycastle, L. L., Bornstein, S. R., Brown, M. J., Buchanan, T. A., Busonero, F., Campbell, H., Cappuccio, F. P., Cavalcanti-Proenca, C., Chen, Y. D. I., Chen, C. M., Chines, P. S., Clarke, R., Coin, L., Connell, J., Day, I. N. M., den Heijer, M., Duan, J. B., Ebrahim, S., Elliott, P., Elosua, R., Eiriksdottir, G., Erdos, M. R., Eriksson, J. G., Facheris, M. F., Felix, S. B., Fischer-Posovszky, P., Folsom, A. R., Friedrich, N., Freimer, N. B., Fu, M., Gaget, S., Gejman, P. V., Geus, E. J. C., Gieger, C., Gjesing, A. P., Goel, A., Goyette, P., Grallert, H., Grassler, J., Greenawalt, D. M., Groves, C. J., Gudnason, V., Guiducci, C., Hartikainen, A. L., Hassanali, N., Hall, A. S., Havulinna, A. S., Hayward, C., Heath, A. C., Hengstenberg, C., Hicks, A. A., Hinney, A., Hofman, A., Homuth, G., Hui, J., Igl, W., Iribarren, C., Isomaa, B., Jacobs, K. B., Jarick, I., Jewell, E., John, U., Jorgensen, T., Jousilahti, P., Jula, A., Kaakinen, M., Kajantie, E., Kaplan, L. M., Kathiresan, S., Kettunen, J., Kinnunen, L., Knowles, J. W., Kolcic, I., Konig, I. R., Koskinen, S., Kovacs, P., Kuusisto, J., Kraft, P., Kvaloy, K., Laitinen, J., Lantieri, O., Lanzani, C., Launer, L. J., Lecoeur, C., Lehtimaki, T., Lettre, G., Liu, J. J., Lokki, M. L., Lorentzon, M., Luben, R. N., Ludwig, B., Manunta, P., Marek, D., Marre, M., Martin, N. G., McArdle, W. L., McCarthy, A., McKnight, B., Meitinger, T., Melander, O., Meyre, D., Midthjell, K., Montgomery, G. W., Morken, M. A., Morris, A. P., Mulic, R., Ngwa, J. S., Nelis, M., Neville, M. J., Nyholt,

D. R., O'Donnell, C. J., O'Rahilly, S., Ong, K. K., Oostra, B., Pare, G., Parker, A. N., Perola, M., Pichler, I., Pietilainen, K. H., Platou, C. G. P., Polasek, O., Pouta, A., Rafelt, S., Raitakari, O., Rayner, N. W., Ridderstrale, M., Rief, W., Ruokonen, A., Robertson, N. R., Rzehak, P., Salomaa, V., Sanders, A. R., Sandhu, M. S., Sanna, S., Saramies, J., Savolainen, M. J., Scherag, S., Schipf, S., Schreiber, S., Schunkert, H., Silander, K., Sinisalo, J., Siscovick, D. S., Smit, J. H., Soranzo, N., Sovio, U., Stephens, J., Surakka, I., Swift, A. J., Tammesoo, M. L., Tardif, J. C., Teder-Laving, M., Teslovich, T. M., Thompson, J. R., Thomson, B., Tonjes, A., Tuomi, T., van Meurs, J. B. J., van Ommen, G. J., Vatin, V., Viikari, J., Visvikis-Siest, S., Vitart, V., Vogel, C. I. G., Voight, B. F., Waite, L. L., Wallaschofski, H., Walters, G. B., Widen, E., Wiegand, S., Wild, S. H., Willemsen, G., Witte, D. R., Witteman, J. C., Xu, J. F., Zhang, Q. Y., Zgaga, L., Ziegler, A., Zitting, P., Beilby, J. P., Farooqi, I. S., Hebebrand, J., Huikuri, H. V., James, A. L., Kahonen, M., Levinson, D. F., Macciardi, F., Nieminen, M. S., Ohlsson, C., Palmer, L. J., Ridker, P. M., Stumvoll, M., Beckmann, J. S., Boeing, H., Boerwinkle, E., Boomsma, D. I., Caulfield, M. J., Chanock, S. J., Collins, F. S., Cupples, L. A., Smith, G. D., Erdmann, J., Froguel, P., Gronberg, H., Gyllensten, U., Hall, P., Hansen, T., Harris, T. B., Hattersley, A. T., Hayes, R. B., Heinrich, J., Hu, F. B., Hveem, K., Illig, T., Jarvelin, M. R., Kaprio, J., Karpe, F., Khaw, K. T., Kiemeney, L. A., Krude, H., Laakso, M., Lawlor, D. A., Metspalu, A., Munroe, P. B., Ouwehand, W. H., Pedersen, O., Penninx, B. W., Peters, A., Pramstaller, P. P., Quertermous, T., Reinehr, T., Rissanen, A., Rudan, I., Samani, N. J., Schwarz, P. E. H., Shuldiner, A. R., Spector, T. D., Tuomilehto, J., Uda, M., Uitterlinden, A., Valle, T. T., Wabitsch, M., Waeber, G., Wareham, N. J., Watkins, H., Wilson, J. F., Wright, A. F., Zillikens, M. C., Chatterjee, N., McCarroll, S. A., Purcell, S., Schadt, E. E., Visscher, P. M., Assimes, T. L., Borecki, I. B., Deloukas, P., Fox, C. S., Groop, L. C., Haritunians, T., Hunter, D. J., Kaplan, R. C., Mohlke, K. L., O'Connell, J. R., Peltonen, L., Schlessinger, D., Strachan, D. P., van Duijn, C. M., Wichmann, H. E., Frayling, T. M., Thorsteinsdottir, U., Abecasis, G. R., Barroso, I., Boehnke, M., Stefansson, K., North, K. E., McCarthy, M. I., Hirschhorn, J. N., Ingelsson, E., Loos, R. J. F. and Magic, *Association analyses of 249,796 individuals reveal 18 new loci*

associated with body mass index. Nature Genetics, 2010. **42**(11): p. 937-U953.

32. Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., Johansen, C. T., Fouchier, S. W., Isaacs, A., Peloso, G. M., Barbalic, M., Ricketts, S. L., Bis, J. C., Aulchenko, Y. S., Thorleifsson, G., Feitosa, M. F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X. H., Guo, X. Q., Li, M. Y., Cho, Y. S., Go, M. J., Kim, Y. J., Lee, J. Y., Park, T., Kim, K., Sim, X., Ong, R. T. H., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Zhao, J. H., Yuan, X., Luan, J. A., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y. L., Wright, A. F., Witteman, J. C. M., Wilson, J. F., Willemsen, G., Wichmann, H. E., Whitfield, J. B., Waterworth, D. M., Wareham, N. J., Waeber, G., Vollenweider, P., Voight, B. F., Vitart, V., Uitterlinden, A. G., Uda, M., Tuomilehto, J., Thompson, J. R., Tanaka, T., Surakka, I., Stringham, H. M., Spector, T. D., Soranzo, N., Smit, J. H., Sinisalo, J., Silander, K., Sijbrands, E. J. G., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruokonen, A., Rudan, I., Rose, L. M., Roberts, R., Rieder, M., Psaty, B. M., Pramstaller, P. P., Pichler, I., Perola, M., Penninx, B. W. J. H., Pedersen, N. L., Pattaro, C., Parker, A. N., Pare, G., Oostra, B. A., O'Donnell, C. J., Nieminen, M. S., Nickerson, D. A., Montgomery, G. W., Meitinger, T., McPherson, R., McCarthy, M. I., McArdle, W., Masson, D., Martin, N. G., Marroni, F., Mangino, M., Magnusson, P. K. E., Lucas, G., Luben, R., Loos, R. J. F., Lokki, M. L., Lettre, G., Langenberg, C., Launer, L. J., Lakatta, E. G., Laaksonen, R., Kyvik, K. O., Kronenberg, F., Konig, I. R., Khaw, K. T., Kaprio, J., Kaplan, L. M., Johansson, A., Jarvelin, M. R., Janssens, A. C. J. W., Ingelsson, E., Igi, W., Hovingh, G. K., Hottenga, J. J., Hofman, A., Hicks, A. A., Hengstenberg, C., Heid, I. M., Hayward, C., Havulinna, A. S., Hastie, N. D., Harris, T. B., Haritunians, T., Hall, A. S., Gyllensten, U., Guiducci, C., Groop, L. C., Gonzalez, E., Gieger, C., Freimer, N. B., Ferrucci, L., Erdmann, J., Elliott, P., Ejebe, K. G., Doering, A., Dominiczak, A. F., Demissie, S., Deloukas, P., de Geus, E. J. C., de Faire, U., Crawford, G., Collins, F. S., Chen, Y. D. I., Caulfield, M. J., Campbell, H., Burtt, N. P., Bonnycastle, L. L., Boomsma, D. I.,

Boekholdt, S. M., Bergman, R. N., Barroso, I., Bandinelli, S., Ballantyne, C. M., Assimes, T. L., Quertermous, T., Altshuler, D., Seielstad, M., Wong, T. Y., Tai, E. S., Feranil, A. B., Kuzawa, C. W., Adair, L. S., Taylor, H. A., Borecki, I. B., Gabriel, S. B., Wilson, J. G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R. M., Mohlke, K. L., Ordovas, J. M., Munroe, P. B., Kooner, J. S., Tall, A. R., Hegele, R. A., Kastelein, J. J. P., Schadt, E. E., Rotter, J. I., Boerwinkle, E., Strachan, D. P., Mooser, V., Stefansson, K., Reilly, M. P., Samani, N. J., Schunkert, H., Cupples, L. A., Sandhu, M. S., Ridker, P. M., Rader, D. J., van Duijn, C. M., Peltonen, L., Abecasis, G. R., Boehnke, M. and Kathiresan, S., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. **466**(7307): p. 707-713.

33. Franke, A., McGovern, D. P., Barrett, J. C., Wang, K., Radford-Smith, G. L., Ahmad, T., Lees, C. W., Balschun, T., Lee, J., Roberts, R., Anderson, C. A., Bis, J. C., Bumpstead, S., Ellinghaus, D., Festen, E. M., Georges, M., Green, T., Haritunians, T., Jostins, L., Latiano, A., Mathew, C. G., Montgomery, G. W., Prescott, N. J., Raychaudhuri, S., Rotter, J. I., Schumm, P., Sharma, Y., Simms, L. A., Taylor, K. D., Whiteman, D., Wijmenga, C., Baldassano, R. N., Barclay, M., Bayless, T. M., Brand, S., Buning, C., Cohen, A., Colombel, J. F., Cottone, M., Stronati, L., Denson, T., De Vos, M., D'Inca, R., Dubinsky, M., Edwards, C., Florin, T., Franchimont, D., Gearry, R., Glas, J., Van Gossum, A., Guthery, S. L., Halfvarson, J., Verspaget, H. W., Hugot, J. P., Karban, A., Laukens, D., Lawrance, I., Lemann, M., Levine, A., Libioulle, C., Louis, E., Mowat, C., Newman, W., Panes, J., Phillips, A., Proctor, D. D., Regueiro, M., Russell, R., Rutgeerts, P., Sanderson, J., Sans, M., Seibold, F., Steinhart, A. H., Stokkers, P. C., Torkvist, L., Kullak-Ublick, G., Wilson, D., Walters, T., Targan, S. R., Brant, S. R., Rioux, J. D., D'Amato, M., Weersma, R. K., Kugathasan, S., Griffiths, A. M., Mansfield, J. C., Vermeire, S., Duerr, R. H., Silverberg, M. S., Satsangi, J., Schreiber, S., Cho, J. H., Annese, V., Hakonarson, H., Daly, M. J. and Parkes, M., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nat Genet, 2010. **42**(12): p. 1118-1125.

34. Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., Ruderfer, D. M., McQuillin, A., Morris, D.

W., O'Dushlaine, C. T., Corvin, A., Holmans, P. A., Macgregor, S., Gurling, H., Blackwood, D. H. R., Corvin, A., Craddock, N. J., Gill, M., Hultman, C. M., Kirov, G. K., Lichtenstein, P., Muir, W. J., Owen, M. J., Pato, C. N., Scolnick, E. M., St Clair, D., Craddock, N. J., Holmans, P. A., Williams, N. M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Hultman, C. M., Lichtenstein, P., Thelander, E. F., Sullivan, P., Kenny, E., Quinn, E. M., Gill, M., Corvin, A., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Crombie, C., Fraser, G., Kuan, S. L., Walker, N., Blackwood, D. H. R., Muir, W. J., McGhee, K. A., Pickard, B., Malloy, P., Maclean, A. W., Van Beck, M., Wray, N. R., Macgregor, S., Visscher, P. M., Pato, M. T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J. A., Ferreira, C. P., Macedo, A., Azevedo, M. H., Kirby, A. N., Ferreira, M. A. R., Daly, M. J., Chambert, K., Kuruvilla, F., Gabriel, S. B., Ardlie, K., Moran, J. L., Daly, M. J. and Scolnick, E. M., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-752.

35. Yang, J. A., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. and Visscher, P. M., *Common SNPs explain a large proportion of the heritability for human height.* Nature Genetics, 2010. **42**(7): p. 565-U131.

36. Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., Kraft, P., Chen, R., Kallberg, H. J., Kurreeman, F. A. S., Kathiresan, S., Wijmenga, C., Gregersen, P. K., Alfredsson, L., Siminovitch, K. A., Worthington, J., de Bakker, P. I. W., Raychaudhuri, S., Plenge, R. M., Metaan, D. G. R. and Con, M. I. G., *Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis.* Nature Genetics, 2012. **44**(5): p. 483-+.

37. Lee, S. H., Wray, N. R., Goddard, M. E. and Visscher, P. M., *Estimating Missing Heritability for Disease from Genome-wide Association Studies.* American Journal of Human Genetics, 2011. **88**(3): p. 294-305.

38. Liu, X. X., Kawamura, Y., Shimada, T., Otowa, T., Koishi, S., Sugiyama, T., Nishida, H., Hashimoto, O., Nakagami, R., Tochigi, M.,

Umekage, T., Kano, Y., Miyagawa, T., Kato, N., Tokunaga, K. and Sasaki, T., *Association of the oxytocin receptor (OXTR) gene polymorphisms with autism spectrum disorder (ASD) in the Japanese population.* Journal of Human Genetics, 2010. **55**(3): p. 137-141.

39. Shiota, S., Tochigi, M., Shimada, H., Ohashi, J., Kasai, K., Kato, N., Tokunaga, K. and Sasaki, T., *Association and interaction analyses of NRG1 and ERBB4 genes with schizophrenia in a Japanese population.* Journal of Human Genetics, 2008. **53**(10): p. 929-935.
40. Tochigi, M., Zhang, X., Ohashi, J., Hibino, H., Otowa, T., Rogers, M., Kato, T., Okazaki, Y., Kato, N., Tokunaga, K. and Sasaki, T., *Association Study Between the TNXB Locus and Schizophrenia in a Japanese Population.* American Journal of Medical Genetics Part B (Neuropsychiatric Genetics), 2007. **144B**(3): p. 305-309.
41. Otowa, T., Kawamura, Y., Nishida, N., Sugaya, N., Koike, A., Yoshida, E., Inoue, K., Yasuda, S., Nishimura, Y., Liu, X., Konishi, Y., Nishimura, F., Shimada, T., Kuwabara, H., Tochigi, M., Kakiuchi, C., Umekage, T., Miyagawa, T., Miyashita, A., Shimizu, E., Akiyoshi, J., Someya, T., Kato, T., Yoshikawa, T., Kuwano, R., Kasai, K., Kato, N., Kaiya, H., Tokunaga, K., Okazaki, Y., Tanii, H. and Sasaki, T., *Meta-analysis of genome-wide association studies for panic disorder in the Japanese population.* Translational Psychiatry, 2012. **2**.
42. Adachi, S., Tajima, A., Quan, J., Haino, K., Yoshihara, K., Masuzaki, H., Katabuchi, H., Ikuma, K., Suginami, H., Nishida, N., Kuwano, R., Okazaki, Y., Kawamura, Y., Sasaki, T., Tokunaga, K., Inoue, I. and Tanaka, K., *Meta-analysis of genome-wide association scans for genetic susceptibility to endometriosis in Japanese population.* J Hum Genet, 2010. **55**(12): p. 816-821.
43. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. and Sham, P. C., *PLINK: A tool set for whole-genome association and population-based linkage analyses.* American Journal of Human Genetics, 2007. **81**(3): p. 559-575.
44. Patterson, N., Price, A. L. and Reich, D., *Population structure and eigenanalysis.* Plos Genetics, 2006. **2**(12): p. 2074-2093.
45. International HapMap, C., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-796.

46. Kundu, S., Aulchenko, Y. S., van Duijn, C. M. and Janssens, A. C., *PredictABEL: an R package for the assessment of risk prediction models.* Eur J Epidemiol, 2011. **26**(4): p. 261-264.
47. Austin, G., *The New S Language - a Programming Environment for Data-Analysis and Graphics.* Economic Journal, 1990. **100**(401): p. 650-652.
48. Pollmacher, T., Schulz, H., Geisler, P., Kiss, E., Albert, E. D. and Schwarzfischer, F., *Dr2-Positive Monozygotic Twins Discordant for Narcolepsy.* Sleep, 1990. **13**(4): p. 336-343.
49. Barclay, N. L. and Gregory, A. M., *Quantitative genetic research on sleep: A review of normal sleep, sleep disturbances and associated emotional, behavioural, and health-related difficulties.* Sleep Medicine Reviews, 2013. **17**(1): p. 29-40.
50. Lencz, T., Knowles, E., Davies, G., Guha, S., Liewald, D. C., Starr, J. M., Djurovic, S., Melle, I., Sundet, K., Christoforou, A., Reinvang, I., Mukherjee, S., DeRosse, P., Lundervold, A., Steen, V. M., John, M., Espeseth, T., Raikkonen, K., Widen, E., Palotie, A., Eriksson, J. G., Giegling, I., Konte, B., Ikeda, M., Roussos, P., Giakoumaki, S., Burdick, K. E., Payton, A., Ollier, W., Horan, M., Donohoe, G., Morris, D., Corvin, A., Gill, M., Pendleton, N., Iwata, N., Darvasi, A., Bitsios, P., Rujescu, D., Lahti, J., Hellard, S. L., Keller, M. C., Andreassen, O. A., Deary, I. J., Glahn, D. C. and Malhotra, A. K., *Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consorTium (COGENT).* Molecular Psychiatry, 2014. **19**(2): p. 168-174.
51. Gengler, N., VanVleck, L. D., MacNeil, M. D., Misztal, I. and Pariacote, F. A., *Influence of dominance relationships on the estimation of dominance variance with sire-dam subclass effects.* Journal of Animal Science, 1997. **75**(11): p. 2885-2891.
52. Palucci, V., Schaeffer, L. R., Miglior, F. and Osborne, V., *Non-additive genetic effects for fertility traits in Canadian Holstein cattle (Open Access publication).* Genetics Selection Evolution, 2007. **39**(2): p. 181-193.
53. Norris, D., Varona, L., Ngambi, J. W., Visser, D. P., Mbajiorgu, C. A. and Voordewind, S. F., *Estimation of the additive and dominance variances in SA Duroc pigs.* Livestock Science, 2010. **131**(1): p.

144-147.

54. Kirino, Y., Bertsias, G., Ishigatsubo, Y., Mizuki, N., Tugal-Tutkun, I., Seyahi, E., Ozyazgan, Y., Sacli, F. S., Erer, B., Inoko, H., Emrence, Z., Cakar, A., Abaci, N., Ustek, D., Satorius, C., Ueda, A., Takeno, M., Kim, Y., Wood, G. M., Ombrello, M. J., Meguro, A., Gul, A., Remmers, E. F. and Kastner, D. L., *Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1.* Nat Genet, 2013. **45**(2): p. 202-207.

55. Consortium, T. I. S., *Rare chromosomal deletions and duplications increase risk of schizophrenia.* Nature, 2008. **455**(7210): p. 237-241.

56. Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., Zhang, H., Estes, A., Brune, C. W., Bradfield, J. P., Imielinski, M., Frackelton, E. C., Reichert, J., Crawford, E. L., Munson, J., Sleiman, P. M., Chiavacci, R., Annaiah, K., Thomas, K., Hou, C., Glaberson, W., Flory, J., Otieno, F., Garris, M., Soorya, L., Klei, L., Piven, J., Meyer, K. J., Anagnostou, E., Sakurai, T., Game, R. M., Rudd, D. S., Zurawiecki, D., McDougle, C. J., Davis, L. K., Miller, J., Posey, D. J., Michaels, S., Kolevzon, A., Silverman, J. M., Bernier, R., Levy, S. E., Schultz, R. T., Dawson, G., Owley, T., McMahon, W. M., Wassink, T. H., Sweeney, J. A., Nurnberger, J. I., Coon, H., Sutcliffe, J. S., Minshew, N. J., Grant, S. F., Bucan, M., Cook, E. H., Buxbaum, J. D., Devlin, B., Schellenberg, G. D. and Hakonarson, H., *Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.* Nature, 2009. **459**(7246): p. 569-573.

57. Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Almeida, J., Bacchelli, E., Bader, G. D., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bolte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian, J., Bryson, S. E., Carson, A. R., Casallo, G., Casey, J., Chung, B. H., Cochrane, L., Corsello, C., Crawford, E. L., Crossett, A., Cytrynbaum, C., Dawson, G., de Jonge, M., Delorme, R., Drmic, I., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B. A., Folstein, S. E., Fombonne, E., Freitag, C. M., Gilbert, J., Gillberg, C., Glessner, J. T., Goldberg, J., Green, A., Green, J., Guter, S. J., Hakonarson, H., Heron, E. A., Hill, M., Holt, R., Howe, J. L., Hughes, G., Hus, V., Igliozzi, R., Kim, C., Klauck, S. M., Kolevzon, A.,

Korvatska, O., Kustanovich, V., Lajonchere, C. M., Lamb, J. A., Laskawiec, M., Leboyer, M., Le Couteur, A., Leventhal, B. L., Lionel, A. C., Liu, X. Q., Lord, C., Lotspeich, L., Lund, S. C., Maestrini, E., Mahoney, W., Mantoulan, C., Marshall, C. R., McConachie, H., McDougle, C. J., McGrath, J., McMahon, W. M., Merikangas, A., Migita, O., Minshew, N. J., Mirza, G. K., Munson, J., Nelson, S. F., Noakes, C., Noor, A., Nygren, G., Oliveira, G., Papanikolaou, K., Parr, J. R., Parrini, B., Paton, T., Pickles, A., Pilorge, M., Piven, J., Ponting, C. P., Posey, D. J., Poustka, A., Poustka, F., Prasad, A., Ragoussis, J., Renshaw, K., Rickaby, J., Roberts, W., Roeder, K., Roge, B., Rutter, M. L., Bierut, L. J., Rice, J. P., Salt, J., Sansom, K., Sato, D., Segurado, R., Sequeira, A. F., Senman, L., Shah, N., Sheffield, V. C., Soorya, L., Sousa, I., Stein, O., Sykes, N., Stoppioni, V., Strawbridge, C., Tancredi, R., Tansey, K., Thiruvahindrapduram, B., Thompson, A. P., Thomson, S., Tryfon, A., Tsiantis, J., Van Engeland, H., Vincent, J. B., Volkmar, F., Wallace, S., Wang, K., Wang, Z., Wassink, T. H., Webber, C., Weksberg, R., Wing, K., Wittemeyer, K., Wood, S., Wu, J., Yaspan, B. L., Zurawiecki, D., Zwaigenbaum, L., Buxbaum, J. D., Cantor, R. M., Cook, E. H., Coon, H., Cuccaro, M. L., Devlin, B., Ennis, S., Gallagher, L., Geschwind, D. H., Gill, M., Haines, J. L., Hallmayer, J., Miller, J., Monaco, A. P., Nurnberger, J. I., Jr., Paterson, A. D., Pericak-Vance, M. A., Schellenberg, G. D., Szatmari, P., Vicente, A. M., Vieland, V. J., Wijsman, E. M., Scherer, S. W., Sutcliffe, J. S. and Betancur, C., *Functional impact of global rare copy number variation in autism spectrum disorders.* Nature, 2010. **466**(7304): p. 368-372.

58. Williams, N. M., Zaharieva, I., Martin, A., Langley, K., Mantripragada, K., Fossdal, R., Stefansson, H., Stefansson, K., Magnusson, P., Gudmundsson, O. O., Gustafsson, O., Holmans, P., Owen, M. J., O'Donovan, M. and Thapar, A., *Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis.* Lancet, 2010. **376**(9750): p. 1401-1408.

59. Jarick, I., Volckmar, A. L., Putter, C., Pechlivanis, S., Nguyen, T. T., Dauvermann, M. R., Beck, S., Albayrak, O., Scherag, S., Gilsbach, S., Cichon, S., Hoffmann, P., Degenhardt, F., Nothen, M. M., Schreiber, S., Wichmann, H. E., Jockel, K. H., Heinrich, J., Tiesler, C. M. T., Faraone, S. V., Walitza, S., Sinzig, J., Freitag, C., Meyer, J.,

Herpertz-Dahlmann, B., Lehmkuhl, G., Renner, T. J., Warnke, A., Romanos, M., Lesch, K. P., Reif, A., Schimmelmann, B. G., Hebebrand, J., Scherag, A. and Hinney, A., *Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder.* Molecular Psychiatry, 2014. **19**(1): p. 115-121.

60. Wang, L. Y., Nuytemans, K., Bademci, G., Jauregui, C., Martin, E. R., Scott, W. K., Vance, J. M. and Zuchner, S., *High-Resolution Survey in Familial Parkinson Disease Genes Reveals Multiple Independent Copy Number Variation Events in PARK2.* Human Mutation, 2013. **34**(8): p. 1071-1074.
61. Mitsui, J., Takahashi, Y., Goto, J., Tomiyama, H., Ishikawa, S., Yoshino, H., Minami, N., Smith, D. I., Lesage, S., Aburatani, H., Nishino, I., Brice, A., Hattori, N. and Tsuji, S., *Mechanisms of Genomic Instabilities Underlying Two Common Fragile-Site-Associated Loci, PARK2 and DMD, in Germ Cell and Cancer Cell Lines.* American Journal of Human Genetics, 2010. **87**(1): p. 75-89.
62. Scheuerle, A. and Wilson, K., *PARK2 copy number aberrations in two children presenting with autism spectrum disorder: further support of an association and possible evidence for a new microdeletion/microduplication syndrome.* Am J Med Genet B Neuropsychiatr Genet, 2011. **156B**(4): p. 413-420.
63. Roberts, J. L., Hovanes, K., Dasouki, M., Manzardo, A. M. and Butler, M. G., *Chromosomal microarray analysis of consecutive individuals with autism spectrum disorders or learning disability presenting for genetic services.* Gene, 2014. **535**(1): p. 70-78.
64. Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., Chu, S. H., Moreau, M. P., Gupta, A. R., Thomson, S. A., Mason, C. E., Bilguvar, K., Celestino-Soper, P. B. S., Choi, M., Crawford, E. L., Davis, L., Wright, N. R. D., Dhodapkar, R. M., DiCola, M., DiLullo, N. M., Fernandez, T. V., Fielding-Singh, V., Fishman, D. O., Frahm, S., Garagaloyan, R., Goh, G. S., Kammela, S., Klei, L., Lowe, J. K., Lund, S. C., McGrew, A. D., Meyer, K. A., Moffat, W. J., Murdoch, J. D., O'Roak, B. J., Ober, G. T., Pottenger, R. S., Raubeson, M. J., Song, Y., Wang, Q., Yaspan, B. L., Yu, T. W.,

Yurkiewicz, L. R., Beaudet, A. L., Cantor, R. M., Curland, M., Grice, D. E., Gunel, M., Lifton, R. P., Mane, S. M., Martin, D. M., Shaw, C. A., Sheldon, M., Tischfield, J. A., Walsh, C. A., Morrow, E. M., Ledbetter, D. H., Fombonne, E., Lord, C., Martin, C. L., Brooks, A. I., Sutcliffe, J. S., Cook, E. H., Geschwind, D., Roeder, K., Devlin, B. and State, M. W., *Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism.* Neuron, 2011. **70**(5): p. 863-885.

65. Koike, A., Nishida, N., Yamashita, D. and Tokunaga, K., *Comparative analysis of copy number variation detection methods and database construction.* BMC Genet, 2011. **12**: p. 29.
66. Wang, K., Li, M. Y., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., Hakonarson, H. and Bucan, M., *PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data.* Genome Research, 2007. **17**(11): p. 1665-1674.
67. McCarroll, S. A., Kuruvilla, F. G., Korn, J. M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M. H., de Bakker, P. I., Maller, J. B., Kirby, A., Elliott, A. L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P. J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K. W., Rava, R., Daly, M. J., Gabriel, S. B. and Altshuler, D., *Integrated detection and population-genetic analysis of SNPs and copy number variation.* Nat Genet, 2008. **40**(10): p. 1166-1174.
68. Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C. M., Hart, J., Landrum, M. J., McGarvey, K. M., Murphy, M. R., O'Leary, N. A., Pujar, S., Rajput, B., Rangwala, S. H., Riddick, L. D., Shkeda, A., Sun, H. Z., Tamez, P., Tully, R. E., Wallin, C., Webb, D., Weber, J., Wu, W. D., DiCuccio, M., Kitts, P., Maglott, D. R., Murphy, T. D. and Ostell, J. M., *RefSeq: an update on mammalian reference sequences.* Nucleic Acids Research, 2014. **42**(D1): p. D756-D763.
69. Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J. P., Russell, R. K., Sleiman, P. M., Imielinski, M., Glessner, J., Hou, C., Wilson, D. C., Walters, T., Kim, C., Frackelton, E. C., Lionetti, P., Barabino, A., Van Limbergen, J., Guthery, S., Denson, L., Piccoli, D., Li, M., Dubinsky, M., Silverberg, M., Griffiths, A., Grant, S. F.,

Satsangi, J., Baldassano, R. and Hakonarson, H., *Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease.* Am J Hum Genet, 2009. **84**(3): p. 399-405.

70. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. and Mesirov, J. P., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-15550.
71. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. and Consortium, G. O., *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000. **25**(1): p. 25-29.
72. Kanehisa, M. and Goto, S., *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
73. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T., *Cytoscape: A software environment for integrated models of biomolecular interaction networks.* Genome Research, 2003. **13**(11): p. 2498-2504.
74. Zhang, D. D., Qian, Y. D., Akula, N., Alliey-Rodriguez, N., Tang, J. S., Gershon, E. S., Liu, C. Y. and Study, B. G., *Accuracy of CNV Detection from GWAS Data.* Plos One, 2011. **6**(1).
75. Khor, S. S., Miyagawa, T., Toyoda, H., Yamasaki, M., Kawamura, Y., Tanii, H., Okazaki, Y., Sasaki, T., Lin, L., Faraco, J., Rico, T., Honda, Y., Honda, M., Mignot, E. and Tokunaga, K., *Genome-wide association study of HLA-DQB1*06:02 negative essential hypersomnia.* PeerJ, 2013. **1**: p. e66.
76. Crosiers, D., Theuns, J., Cras, P. and Van Broeckhoven, C., *Parkinson disease: insights in clinical, genetic and pathological features of monogenic disease subtypes.* J Chem Neuroanat, 2011. **42**(2): p. 131-141.
77. Mariani, M., Crosti, F., Redaelli, S., Fossati, C., Piras, R., Biondi, A., Dalpra, L. and Selicorni, A., *Partial duplication of the PARK2 gene in a child with developmental delay and her normal mother: A second report.* American Journal of Medical Genetics Part

B-Neuropsychiatric Genetics, 2013. **162B**(5): p. 485-486.

78. Itier, J. M., Ibanez, P., Mena, M. A., Abbas, N., Cohen-Salmon, C., Bohme, G. A., Laville, M., Pratt, J., Corti, O., Pradier, L., Ret, G., Joubert, C., Periquet, M., Araujo, F., Negroni, J., Casarejos, M. J., Canals, S., Solano, R., Serrano, A., Gallego, E., Sanchez, M., Denefle, P., Benavides, J., Tremp, G., Rooney, T. A., Brice, A. and Garcia de Yebenes, J., *Parkin gene inactivation alters behaviour and dopamine neurotransmission in the mouse.* Hum Mol Genet, 2003. **12**(18): p. 2277-2291.
79. Oyama, G., Yoshimi, K., Natori, S., Chikaoka, Y., Ren, Y. R., Funayama, M., Shimo, Y., Takahashi, R., Nakazato, T., Kitazawa, S. and Hattori, N., *Impaired in vivo dopamine release in parkin knockout mice.* Brain Res, 2010. **1352**: p. 214-222.
80. Carlsson, A., *Perspectives on the Discovery of Central Monoaminergic Neurotransmission.* Annual Review of Neuroscience, 1987. **10**: p. 19-40.
81. Mazei-Robison, M. S., Couch, R. S., Shelton, R. C., Stein, M. A. and Blakely, R. D., *Sequence variation in the human dopamine transporter gene in children with attention deficit hyperactivity disorder.* Neuropharmacology, 2005. **49**(6): p. 724-736.
82. Greenwood, T. A., Schork, N. J., Eskin, E. and Kelsoe, J. R., *Identification of additional variants within the human dopamine transporter gene provides further evidence for an association with bipolar disorder in two independent samples.* Molecular Psychiatry, 2006. **11**(2): p. 125-133.
83. Gagnon, J. F., Bedard, M. A., Fantini, M. L., Petit, D., Panisset, M., Rompre, S., Carrier, J. and Montplaisir, J., *REM sleep behavior disorder and REM sleep without atonia in Parkinson's disease.* Neurology, 2002. **59**(4): p. 585-589.
84. Abbott, A., *Neuroscience: While you were sleeping.* Nature, 2005. **437**(7063): p. 1220-1222.
85. Adler, C. H., *Nonmotor complications in Parkinson's disease.* Movement Disorders, 2005. **20**: p. S23-S29.
86. Malhotra, D. and Sebat, J., *CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics.* Cell, 2012. **148**(6): p. 1223-1241.
87. Cvetkovic-Lopes, V., Bayer, L., Dorsaz, S., Maret, S., Pradervand, S.,

Dauvilliers, Y., Lecendreux, M., Lammers, G. J., Donjacour, C. E. H. M., Du Pasquier, R. A., Pfister, C., Petit, B., Hor, H., Muhlethaler, M. and Tafti, M., *Elevated Tribbles homolog 2-specific antibody levels in narcolepsy patients.* Journal of Clinical Investigation, 2010. **120**(3): p. 713-719.

88. Toyoda, H., Tanaka, S., Miyagawa, T., Honda, Y., Tokunaga, K. and Honda, M., *Anti-Tribbles homolog 2 autoantibodies in Japanese patients with narcolepsy.* Sleep, 2010. **33**(7): p. 875-878.
89. Kawashima, M., Lin, L., Tanaka, S., Jennum, P., Knudsen, S., Nevsimalova, S., Plazzi, G. and Mignot, E., *Anti-Tribbles Homo log 2 (TRIB2) Autoantibodies in Narcolepsy are Associated with Recent Onset of Cataplexy.* Sleep, 2010. **33**(7): p. 869-874.
90. Kim, T., Park, J. K., Kim, H. J., Chung, J. H. and Kim, J. W., *Association of histone deacetylase genes with schizophrenia in Korean population.* Psychiatry Res, 2010. **178**(2): p. 266-269.
91. Han, H., Yu, Y., Shi, J., Yao, Y., Li, W., Kong, N., Wu, Y., Wang, C., Wang, S., Meng, X. and Kou, C., *Associations of histone deacetylase-2 and histone deacetylase-3 genes with schizophrenia in a Chinese population.* Asia Pac Psychiatry, 2013. **5**(1): p. 11-16.

Tables

Table 1: Phenotypic variance explained by narcolepsy polygenic risks, including *HLA-DQB1*06:02* effects

		$P < 0.001$	$P < 0.01$	$P < 0.1$	$P < 0.2$	$P < 0.3$	$P < 0.4$	$P < 0.5$
Narcolepsy	Nagelkerke_R^2	57.8%	57.9%	58.1%	58.1%	58.1%	58.1%	**58.1%**
	$P_{HLA\text{-}DQB1*06:02}$	2.31E-48	2.31E-48	2.55E-48	2.36E-48	2.40E-48	2.23E-48	**2.30E-48**
	P_{others}	2.32E-01	2.05E-01	9.79E-02	8.00E-02	7.11E-02	6.78E-02	**6.73E-02**
EHS with *HLA-DQB1*06:02*	Nagelkerke_R^2	39.9%	39.9%	40.1%	40.2%	40.3%	40.3%	**40.4%**
	$P_{HLA\text{-}DQB1*06:02}$	7.91E-14	7.65E-14	7.08E-14	7.02E-14	7.01E-14	7.02E-14	**7.02E-14**
	P_{others}	2.84E-01	2.92E-01	1.99E-01	1.64E-01	1.48E-01	1.38E-01	**1.34E-01**

*HLA-DQB1*06:02* effects were reflected by including SNP rs7744293, which was in the highest correlation with *HLA-DQB1*06:02* (r^2 = ~0.8), into the regression model. $P_{HLA\text{-}DQB1*06:02}$ and P_{others} were P values for regression coefficients of *HLA-DQB1*06:02* and individual polygenic risk scores other than the *HLA* region, respectively. Nagelkerkes' pseudo R^2 reflected coefficient of determination. Nagelkerkes' pseudo R^2 and P values were the averages of one thousand-permutation analysis, randomizing sample division of narcoleptic patients and healthy controls into discovery- and test-stage. These results were based on the pruning criteria of $r^2 = 0.25$ for SNP inclusion.

Table 2: Phenotypic variance explained by narcolepsy polygenic risks for each disease

		$P < 0.001$	$P < 0.01$	$P < 0.1$	$P < 0.2$	$P < 0.3$	$P < 0.4$	$P < 0.5$
Narcolepsy	Nagelkerke_R^2	0.7%	0.7%	1.1%	1.2%	1.3%	1.3%	**1.3%**
	P_{others}	1.66E-01	1.41E-01	4.78E-02	3.37E-02	2.72E-02	2.51E-02	**2.43E-02**
EHS with *HLA-DQB1*06:02*	Nagelkerke_R^2	1.0%	1.0%	1.1%	1.2%	1.3%	1.4%	**1.4%**
	P_{others}	2.93E-01	3.03E-01	2.34E-01	1.93E-01	1.74E-01	1.61E-01	**1.56E-01**
EHS without *HLA-DQB1*06:02*	Nagelkerke_R^2	0.3%	0.4%	0.4%	0.4%	0.4%	0.4%	**0.4%**
	P_{others}	4.50E-01	3.70E-01	3.15E-01	3.12E-01	3.07E-01	3.08E-01	**3.06E-01**
Panic disorder	Nagelkerke_R^2	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	**0.1%**
	P_{others}	5.04E-01	4.98E-01	4.95E-01	5.02E-01	5.12E-01	5.20E-01	**5.20E-01**
Autism	Nagelkerke_R^2	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	**0.2%**
	P_{others}	5.09E-01	5.31E-01	5.45E-01	5.17E-01	4.95E-01	4.81E-01	**4.75E-01**
Schizophrenia	Nagelkerke_R^2	0.3%	0.3%	0.3%	0.4%	0.4%	0.5%	**0.5%**
	P_{others}	4.83E-01	5.08E-01	4.90E-01	4.40E-01	4.11E-01	3.88E-01	**3.79E-01**

*HLA-DQB1*06:02* effects were not taken into account. P_{others} was P values for regression coefficients of individual polygenic risk scores other than the *HLA* region. Nagelkerkes' pseudo R^2 and P values were the averages of one thousand-permutations, randomizing sample division of narcolepsy and healthy controls into discovery- and test-stage. These results were based on the pruning criteria of $r^2 = 0.25$ for SNP inclusion.

Table 3: Narcolepsy heritability explained by the region other than *HLA* and by *CPT1B, TRA@* and *P2RY11*, estimated based on different pruning criteria.

pruning criteria		Narcolepsy heritability explained by the region other than *HLA*					Narcoepsy heritability explained by *CPT1B, TRA@ and P2RY11*				
		P < 0.1	*P* < 0.2	*P* < 0.3	*P* < 0.4	*P* < 0.5	*P* < 0.1	*P* < 0.2	*P* < 0.3	*P* < 0.4	*P* < 0.5
r^2<0.25	Nagelkerke_R^2	1.1%	1.2%	1.3%	1.3%	**1.3%**	0.9%	0.9%	0.8%	0.8%	**0.8%**
	P_{others}	4.78E-02	3.37E-02	2.72E-02	2.51E-02	**2.43E-02**	8.28E-02	8.72E-02	9.33E-02	9.52E-02	**9.74E-02**
r^2<0.5	Nagelkerke_R^2	1.0%	1.2%	1.2%	1.2%	**1.3%**	1.0%	0.9%	0.9%	0.9%	**0.8%**
	P_{others}	3.03E-02	2.03E-02	2.49E-02	1.73E-02	**1.77E-02**	6.55E-02	7.02E-02	7.70E-02	7.89E-02	**8.02E-02**
r^2<0.75	Nagelkerke_R^2	1.8%	2.0%	2.0%	2.0%	**2.1%**	1.4%	1.3%	1.2%	1.2%	**1.1%**
	P_{others}	7.16E-03	4.91E-03	4.06E-03	3.69E-03	**3.54E-03**	2.28E-02	2.94E-02	3.48E-02	3.70E-02	**3.82E-02**
r^2<1	Nagelkerke_R^2	1.3%	1.5%	1.6%	1.6%	**1.7%**	0.9%	0.8%	0.7%	0.7%	**0.7%**
	P_{others}	2.67E-02	1.45E-02	1.06E-02	9.01E-03	**8.31E-03**	8.79E-02	1.02E-01	1.11E-01	1.16E-01	**1.19E-01**

SNP inclusion criteria of r^2 were assessed, ranging from 0.25 to 1. *HLA-DQB1*06:02* effects were not taken into account. P_{others} was P values for regression coefficients of individual polygenic risk scores other than the *HLA* region. Narcolepsy heritability explained by *CPT1B*, *TRA@* and *P2RY11*, were calculated, using only SNPs residing within 500kb up- and down-stream of the narcolepsy susceptibility SNPs.

Table 4: The characteristics of common small-effect SNPs in the GWAS, using 393 narcolepsy vs 1,582 healthy controls

	OR of SNPs		
Cut off *P* value	# of SNPs	Median	Mean
0.001	1364	1.54	1.80
0.01	6222	1.32	1.44
0.1	53106	1.21	1.25
0.2	103465	1.18	1.21
0.3	153086	1.16	1.18
0.4	202955	1.14	1.16
0.5	252106	1.12	1.15

The summary statistics of ORs of SNPs were calculated based on the GWAS, utilizing all 393 narcoleptic patients and all 1,582 healthy controls.

Table 5: Narcolepsy global burden of rare large-scale CNVs by PennCNV

CNV Type	Size range	No. of CNVs				Average ratio (case /control)	Permuted *P* value
		Case (*n* = 327)		Control (*n* = 459)			
		Total	Average	Total	Average		
Deletions and duplications	All	555	1.70	506	1.10	1.54	5.00×10^{-4}
	100kb-200kb	277	0.85	242	0.53	1.61	4.40×10^{-4}
	200kb-500kb	209	0.64	186	0.41	1.58	2.85×10^{-3}
	500kb-1Mb	50	0.15	46	0.10	1.53	5.35×10^{-2}
	>1Mb	19	0.06	32	0.07	0.83	7.55×10^{-1}
Deletions	All	265	0.81	239	0.52	1.56	4.35×10^{-2}
	100kb-200kb	138	0.42	141	0.31	1.37	8.34×10^{-2}
	200kb-500kb	96	0.29	79	0.17	1.71	4.52×10^{-2}
	500kb-1Mb	20	0.06	9	0.02	3.12	2.18×10^{-2}
	>1Mb	11	0.03	10	0.02	1.54	2.80×10^{-1}
Duplications	All	290	0.89	267	0.58	1.52	4.20×10^{-4}
	100kb-200kb	139	0.43	101	0.22	1.93	2.00×10^{-5}
	200kb-500kb	113	0.35	107	0.23	1.48	8.52×10^{-3}
	500kb-1Mb	30	0.09	37	0.08	1.14	3.45×10^{-1}
	>1Mb	8	0.02	22	0.05	0.51	9.73×10^{-1}

The table shows total burden for rare large-scale CNVs in narcoleptic patients (n = 327) compared to controls (n = 459). *P* values were estimated in 100,000 permutations.

Table 6: Narcolepsy global burden of genes disrupted by rare large-scale CNVs by PennCN

CNV Type	Size range	No. of genes disrupted by CNVs Case (*n* = 327) Total	Case (*n* = 327) Average	Control (*n* = 459) Total	Control (*n* = 459) Average	Average ratio (case /control)	Permuted *P* value
Deletions and duplications	All	1170	4.70	768	2.15	2.18	5.75×10^{-3}
	100kb-200kb	411	1.43	260	0.67	2.13	2.42×10^{-3}
	200kb-500kb	569	2.14	265	0.72	2.98	3.60×10^{-4}
	500kb-1Mb	216	0.78	150	0.36	2.14	8.90×10^{-2}
	>1Mb	109	0.36	151	0.40	0.89	5.64×10^{-1}
Deletions	All	528	2.02	310	0.81	2.49	8.44×10^{-2}
	100kb-200kb	186	0.62	146	0.36	1.74	1.19×10^{-1}
	200kb-500kb	240	0.86	82	0.23	3.73	4.19×10^{-2}
	500kb-1Mb	124	0.43	37	0.08	5.25	7.88×10^{-2}
	>1Mb	34	0.10	60	0.14	0.75	6.61×10^{-1}
Duplications	All	763	2.67	502	1.34	2.00	6.85×10^{-3}
	100kb-200kb	230	0.80	115	0.31	2.59	1.90×10^{-4}
	200kb-500kb	374	1.28	189	0.49	2.62	9.70×10^{-4}
	500kb-1Mb	104	0.34	113	0.28	1.22	3.53×10^{-1}
	>1Mb	83	0.25	116	0.26	0.97	4.99×10^{-1}

The table shows total burden for genes disrupted by rare large-scale CNVs in narcoleptic patients (n = 327) compared to controls (n = 459). Genes were defined by glist-hg18 file, based on NCBI Build 36 and provided in PLINK. *P* values were estimated in 100,000 permutations.

Table 7: EHS with *HLA-DQB1*06:02* global burden of rare large-scale CNVs by PennCNV

CNV Type	Size range	No. of CNVs				Average ratio (case /control)	Permuted *P* value
		Case (*n* = 33)		Control (*n* = 459)			
		Total	Average	Total	Average		
Deletions and duplications	All	52	1.58	496	1.08	1.46	3.24×10^{-2}
	100kb-200kb	30	0.91	245	0.53	1.70	1.57×10^{-2}
	200kb-500kb	17	0.52	174	0.38	1.36	1.59×10^{-1}
	500kb-1Mb	2	0.06	48	0.10	0.58	8.53×10^{-1}
	>1Mb	3	0.09	29	0.06	1.44	3.63×10^{-1}
Deletions	All	29	0.88	228	0.50	1.77	4.67×10^{-2}
	100kb-200kb	22	0.67	144	0.31	2.13	1.35×10^{-2}
	200kb-500kb	7	0.21	68	0.15	1.43	2.26×10^{-1}
	500kb-1Mb	0	0.00	9	0.02	0.00	1
	>1Mb	0	0.00	7	0.02	0.00	1
Duplications	All	23	0.70	268	0.58	1.19	2.51×10^{-1}
	100kb-200kb	8	0.24	101	0.22	1.10	4.49×10^{-1}
	200kb-500kb	10	0.30	106	0.23	1.31	2.63×10^{-1}
	500kb-1Mb	2	0.06	39	0.08	0.71	7.68×10^{-1}
	>1Mb	3	0.09	22	0.05	1.90	2.29×10^{-1}

The table shows total burden for rare large-scale CNVs in EHS with *HLA-DQB1*06:02* patients (n = 33) compared to controls (n = 459). *P* values were estimated in 100,000 permutations.

Table 8: EHS with *HLA-DQB1*06:02* global burden of genes disrupted by rare large-scale CNVs by PennCNV

CNV Type	Size range	No. of genes disrupted by CNVs				Average ratio (case /control)	Permuted *P* value
		Case (*n* = 33)		Control (*n* = 459)			
		Total	Average	Total	Average		
Deletions and duplications	All	80	2.58	782	2.19	1.18	2.98×10^{-1}
	100kb-200kb	27	0.85	261	0.68	1.25	2.98×10^{-1}
	200kb-500kb	35	1.06	268	0.71	1.49	1.66×10^{-1}
	500kb-1Mb	11	0.33	152	0.37	0.91	3.96×10^{-1}
	>1Mb	11	0.33	159	0.43	0.77	5.29×10^{-1}
Deletions	All	32	1.06	311	0.81	1.31	2.63×10^{-1}
	100kb-200kb	25	0.79	148	0.39	2.01	1.19×10^{-1}
	200kb-500kb	9	0.27	81	0.21	1.32	2.74×10^{-1}
	500kb-1Mb	0	0.00	37	0.08	0.00	1
	>1Mb	0	0.00	58	0.13	0.00	1
Duplications	All	48	1.52	520	1.38	1.10	3.56×10^{-1}
	100kb-200kb	2	0.06	114	0.29	0.21	9.52×10^{-1}
	200kb-500kb	26	0.79	196	0.51	1.56	1.71×10^{-1}
	500kb-1Mb	11	0.33	115	0.29	1.17	3.26×10^{-1}
	>1Mb	11	0.33	127	0.31	1.09	3.93×10^{-1}

The table shows total burden for genes disrupted by rare large-scale CNVs in EHS with *HLA-DQB1*06:02* patients (n = 33) compared to controls (n = 459). Genes were defined by glist-hg18 file, based on NCBI Build 36 and provided in PLINK. *P* values were estimated in 100,000 permutations.

Table 9: EHS without *HLA-DQB1*06:02* global burden of rare large-scale CNVs by PennCNV

CNV Type	Size range	No. of CNVs				Average ratio (case /control)	Permuted *P* value
		Case (*n* = 104)		Control (*n* = 459)			
		Total	Average	Total	Average		
Deletions and duplications	All	166	1.60	512	1.12	1.43	2.17×10^{-3}
	100kb-200kb	94	0.90	246	0.54	1.69	3.30×10^{-4}
	200kb-500kb	56	0.54	187	0.41	1.32	6.29×10^{-2}
	500kb-1Mb	10	0.10	49	0.11	0.90	6.70×10^{-1}
	>1Mb	6	0.06	30	0.07	0.88	6.86×10^{-1}
Deletions	All	87	0.84	238	0.52	1.61	8.79×10^{-3}
	100kb-200kb	56	0.54	146	0.32	1.69	9.58×10^{-3}
	200kb-500kb	24	0.23	75	0.16	1.41	1.18×10^{-1}
	500kb-1Mb	4	0.04	10	0.02	1.77	2.48×10^{-1}
	>1Mb	3	0.03	7	0.02	1.89	2.78×10^{-1}
Duplications	All	79	0.76	274	0.60	1.27	7.18×10^{-2}
	100kb-200kb	38	0.37	100	0.22	1.68	7.35×10^{-3}
	200kb-500kb	32	0.31	112	0.24	1.26	1.81×10^{-1}
	500kb-1Mb	6	0.06	39	0.08	0.68	8.60×10^{-1}
	>1Mb	3	0.03	23	0.05	0.58	8.89×10^{-1}

The table shows total burden for rare large-scale CNVs in EHS without *HLA-DQB1*06:02* patients (n = 104) compared to controls (n = 459). *P* values were estimated in 100,000 permutations.

Table 10: EHS without *HLA-DQB1*06:02* global burden of genes disrupted by rare large-scale CNVs by PennCNV

CNV Type	Size range	No. of genes disrupted by CNVs Case (*n* = 104)		Control (*n* = 459)		Average ratio (case /control)	Permuted *P* value
		Total	Average	Total	Average		
Deletions and duplications	All	404	4.07	784	2.26	1.80	3.25×10^{-2}
	100kb-200kb	135	1.32	263	0.69	1.91	1.16×10^{-2}
	200kb-500kb	182	1.82	270	0.76	2.40	4.21×10^{-2}
	500kb-1Mb	36	0.35	158	0.38	0.91	5.02×10^{-1}
	>1Mb	61	0.59	159	0.43	1.36	2.90×10^{-1}
Deletions	All	101	0.97	314	0.85	1.14	3.49×10^{-1}
	100kb-200kb	62	0.60	150	0.40	1.48	1.77×10^{-1}
	200kb-500kb	29	0.28	84	0.24	1.16	3.78×10^{-1}
	500kb-1Mb	4	0.04	37	0.08	0.46	5.90×10^{-1}
	>1Mb	6	0.06	58	0.13	0.46	7.11×10^{-1}
Duplications	All	308	3.10	521	1.41	2.20	3.30×10^{-2}
	100kb-200kb	73	0.72	115	0.29	2.51	8.02×10^{-3}
	200kb-500kb	156	1.54	196	0.52	2.98	4.12×10^{-2}
	500kb-1Mb	32	0.31	121	0.30	1.03	4.23×10^{-1}
	>1Mb	55	0.53	127	0.31	1.73	2.23×10^{-1}

The table shows total burden for genes disrupted by rare large-scale CNVs in EHS without *HLA-DQB1*06:02* patients (n = 104) compared to controls (n = 459). Genes were defined by glist-hg18 file, based on NCBI Build 36 and provided in PLINK. *P* values were estimated in 100,000 permutations.

Table 11: Gene-sets significantly associated with narcolepsy

No.	Pathway name	No. of genes in pathways	Fisher's *P* value	No. of sample	
				Case	Control
1	regulation of immune response	667	1.93×10^{-4}	26	10
2	peptidyl-lysine modification	148	1.94×10^{-4}	12	1
3	cell cycle arrest	421	2.00×10^{-4}	19	5
4	peptidyl-amino acid modification	639	3.47×10^{-4}	20	6
5	G2/M transition checkpoint	40	3.50×10^{-4}	9	0
6	synaptic vesicle	104	3.50×10^{-4}	9	0
7	negative regulation of cell differentiation	422	3.67×10^{-4}	15	3
8	positive regulation of immune response	445	3.84×10^{-4}	18	5
9	covalent chromatin modification	302	4.01×10^{-4}	17	4
10	histone modification	294	4.01×10^{-4}	17	4
11	negative regulation of developmental process	532	4.01×10^{-4}	17	4
12	response to molecule of bacterial origin	214	4.40×10^{-4}	11	1
13	protein acetylation	132	4.40×10^{-4}	11	1
14	response to lipopolysaccharide	194	4.40×10^{-4}	11	1
15	protein acylation	145	4.40×10^{-4}	11	1
16	rhythmic process	206	4.40×10^{-4}	11	1
17	negative regulation of cell cycle	505	6.03×10^{-4}	20	7

18	negative regulation of transcription from RNA polymerase II promoter	468	6.28×10^{-4}	19	6
19	sex differentiation	283	7.40×10^{-4}	14	3
20	cellular aromatic compound metabolic process	308	7.47×10^{-4}	16	4
21	response to abiotic stimulus	618	8.45×10^{-4}	27	13
22	negative regulation of cell development	109	8.53×10^{-4}	8	0
23	response to hydrogen peroxide	81	8.53×10^{-4}	8	0
24	cellular response to fibroblast growth factor stimulus	187	8.53×10^{-4}	8	0
25	response to fibroblast growth factor stimulus	187	8.53×10^{-4}	8	0
26	protein transporter activity	96	8.53×10^{-4}	8	0
27	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	194	9.90×10^{-4}	10	1
28	internal protein amino acid acetylation	120	9.90×10^{-4}	10	1
29	regulation of cell morphogenesis involved in differentiation	173	9.90×10^{-4}	10	1
30	histone acetylation	113	9.90×10^{-4}	10	1
31	internal peptidyl-lysine acetylation	115	9.90×10^{-4}	10	1
32	peptidyl-lysine acetylation	118	9.90×10^{-4}	10	1

The table shows the list of 32 significant gene-sets from both KEGG and the Gene Ontology with FDR 5%. *P* value was calculated using I(*i,j*) score. Define indicator I(*i,j*) = 1 if a sample *i* carries a CNV overlapping at least one gene in gene-set *j*, and zero otherwise. The total I(*i,j*) score for gene set *j* is obtained by summing over *i*. For testing, this sum is partitioned between cases and controls, shown in the column of the number of sample. The number in the first column is index used in gene-set index in Figure 14.

Figures

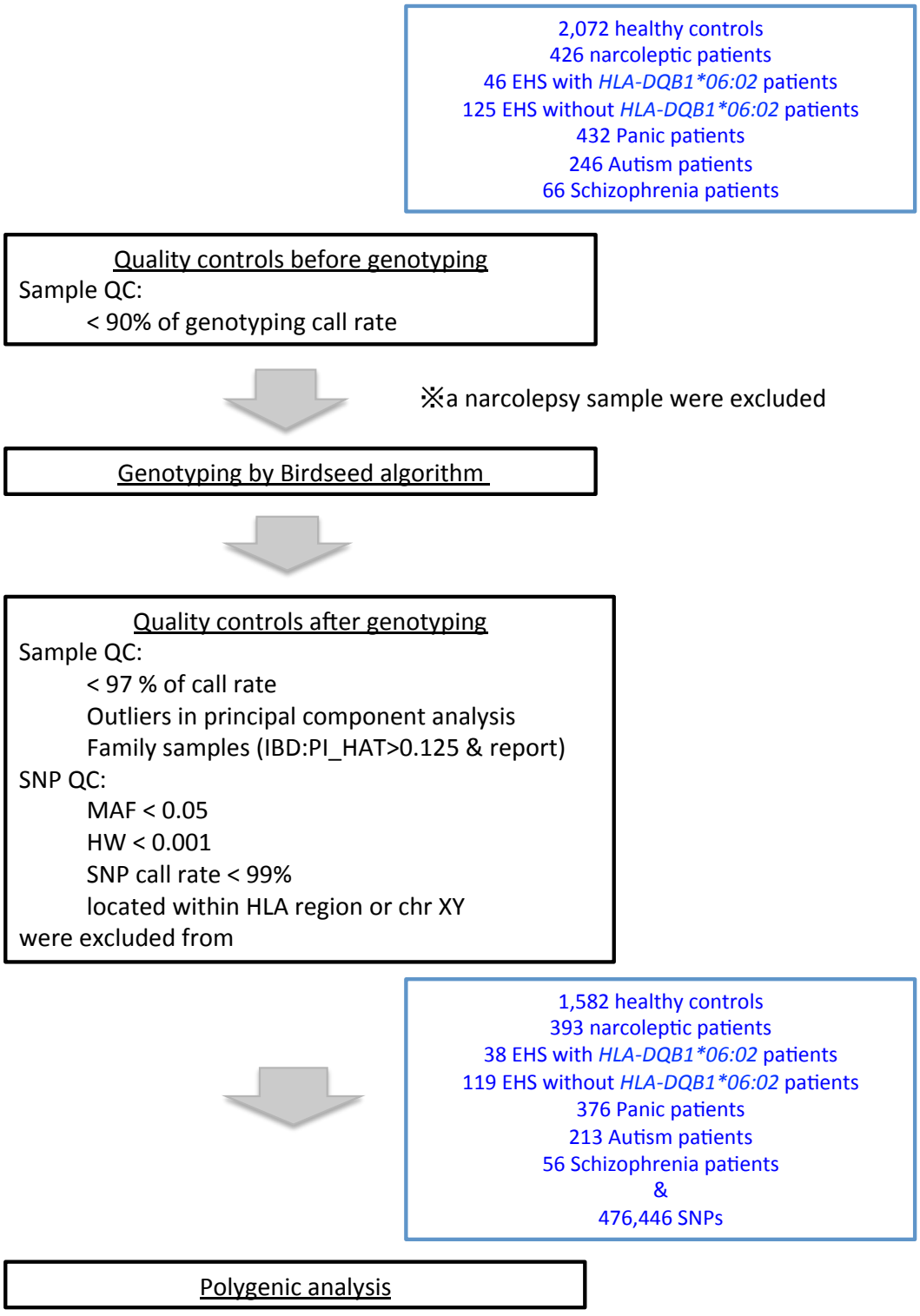

Figure 1: Overflow of quality controls and genotyping. Blue number indicates samples remained after the each step.

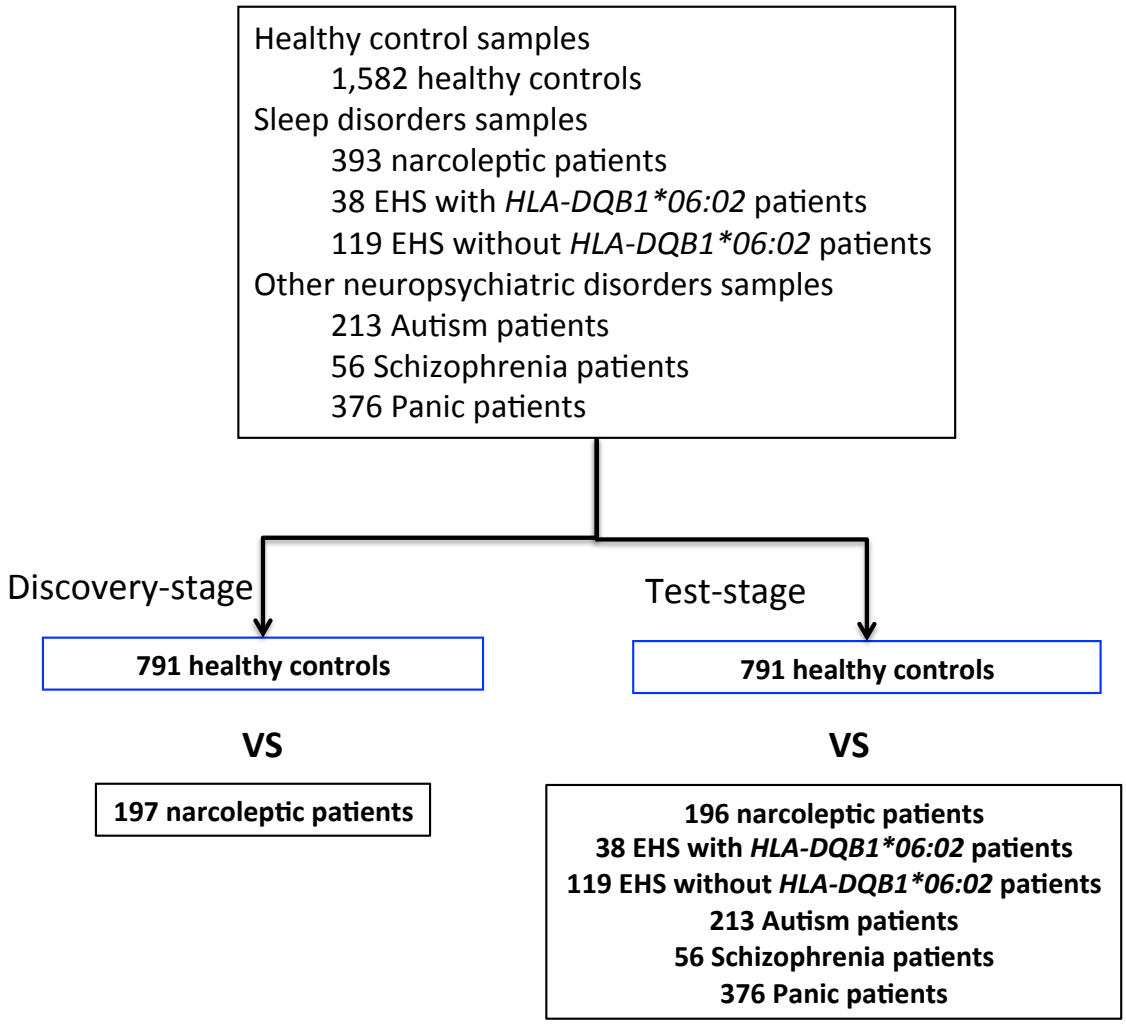

Figure 2: Study design and sample division. One group (discovery-stage) was used for a genome-wide association analysis in order to calculate risk-effects of each common variant. The other group (test-stage) was used for estimation of the capability of common variants to explain phenotypic variance, using collective effects of narcolepsy polygenic risks. Narcoleptic patients and healthy controls were randomly divided into two groups from each permutation. Healthy controls in the test-stage were same across different phenotypes within one permutation.

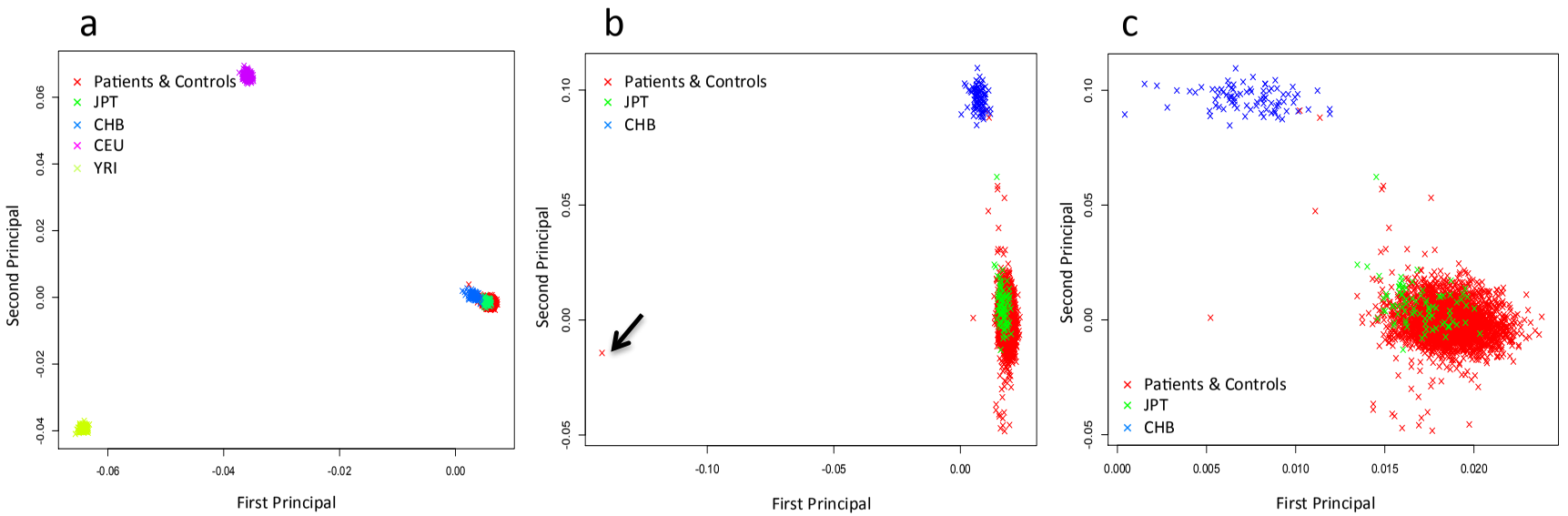

Figure 3: Quality control by a principal components analysis, excluding an outlier. 91 JPT, 90 CHB, 180 CEU and 180 YRI were derived from Hapmap3. (a) The plot of all samples and HapMap samples in the first two principal components. (red: sample populations, light green: JPT, blue: CHB, purple: CEU, mint green: YRI) (b) The plot of all samples and JPT + CHB in the first two principal components. (red: sample populations, light green: JPT, blue: CHB) Black arrow is a sample excluded in the analysis. (c) After excluding an outlier, the plot of all samples and JPT + CHB in the first two principal components. (red: sample populations, light green: JPT, blue: CHB)

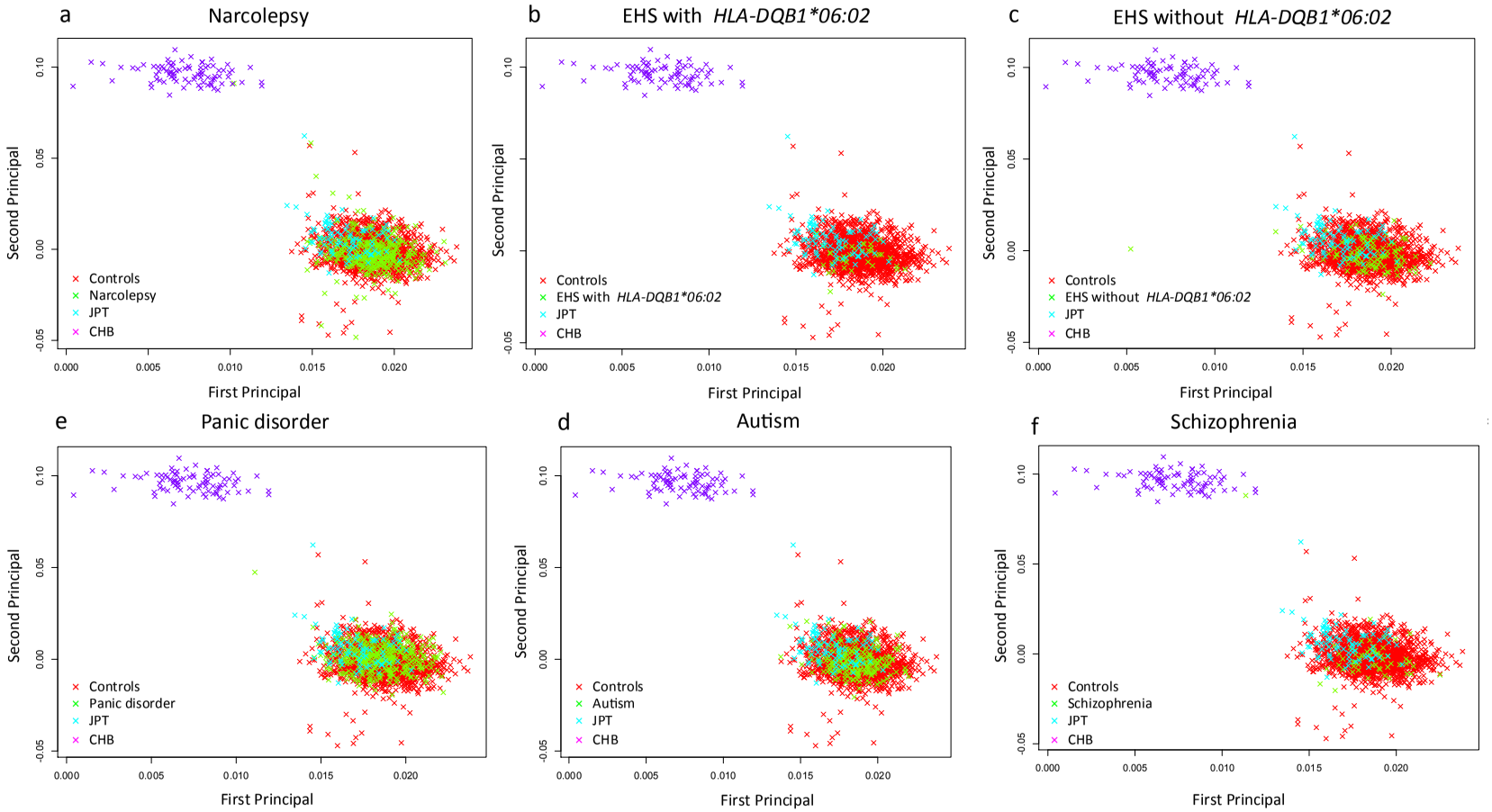

Figure 4: A principal component analysis for each disease. Individuals with each disease and healthy controls were plotted. (a) The result of narcolepsy, control + JPT & CHB. (b) The result of EHS with *HLA-DQB1*06:02*, control + JPT & CHB. (c) The first two principal components of EHS without *HLA-DQB1*06:02*, control + JPT & CHB. (d) The result of panic disorder, control + JPT & CHB. (d) The result of autism, control + JPT & CHB. (f) The result of schizophrenia, control + JPT & CHB

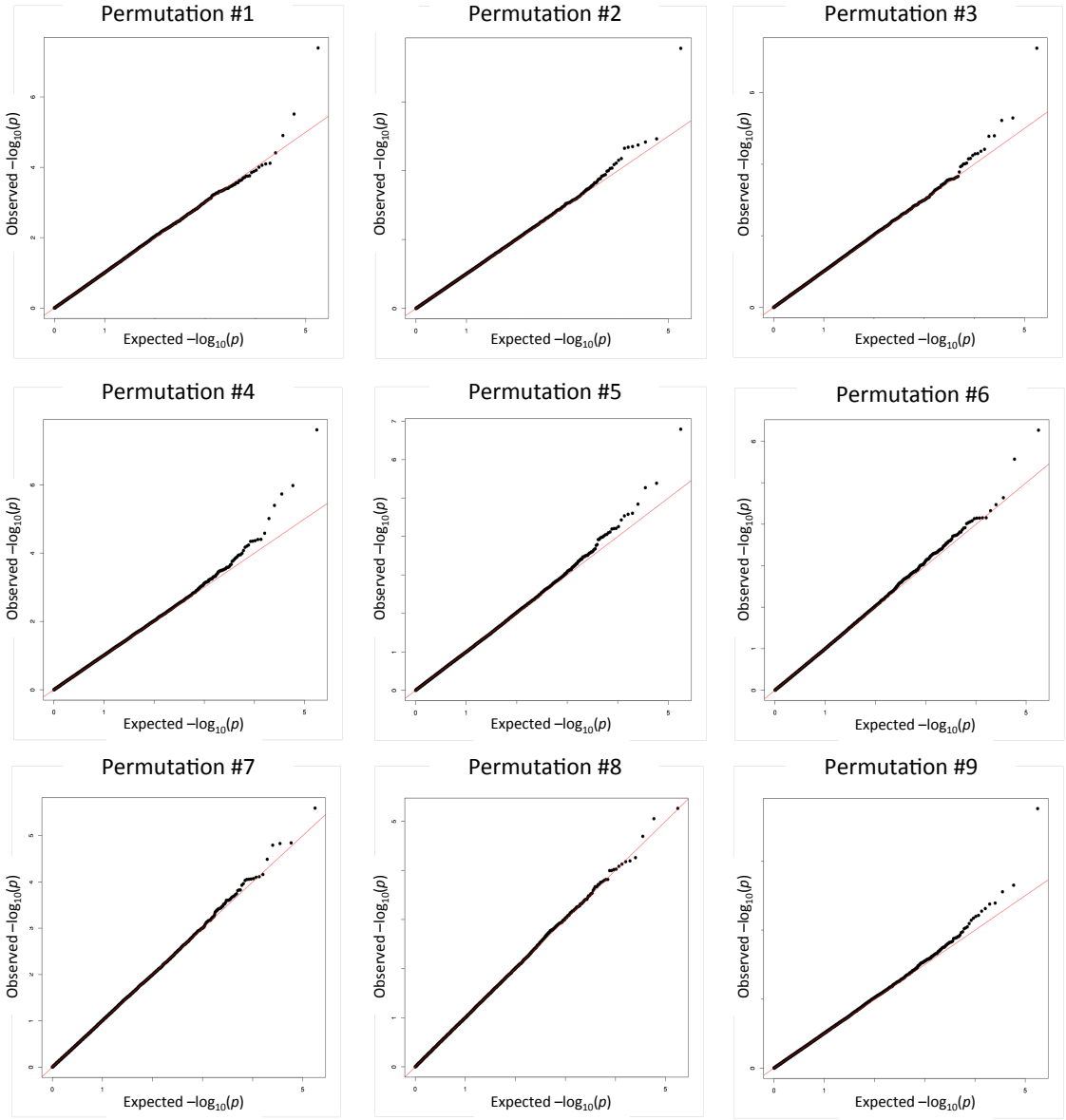

Figure 5: Examples of quantile-quantile plots for 1,000 permutation GWASs in discovery-stage. The GWASs were after quality controls, without the *HLA* region and based on the pruning criterion of $r^2 < 0.25$. The average lamda of 1,000 permutation GWAS was 1.004, showing no population stratifications.

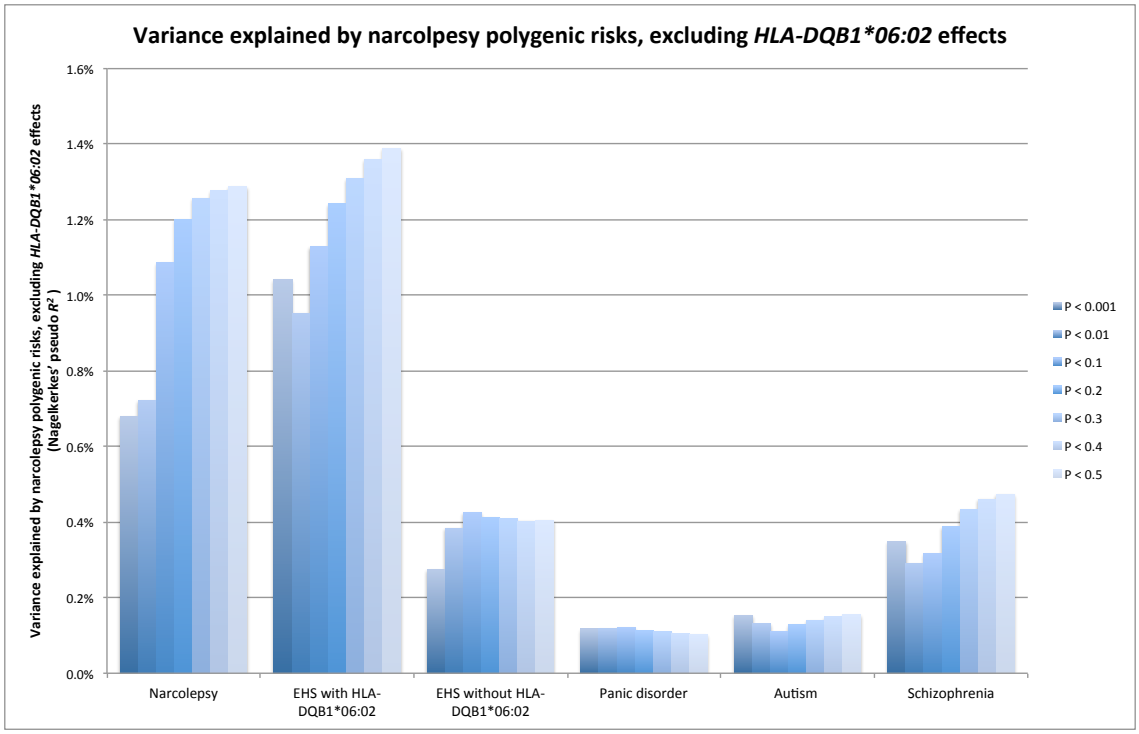

Figure 6: Phenotypic variance explained by narcolepsy polygenic risks, excluding *HLA-DQB1*06:02* effects. Nagelkerkes' pseudo R^2 was utilized to reflect the capability of polygenic risks to explain disease onset.

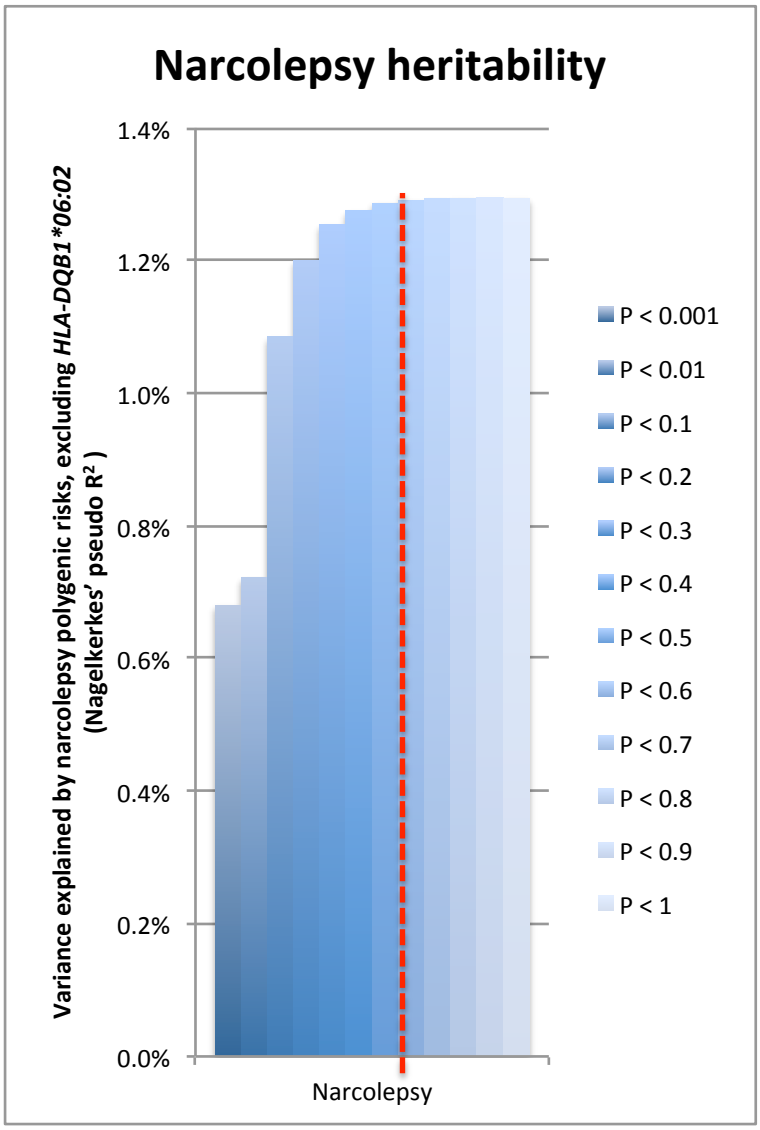

Figure 7: Polygenic feature of narcolepsy. Red dots line was 0.5 of *P* value. Nagelkerkes' pseudo R^2 was utilized to reflect the capability of polygenic risks to explain disease onset.

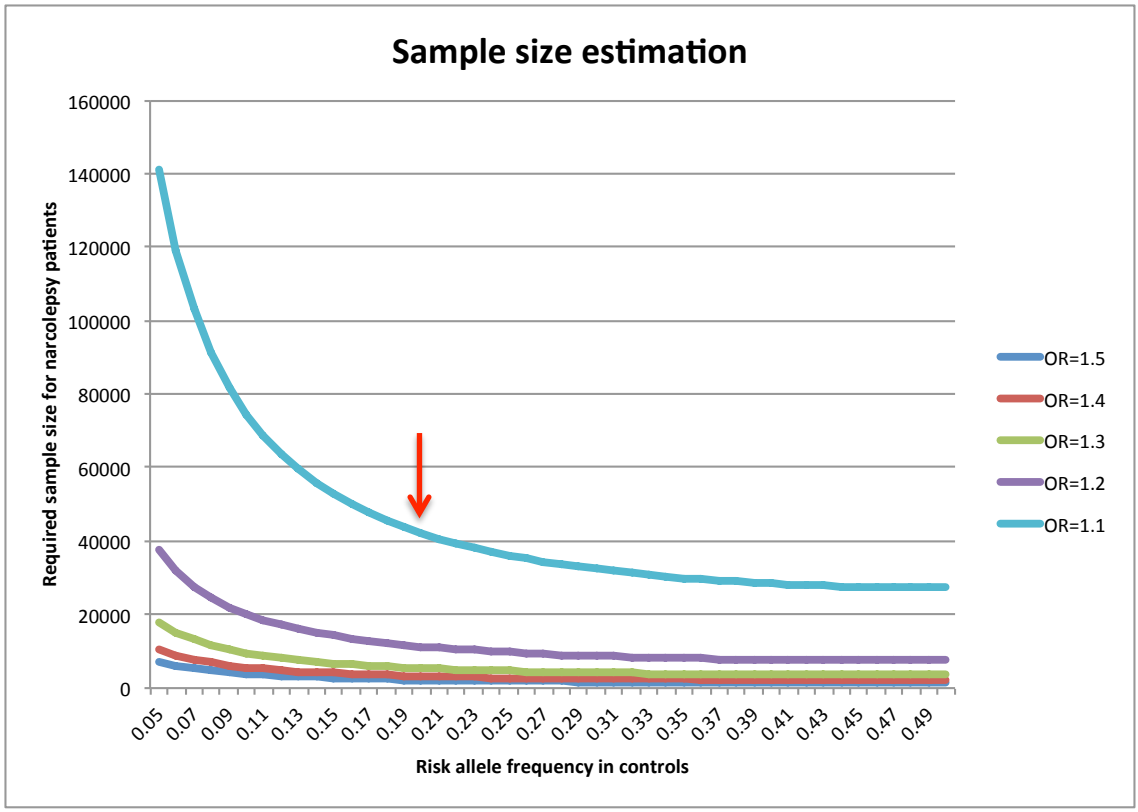

Figure 8: Sample size estimation for discovery of common small-effect SNPs. Over 42,000 narcoleptic patients were required, when ORs of SNPs = 1.1; risk allele frequency in controls = 0.2; case-controls ratio = 0.25; α-level $P = 5\times10^{-8}$, shown in red arrow.

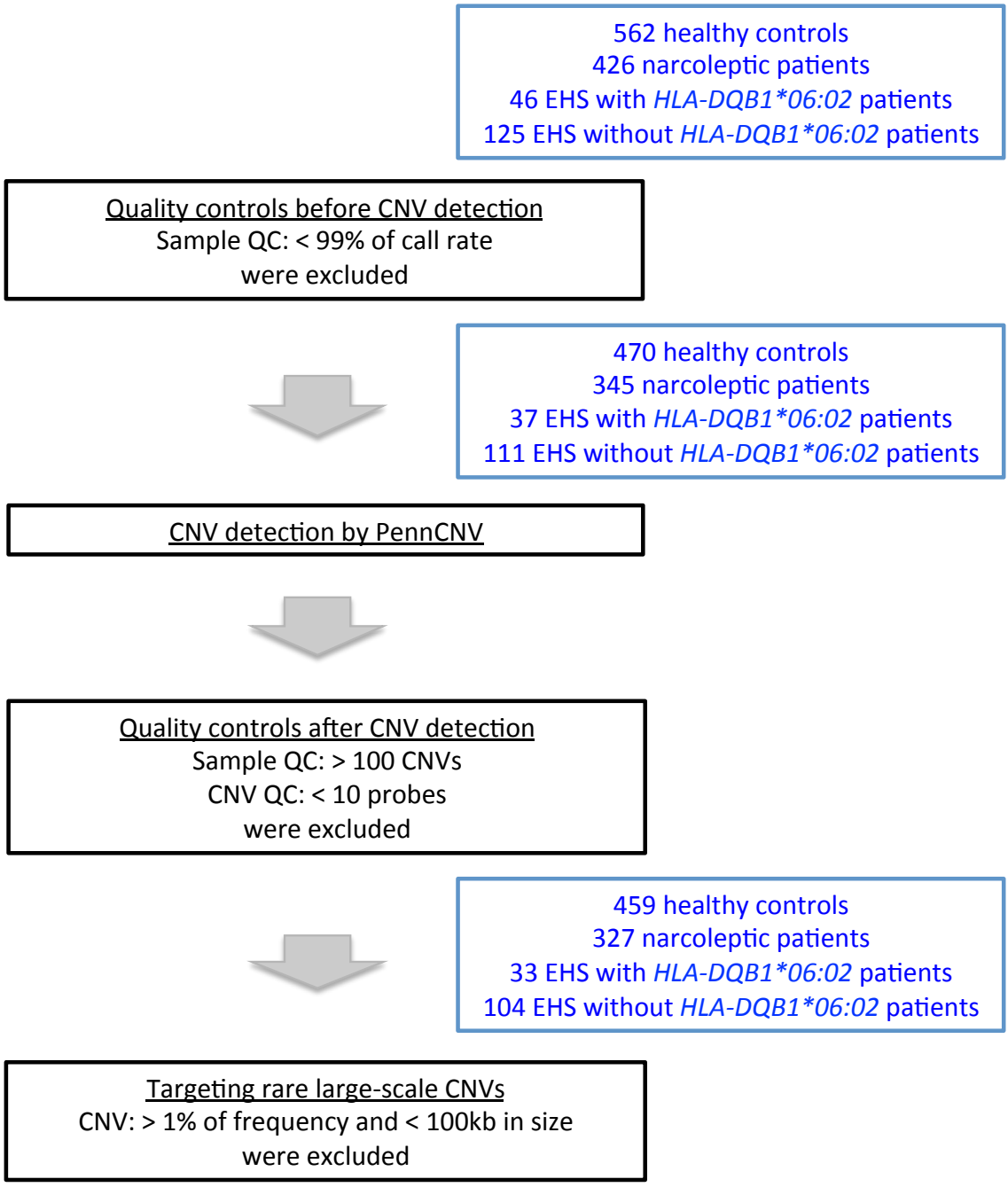

Figure 9: Quality controls and CNV detection flow. Blue number indicates samples remained after the each step.

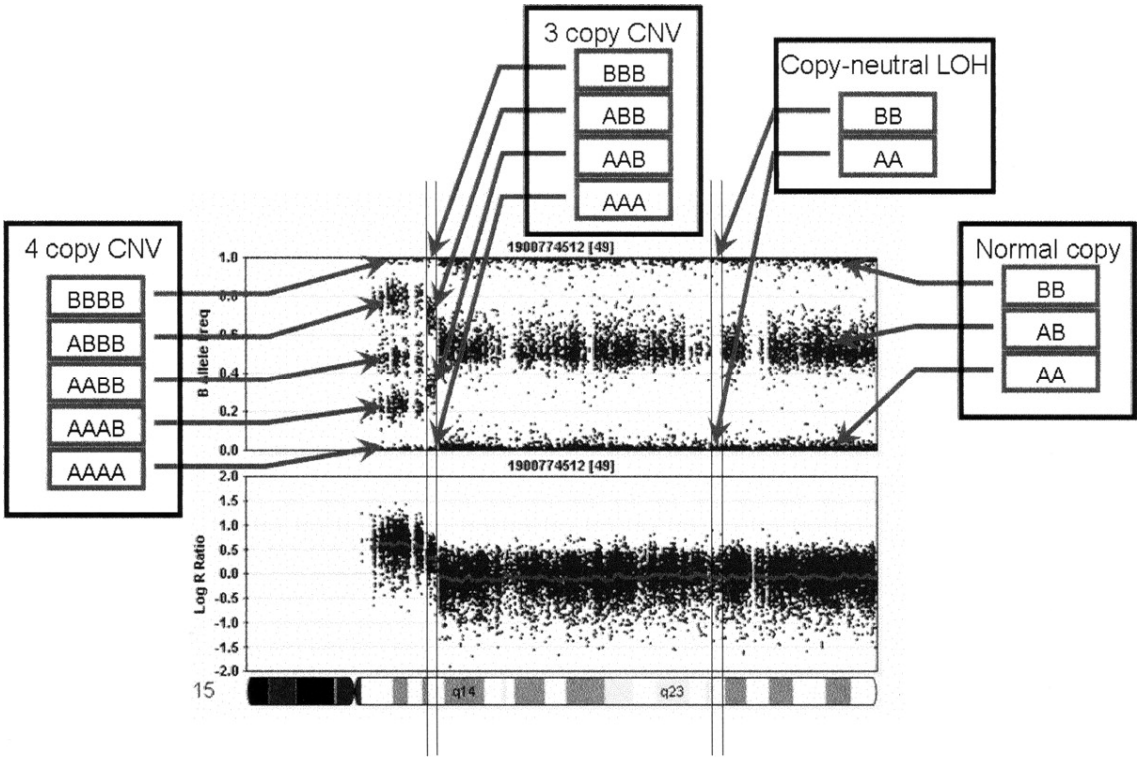

Figure 10: An illustration of CNV detection by PennCNV using log R ratio (LRR) and B allele freq (BAF) values of a sample. LRR is a measure of normalized total signal intensity relative to expected value, (= $\log2(R_{observed}/R_{expected})$, R is a sum of probe intensities). BAF is a normalized measure of relative signal intensity ratio of the B and A alleles in the SNP array. A normal chromosome region has three BAF genotype clusters, as represented as AA, AB, and BB genotypes in boxes, and with LRR values centered around zero. The copy-neutral LOH (loss of heterozygosity) region has normal LRR values, but without the AB genotype cluster. The increased copy number for a CNV region can be detected based on an increased number of clusters in the BAF distribution, as well as increased LRR values. The patterns of LRR and BAF for different CNV regions, normal regions, and copy-neutral LOH regions are distinct from each other, thus the combination of LRR and BAF can be used to generate CNV calls.

Modified from *PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data*. Genome Research, 2007. 17(11): p. 1665-1674.

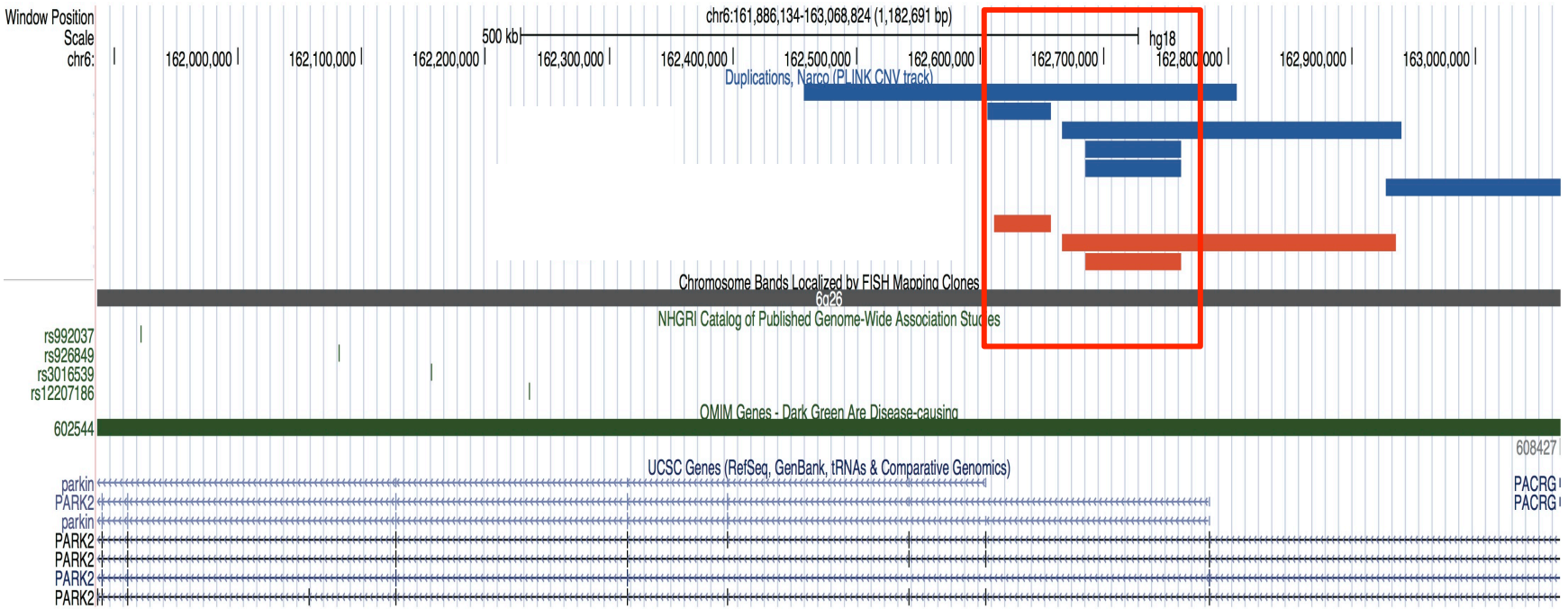

Figure 11: The *PARK2* region significantly associated with narcolepsy and EHS without *HLA-DQB1*06:02*. (Chr 6:162,685,167-162,762,467).

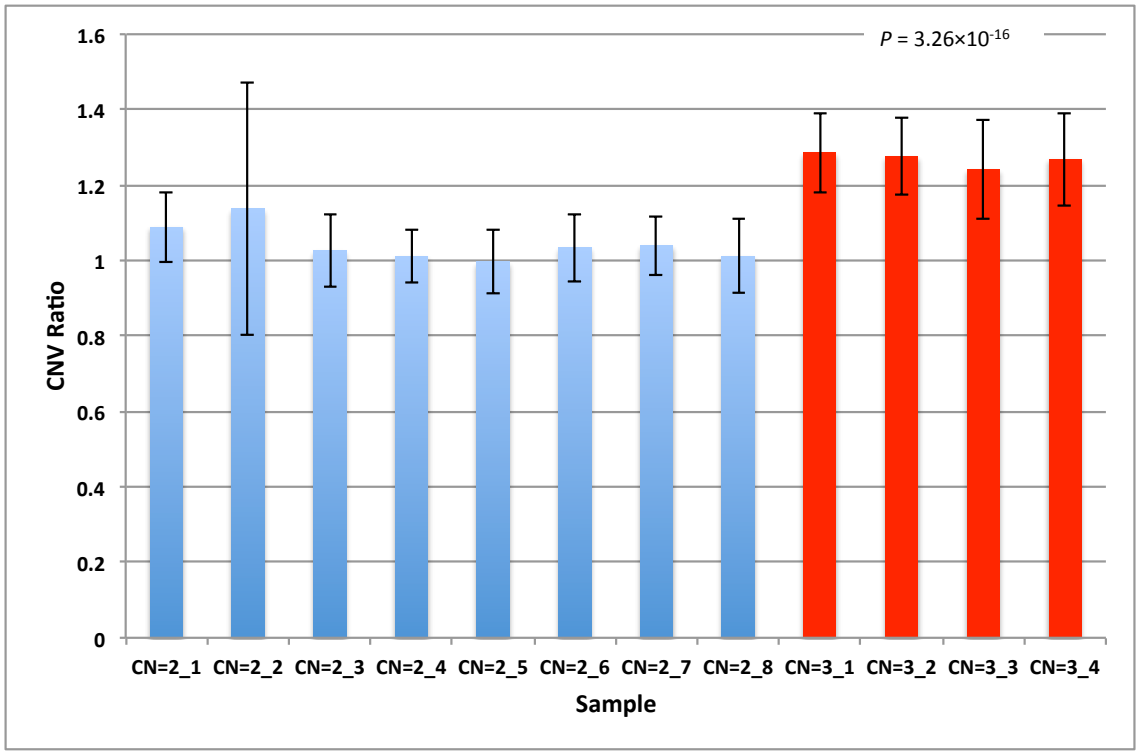

Figure 12: Validation of duplications in *PARK2* (chr6:162,722,335) by quantitative PCR. (Four narcolepsy patients vs eight healthy controls)

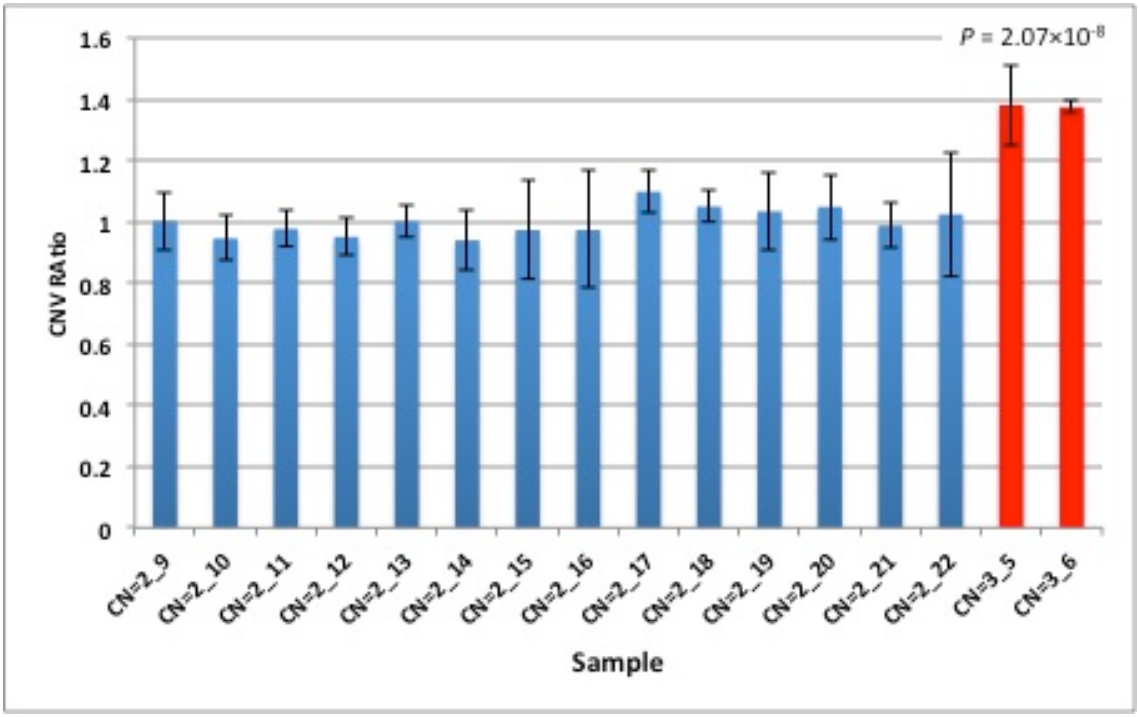

Figure 13: Validation of duplications in *PARK2* (chr6:162,722,335) by quantitative PCR. (Two EHS without *HLA-DQB1*06:02* patients vs 14 healthy controls)

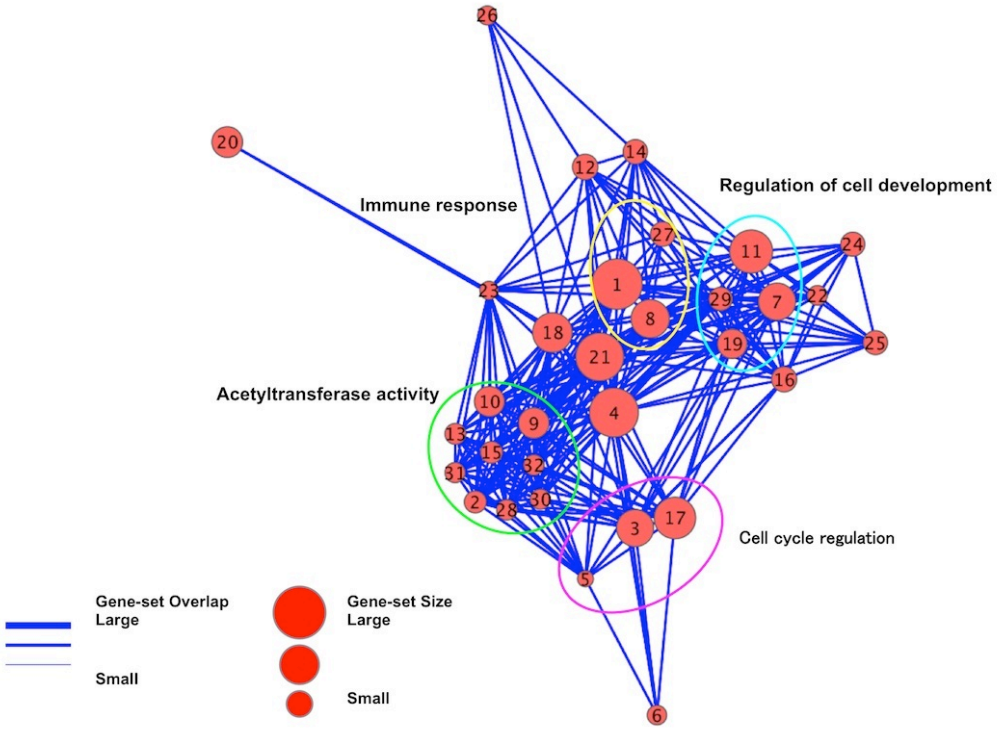

Figure 14: Functional clusters of the 32 gene sets that were found to be significantly associated with narcolepsy. The figure shows a network of gene sets (nodes) that are related by mutual overlap (edges). Node size is proportional to the total number of genes in each gene set. The number inside a node is the index used in the first column in Table 11. Edge width represents the number of overlapping genes between two gene sets.