博士論文

論文題目　Genome-wide association studies identified new candidate regions for essential hypersomnia

（ゲノムワイド関連解析による真性過眠症の新規疾患感受性領域の同定）

氏　　名　　許　昔舜

Khor Seik Soon

Table of Contents **Page**

Abstract

Essential hypersomnia (EHS), a sleep disorder characterized by excessive daytime sleepiness (EDS), can be divided into two broad classes based on the presence or absence of the Human Leukocyte Antigen (*HLA)-DQB1*06:02* allele. *HLA-DQB1*06:02* positive EHS and narcolepsy with cataplexy are associated with the same susceptibility genes and orexin (hyporetin) deficiency. In contrast, there are fewer studies of *HLA-DQB1*06:02* negative EHS which, I hypothesized, involves different pathophysiological pathways from narcolepsy with cataplexy. Previous studies have shown that *HLA-DQB1*06:02* negative EHS shows normal orexin levels and is not associated with orexin deficiency. In order to identify susceptibility genes associated with *HLA-DQB1*06:02* negative EHS, I conducted a genome-wide association study (GWAS) of 125 unrelated Japanese EHS patients lacking the *HLA-DQB1*06:02* allele and 562 Japanese healthy controls. A comparative study was also performed on 268 *HLA-DQB1*06:02* negative Caucasian hypersomnia patients and 1761 *HLA-DQB1*06:02* negative Caucasian healthy controls. I identified three SNPs that each represented a unique locus, rs16826005 (*P* = 1.02E-07; *NCKAP5*), rs11854769 (*P* = 6.69E-07; *SPRED1*), and rs10988217 (*P* = 3.43E-06; *CRAT*) that were associated with an increased risk of EHS in this Japanese population. Interestingly, rs10988217 showed a similar tendency in its association with both *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy in both Japanese and Caucasian populations. In order to identify SNPs marker that can be used to differentiate between EHS and narcolepsy with cataplexy, I conducted a comparison study in which *MBP* (*P* = 2.74E-06) and *QKI* (*P* = 2.22E-05) showed associations. Quantitative enzyme-linked immunosorbent assay (ELISA) studies for MBP level in the cerebrospinal fluid (CSF) of 21 low orexin hypersomnias (represents narcolepsy with cataplexy) and 23 normal orexin hypersomnias (represents EHS) showed that normal orexin hypersomnias group had a significant elevated level of CSF MBP(*P* = 2.50E-02) suggesting demyelination might occur in EHS patients.

Besides looking at single point analyses, I performed a pathway analysis and identified the glutamate metabolism to be significant (P = 1.70E-03) with the risk of having *HLA-DQB1*06:02* negative EHS. High-performance liquid chromatography (HPLC) analysis on 21 low orexin hypersomnias and 23 normal orexin hypersomnias showed gamma-aminobutyric acid (GABA) level is significantly increased in the low orexin hypersomnias group (P = 3.15E-03) parallel with severity of EDS in narcolepsy with cataplexy.

Abbreviation

bp	Base-pair
CHB	Chinese population
CI	Confidence Interval
CNV	Copy Number Variation
CSF	Cerebrospinal fluid
EDS	Excessive Daytime Sleepiness
EHS	Essential Hypersomnia
ELISA	Enzyme-Linked ImmunoSorbent Assay
ESS	Epworth Sleepiness Scale
FDR	False Discovery Rate
GWAS	Genome-Wide Association Study
HLA	Human Leukocyte Antigen
HPLC	High-performance liquid chromatography
HWE	Hardy-Weinberg equilibrium
ICSD	International Classification of Sleep Disorders
JPT	Japanese population
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency

MAPK	Mitogen-activated protein (MAP) kinases
MSLT	Multiple Sleep Latency Test
NA	Not applicable
NMDA	N-methyl-D-aspartate receptor
NC	No Call number for genotyped SNPs
OD	Optical Density
EEG	Electroencephalogram
OR	Odd ratio
QQ	Quantile-quantile
QC	Quality Check
Ras	RAt Sarcoma
REM	Rapid Eye Movement
NREM	Non-Rapid Eye Movement
SNPs	Single Nucleotide Polymorphisms
SOREMPs	Sleep Onset Rapid Eye Movement Periods

1. General Introduction

Before the discovery of rapid eye movement (REM) in sleep and duality in sleep, sleep was universally regarded as an inactive state of the brain. Sleep was thought to be induced as a result of reduced sensory input as a consequent of reducing of brain activity. Wakefulness and staying awake are defined by the reversal of this process mainly as a result by environmental inducement. We have learned much more about sleep since 1834 and it is best summarized by modern investigator, J.Allan Hobson who has mentioned in his book *Sleep* [1] published in 1989 that 'In the short period of time, researchers have discovered that sleep is a dynamic behavior. Not simply the absence of waking, sleep is a special activity of the brain, controlled by elaborate and precise mechanism'.

It is perhaps the most exciting development in the understanding of sleep that REM sleep was discovered and defined by Nathaniel Kleitman [2, 3] in the year 1951 and it is the first time that REM sleep has been linked to the event of dreaming. With the development of electroencephalogram (EEG), the duality of sleep is finally defined as what we know now as REM sleep and non-rapid eye movement (NREM). In normal healthy mature individuals, sleep begins in NREM and progresses through deeper NREM (stages 2,3 and 4 with the classical definition or stages N2 and N3 with updated definition) before the occurring of the first episode of REM sleep which approximately 80 minutes later. Thus, one full cycle of NREM and REM sleep takes around 90 minutes. REM sleep episodes are known to be lengthening throughout the cycle of sleep. However, there are some variations even in normal individuals, for example, in newborn human, sleep begins in the REM sleep period before proceeding to NREM and they have shorter sleep cycle (approximately 50 minutes/cycle) [4]. Besides that, sleep stages can be affected by factors such as prior sleep history (whether ones suffer from sleep loss from the night before) [5], circadian rhythms [6], drug ingestion that normally prescribed for treatment of sleep disorder, and disease

pathology (for example: narcolepsy, sleep apnea syndromes, sleep fragmentation). Thus, the characteristics of REM sleeps were often used as a diagnostic marker for the diagnosis of sleep disorders.

1.1 The measurement of sleepiness

Like any other functions in human body, sleep suffers from a number of problems which can be largely classified as insomnia (having trouble of having adequate sleep) and hypersomnia (having problem of excessive sleep period). Around 10% to 25% of population is reported to suffer from the problem of sleepiness depending on the definition of sleepiness, the method used to measure sleepiness scale and the population sampled. Traditional way of sleepiness survey studies are usually carried out in the way of surveys and questionnaires that related to the feeling state during sleep attacks, the tiredness when sleep attacks, and the effort needed to fight the sleep attacks [7]. However, these data are mainly unreliable since participants may answer subjectively depending on the mood when the questionnaire is carried out. In order to overcome this problem, many scientific measurement equipment and methods had been developed and in particular Multiple Sleep Latency Test (MSLT) [8] is accepted widely as the standard method of quantifying sleepiness.

Epworth Sleepiness Scale (ESS) is carried out in the method of short questionnaire in which the subjects are to answer the probability of them falling asleep with increasing probability from a scale of 0 to 3 in eight different situations that occurred in daily lives. The scores from the eight questions are then added to obtain a final number. A number of 0-9 is considered as normal; 10-24 indicates expert medical advice is needed. For example: 11-15 indicate mild to moderate sleep apnea and a score above 16 indicated the possibility of severe sleep apnea or narcolepsy [9].

MSLT is a standard physiological measurement that measure the time elapsed from the start of a daytime nap period to the first signs of sleep, defined as sleep latency. The test is carried out using standard polysomnographic techniques and the test is carried out at 2-hour intervals throughout the day while lying in a quiet and dark bedroom. MSLT explored the assumption that sleepiness is a physiological need state that leads to an increase needs to fall asleep. The measurement system used to express has been the average daily sleep latency (the mean of four or five tests conducted). The mean sleep latency on MSLT on narcolepsy with cataplexy is less than or equal to 8 min according to the international classification of sleep disorders second edition (ICSD-2) [10]. The MSLT test is used as a standard measurement in the diagnosis of narcolepsy especially to distinguish between physical tiredness and true excessive daytime sleepiness (EDS) [11].

1.2 Essential hypersomnia and narcolepsy with cataplexy

Essential hypersomnia (EHS) is a sleep disorder characterized by excessive daytime sleepiness (intolerable sleepiness in the daytime). However, EHS patients do not display cataplexy (sudden loss of muscle tone in response to strong emotion) characteristic. Narcolepsy with cataplexy which has been well known as a similar hypersomnia exhibit both excessive daytime sleepiness and cataplexy. Around the year 1986, a group of narcolepsy-like patients without cataplexy but showed excessive daytime sleepiness features which are indistinguishable from those of narcolepsy with cataplexy such as shorter episodes of irresistible daytime sleepiness, feeling of refreshment after short naps and no prolonged night-time sleep period were discovered. These patients showed frequent sleep-onset rapid eye movement (REM) periods (SOREMPs) when MSLT is performed [12] . In additions, a long-term outcome measurement of EDS in narcolepsy with cataplexy and EHS without any treatment showed that 40% of all EHS patients showed improvement in EDS but similar improvement only observed in 14% of narcoleptic patients[13].

Initially, these patients are defined as idiopathic hypersomnia according to the classification of ICSD first edition [12], but later discovery showed that symptomatic characteristics of these patients are distinctly different from those of classical idiopathic hypersomnia syndrome (IHS), in which long naps and prolonged nocturnal sleep are reported to exist [14].

Narcolepsy with cataplexy usually begins during adolescence and affects 0.16–0.18% of the general population in Japan and 0.02–0.06% of the population in the United States and Europe, with men and women equally affected [15, 16]13]. The exact prevalence of EHS is largely unknown. EHS is at least of similar frequency as narcolepsy with cataplexy, but few research data are available. EHS usually occurs between ages 15-30 and get severe over the years [17] while narcolepsy with cataplexy occurrence major peak at the age of 15 years and a second smaller peak at 36 years old [18].

Around 40% of EHS patients carried Human Leukocyte Antigen *(HLA)-DQB1*06:02* compared to 100% prevalence in narcolepsy with cataplexy patients but higher prevalence than general population which is around 12% (Table 1.1) [19]. Previous reports have indicated the relation of *HLA-DQB1*06:02* with the pathogenesis of narcolepsy [20], thus a different pathogenesis pathway was suggested to be involved for EHS. In addition, EHS would more likely to be a subgroup of hypersomnia based on their different genetic composition, milder disease severity and better treatment response from narcolepsy with cataplexy (Table 1.1) [19]. Epworth Sleepiness Scale (ESS) and MSLT on narcolepsy with cataplexy and EHS have also showed that the severity of EDS was significantly milder in EHS compared to those of narcolepsy with cataplexy (Table 1.1) [21].

Orexin (hypocretin) level is reported to be low in the cerebrospinal fluid (CSF) of narcolepsy with cataplexy patients and post-mortem of narcolepsy with cataplexy patients discovered 95% loss of orexin producing cells in the brain [22, 23]. However, CSF orexin level in EHS are mainly dominated by normal orexin level (< 200pg/mL) patients which are indistinguishable from normal healthy individual (Table 1.1) [18].

Both genetics and environmental factors are known to trigger or cause the onset of narcolepsy with cataplexy [15]. In recent years, a lot of genetics studies of different scales and populations have been published with different results. Here, I introduced only the papers where the results with outstanding *P*-values and replicated in different populations. Narcolepsy has constantly been reported to be strongly associated with *HLA-DQB1*06:02* [20, 24]. Our group has conducted a genome-wide association study (GWAS) in Japanese narcolepsy with cataplexy and identified a novel association located in carnitine palmitoyltransferase 1B (*CPT1B*) [25]. Besides that, Stanford university sleep research group has also conducted a GWAS in Caucasian and identified two novel associations in T-cell receptor α-locus (*TCRA*) [26] and purinergic receptor P2Y (*P2RY11*) [27]. Recently, multiple research groups focusing on different ethnic groups have reported a relationship between H1N1flu pandemic with the onset of narcolepsy with cataplexy especially in the children [28, 29]. Amongst the susceptible genes identified in narcolepsy with cataplexy, *TCRA* [30] gene was replicated in *HLA-DQB1*06:02* positive EHS group but not *HLA-DQB1*06:02* negative EHS group, indicating that *HLA-DQB1*06:02* carrying hypersomnia groups shared the same susceptibility genes. While both associations with *HLA* and *TCRA* defined narcolepsy with cataplexy as an immunological related disease, initial experiments findings demonstrated that *HLA-DQB1*06:02* negative EHS is not caused by dysfunctional in the immunological system [31].

Based on the knowledge that *HLA-DQB1*06:02* negative EHS might possess different etiological pathways from narcolepsy with cataplexy, I have pursued on studying the genetic background of EHS. Thus, the objective of my study is to identify common genetic variations associated with the susceptibility to *HLA-DQB1*06:02* negative EHS in the Japanese population. Associated susceptibility genes were further confirmed by functional studies. Besides looking at single nucleotide polymorphism (SNP) associations, a pathway analysis was performed where SNPs were grouped based on their function registered in database.

2. Materials and methods

PART1: GWAS of Japanese *HLA-DQB1*06:02* negative EHS

2.1 Subjects of study

A collaborative study has been established between The University of Tokyo, and Neuropsychiatric Research Institute of Japan (Tokyo, Japan), in order to identify common genetic variants associated with the susceptibility to EHS in the Japanese population. Diagnostic criteria for EHS are based on (i) recurrent daytime sleep episodes that occur basically everyday over a period of at least 6 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 [19, 21, 31, 32]

I focused our studies on patients with EHS, but lacking *HLA-DQB1*06:02*, because previous studies indicated that *HLA-DQB1*06:02* negative EHS is essentially different from HLA-positive EHS and narcolepsy with cataplexy [19, 30, 31]. Here, I report a GWAS to identify new susceptible genes for *HLA-DQB1*06:02* negative EHS. In this GWAS, I recruited 125 individuals who were given a diagnosis of EHS at the clinic affiliated with Neuropsychiatric Research Institute of Japan and 562 Japanese individuals as healthy controls. All genomic DNA samples were genotyped using Affymetrix Genome-Wide SNP Array 6.0 platform. This research was approved by the research ethics review committees of the University of Tokyo and Neuropsychiatric Research Institute of Japan.

2.2 Measurement of DNA concentration and pooling of DNA

The genomic DNA was extracted and was measured using Nanodrop ND-1000 spectrophotometer with ratio of absorbance of 260/280nm with optical density (OD) meter. The purity of genomic DNA extracted was assessed by the ratio of A260/280 with a ratio of 1.8 for pure DNA. Genomic DNA was adjusted to50 ng/μl.

2.3 HLA Genotyping

HLA-DQB1 genotyping for EHS samples was performed using Luminex Multi-Analyte Profiling System (xMAP) and WAKFlow HLA typing kit (Wakunaga, Hiroshima, Japan). Briefly, target DNA was amplified by PCR (polymerase chain reaction) with biotinylated primers. The PCR amplicon was then denatured and hybridized to complementary oligonucleotide probes immobilized on fluorescent coded microsphere beads. In the meanwhile, biotinylated PCR products were labeled with phycoerythrin-conjugated streptavidin, and finally HLA typing was examined by Luminex 100 (Luminex, Austin, TX).

2.4 SNP genotyping

Genotyping by Affymetrix Genome-Wide SNP Array 6.0 was performed as part of my master degree program in the University of Tokyo. Affymetrix Genome-Wide SNP Array 6.0 contains more than 906,600 SNPs and more than 946,000 probes for detection of copy number variation (CNV). SNPs on the array are presented on 200 to 1,100 base pairs (bp) Nsp I or Sty I digested fragments in the human genome, and Genome-Wide Human SNP Nsp/Sty Assay kit 6.0 was utilized for amplifications of these fragments. Briefly, total genomic DNA (500ng) was digested with Nsp I and Sty I restriction enzymes and ligated to adaptors that recognize the cohesive

4 bp overhangs. A generic primer that recognized the adaptor sequence was used to facilitate the amplification of adaptor-ligated DNA fragments. Subsequently, PCR amplified products were combined and purified using polystyrene beads. The amplified DNA was then randomly fragmented, labeled and hybridized to a Genome-Wide Human SNP Array 6.0 chip. (Figure 2.1)

2.5 Genotype calling and data cleaning

Genotype calling was performed using Affymetrix Genotyping Console 4.0, which employs the Birdseed genotype calling algorithm for Affymetrix Genome-Wide SNP Array 6.0 [33]. QC call rate is well-correlated with the clustering performance and is an efficient single-sample metric before downstream clustering. Samples with a low quality control call rate (typically <95%) were excluded before performing the full clustering analysis of genotypes. The result was later exported into PLINK format (ped and bed files) for downstream data cleaning and association studies. PLINK is a free, open-source whole genome association analysis toolset, designed to perform a range of basic and large-scale analyses in a computationally efficient manner [34]. Schematic presentation of statistical analysis performed is showed in Figure 2.2.

Raw genotyping data of Affymetrix Genome-Wide SNP Array 6.0 are reported to be sometimes associated with spurious results, thus stringent data cleaning is of paramount importance to exclude these spurious results [35]. The following four parameters were used to exclude low quality samples:

1. SNP call rate: For the assessment of appropriate data cleaning threshold, total SNP call rate threshold of ≥ 95%, ≥ 97%, and ≥ 99% were performed and its efficiencies were later confirmed by visual checking of clustering graph.

2. Hardy-Weinberg equilibrium (HWE): Deviation from HWE can occur due to multiple possibilities such as deletion polymorphism, population stratification, inbreeding and CNV [36]. However, genotyping errors may cause deviation from HWE as well. Associated SNPs presented in cases might be deviated from HWE due to imbalance of susceptibility or protective allele, thus HWE cleaning criteria only applied to control samples with a cut-off threshold of $P < 0.001$.

3. Minor allele frequency (MAF): Uncommon SNPs (MAF < 5%) are reported to exhibit higher false positive rate compared to common SNPs [37]. Therefore, SNPs with MAF < 5% were excluded from subsequent analyses.

4. SNPs located on chromosome X were excluded from downstream analyses as this study did not performed sex-matched design.

2.6 Statistical analysis

Association analyses of SNPs were analyzed using an allelic model, a dominant model, a recessive model, or a Cochran-Armitage trend test using PLINK (v1.07) [34]. Population stratification within the Japanese patients with EHS was evaluated based on the genomic inflation factor (λ), which was calculated from the median of the Cochran-Armitage trend test. The quantile-quantile (Q-Q) plot was plotted with expected distribution of association test statistics under null distribution across the Cochran-Armitage trend observed *P*-values with R statistical environment version 2.9.0. Peta odd ratio was used to calculate odd ratio for any contingency column with counting of 0, online calculator is available at http://www.hutchon.net/ConfidORnulhypo.htm . Manhattan-plot was generated using Haploview (v4.1) [38]. For the Caucasian comparative study, an allelic, a dominant, and a recessive model were each assessed using a 2-tailed chi-square test. The eQTL analyses were performed based on data from the Sanger Institute GENEVAR project

[39]; these data are based on three cell types (fibroblast, lymphoblastoid cell line and T-cell) of 75 unrelated Western European origin individuals [40]. The SNPExpress Database [41], which is based on 93 autopsy-collected cortical brain tissues with no defined neuropsychiatric condition and 80 peripheral blood mononucleated cell (PMBC) samples collected from healthy donors, was also used as a reference for eQTL analyses. Relationships between the genotypes of candidate loci and the expression levels of nearby genes and transcripts were also examined.

2.7 Whole-genome imputation

High-throughput genotyping technology of Affymetrix Genome-Wide SNP Array 6.0 has benefited biological research in many ways, including improved power for genetic association analyses. However, limitation still exists; for example: Affymetrix Genome-Wide SNP Array 6.0 has the capability of screening approximately 906,600 SNPs, but this only represents 30% of approximately three millions SNPs detected by HapMap. Furthermore, as shown above, many of these 906,600 SNPs in Affymetrix Genome-Wide SNP Array 6.0 are not available for use in association analyses due to deviation from HWE, rare alleles and low call rates. As a result, genotype data are often missing and the missing data reduce the power of the study. A potential solution for this problem is to impute untyped SNPs from their nearby genotyped SNPs through their (Linkage Disequilibrium) LD pattern [42, 43] . Several imputation methods use various statistical models such as the haplotype-clustering algorithm [44] , the hidden Markov model (HMM)[45] and the Markov Chain model [46]. Generally, imputation accuracy increases with stronger LD, lower MAF and higher marker density [42]. These model robustness to impute has also been reviewed in several papers [42, 43, 47, 48], among all programs available; MACH [46] was chosen for its advantages for its accuracy regardless of marker density and size of reference samples.

The majority of published GWAS only performed a regional based imputation surrounding the top associated regions. In this study, I performed a genome-wide imputation in order to identify novel SNPs that might exist in regions that are not top associated regions. Imputed data are useful for downstream analysis. Besides, I sought to build a standard system in which a genome-wide imputation can be performed on other studies or when new references are published. As it was not computer power feasible to perform a whole-genome imputation in one batch, I divided the GWAS data to 22 parts according to 22 autosomes. The imputation was performed in parallel in computer server node. The efficacy of the imputation was investigated across 22 autosomes based on the number of SNPs in the reference in comparison to successfully imputed SNP after QC. Computer power and time needed to perform the imputation on each chromosome are also reported.

Briefly, for the inference of untyped and missing genotypes around candidate chromosomal loci, genotypes from our own samples together with haplotypes for HapMap reference samples (Japanese from Tokyo, JPT) [49] (http://hapmap.ncbi.nlm.nih.gov/) were prepared. MACH (v1.0) [46] was used to estimate haplotypes, map crossover and error rates using 50 iterations of the Markov chain Monte Carlo algorithm. By employing genotype information from HapMap database, maximum likelihood genotypes were generated (Figure 2.3). For the quality control of the imputed genotypes, I retained imputed genotypes with the estimated $r^2 > 0.3$. The imputed genotypes were re-analyzed by allelic, dominant, recessive models and Cochran-Armitage Trend Test utilizing PLINK 1.7 [34]. Regional association plots were drawn using LocusZoom(v1.1) [50]. In order to validate the accuracy of the top SNPs from the Affymetrix Genome-Wide SNP Array 6.0 platform, three SNPs (rs11854769, rs12471007 and rs10988217) were genotyped by Taqman assay.

2.8 SNP selection for comparative study

In order to confirm the effects of susceptibility SNPs found in Japanese *HLA-DQB1*06:02* negative EHS GWAS samples, I conducted a comparative study. This was a collaboration study with the Center for Narcolepsy, Stanford University School of Medicine. The comparative study consisted of 268 *HLA-DQB1*06:02* negative Caucasian hypersomnias and 1,761 *HLA-DQB1*06:02* negative Caucasian healthy controls. All subjects had given written informed consent for their participation in these studies in accordance with the process approved by ethics committees of the University of Tokyo and Stanford University.

Regarding the SNP selection for the comparative study, after imputation of the top regions, normally SNPs which has the lowest SNPs will be chosen from a particular LD block. In this study, I selected SNPs with P_{min} < 1.0E-06 for replication study. I focused on genetic loci that had a couple of SNPs which are significant in a particular LD block. MAF in Caucasian of the SNPs I chosen were also investigated using the HapMap database[49] as SNPs with extremely low MAF (<0.05) would be hard to be replicated in Caucasian because large number of samples will be needed to reach significance. Finally, when I found an associated SNP in the patients with the Caucasian hypersomnias, the SNP was tested in narcolepsy with cataplexy patients in Japanese (n=422) and Caucasian populations (n=1,093).

PART2: Comparison study of EHS GWAS and narcolepsy with cataplexy GWAS

3.1 GWAS of narcolepsy with cataplexy

Hiromi Toyoda from our laboratory has performed a GWAS on narcolepsy with cataplexy using the same experimental and analysis as GWAS on EHS. All narcolepsy with cataplexy patients are diagnosed according to the ICSD-2 [10]. I recruited 422 narcolepsy with cataplexy patients Neuropsychiatric Research Institute of Japan and 562 Japanese individuals as healthy controls. All the narcolepsy with cataplexy samples were *HLA-DQB1*06:02* positive. All genomic DNA samples were genotyped using Affymetrix Genome-Wide SNP Array 6.0 platform according to the manufacture's protocol.

3.2 Comparison between *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy utilizing GWAS results

Previous studies have shown that these two sleep disorders might have their independent pathological pathways even though some of their symptoms such as EDS are overlapped. In order to identify SNP markers those are able to use to differentiate these two sleep disorders, I have performed a case-case GWAS comparison study between these two sleep disorders using PLINK (v1.7) [34]. Association analyses of SNPs were analyzed using Cochran-Armitage trend test.

3.3 SNPs selection for further studies

The HLA region was excluded from the case-case GWAS comparison. The top results were mainly dominated by SNPs that were significant in their perspective case-control GWAS, these results also were removed from further analysis. The remaining SNP candidates were chosen after comparing the *P*-value of the original case-control GWAS and case-case GWAS, priority will be

given to SNPs that showed stronger case-case *P*-values compared to their perspective case-control GWAS. Besides, the location of the SNPs (whether it located in the exon, intron and 5' promoter) and the function of the nearest gene is functionally related to the development of reported neurological related studies.

3.4 Lumbar CSF subjects and orexin measurement

Lumber CSF was collected between 10 a.m. and 3 p.m. and samples were kept at −80 °C freezer. Orexin level measurements were carried out at Akita University. Orexin was measured using radioimmunoassay kits (Phoenix Pharmaceuticals, Belmont, CA, USA) as previously reported [51]. The detection limit was 40pg/mL. The statistical significance was analyzed using the Kruskal–Wallis, with post hocBonferroni/Dunn test. In total, I have recruited 21 low orexin hypersomnias (≤ 200 pg/mL) and 23 normal orexin hypersomnias (> 200 pg/mL). All CSF samples were obtained by lumbar puncture and were stored at -80°C until assayed.

3.5 Enzyme-Linked ImmunoSorbent Assay (ELISA)

Quantification of MBP was carried out in the CSF samples of normal orexin hypersomnia patients and low orexin hypersomnia patients by quantitative MBP ELISA from Beckman Coulter Inc (DSL-10-58300) which specialized for quantitative measurement of MBP in CSF. I described the CSF samples as normal and low orexin hypersomnia patients because the subjects in this study have not been exactly diagnosed. The subjects have shown the symptom of EDS like narcolepsy and EHS and have currently been diagnosed as CNS hypersomnia by medical doctors. HLA of the subjects were not typed. However, *HLA-DQB1*06:02* negative EHS displayed normal level of CSF orexins while most of the narcolepsy with cataplexy patients carried *HLA-DQB1*06:02* result in undetectable or low level of CSF orexin. Normal and low orexin level hypersomnia patients would

correspond to EHS and narcolepsy with cataplexy, respectively. In total, measurements were carried out on 21 low orexin hypersomnias and 23 normal orexin hypersomnias. Experiments were carried out according to manufacturer's instructions. All measurements were performed in duplicate manners including standards, positive controls with known MBP concentration and samples. Inter and intra assays coefficients of variation were calculated to ensure the reproducibility of data. ELISA data were analyzed by Mann-Whitney U Test.

PART3: Pathway analysis

I applied the Gene Set Enrichment Analysis (GSEA) approach to genetic variant association data [52]. Analysis of GWAS often focuses on identifying individual SNPs that modify the risk of a phenotype, assuming the underlying association of an individual SNP without considering the involvement of any other SNPs [53]. Analyses of the effect of interaction between two SNPs or more (epistasis) either within the gene [53] or between different genes [34] have been reported, but these methods did not take the underlying biological sense into consideration, thus the results are often difficult to interpret. Here, GSEA method is used to test whether predefined biological processes or gene sets are enriched for genes with coordinate modest expression differences between two samples, these differences are normally hard to detect when inspecting each gene separately [54].

4.1 Statistical analysis

The outline of the methodology can be found in Figure **4.1**. The analysis is packed into a package (MAGENTA) that runs under Matlab with statistical toolbox installed. All data processing and analysis were performed offline using a commercial software package (MATLAB 7.12.0(R2011a), The MathWorks Inc., Natick, MA, 2000). In general, the GSEA based pathway analysis can be divided into four major steps. Step 1: All the SNPs from *HLA-DQB1*06:02* negative EHS GWAS were mapped onto genes in the genomes with the boundaries setting of 110kb upstream and 40kb downstream. Each gene in the pathway was assigned to a set of SNPs located in the gene boundaries I defined. Genes in the HLA region were removed due to high LD and high gene density. In Step2: Each SNP mapped to the genes was given a score, in this case, the more significant the *P*-value, the higher score this particular SNP will be scored. Step3: The confounding

effects such as gene size (in the case where two genes are assigned the same best SNP *P*-value, the smaller gene tends to be more significant than the large gene) was corrected using a step-wise multivariate linear regression analysis. Step4: the final step was to calculate a gene set enrichment P-value for each biological pathway or a gene set of interest. Non-parametric statistical test was applied to test whether for all genes in gene set were enriched for highly ranked gene scores more than would be expected by chance, compared to randomly sampled gene sets of identical size from the genome. False discovery rate (FDR) was calculated to eliminate false positive pathways. Only FDR of less than 0.2 was include in further analysis as recommended by the authors [54].

Two different gene ontology databases was used in this analysis: Kyoto Encyclopedia of Genes and Genomes (KEGG) [55, 56] and Ingenuity pathway (Ingenuity® Systems, www.ingenuity.com). In total, there are 136 pathways in the KEGG database (Table 4.1) as of February 2012 and 82 pathways in the Ingenuity pathway (Table 4.2).

4.2 Literature review of the associated pathway

After GSEA based pathway analysis was performed and filtered by FDR, the genes in the associated pathway were studied in details by classifying the genes according to the published function in NCBI database. A relationship graph was drawn for illustration purposes in the results section.

4.3 High-performance liquid chromatography (HPLC)

HPLC was carried out in collaboration with Tokyo Metropolitan Institute of Medical Science. The CSF concentration of GABA, glutamate and glutamine in the perfusate was measured by HPLC (EDT-300, Eicom), with fluorescence detection (Soma S-3350; excitation at 340 nm,

emission at 440 nm), and quantified with a PowerChrom (AD Instruments, Sydney, Australia) using external amino acid standards (Sigma, St. Louis, MO). Precolumn derivatization was performed with o-phthaldialdehyde/2-mercaptoethanol by an autosampler (Model 231XL; Gilson) at 10°C for 3 min. The derivatives were then separated in a liquid chromatography column (MA-5ODS; Eicom) at 30°C with 30% methanol in 0.1 M phosphate buffer (pH 6.0), being degassed by an on-line degasser (DG-100; Eicom). The detection limits for glutamate, GABA, and glutamine were 20 femtomoles (fmol). HPLC data were analyzed by Mann-Whitney U Test.

3. Results

PART1: GWAS of Japanese HLA-DQB1*06:02 negative EHS

Affymetrix Genome-Wide SNP Array 6.0 platform was utilized to perform genotyping of 125 *HLA-DQB1*06:02* negative Japanese EHS cases, 562 Japanese controls (Table 5.1). Before genotype calling on Affymetrix Genotyping Console 4.0, 8 samples with QC call rate less than 95% was excluded from analysis. Genotyping calling was then performed and exported to PLINK (v1.07)[34]. Data cleaning was performed at SNP call rate 95%, 97% and 99%, HWE P-value < 0.001 and MAF < 5% (Figure 2.2). After data cleaning of SNP call rate at 95%, 97% and 99%, total SNPs of 905,486 were filtered to 596,734 SNPs, 580,229 SNPs and 508,367 SNPs respectively. To evaluate the efficiency of each SNP call rates, cluster graph of top 100 SNPs were checked visually. Examples of cluster graphs (good and bad) were shown in Figure 5.1. Most of the 100 top SNPs in 95% level were spurious association, moderately in 97% and only 1 badly clustered SNP at call rate of 99%. SNP call rate of 99% was applied and 1 badly cluster SNP was removed from subsequent analysis (Table 5.2). After SNPs QC, an association analysis was performed using 508,366 SNPs on the basis of allelic, dominant, recessive models and Cochran-Armitage Trend Test for each individual. (Table 5.3)

Quantile-quantile plot (Q-Q plot) is a diagnostic plot that compares the distribution of observed test statistics with the distribution expected under the null. Population stratification was accessed by calculating the genomic inflation factor (λ); the λ of this data set was 1.008; this finding indicated that errors resulting from population stratification, cryptic relatedness, or both were unlikely (Figure 5.2). A genome-wide Manhattan plot was drawn using the chromosomal positions of individual SNPs (x-axis) and the negative logarithm of P values calculated with the Cochran-

Armitage trend test (y-axis) (Figure 5.3). Multiple regions were shown to be suggestively associated with *HLA-DQB1*06:02* negative EHS in Japanese population (Table 5.3) especially *NCKAP5* region on chromosome 2 and *SPRED1* region on chromosome 15. Additionally, I chose to further investigate *CRAT* region on chromosome 9 because of its functional importance.

A whole genome imputation was performed on *HLA-DQB1*06:02* negative Japanese EHS GWAS data using Japanese HapMap 3 data as reference. The number of SNPs ratio of *HLA-DQB1*06:02* negative Japanese EHS GWAS SNPs in comparison to Japanese HapMap 3 SNPs was roughly around 2.5-3.0 across all the 22 autosomes. The overall imputation efficacy after QC across 22 autosomes was good (>95%) except chromosome 19 with an accuracy of 88% (Figure 5.4). Imputation efficacy is known to be severely disturbed by the GC content; thus, GC content across 22 autosomes was calculated based on Japanese HapMap3 data build 37 (Table 5.4). The GC content in chromosome 19 is the highest among all the 22 autosomes with a ratio of 0.48. Besides, time and computer memory consumption for imputing each chromosome was shown in Figure 5.5. In general, the time and computer memory consumption increased as the number of SNPs involved in the imputation increased, both in the reference data or our own data. MACH[46] imputation was divided into two steps: the haplotyping step and the imputation step. In general, the imputation step required more computer memory consumption than the haplotyping step but the difference was not vast (Figure 5.5). On the other hand, the imputation steps required much more time for performing the imputation than the haplotyping steps, for example: chromosome 1 required only ~28 hours to perform the haplotyping step compared to more than ~80 hours in the imputation step. Thus the total hour needed for performing a full imputation on chromosome 1 was equivalent to ~108hours. If all the imputation was to perform by step-by-step wise (only 1 chromosome imputation/run), it would required ~1333 hours (55.5days) to complete the whole genome imputation. However, I

performed the whole genome imputation using the parallel system in computer server node; I can significantly reduce the time needed for performing the imputation by nearly 75%. I could not identify more significant SNPs than the GWAS SNPs after the whole genome imputation, but I have successfully imputed several marginal SNPs around the top SNPs region which will help in selecting SNPs for replication study (Figure 5.6).

The SNP rs16826005 had the most significant *P*-value (1.02E-07, per-allele odds ratio (OR) of 1.89 with 95% confidence interval (CI) of 1.43–2.50) of all SNPs assessed; rs16826005 was located within a 19-kb LD block on chromosome 2q21.2 (Figure 5.6a, Table 5.3). This LD-block covered the intronic region of *NCKAP5* (NCK-associated protein 5) gene. An imputation analysis of this region revealed modest associations between *HLA-DQB1*06:02* negative EHS and SNPs in the *NCKAP5* gene. This finding indicated that *NCKAP5* may play a causative role in EHS pathogenesis. Expression data for *NCKAP5* is not readily available in the Gene Variation Database (GeneVar) [39] or the SNPExpress database [41].

The SNP rs11854769 had the second most significant *P*-value by regional classification (6.69E-07, per-allele OR of 2.27 with 95% CI of 1.64–3.14) in this analysis. This SNP, rs11854769, resided within a 10 kb LD block on chromosome 15q14 (Figure 5.6b Table 5.3) and was located 42 kb upstream of *SPRED1* (sprouty-related, EVH1 domain containing 1). An imputation analysis of this region did not reveal any additional SNPs that were more significantly associated with *HLA-DQB1*06:02* negative EHS (Figure 5.6b). An eQTL analysis showed that rs11854769 did not affect the expression level of *SPRED1* in either the GeneVar database[39] (Figure 5.7a) or the SNPExpress database [41] (Figure 5.8a).

The SNP rs10988217 (*P*-value of 3.43E-06, per-allele OR of 1.52 with 95% CI of 1.13–2.04), which was located within a 63 kb LD block on chromosome 9q34.11, was also of interest (Figure 5.6c Table 5.3). The associated SNP is located in the intronic region of *PPP2R4* (protein phosphatase 2A activator, regulatory subunit 4) and *CRAT* (carnitine O-acetyltransferase). An imputation analysis of this region revealed modest associations between *HLA-DQB1*06:02* negative EHS and SNPs in the *CRAT/PPP2R4* region (Figure 5.6c).The eQTL analysis revealed that rs10988217 affected the expression of *CRAT* transcripts (Figure 5.8b), but not *PPP2R4* transcripts (Figure 5.8c). Similar findings were observed in the GeneVar database [39]; rs10988217 was associated with *CRAT* expression levels ($P < 0.05$) in three cell types (fibroblast, lymphoblastoid cell line and T-cell) (Figure 5.7b), but not with any changes in *PPP2R4* expression (Figure 5.7c).

A comparative study of Caucasian patients with *HLA-DQB1*06:02* negative hypersomnia revealed that rs10988217, the SNP in the *PPP2R4-CRAT* region, was significantly associated with *HLA-DQB1*06:02* negative hypersomnia in this population (*P*-value of 2.51E-02, per-allele OR of 1.25 with 95% CI of 1.03–1.52) (Table 5.5); these finds were similar to those from the GWAS of Japanese patients with *HLA-DQB1*06:02* negative EHS (*P*-value of 3.43E-06, per-allele OR of 1.52 with 95% CI of 1.13–2.04) (Table 5.3). To investigate possible contributions of this SNP to other forms of hypersomnia, I tested the association of rs10988217 with narcolepsy with cataplexy in Japanese and Caucasian patients [26]. Significant associations were observed in Japanese narcolepsy (*P*-value of 2.20E-02, per-allele OR of 1.22 with 95% CI of 1.03–1.45) and Caucasian narcolepsy (*P*-value of 2.82E-02, per-allele OR of 1.13 with 95% CI of 1.01–1.27) (Table 5.6). Other SNPs that showed associations with an increased risk of EHS in the Japanese population were

also genotyped in the samples from Caucasian patients with *HLA-DQB1*06:02* negative hypersomnia, but no significant associations were found (Table 5.3).

PART2: Comparison study of EHS and narcolepsy with cataplexy utilizing GWAS results

The GWAS between 125 *HLA-DQB1*06:02* negative EHS patients and 562 healthy controls indicated that four SNPs in strong LD headed by rs9283864 (*P* = 7.32E-06; OR = 1.89; 95% CI = 1.40-2.54) located approximately 140 kb upstream of quaking (*QKI*) gene were associated with susceptibility of *HLA-DQB1*06:02* negative EHS (Table 5.7, Figure 5.9). In addition, rs4890893 located in the second intron of myelin basic protein (*MBP*) was also associated with *HLA-DQB1*06:02* negative EHS (*P* = 1.95E-04; OR= 2.05; CI= 1.40-3.00). Direct involvement of *QKI* in demyelination and presents of QRE (QKI response elements) in both human and mice *MBP* [57] make it an interesting candidate for further investigation.

A case-case association study was performed between 125 *HLA-DQB1*06:02* negative EHS and 422 *HLA-DQB1*06:02* positive narcolepsy with cataplexy patients to compare SNPs in *QKI* and *MBP* regions between these two group of hypersomnias (Table 5.7). SNPs in these regions which were associated with *HLA-DQB1*06:02* negative EHS in the case control GWAS also showed significant differences between these two group of hypersomnias (Table 5.7). The P-value of rs9283864 in *QKI* was (*P*= 2.22E-05; OR= 1.89; CI=1.39-2.56) (Table 5.8, Figure 5.9). Interestingly, an analysis of rs4890893 in *MBP* provided *P* = 2.74E-06 with an OR= 2.51, CI=1.71-3.70. The P-value was more significant than those of the case control comparisons GWAS for narcolepsy with cataplexy (*P* = 4.74E-02; OR = 0.82; 95% CI = 0.67-1.00) and *HLA-DQB1*06:02* negative EHS (*P* = 1.95E-04; OR = 2.05; 95% CI = 1.40-3.00) respectively (Table 5.7, Figure 5.9).

Quantitative ELISA studies were performed to evaluate the differences of MBP level in the CSF of 21 low orexin hypersomnia groups and 23 normal orexin hypersomnia groups. Normal and low orexin level hypersomnia patients would correspond to EHS and narcolepsy with cataplexy

respectively (details in Methods 3.5). The normal orexin hypersomnia group had a significant increased mean of CSF MBP concentration (0.52 ± 0.04ng/mL) in comparison to low orexin hypersomnia group with mean CSF MBP concentration of 0.43 ± 0.06 ng/mL (Figure 5.10). Mann-Whitney two-tailed U test showed a *P*-value of 0.025 between these two groups. Inter and intra assay coefficients of variation were both less than 15%.

PART3: Pathway analysis

A pathway analysis was performed using the Mathlab package named MAGENTA[54]. Two databases: KEGG [55, 56] and Ingenuity pathways (Ingenuity® Systems, www.ingenuity.com) were used as gene-set references. After QC filtering at FDR< 0.2 in both databases results (Table 5.9, Table 5.10), only one pathway (glutamate metabolism) was found to be significant with nominal GSEA *P*-value of 1.70E-03 and FDR of 0.15. However, the default SNP pruning (based on LD block) of the software was performed on Caucasian, so I investigated the effect of SNP pruning by building a new reference using CHB+JPT (Chinese + Japanese) HapMap 3 database. SNP pruning was performed using PLINK (V1.7) [34] . After SNP pruning using CHB+JPT HapMap, nominal GSEA *P*-value of glutamate metabolism pathway did not differ much from the original of 1.70E-03 to 2.80E-03 (Table 5.11).

The genes involved in the significantly associated glutamate pathway were investigated in details and listed in Table 5.12. The gene *P*-value of each gene calculated in the pathway, the number of SNPs analyzed in the pathway, the best SNPs in the genes and its *P*-value are also shown in the table. All the best SNPs listed were checked for their correlation of their respective gene whether the SNP interfere with the expression of the SNP in Genevar expression database [39] but no conclusive pattern could be observed.

In order to better illustrate the result, I performed an intensive background study of the gene listed in Table 5.12. The genes were then categorized by their general function and their relationships were also investigated. The gene components of glutamate metabolism pathway were divided to neurotransmission, protein synthesis & degradation, energy production and ammonia disposal.

After pinning down the candidate amino acid to three candidates, glutamate, glutamine, and GABA, HPLC (Figure 5.11) was utilized for accurate quantification of the three amino acids concentration in the CSF of normal orexin hypersomnia groups and low orexin hypersomnia groups. Among the three amino acids, GABA levels were significantly higher (Mann-Whitney two-tailed U test *P*-value = 0.003) in low orexin hypersomnia groups (1.21μM ± 0.21) in comparison with normal orexin hypersomnia (0.56μM ± 0.09) (Figure 5.12). Glutamate (Mann-Whitney two-tailed U test *P*-value = 0.44) and glutamine (Mann-Whitney two-tailed U test *P*-value = 0.62) levels were both not significantly different between normal orexin hypersomnia and low orexin hypersomnia.

4. Discussions

PART1: GWAS of Japanese *HLA-DQB1*06:02* negative EHS

This study represents the first GWAS designed to identify common genetic variants associated with *HLA-DQB1*06:02* negative EHS in a Japanese population. In total, 125 individuals with *HLA-DQB1*06:02* negative EHS and 562 healthy controls participated were recruited. I identified novel candidate regions associated with an increased risk of *HLA-DQB1*06:02* negative EHS.

For the data cleaning of the GWAS, I have applied three different criteria: SNP call rate, HWE and MAF respectively. Particularly, SNP call rate was performed at three thresholds, 95%, 97% and 99%. I observed that accuracy of genotyping was highly influenced by the SNP call rate with an immense decrease of badly clustered SNP at SNP call rate of 99%. This was most likely due to the fact that poorly genotyped SNPs across individuals' samples are more likely to generate poor intensity signals which directly influence the clustering of SNPs.

Whole genome imputation efficacy was calculated across 22 autosomes and chromosome 19 showed the lowest efficacy at 88% (Figure 5.4). It is most likely due to the high GC content in chromosome 19 compared to other chromosomes (Table 5.4). High GC content can cause drastic decrease of imputation efficacy for low genotyping call or miscall of genotypes around high GC content regions.

Several SNPs located in an intron of *NCKAP5* gene showed associations with an increased risk of *HLA-DQB1*06:02* negative EHS in this study. *NCKAP5* variants are reported to be associated with bipolar disorder [58], attention deficit hyperactivity disorders [59], and multiple

sclerosis [60]. A further meta-analysis of combination of schizophrenia and bipolar disorder confirmed the association between *NCKAP5* variants and both schizophrenia and bipolar disorder [61]. Currently, the function of *NCKAP5* is unknown.

SPRED1 was the gene closest to the SNP with the second most significant *P*-value by regional based; *SPRED1* is a member of the Sprouty family of proteins and is known to be phosphorylated by a tyrosine kinase in response to several growth factors [62]. Proteins in the *SPRED1* family act as negative regulators of RAS-RAF interactions and of the mitogen-activated protein kinase (MAPK) signaling pathway [63]. RAS/MAPK signaling has been implicated in the mediation of reversible circadian outputs [64] and sleep/wake condition mechanism [65] of the brain. Some narcolepsy without cataplexy patients have been reported to exhibit down-regulation of lumbar cerebrospinal fluid orexin level [66] and hypothalamic peptides orexin has been reported to contribute to the intrusion of REM sleep behaviors into wakefulness by coordinating the activity of RAS through orexin neuron during waking stage [67]. Since *SPRED1* is an important component in the RAS pathway, *SPRED1* might play a crucial role in regulating REM sleep behaviors. In addition, germline mutations in genes involved in the RAS pathway lead to neuro-cardio-facial-cutaneous (NCFC) syndromes (ex: neurofibromatosis 1 (NF1, OMIM 162200) [63], Noonan syndrome (NS, OMIM 163950), LEOPARD syndrome (LS, OMIM 151100), cardio-facio-cutaneous syndrome (CFC, OMIM 115150), and Costello syndrome (CS, OMIM 218040) [68]). The SNP rs11854769 was identified in a recent GWAS of bipolar disorder [69]; this finding may indicate that this *SPRED1* variant may confer a genetic predisposition for multiple neuropsychiatric diseases.

Association studies for rs10988217 in *CRAT* showed significant associations with both *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy patients. In addition, these

associations were observed not only in Japanese but also in Caucasian (Table 5.6). These results indicated that *CRAT* may be a susceptibility gene for different types of hypersomnias. The eQTL analysis of rs10988217 revealed that the SNP was associated with alterations in the expression of *CRAT* (Figure 5.7b, 5.8b). Besides the *CRAT* association in this study, interestingly, a recent GWAS [25] for narcolepsy with cataplexy in a Japanese population identified a significant association between a SNP adjacent to *CPT1B* (carnitine palmitoyltransferase 1B). This SNP was also associated with changes in *CPT1B* expression levels [25]. Furthermore, a subsequent association study demonstrated an association between *CPT1B* and EHS [32]. Both *CRAT* and *CPT1B* are involved in the β-oxidation of long-chain fatty acid. *CRAT* gene encodes the carnitine acetyltransferase protein, which is a key enzyme in the β-oxidation pathway in mitochondria, peroxisomes, and the endoplasmic reticulum. *CRAT* is responsible for catalyzing the reversible transfer of acyl compartments groups from an acyl-CoA thioester to carnitine, and this enzyme regulates the ratio of acylCoA/CoA in the subcellular compartments [70, 71] (Figure 6.1). *CPT1B* is the rate-controlling enzyme of long-chain fatty acid β-oxidation in the mitochondria of muscle tissue. *CPT1B* catalyzes the transport of long-chain fatty acyl-CoAs from the cytoplasm into the mitochondria through the carnitine shuttle (Figure 6.1). Deficiency of short-chain acyl-coenzyme A dehydrogenase in a mouse model resulted in slowing of theta frequency during REM sleep [72]. Additionally, acetyl-L-carnitine (ALCAR) is a potential treatment for neurological diseases such as Parkinson's disease [73] and Alzheimer's disease [74]; it is also known to restore β-oxidation of fatty acids in the mitochondria and rescued the slow theta frequency in REM sleep of mice lacking short-chain acyl-coenzyme A dehydrogenase [72]. Besides, our group has recently reported a clinical trial of oral L-carnitine for narcolepsy with cataplexy and the results suggested that oral L-carnitine can be a treatment for narcolepsy with cataplexy [75, 76]. On the basis of these reports, the

results in the study indicated that the pathophysiology of hypersomnias is associated with metabolic alterations [75, 76].

For rs10988217 in *CRAT*, the best P-values in Japanese *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy patients were from the recessive model (Table 5.3). The recessive model was not significant in Caucasian hypersomnia and narcolepsy with cataplexy patients, but the allelic model showed a significant association (Table 5.6). This might be due to a difference between the populations. In addition, the risk allele (G) for rs10988217 was minor in Japanese but major in Caucasian (Table 5.6). As another possibility, rs10988217 is not the primary SNP of *CRAT* region. LD of the primary SNP and rs10988217 might be different between Japanese and Caucasian, contributing to the different significant model. Therefore, a further replication study and re-sequencing should be required to overcome the limitations.

PART2: Comparison study of EHS GWAS and narcolepsy with cataplexy GWAS

Almost all of narcolepsy with cataplexy patients carried *HLA-DQB1*06:02* resulting in undetectable or low level of CSF orexin and those who lack of *HLA-DQB1*06:02* allele displayed normal level of CSF orexin [77, 78]. Reports [77, 79] demonstrated that narcolepsy without cataplexy especially those who lack of *HLA-DQB1*06:02* allele follows a difference etiological pathway than that of deficiency in orexin ligand. Both *MBP* and *QKI* are involved in the formation and stabilization of myelin sheath and the SNPs on/near the genes were significant differences between narcolepsy with cataplexy and *HLA-DQB1*06:02* negative EHS.

The function of *QKI* has been well characterized in mice where a deletion including the 5' regulatory region of the mouse quaking gene (*QKI*) causes tremor and severe demyelination of the CNS [80]. This mutation causes the dysfunction of the oligodendrocytes and reduces the expression of myelin components in CNS [81]. Mouse QKI protein directly regulates several myelin-specific genes including the expression of *MBP* [82] and alternative splicing of myelin-associated glycoprotein (*MAG*). In human, *QKI* had been suggested to be one of the causal genes for schizophrenia [83]. Gene expression studies of schizophrenia patients demonstrated a down regulation of a set of genes expressed only in mature myelin-producing oligodendrocytes, such as *MAG*, *PLP1*, and *MBP* [84]. Evidence of alternative splicing of *QKI* gene is responsible for the down-regulation of myelin-related genes has also been reported [83]. Interestingly, a recent study reported that a mutation in myelin oligodendrocyte glycoprotein (*MOG*) gene, a minor component of the myelin sheath, was detected in familial narcolepsy. However, it is noteworthy that part of the family members of this study carried non-*HLA-DQB1*06:02* alleles and presentation of non typical cataplexy [85].

MBP is the second most abundant protein of the CNS representing about 30% of the total myelin protein and accounted for 10% of myelin weight [86]. CNS myelin sheath is produced by oligodendrocytes which play a cardinal role in extending membrane processes that wrap around the axon. Myelin is an electrical insulator and local circuit travel down the axon through depolarization of the nodes of Ranvier sequentially. This process is much more efficient in contrast to unmyelinated axon which propagates in continuous sequential fashion through adjacent sections of membrane [87]. Multiple sclerosis (MS) is a disease that the underlying factor involved demyelination of the white matter in the spinal cord and brain [88]. Biochemical analysis of MS patients' CSF for *MBP* enhanced the specificity of clinical differential analysis [89]. Recently, MS patients are reported to suffer from sleep problems and frequent yawning [90]. The potential underlying mechanism was described by Morelli, A., *et al* [91, 92] to be related to proton (H+) gradient across myelin membranes which functions in adenosine triphosphate (ATP) synthesis. *MBP* and phospholipids which hold high buffering capacity for proton (H+) can be discharge to produce energy during the wake period. In contrast, sleep would be induced when myelin is fully discharged and wakefulness will occur when myelin is completely recharged. Taken together, it is possible to speculate that demyelination is involved in the excessive daytime sleepiness which is a characteristic of *HLA-DQB1*06:02* negative EHS patients.

PART3: Pathway analysis

After SNP pruning and QC of FDR<0.2, glutamate metabolism pathway was found to be significant with P-value of 1.70E-03 and FDR of 0.15. After dividing the genes involved in the pathway into four major categories based on their function in glutamate metabolism pathway, I decided to do a follow-up study on the neurotransmission pathway and the reversible form of glutamate and glutamine due to their direct correlation to their vital roles in the normal function of neurotransmission in the brain. Glutamate is best known to be the most abundant excitatory neurotransmitter in the vertebrate nervous system [93]. Glutamate also plays a vital role in synaptic plasticity and thus glutamate is essential in normal cognitive functions such as building memory in the brain and learning processes [94]. In normal neurotransmission processes, glutamate will be released to the synaptic cleft where the signal is passed down to the next synapse. However, these glutamates must be removed from the synaptic cleft by glutamate transporters after the signal is passed down. However, in the cases of brain/spinal cord injury or diseases, these glutamates will accumulate outside cells and this will cause the calcium ions enter cells via the N-methyl-D-aspartate (NMDA) receptor. This process will lead to serious neuronal damage and eventually lead to cell death. This whole process is term as glutamate excitotoxicity [95]. Glutamate excitotoxicity has been reported to be related to multiple sclerosis[96, 97], Alzheimer's disease [98] and schizophrenia [99].

Reduced expressions of genes associated with oligodendrocytes and myelin have been reported [100-102] in schizophrenia patients. The density of oligodendrocytes was found to be lower in schizophrenia patients in the grey matter and white matter of the brain [103, 104]. Signals of damaged myelin sheath lamellae were also observed in schizophrenic brain [105].

Multiple sclerosis is a demyelinating disease where the myelin shealth is seriously damaged. In the case model for multiple sclerosis, glutamate levels were found to be elevated in both CSF [106] and the brain of multiple sclerosis patients [107]. The expression of glutamate receptors has also found to be elevated in the brain of multiple sclerosis patients, [108] suggesting that hypo-glutamatergia might occur in the synapse. Indeed, glutamate excitotoxicity of oligodendrocytes through glutamate receptors is thought to be the major factor for demyelination in multiple sclerosis [96].

Given the findings in schizophrenia, multiple sclerosis and my finding in PART 2 of this theses where MBP level is found to be increased in CSF of normal orexin hypersomnias group, I performed a preliminary screening using HPLC analysis on the level of glutamate, glutamine and GABA in the CSF of low orexin hypersomnias group and normal orexin hypersomnia group. Among the three amino acids, only GABA levels in low orexin hypersomnia group (1.21μM ± 0.21) were significantly higher compared with normal orexin hypersomnia group (0.56μM ± 0.09) (*P*-value = 0.003) (Figure 5.12). GABA is a main inhibitory neurotransmitter of the CNS and it is well established that activation of GABA receptors induced sleep[109]. The result is parallel with published reports that sleepiness of narcolepsy with cataplexy (low orexin hypersomnia groups) is more severe compared to EHS (normal orexin hypersomnia groups) based on EDS severity, dosage of psycho stimulant needed and treatment response to the psycho stimulant [19]. Nevertheless, although EHS patients are suffered from a milder form of EDS, SNPs in the glutamate metabolism pathway with small effects might also contribute to EDS of EHS. It is not surprising that I could not detect any differences in the levels of glutamate and glutamine, as they were removed rapidly from the synaptic cleft [110] even though excessive release of glutamate might have occurred beforehand.

5. Conclusions

In PART1, a GWAS was conducted between 125 unrelated Japanese EHS patients lacking the *HLA-DQB1*06:02* allele and 562 Japanese healthy controls. A comparative study was also performed on 268 *HLA-DQB1*06:02* negative Caucasian hypersomnia patients and 1761 *HLA-DQB1*06:02* negative Caucasian healthy controls. I have successfully identified three SNPs that each represented a unique locus, rs16826005 (*P* = 1.02E-07; *NCKAP5*), rs11854769 (*P* = 6.69E-07; *SPRED1*), and rs10988217 (*P* = 3.43E-06; *CRAT*) that were associated with an increased risk of EHS in this Japanese population. Interestingly, rs10988217 showed a similar tendency in its association with both *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy in both Japanese and Caucasian populations. In addition, eQTL analysis showed that risk allele of rs10988217 increased the expression of *CRAT*. Regulatory SNP in *CRAT* gene, together with previously identified *CPT1B* as susceptibility gene for narcolepsy through the long-chain fatty acid β-oxidation pathway contribute to the risk of having *HLA-DQB1*06:02* negative EHS. However, follow-up studies such as knock-out of genes identified in mice model would provide further insight on the pathological mechanism of hypersomnias.

In PART2, with the hypothesis that *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy might follow different etiological pathway based on their different responses to psycho stimulant, I have conducted a comparison study between *HLA-DQB1*06:02* negative EHS and narcolepsy with cataplexy. Results showed that myelin related genes such as *MBP* (*P* = 2.74E-06) and *QKI* (*P* = 2.22E-05) are potential genetic markers which can be used to differentiate the two hypersomnias. Furthermore, I demonstrated that mean MBP concentration in CSF was significantly higher in normal orexin hypersomnias group (represents the EHS) in contrast to low orexin hypersomnias group (represents narcolepsy with cataplexy). Previous study have shown that myelin

play an important role in the sleep-wake mechanism in which proton (H+) across the myelin sheath is discharged to produce energy during wake cycle. On the other hands, sleep induction occurs when the proton is depleted [91, 92]. The underlying confounding factors for demyelination in normal orexin hypersomnia can either be the susceptibility SNPs in the *QKI* gene (which contributes to demyelination of the CNS myelin) or SNPs located in *MBP* gene (which cause the instability of the MBP protein). However, this study represents a preliminary screening of MBP protein and the exact mechanism of demyelination in normal orexin hypersomnia remains to be elucidated. It is important to perform additional experiments in a variety of sub-classified hypersomnias to evaluate the effectiveness of MBP as a biological marker to increase the clinical diagnosis specificity of *HLA-DQB1*06:02* negative hypersomnias.

In PART3, besides looking at single point analyses, I performed a pathway analysis and identified the glutamate metabolism to be significantly (P = 1.70E-03) related to the risk of EHS in Japanese population. After literature based categorizing, I focused on the neurotransmitter pathway within the glutamate metabolism pathway for further analysis. HPLC analysis (glutamate, glutamine and GABA) on 21 low orexin hypersomnias and 23 normal orexin hypersomnias showed that GABA level is significantly increased in the low orexin hypersomnias group (P = 3.15E-03) parallel with the more severe EDS symptom in narcolepsy with cataplexy. However, this study represent a preliminarily screening of the level of neurotransmitters in these two hypersomnia groups, further studies should be performed on healthy individual CSF to measure the level of these neurotransmitters for comparison purposes.

All the findings above must improve the understanding of genetic variations' contribution to the pathogenesis for EHS.

6. Acknowledgement

I owe my deepest gratitude to Professor Katsushi Tokunaga (Department of Human Genetics, Graduate School of Medicine, the University of Tokyo), for supervising my study and providing me opportunity to complete my master degree study in the University of Tokyo. I am heartily thankful to my supervisor, Dr. 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, Dr. Hiromi Toyoda, Dr. Yuki Hitomi , Dr. Hiroko Miyadera and Dr. Nao Nishida for their enlightening suggestions and insightful suggestion to this study. I would like to show my gratitude to Dr.Yutaka Honda, Dr. Makoto Honda (Japan Somnology Centre, Neuropsychiatric Research Institute, Tokyo, Japan), Professor Emmanuel Mignot, Dr.Ling Lin, Dr. Minae Kawashima (Stanford University), associate professor Takashi Kanbayashi and Dr.Aya Imanishi (Akita University). A special thank you is dedicated to Dr.Tohru Kodama (Tokyo Metropolitan Institute of Medical Science) for his guide in HPLC technology and I am indebted to my many of colleagues in our laboratory.

References

1. Hobson JA: **Sleep**. New York: Scientific American Library : Distributed by W.H. Freeman; 1989.
2. ASERINSKY E, KLEITMAN N: **Regularly occurring periods of eye motility, and concomitant phenomena, during sleep**. *Science* 1953, **118**(3062):273-274.
3. ASERINSKY E, KLEITMAN N: **Two types of ocular motility occurring in sleep**. *J Appl Physiol* 1955, **8**(1):1-10.
4. Kryger MH, Roth T, Dement WC: **Principles and practice of sleep medicine**, 5th edn. Philadelphia, PA: Saunders/Elsevier; 2011.
5. Agnew HW, Webb WB, Williams RL: **The first night effect: an EEG study of sleep**. *Psychophysiology* 1966, **2**(3):263-266.
6. Czeisler CA, Weitzman E, Moore-Ede MC, Zimmerman JC, Knauer RS: **Human sleep: its duration and organization depend on its circadian phase**. *Science* 1980, **210**(4475):1264-1267.
7. Bixler EO, Kales A, Soldatos CR, Kales JD, Healey S: **Prevalence of sleep disorders in the Los Angeles metropolitan area**. *Am J Psychiatry* 1979, **136**(10):1257-1262.
8. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S: **Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness**. *Sleep* 1986, **9**(4):519-524.
9. Johns MW: **Sleepiness in different situations measured by the Epworth Sleepiness Scale**. *Sleep* 1994, **17**(8):703-710.

10. American Academy of Sleep Medicine.: **The international classification of sleep disorders : diagnostic and coding manual**, 2nd edn. Westchester, Ill.: American Academy of Sleep Medicine; 2005.

11. Richardson GS, Carskadon MA, Flagg W, Van den Hoed J, Dement WC, Mitler MM: **Excessive daytime sleepiness in man: multiple sleep latency measurement in narcoleptic and control subjects**. *Electroencephalogr Clin Neurophysiol* 1978, **45**(5):621-627.

12. **The international classification of sleep disorders**. Rochester, MN: American Academy of Sleep Medicine; 2000.

13. Meier-Ewert K: **Sleep-wake disorders**. New York: Plenum Press; 1997.

14. Roth B: **Narcolepsy and hypersomnia: review and classification of 642 personally observed cases.** *Schweiz Arch Neurol Neurochir Psychiatr* 1976, **119**(1):31-41.

15. Mignot E: **Genetic and familial aspects of narcolepsy.** *Neurology* 1998, **50**(2 Suppl 1):S16-22.

16. Overeem S, Mignot E, van Dijk JG, Lammers GJ: **Narcolepsy: clinical features, new pathophysiologic insights, and future perspectives.** *J Clin Neurophysiol* 2001, **18**(2):78-105.

17. Dauvilliers Y, Arnulf I, Mignot E: **Narcolepsy with cataplexy.** *Lancet* 2007, **369**(9560):499-511.

18. Dauvilliers Y, Arnulf I, Mignot E: **Narcolepsy with cataplexy**. *Lancet* 2007, **369**(9560):499-511.

19. Mukai J, Inoue Y, Honda Y, Takahashi Y, Ishii A, Saitoh K, Nanba K: **Clinical characteristics of essential hypersomnia syndrome** *Sleep and Biological Rhythms* 2003, **1**:229-231.

20. 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, Risch N: **Complex HLA-DR and -DQ**

interactions confer risk of narcolepsy-cataplexy in three ethnic groups. *Am J Hum Genet* 2001, **68**(3):686-699.

21. Komada Y, Inoue Y, Mukai J, Shirakawa S, Takahashi K, Honda Y: **Difference in the characteristics of subjective and objective sleepiness between narcolepsy and essential hypersomnia.** *Psychiatry Clin Neurosci* 2005, **59**(2):194-199.

22. 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 GJ, Bouras C, Kucherlapati R, Nishino S, 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):991-997.

23. Thannickal TC, Moore RY, Nienhuis R, Ramanathan L, Gulyani S, Aldrich M, Cornford M, Siegel JM: **Reduced number of hypocretin neurons in human narcolepsy**. *Neuron* 2000, **27**(3):469-474.

24. Juji T, Satake M, Honda Y, Doi Y: **HLA antigens in Japanese patients with narcolepsy. All the patients were DR2 positive.** *Tissue Antigens* 1984, **24**(5):316-319.

25. Miyagawa T, Kawashima M, Nishida N, Ohashi J, Kimura R, Fujimoto A, Shimada M, Morishita S, Shigeta T, Lin L, Hong SC, Faraco J, Shin YK, Jeong JH, Okazaki Y, Tsuji S, Honda M, Honda Y, Mignot E, Tokunaga K: **Variant between CPT1B and CHKB associated with susceptibility to narcolepsy.** *Nat Genet* 2008, **40**(11):1324-1328.

26. Hallmayer J, Faraco J, Lin L, Hesselson S, Winkelmann J, Kawashima M, Mayer G, Plazzi G, Nevsimalova S, Bourgin P, Hong SC, Hong SS, Honda Y, Honda M, Högl B, Longstreth WT, Montplaisir J, Kemlink D, Einen M, Chen J, Musone SL, Akana M, Miyagawa T, Duan J, Desautels A, Erhardt C, Hesla PE, Poli F, Frauscher B, Jeong JH, Lee SP, Ton TG, Kvale M,

Kolesar L, Dobrovolná M, Nepom GT, Salomon D, Wichmann HE, Rouleau GA, Gieger C, Levinson DF, Gejman PV, Meitinger T, Young T, Peppard P, Tokunaga K, Kwok PY, Risch N, Mignot E: **Narcolepsy is strongly associated with the T-cell receptor alpha locus**. *Nat Genet* 2009, **41**(6):708-711.

27. Kornum BR, Kawashima M, Faraco J, Lin L, Rico TJ, Hesselson S, Axtell RC, Kuipers H, Weiner K, Hamacher A, Kassack MU, Han F, Knudsen S, Li J, Dong X, Winkelmann J, Plazzi G, Nevsimalova S, Hong SC, Honda Y, Honda M, Högl B, Ton TG, Montplaisir J, Bourgin P, Kemlink D, Huang YS, Warby S, Einen M, Eshragh JL, Miyagawa T, Desautels A, Ruppert E, Hesla PE, Poli F, Pizza F, Frauscher B, Jeong JH, Lee SP, Strohl KP, Longstreth WT, Kvale M, Dobrovolna M, Ohayon MM, Nepom GT, Wichmann HE, Rouleau GA, Gieger C, Levinson DF, Gejman PV, Meitinger T, Peppard P, Young T, Jennum P, Steinman L, Tokunaga K, Kwok PY, Risch N, Hallmayer J, Mignot E: **Common variants in P2RY11 are associated with narcolepsy**. *Nat Genet* 2011, **43**(1):66-71.

28. Dauvilliers Y, Montplaisir J, Cochen V, Desautels A, Einen M, Lin L, Kawashima M, Bayard S, Monaca C, Tiberge M, Filipini D, Tripathy A, Nguyen BH, Kotagal S, Mignot E: **Post-H1N1 narcolepsy-cataplexy**. *Sleep* 2010, **33**(11):1428-1430.

29. Dauvilliers Y, Arnulf I, Lecendreux M, Monaca Charley C, Franco P, Drouot X, d'Ortho MP, Launois S, Lignot S, Bourgin P, Nogues B, Rey M, Bayard S, Scholz S, Lavault S, Tubert-Bitter P, Saussier C, Pariente A, group N-Vs: **Increased risk of narcolepsy in children and adults after pandemic H1N1 vaccination in France**. *Brain* 2013, **136**(Pt 8):2486-2496.

30. Miyagawa T, Honda M, Kawashima M, Shimada M, Tanaka S, Honda Y, 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):63-65.

31. Honda Y, Juji T, Matsuki K, Naohara T, Satake M, Inoko H, Someya T, Harada S, Doi Y: **HLA-DR2 and Dw2 in narcolepsy and in other disorders of excessive somnolence without cataplexy.** *Sleep* 1986, **9**(1 Pt 2):133-142.

32. Miyagawa T, Honda M, Kawashima M, Shimada M, Tanaka S, Honda Y, 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):e5394.

33. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D: **Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs.** *Nat Genet* 2008, **40**(10):1253-1260.

34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**(3):559-575.

35. Miyagawa T, Nishida N, Ohashi J, Kimura R, Fujimoto A, Kawashima M, Koike A, Sasaki T, Tanii H, Otowa T, Momose Y, Nakahara Y, Gotoh J, Okazaki Y, Tsuji S, Tokunaga K: **Appropriate data cleaning methods for genome-wide association study.** *J Hum Genet* 2008, **53**(10):886-893.

36. Balding DJ: **A tutorial on statistical methods for population association studies.** *Nat Rev Genet* 2006, **7**(10):781-791.

37. Tabangin ME, Woo JG, Martin LJ: **The effect of minor allele frequency on the likelihood of obtaining false positives.** *BMC Proc* 2009, **3 Suppl 7**:S41.

38. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**(2):263-265.

39. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET: **Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies.** *Bioinformatics* 2010, **26**(19):2474-2476.

40. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE: **Common regulatory variation impacts gene expression in a cell type-dependent manner.** *Science* 2009, **325**(5945):1246-1250.

41. Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulette CM, Denny TN, Goldstein DB: **Tissue-specific genetic control of splicing: implications for the study of complex traits.** *PLoS Biol* 2008, **6**(12):e1.

42. Pei YF, Li J, Zhang L, Papasian CJ, Deng HW: **Analyses and comparison of accuracy of different genotype imputation methods.** *PLoS One* 2008, **3**(10):e3551.

43. Pei YF, Zhang L, Li J, Deng HW: **Analyses and comparison of imputation-based association methods.** *PLoS One* 2010, **5**(5):e10827.

44. Scheet P, Stephens M: **A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.** *Am J Hum Genet* 2006, **78**(4):629-644.

45. Marchini J, Howie B, Myers S, McVean G, Donnelly P: **A new multipoint method for genome-wide association studies by imputation of genotypes.** *Nat Genet* 2007, **39**(7):906-913.

46. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: **MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.** *Genet Epidemiol* 2010, **34**(8):816-834.

47. Anderson CA, Pettersson FH, Barrett JC, Zhuang JJ, Ragoussis J, Cardon LR, Morris AP: **Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms.** *Am J Hum Genet* 2008, **83**(1):112-119.

48. Marchini J, Howie B: **Genotype imputation for genome-wide association studies.** *Nat Rev Genet* 2010, **11**(7):499-511.

49. Consortium IH: **A haplotype map of the human genome.** *Nature* 2005, **437**(7063):1299-1320.

50. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: **LocusZoom: regional visualization of genome-wide association scan results.** *Bioinformatics* 2010, **26**(18):2336-2337.

51. Nishino S, Ripley B, Overeem S, Nevsimalova S, Lammers GJ, Vankova J, Okun M, Rogers W, Brooks S, Mignot E: **Low cerebrospinal fluid hypocretin (Orexin) and altered energy homeostasis in human narcolepsy**. *Ann Neurol* 2001, **50**(3):381-388.

52. Wang K, Li M, Bucan M: **Pathway-based approaches for analysis of genomewide association studies**. *Am J Hum Genet* 2007, **81**(6):1278-1283.

53. Dinu I, Mahasirimongkol S, Liu Q, Yanai H, Sharaf Eldin N, Kreiter E, Wu X, Jabbari S, Tokunaga K, Yasui Y: **SNP-SNP interactions discovered by logic regression explain Crohn's disease genetics**. *PLoS One* 2012, **7**(10):e43035.

54. Segrè AV, Groop L, Mootha VK, Daly MJ, Altshuler D, Consortium D, investigators M: **Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits**. *PLoS Genet* 2010, **6**(8).

55. Kanehisa M: **Molecular network analysis of diseases and drugs in KEGG**. *Methods Mol Biol* 2013, **939**:263-275.

56. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 2000, **28**(1):27-30.

57. Aberg K, Saetre P, Lindholm E, Ekholm B, Pettersson U, Adolfsson R, Jazin E: **Human QKI, a new candidate gene for schizophrenia involved in myelination**. *Am J Med Genet B Neuropsychiatr Genet* 2006, **141B**(1):84-90.

58. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zöllner S, Schork NJ, Kelsoe JR: **Genome-wide association study of bipolar disorder in European American and African American individuals.** *Mol Psychiatry* 2009, **14**(8):755-763.

59. Lasky-Su J, Neale BM, Franke B, Anney RJ, Zhou K, Maller JB, Vasquez AA, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen HC, Taylor E, Daly M, Laird N, Lange C, Faraone SV: **Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations.** *Am J Med Genet B Neuropsychiatr Genet* 2008, **147B**(8):1345-1354.

60. Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, Radue EW, Lindberg RL, Uitdehaag BM, Johnson MR, Angelakopoulou A, Hall L, Richardson JC, Prinjha RK, Gass A,

Geurts JJ, Kragt J, Sombekke M, Vrenken H, Qualley P, Lincoln RR, Gomez R, Caillier SJ, George MF, Mousavi H, Guerrero R, Okuda DT, Cree BA, Green AJ, Waubant E, Goodin DS, Pelletier D, Matthews PM, Hauser SL, Kappos L, Polman CH, Oksenberg JR: **Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis.** *Hum Mol Genet* 2009, **18**(4):767-778.

61. Wang KS, Liu XF, Aragam N: **A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder.** *Schizophr Res* 2010, **124**(1-3):192-199.

62. Cabrita MA, Christofori G: **Sprouty proteins, masterminds of receptor tyrosine kinase signaling.** *Angiogenesis* 2008, **11**(1):53-62.

63. Brems H, Chmara M, Sahbatou M, Denayer E, Taniguchi K, Kato R, Somers R, Messiaen L, De Schepper S, Fryns JP, Cools J, Marynen P, Thomas G, Yoshimura A, Legius E: **Germline loss-of-function mutations in SPRED1 cause a neurofibromatosis 1-like phenotype.** *Nat Genet* 2007, **39**(9):1120-1126.

64. Williams JA, Su HS, Bernards A, Field J, Sehgal A: **A circadian output in Drosophila mediated by neurofibromatosis-1 and Ras/MAPK.** *Science* 2001, **293**(5538):2251-2256.

65. Evans BM: **Sleep, consciousness and the spontaneous and evoked electrical activity of the brain. Is there a cortical integrating mechanism?** *Neurophysiol Clin* 2003, **33**(1):1-10.

66. Bourgin P, Zeitzer JM, Mignot E: **CSF hypocretin-1 assessment in sleep and neurological disorders.** *Lancet Neurol* 2008, **7**(7):649-662.

67. Burlet S, Tyler CJ, Leonard CS: **Direct and indirect excitation of laterodorsal tegmental neurons by Hypocretin/Orexin peptides: implications for wakefulness and narcolepsy.** *J Neurosci* 2002, **22**(7):2862-2872.

68. Pasmant E, Sabbagh A, Hanna N, Masliah-Planchon J, Jolly E, Goussard P, Ballerini P, Cartault F, Barbarot S, Landman-Parker J, Soufir N, Parfait B, Vidaud M, Wolkenstein P, Vidaud D, France RN: **SPRED1 germline mutations caused a neurofibromatosis type 1 overlapping phenotype.** *J Med Genet* 2009, **46**(7):425-430.

69. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre DJ, MacLean AW, St CD, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N, Consortium WTCC: **Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder.** *Nat Genet* 2008, **40**(9):1056-1058.

70. Wu D, Govindasamy L, Lian W, Gu Y, Kukar T, Agbandje-McKenna M, McKenna R: **Structure of human carnitine acetyltransferase. Molecular basis for fatty acyl transfer**. *J Biol Chem* 2003, **278**(15):13159-13165.

71. van der Leij FR, Huijkman NC, Boomsma C, Kuipers JR, Bartelds B: **Genomics of the human carnitine acyltransferase genes.** *Mol Genet Metab* 2000, **71**(1-2):139-153.

72. Tafti M, Petit B, Chollet D, Neidhart E, de Bilbao F, Kiss JZ, Wood PA, Franken P: **Deficiency in short-chain fatty acid beta-oxidation affects theta oscillations during sleep.** *Nat Genet* 2003, **34**(3):320-325.

73. Beal MF: **Bioenergetic approaches for neuroprotection in Parkinson's disease.** *Ann Neurol* 2003, **53 Suppl 3**:S39-47; discussion S47-38.

74. Montgomery SA, Thal LJ, Amrein R: **Meta-analysis of double blind randomized controlled clinical trials of acetyl-L-carnitine versus placebo in the treatment of mild cognitive impairment and mild Alzheimer's disease.** *Int Clin Psychopharmacol* 2003, **18**(2):61-71.

75. Miyagawa T, Kawamura H, Obuchi M, Ikesaki A, Ozaki A, Tokunaga K, Inoue Y, Honda M: **Effects of oral L-carnitine administration in narcolepsy patients: a randomized, double-blind, cross-over and placebo-controlled trial**. *PLoS One* 2013, **8**(1):e53707.

76. Miyagawa T, Miyadera H, Tanaka S, Kawashima M, Shimada M, Honda Y, Tokunaga K, Honda M: **Abnormally low serum acylcarnitine levels in narcolepsy patients**. *Sleep* 2011, **34**(3):349-353A.

77. Kanbayashi T, Inoue Y, Chiba S, Aizawa R, Saito Y, Tsukamoto H, Fujii Y, Nishino S, Shimizu T: **CSF hypocretin-1 (orexin-A) concentrations in narcolepsy with and without cataplexy and idiopathic hypersomnia**. *J Sleep Res* 2002, **11**(1):91-93.

78. Nishino S, Ripley B, Overeem S, Lammers GJ, Mignot E: **Hypocretin (orexin) deficiency in human narcolepsy**. *Lancet* 2000, **355**(9197):39-40.

79. Honda Y, Juji T: **HLA in narcolepsy**. Berlin ; New York: Springer-Verlag; 1988.

80. SIDMAN RL, DICKIE MM, APPEL SH: **MUTANT MICE (QUAKING AND JIMPY) WITH DEFICIENT MYELINATION IN THE CENTRAL NERVOUS SYSTEM**. *Science* 1964, **144**(3616):309-311.

81. Hardy RJ, Loushin CL, Friedrich VL, Chen Q, Ebersole TA, Lazzarini RA, Artzt K: **Neural cell type-specific expression of QKI proteins is altered in quakingviable mutant mice**. *J Neurosci* 1996, **16**(24):7941-7949.

82. Ebersole TA, Chen Q, Justice MJ, Artzt K: **The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins**. *Nat Genet* 1996, **12**(3):260-265.

83. Aberg K, Saetre P, Jareborg N, Jazin E: **Human QKI, a potential regulator of mRNA expression of human oligodendrocyte-related genes involved in schizophrenia**. *Proc Natl Acad Sci U S A* 2006, **103**(19):7482-7487.

84. Davis KL, Stewart DG, Friedman JI, Buchsbaum M, Harvey PD, Hof PR, Buxbaum J, Haroutunian V: **White matter changes in schizophrenia: evidence for myelin-related dysfunction**. *Arch Gen Psychiatry* 2003, **60**(5):443-456.

85. Hor H, Bartesaghi L, Kutalik Z, Vicário JL, de Andrés C, Pfister C, Lammers GJ, Guex N, Chrast R, Tafti M, Peraita-Adrados R: **A missense mutation in myelin oligodendrocyte glycoprotein as a cause of familial narcolepsy with cataplexy**. *Am J Hum Genet* 2011, **89**(3):474-479.

86. Siegel GJ: **Basic neurochemistry : molecular, cellular, and medical aspects**, 7th edn. Amsterdam ; Boston: Elsevier; 2006.

87. Lazzarini RA: **Myelin biology and disorders**. Sand Diego, Calif.: Elsevier Academic Press; 2004.

88. Compston A, Coles A: **Multiple sclerosis**. *Lancet* 2008, **372**(9648):1502-1517.

89. Greene DN, Schmidt RL, Wilson AR, Freedman MS, Grenache DG: **Cerebrospinal fluid myelin basic protein is frequently ordered but has little value: a test utilization study**. *Am J Clin Pathol* 2012, **138**(2):262-272.

90. Gallup AC, Gallup GG, Feo C: **Yawning, sleep, and symptom relief in patients with multiple sclerosis**. *Sleep Med* 2010, **11**(3):329-330.

91. Morelli A, Ravera S, Panfoli I: **Hypothesis of an energetic function for myelin**. *Cell Biochem Biophys* 2011, **61**(1):179-187.

92. Morelli A, Ravera S, Panfoli I: **Myelin sheath: a new possible role in sleep mechanism**. *Sleep Med* 2011, **12**(2):199.

93. Meldrum BS: **Glutamate as a neurotransmitter in the brain: review of physiology and pathology**. *J Nutr* 2000, **130**(4S Suppl):1007S-1015S.

94. McEntee WJ, Crook TH: **Glutamate: its role in learning, memory, and the aging brain**. *Psychopharmacology (Berl)* 1993, **111**(4):391-401.

95. Manev H, Favaron M, Guidotti A, Costa E: **Delayed increase of Ca2+ influx elicited by glutamate: role in neuronal death**. *Mol Pharmacol* 1989, **36**(1):106-112.

96. Pitt D, Werner P, Raine CS: **Glutamate excitotoxicity in a model of multiple sclerosis**. *Nat Med* 2000, **6**(1):67-70.

97. Pitt D, Nagelmeier IE, Wilson HC, Raine CS: **Glutamate uptake by oligodendrocytes: Implications for excitotoxicity in multiple sclerosis**. *Neurology* 2003, **61**(8):1113-1120.

98. Hynd MR, Scott HL, Dodd PR: **Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease**. *Neurochem Int* 2004, **45**(5):583-595.

99. Glantz LA, Gilmore JH, Lieberman JA, Jarskog LF: **Apoptotic mechanisms and the synaptic pathology of schizophrenia**. *Schizophr Res* 2006, **81**(1):47-63.

100. Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, Haroutunian V, Fienberg AA: **Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia**. *Proc Natl Acad Sci U S A* 2001, **98**(8):4746-4751.

101. Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB, Starkey M, Webster MJ, Yolken RH, Bahn S: **Oligodendrocyte dysfunction in schizophrenia and bipolar disorder**. *Lancet* 2003, **362**(9386):798-805.

102. Aston C, Jiang L, Sokolov BP: **Microarray analysis of postmortem temporal cortex from patients with schizophrenia**. *J Neurosci Res* 2004, **77**(6):858-866.

103. Hof PR, Haroutunian V, Friedrich VL, Byne W, Buitron C, Perl DP, Davis KL: **Loss and altered spatial distribution of oligodendrocytes in the superior frontal gyrus in schizophrenia**. *Biol Psychiatry* 2003, **53**(12):1075-1085.

104. Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI: **Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium**. *Schizophr Res* 2004, **67**(2-3):269-275.

105. Uranova N, Orlovskaya D, Vikhreva O, Zimina I, Kolomeets N, Vostrikov V, Rachmanova V: **Electron microscopy of oligodendroglia in severe mental illness**. *Brain Res Bull* 2001, **55**(5):597-610.

106. Sarchielli P, Greco L, Floridi A, Gallai V: **Excitatory amino acids and multiple sclerosis: evidence from cerebrospinal fluid**. *Arch Neurol* 2003, **60**(8):1082-1088.

107. Cianfoni A, Niku S, Imbesi SG: **Metabolite findings in tumefactive demyelinating lesions utilizing short echo time proton magnetic resonance spectroscopy**. *AJNR Am J Neuroradiol* 2007, **28**(2):272-277.

108. Newcombe J, Uddin A, Dove R, Patel B, Turski L, Nishizawa Y, Smith T: **Glutamate receptor expression in multiple sclerosis lesions**. *Brain Pathol* 2008, **18**(1):52-61.

109. Gottesmann C: **GABA mechanisms and sleep**. *Neuroscience* 2002, **111**(2):231-239.

110. Auger C, Attwell D: **Fast removal of synaptic glutamate by postsynaptic transporters**. *Neuron* 2000, **28**(2):547-558.

Tables

Table 1.1 Characteristics comparison between narcolepsy and EHS

Characteristics	**Narcolepsy**	**EHS**
Symptoms		
Cataplexy	Present	Absent
Excessive daytime sleepiness	> 3months	> 3months
Nocturnal sleep time	6> hour >10	6> hour >10
Arousal during sleep	Frequent	Frequent
Length of daytime nap	short (<30min)	short (<30min)
Genetic background		
HLA-DQB1*0602 association (12% in normal Japanese population)	100%	30-40%
Scientific measurement		
CSF orexin level	≤ 200 pg/mL	mostly > 200pg/mL
Sleep latencies on multiple sleep latency test (MSLT)	short	intermediate
Epworth Sleepiness Scale (ESS) score	high	moderate
sleep-onset rapid eye movement periods (SOREMS)	≥ 2	≥ 2
Treatment		
Dosages of psycho-stimulants for suppressing daytime sleepiness	High	Low
Treatment response of psycho-stimulants	Normal	Better

Table 4.1: List of pathways tested in the KEGG pathways [55, 56]

No	KEGG pathways
1	BILE_ACID_BIOSYNTHESIS
2	GANGLIOSIDE_BIOSYNTHESIS
3	ALANINE_AND_ASPARTATE_METABOLISM
4	PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM
5	ECMRECEPTOR_INTERACTION
6	GLYCINE,_SERINE_AND_THREONINE_METABOLISM
7	KERATAN_SULFATE_BIOSYNTHESIS
8	RIBOFLAVIN_METABOLISM
9	PROPANOATE_METABOLISM
10	ETHYLBENZENE_DEGRADATION
11	DORSOVENTRAL_AXIS_FORMATION
12	ASCORBATE_AND_ALDARATE_METABOLISM
13	OGLYCAN_BIOSYNTHESIS
14	PHENYLALANINE,_TYROSINE_AND_TRYPTOPHAN_BIOSYNTHESIS
15	SULFUR_METABOLISM
16	NOTCH_SIGNALING_PATHWAY
17	PANTOTHENATE_AND_COA_BIOSYNTHESIS
18	STARCH_AND_SUCROSE_METABOLISM
19	OXIDATIVE_PHOSPHORYLATION
20	REGULATION_OF_ACTIN_CYTOSKELETON
21	CUSTOM_CALCIUM_CHANNEL
22	HEDGEHOG_SIGNALING_PATHWAY
23	T_CELL_RECEPTOR_SIGNALING_PATHWAY
24	LONGTERM_POTENTIATION
25	AMINOSUGARS_METABOLISM
26	NEURODEGENERATIVE_DISORDERS

No	KEGG pathways
69	NGLYCAN_BIOSYNTHESIS
70	PHENYLALANINE_METABOLISM
71	CALCIUM_SIGNALING_PATHWAY
72	AMINOACYLTRNA_SYNTHETASES
73	ALKALOID_BIOSYNTHESIS_II
74	FC_EPSILON_RI_SIGNALING_PATHWAY
75	CHONDROITIN_SULFATE_BIOSYNTHESIS
76	SELENOAMINO_ACID_METABOLISM
77	AXON_GUIDANCE
78	FRUCTOSE_AND_MANNOSE_METABOLISM
79	CELL_ADHESION_MOLECULES_CAMS
80	CYSTEINE_METABOLISM
81	GLUTAMATE_METABOLISM
82	ARGININE_AND_PROLINE_METABOLISM
83	ANTIGEN_PROCESSING_AND_PRESENTATION
84	NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
85	RIBOSOME
86	ABC_TRANSPORTERS__GENERAL
87	TRYPTOPHAN_METABOLISM
88	PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS
89	GALACTOSE_METABOLISM
90	FOLATE_BIOSYNTHESIS
91	MATURITY_ONSET_DIABETES_OF_THE_YOUNG
92	SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT
93	MAPK_SIGNALING_PATHWAY
94	TIGHT_JUNCTION

27 UREA_CYCLE_AND_METABOLISM_OF_AMINO_GROUPS
28 GLYCOSPHINGOLIPID_METABOLISM
29 TYPE_II_DIABETES_MELLITUS
30 CAPROLACTAM_DEGRADATION
31 LONGTERM_DEPRESSION
32 PYRUVATE_METABOLISM
33 CYTOKINECYTOKINE_RECEPTOR_INTERACTION
34 RNA_POLYMERASE
35 APOPTOSIS
36 LINOLEIC_ACID_METABOLISM
37 COMPLEMENT_AND_COAGULATION_CASCADES
38 BUTANOATE_METABOLISM
39 CITRATE_CYCLE_TCA_CYCLE
40 GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM
41 PORPHYRIN_AND_CHLOROPHYLL_METABOLISM
42 GLYCOSYLPHOSPHATIDYLINOSITOLGPIANCHOR_BIOSYNTHESIS
43 CELL_CYCLE
44 ADHERENS_JUNCTION
45 ADIPOCYTOKINE_SIGNALING_PATHWAY
46 GLUTATHIONE_METABOLISM
47 BLOOD_GROUP_GLYCOLIPID_BIOSYNTHESISLACTOSERIES
48 REDUCTIVE_CARBOXYLATE_CYCLE_CO2_FIXATION
49 STREPTOMYCIN_BIOSYNTHESIS
50 GLOBOSIDE_METABOLISM
51 NGLYCAN_DEGRADATION
52 PYRIMIDINE_METABOLISM
53 LIMONENE_AND_PINENE_DEGRADATION
54 LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION
55 METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450
95 PROTEASOME
96 ATP_SYNTHESIS
97 FOCAL_ADHESION
98 METHIONINE_METABOLISM
99 DNA_POLYMERASE
100 INSULIN_SIGNALING_PATHWAY
101 NICOTINATE_AND_NICOTINAMIDE_METABOLISM
102 PURINE_METABOLISM
103 FATTY_ACID_ELONGATION_IN_MITOCHONDRIA
104 WNT_SIGNALING_PATHWAY
105 HEPARAN_SULFATE_BIOSYNTHESIS
106 TYPE_I_DIABETES_MELLITUS
107 FATTY_ACID_METABOLISM
108 PROTEIN_EXPORT
109 BENZOATE_DEGRADATION_VIA_COA_LIGATION
110 BLOOD_GROUP_GLYCOLIPID_BIOSYNTHESISNEOLACTOSERIES
111 GLYCEROLIPID_METABOLISM
112 C21STEROID_HORMONE_METABOLISM
113 ONE_CARBON_POOL_BY_FOLATE
114 NITROBENZENE_DEGRADATION
115 GAMMAHEXACHLOROCYCLOHEXANE_DEGRADATION
116 TGFBETA_SIGNALING_PATHWAY
117 UBIQUITIN_MEDIATED_PROTEOLYSIS
118 GAP_JUNCTION
119 VALINE,_LEUCINE_AND_ISOLEUCINE_DEGRADATION
120 B_CELL_RECEPTOR_SIGNALING_PATHWAY
121 CIRCADIAN_RHYTHM
122 METHANE_METABOLISM
123 JAKSTAT_SIGNALING_PATHWAY

56 NEUROACTIVE_LIGANDRECEPTOR_INTERACTION
57 ARACHIDONIC_ACID_METABOLISM
58 1_AND_2METHYLNAPHTHALENE_DEGRADATION
59 AMINOPHOSPHONATE_METABOLISM
60 INOSITOL_PHOSPHATE_METABOLISM
61 BETAALANINE_METABOLISM
62 ANDROGEN_AND_ESTROGEN_METABOLISM
63 BASAL_TRANSCRIPTION_FACTORS
64 TYROSINE_METABOLISM
65 CELL_COMMUNICATION
66 BIOSYNTHESIS_OF_STEROIDS
67 BISPHENOL_A_DEGRADATION
68 GLYCOSAMINOGLYCAN_DEGRADATION
124 CARBON_FIXATION
125 CUSTOM_CALCIUM_CHANNEL_A1
126 GLYCEROPHOSPHOLIPID_METABOLISM
127 HEMATOPOIETIC_CELL_LINEAGE
128 HISTIDINE_METABOLISM
129 NUCLEOTIDE_SUGARS_METABOLISM
130 NITROGEN_METABOLISM
131 GLYCOLYSIS__GLUCONEOGENESIS
132 LYSINE_DEGRADATION
133 VALINE,_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS
134 TOLLLIKE_RECEPTOR_SIGNALING_PATHWAY
135 PENTOSE_PHOSPHATE_PATHWAY

Table 4.2: List of pathways tested in the Ingenuity pathways (Ingenuity® Systems, www.ingenuity.com)

No	Ingenuity Pathway	No	Ingenuity Pathway
1	VEGF.Signaling	42	PPARa.RXRa.Activation
2	Leukocyte.Extravasation.Signaling	43	Fc.Epsilon.RI.Signaling
3	VDR.RXR.Activation	44	Xenobiotic.Metabolism.Signaling
4	T.Cell.Receptor.Signaling	45	Glucocorticoid.Receptor.Signaling
5	TR.RXR.Activation	46	Aryl.Hydrocarbon.Receptor.Signaling
6	LXR.RXR.Activation	47	PTEN.Signaling
7	Antigen.Presentation.Pathway	48	Sonic.Hedgehog.Signaling
8	Eicosanoid.Signaling	49	Natural.Killer.Cell.Signaling
9	Role.of.BRCA1.in.DNA.Damage.Response	50	Interferon.Signaling
10	LPS.IL-1.Mediated.Inhibition.of.RXR.Function	51	Hepatic.Cholestasis
11	NFKB.Signaling	52	BMP.Signaling.pathway
12	Neuregulin.Signaling	53	Chemokine.Signaling
13	PPAR.Signaling	54	Tight.Junction.Signaling
14	EGF.Signaling	55	Coagulation.System
15	Cell.Cycle.G2.M.DNA.Damage.Checkpoint.Regulation	56	Hypoxia.Signaling.in.the.Cardiovascular.System
16	IL-6.Signaling	57	Integrin.Signaling
17	IL-2.Signaling	58	Toll-like.Receptor.Signaling
18	GM-CSF.Signaling	59	Erythropoietin.Signaling
19	Mitochondrial.Dysfunction	60	Phototransduction.Pathway
20	JAK.Stat.Signaling	61	Serotonin.Receptor.Signaling
21	Calcium.Signaling	62	PDGF.Signaling
22	G-Protein.Coupled.Receptor.Signaling	63	Ephrin.Receptor.Signaling
23	IGF-1.Signaling	64	Protein.Ubiquitination.Pathway
24	Complement.System	65	Cell.Cycle.G1.S.Checkpoint.Regulation
25	Nitric.Oxide.Signaling.in.the.Cardiovascular.System	66	Hepatic.Fibrosis.Hepatic.Stellate.Cell.Activation
26	Estrogen.Receptor.Signaling	67	RAR.Activation

27 Ceramide.Signaling
28 Amyotrophic.Lateral.Sclerosis.Signaling
29 Huntingtons.Disease.Signaling
30 NRF2-mediated.Oxidative.Stress.Response
31 Neurotrophin.TRK.Signaling
32 B.Cell.Receptor.Signaling
33 FGF.Signaling
34 Axonal.Guidance.Signaling
35 Apoptosis.Signaling
36 Insulin.Receptor.Signaling
37 PI3K.AKT.Signaling
38 Death.Receptor.Signaling
39 TGF-beta.Signaling
40 p38.MAPK.Signaling
41 Endoplasmic.Reticulum.Stress.Pathway
68 Dopamine.Receptor.Signaling
69 Actin.Cytoskeleton.Signaling
70 IL-4.Signaling
71 Acute.Phase.Response.Signaling
72 PXR.RXR.Activation
73 Clutamate.Receptor.Signaling
74 SAPK.JNK.Signaling
75 14-3-3-mediated.Signaling
76 p53.Signaling
77 IL-10.Signaling
78 Wnt.beta-catenin.Signaling
79 cAMP-mediated.Signaling
80 ERK.MAPK.Signaling
81 FXR.RXR.Activation

Table 5.1: Subject characteristics of GWAS for Japanese *HLA-DRB1*15:01-DQB1*06:02* negative EHS

Category	**Case**		**Control**	
Total number of samples, n (%)	125	(100)	562	(100)
Sex, n (%)				
Male	64	(51.2)	280	(49.8)
Female	61	(48.8)	282	(50.2)

Table 5.2: Number of SNPs excluded after data cleaning

Data cleaning criteria	**Number of SNPs**
SNP Call Rate < 0.99	139,497
HWE < 0.001 in controls	8,171
MAF < 0.05	212,041
Visual cluster removal	1
Total SNPs remaining	508,366

Table 5.3 List of SNPs that show associations with increase risk of *HLA-DQB1*06:02* negative EHS in the Japanese population

CHR	SNP	Risk	RAF		Allelic				Dominant				Recessive				Pmin	Nearby Gene
		Allele	Case	Control	P-values	OR	L95	U95	P-values	OR	L95	U95	P-values	OR	L95	U95		
2	rs16826005	G	0.472	0.319	3.97E-06	1.89	1.43	2.50	7.51E-03	1.74	1.16	2.62	1.02E-07	3.52	2.17	5.70	1.02E-07	*NCKAP5*
2	rs12471007*#	C	0.456	0.312	1.33E-05	1.86	1.41	2.46	1.86E-02	1.72	1.15	2.56	1.09E-07	4.11	2.55	6.60	1.09E-07	*NCKAP5*
15	rs11854769*	T	0.316	0.177	6.69E-07	2.27	1.64	3.14	7.15E-06	2.43	1.64	3.59	3.61E-04	4.07	1.78	9.31	6.69E-07	*SPRED1*
15	rs2174009	C	0.335	0.193	9.34E-07	2.24	1.62	3.10	1.02E-05	2.40	1.62	3.56	2.87E-04	3.90	1.78	8.56	9.34E-07	*SPRED1*
15	rs16966389	G	0.804	0.642	7.52E-07	2.28	1.63	3.20	6.56E-07	2.73	1.82	4.10	1.07E-02	2.91	1.24	6.86	6.56E-07	*SPRED1*
2	rs359268	C	0.584	0.431	1.16E-05	1.89	1.42	2.51	1.47E-02	1.79	1.12	2.88	1.00E-06	2.81	1.84	4.30	1.00E-06	*BCL11A*
15	rs2134333	A	0.332	0.194	1.99E-06	2.19	1.58	3.02	2.05E-05	2.32	1.57	3.44	3.52E-04	3.83	1.75	8.40	1.99E-06	*SPRED1*
5	rs7725217	C	0.352	0.223	1.96E-05	2.00	1.46	2.74	1.08E-06	2.66	1.78	3.98	4.08E-01	1.41	0.62	3.19	1.08E-06	*TAS2R1*
9	rs10988217*	G	0.320	0.233	4.05E-03	1.52	1.13	2.04	2.55E-01	1.25	0.85	1.85	3.43E-06	3.85	2.11	7.04	3.43E-06	*PPP2R4/CRAT*
2	rs2043234	T	0.359	0.224	8.22E-06	1.95	1.45	2.64	1.15E-03	1.91	1.29	2.83	4.46E-06	4.05	2.14	7.66	4.46E-06	*NCKAP5*

RAF, risk allele frequency; OR, odd ratio; L95, U95, lower and upper confidence limits; Pmin, minimum *P* -value among three genetic models; NA: not applicable

Recessive model is calculated under risk allele homozygotes versus (heterozygous and non-risk homozygotes)

Dominant model is calculated under (risk allele homozygotes and heterozygotes) versus non-risk homozygotes

*SNP count was reconfirmed by Taqman platform, # SNP count adjusted for taqman platform

Peta odd ratio was used to calculate odd ratio for any contingency column with counting of 0

Table 5.4: List of GC content ratio across 22 autosomes, X, Y and mitocondrial calculated based on Japanese HapMap data build 37

Chromosome number	GC content ratio
1	0.42
2	0.40
3	0.40
4	0.38
5	0.40
6	0.40
7	0.41
8	0.40
9	0.41
10	0.42
11	0.42
12	0.41
13	0.39
14	0.41
15	0.42
16	0.45
17	0.46
18	0.40
19	**0.48**
20	0.44
21	0.41
22	0.48
X	0.39
Y	0.39
Mitochondria	0.44

Table 5.5 List of SNPs of interest for *HLA-DQB1*06:02* negative hypersomnias samples in Caucasian population

SNP(A/B)	risk allele	RAF		HWE P-values		Chi-square 2-tailed P-values			OR	L95	U95	Nearest Gene
		Control	Case	Control	Case	Allelic	Dominant	Recessive				
rs11854769 (C/T)	C	0.731	0.732	0.369	0.263	9.70E-01	3.00E-01	6.20E-01	1.00	0.81	1.23	***SPRED1***
rs16966290(C/A)	A	0.885	0.903	0.947	0.286	2.30E-01	7.70E-01	1.70E-01	1.20	0.88	1.63	***SPRED1***
rs12471007(C/G)	G	0.035	0.040	0.179	0.501	5.50E-01	4.40E-01	4.50E-01	1.15	0.72	1.85	***NCKAP5***
rs10988217(G/A)	G	0.597	0.648	0.581	0.456	2.51E-02	3.00E-02	1.16E-01	1.25	1.03	1.52	***CRAT***

RAF: Risk allele frequency, HWE: Hardy-Weinberg Equilibrium, OR: Odds ratio, L95, U95: lower and upper limits of confidence interval at 95%

Recessive model is calculated under risk allele homozygotes versus (heterozygous and non-risk homozygotes)

Dominant model is calculated under (risk allele homozygotes and heterozygotes) versus non-risk homozygotes

Table 5.6 GWAS and comparative studies for rs10988217 in Japanese and Caucasian populations

Subjects	Case count				Control count				AF (Case)		AF (Control)		Chi-Square 2-tailed P-value			P_{min}	OR	L95	U95
	GG	AG	AA	Total	GG	AG	AA	Total	G	A	G	A	Allelic	Dominant	Recessive				
Japanese *HLA-DQB1*06:02* negative EHS GWAS	21	38	66	125	28	206	328	562	0.320	0.680	0.234	0.770	4.05E-03	2.55E-01	3.43E-06	3.43E-06	1.52	1.13	2.04
Caucasian *HLA-DQB1*06:02* negative hypersomnia	108	126	30	264	632	836	292	1760	0.648	0.352	0.597	0.403	2.51E-02	3.00E-02	1.16E-01	2.51E-02	1.25	1.03	1.52
Japanese narcolepsy	37	161	211	409	92	590	878	1560	0.287	0.713	0.250	0.750	2.22E-02	8.93E-02	2.20E-02	2.20E-02	1.22	1.03	1.45
Caucasian narcolepsy	439	503	151	1093	519	625	235	1379	0.632	0.368	0.600	0.400	3.89E-02	2.82E-02	2.00E-01	2.82E-02	1.13	1.01	1.27

AF:allele frequency; OR:Odds Ratio; L95,U95: lower and upper limits of confidence interval at 95%; Pmin: minimum P-value among three genetic models
Recessive model is calculated under risk allele homozygotes versus (heterozygous and non-risk homozygotes)
Dominant model is calculated under (risk allele homozygotes and heterozygotes) versus non-risk homozygotes

Table 5.7: GWAS results of narcolepsy with cataplexy, *HLA-DQB1*06:02* negative EHS and case-case comparison between narcolepsy with cataplexy and *HLA-DQB1*06:02* negative EHS

CHR	SNP	P_{TREND}			Gene Symbol	Gene Distance	Location
		Narcolepsy vs EHS	Narcolepsy GWAS	EHS GWAS			
22	rs2051582	9.13E-07	9.02E-02	8.06E-05	*IL2RB*	0	intron
1	rs945321	1.78E-06	2.89E-02	3.58E-04	*NPHP4*	0	intron
1	rs1695621	2.17E-06	2.62E-02	4.06E-04	*NPHP4*	0	intron
12	rs7306968	2.52E-06	8.24E-04	7.07E-03	*SRRM4*	8755	upstream
18	rs4890893	2.74E-06	4.74E-02	1.95E-04	*MBP*	0	intron
15	rs16966389	2.87E-06	9.71E-01	1.01E-06	*SPRED1*	80178	upstream
1	rs945323	2.96E-06	2.77E-02	5.47E-04	*NPHP4*	0	intron
2	rs16826005	7.38E-06	8.02E-01	5.82E-06	*NCKAP5*	0	intron
12	rs4767735	8.24E-06	8.08E-04	1.58E-02	*SRRM4*	5598	upstream
2	rs7567323	8.30E-06	2.46E-01	8.56E-05	*LOC150568*	12018	upstream
6	rs10945915	7.16E-05	0.7797	1.40E-05	*QKI*	0	intron

Table 5.8: SNP markers of interest in *HLA-DQB1*06:02* negative EHS, narcolepsy with cataplexy and *HLA-DQB1*06:02* negative EHS versus narcolepsy with cataplexy

	rs4890893 (*MBP*)	rs9283864 (*QKI*)	rs10945915 (*QKI*)	rs9365600 (*QKI*)	rs767380 (*QKI*)
EHS	T	C	A	T	C
Freq controls (n)	0.75 (840)	0.51 (570)	0.50 (561)	0.50 (567)	0.50 (567)
Freq cases (n)	0.86 (215)	0.66 (165)	0.65 (162)	0.65 (163)	0.65 (163)
PTREND	1.95E-04	7.32E-06	1.40E-05	1.49E-05	1.65E-05
OR (95% CI)	2.05 (1.40-3.00)	1.89 (1.40-2.54)	1.85 (1.39-2.46)	1.84 (1.37-2.47)	1.83 (1.37-2.47)
Narcolepsy with cataplexy	T	C	A	T	C
Freq controls (n)	0.75 (840)	0.51 (570)	0.50 (561)	0.50 (567)	0.50 (567)
Freq cases (n)	0.71 (599)	0.51 (428)	0.50 (425)	0.51 (427)	0.51 (433)
PTREND	4.74E-02	0.9997	0.802	0.9477	0.6927
OR (95% CI)	0.82 (0.67-1.00)	1.06 (0.88-1.27)	1.02 (0.85-1.23)	1.01 (0.84-1.21)	1.04 (0.86-1.24)
EHS vs Narcolepsy with cataplexy	T	C	A	T	C
Freq EHS (n)	0.86 (215)	0.66 (165)	0.65 (162)	0.65 (163)	0.65 (163)
Freq Narcolepsy (n)	0.71 (599)	0.51 (428)	0.50 (425)	0.51 (427)	0.51 (433)
PTREND	2.74E-06	2.22E-05	7.16E-05	4.67E-05	1.25E-04
OR (95% CI)	2.51 (1.71-3.70)	1.89 (1.39-2.56)	1.81 (1.35-2.42)	1.83 (1.35-2.48)	1.77 (1.31-2.40)

Notes.

Freq: Frequencies, PTREND: P value of trend model, OR: Odd ratio, CI: confidence interval

Table 5.9: Pathway analysis results using KEGG [55, 56] gene-sets as a reference. Norminal gene set enrichment analysis (GSEA) P-value with 95% confidence interval and false discovery rate (FDR) with 95% confidence interval was shown in the table

Gene-set	NOMINAL_GSEA_PVAL_95PERC_CUTOFF	FDR_95PERC_CUTOFF
GLUTAMATE_METABOLISM	1.70E-03	1.51E-01
TYROSINE_METABOLISM	1.47E-01	8.85E-01
APOPTOSIS	1.18E-01	9.28E-01
ETHYLBENZENE_DEGRADATION	1.89E-01	9.30E-01
HEMATOPOIETIC_CELL_LINEAGE	6.88E-01	9.34E-01
PROTEIN_EXPORT	4.65E-01	9.41E-01
LINOLEIC_ACID_METABOLISM	7.35E-01	9.43E-01
GLYCEROPHOSPHOLIPID_METABOLISM	6.80E-01	9.47E-01
BASAL_TRANSCRIPTION_FACTORS	7.72E-01	9.55E-01
LIMONENE_AND_PINENE_DEGRADATION	4.11E-01	9.56E-01
PENTOSE_PHOSPHATE_PATHWAY	7.30E-01	9.57E-01
VALINE,_LEUCINE_AND_ISOLEUCINE_DEGRADATION	7.30E-01	9.57E-01
CITRATE_CYCLE_TCA_CYCLE	4.01E-01	9.60E-01
BILE_ACID_BIOSYNTHESIS	2.70E-01	9.64E-01
CUSTOM_CALCIUM_CHANNEL	7.14E-01	9.64E-01
METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	6.91E-01	9.64E-01
CHONDROITIN_SULFATE_BIOSYNTHESIS	1.88E-01	9.65E-01
SNARE_INTERACTIONS_IN_VESICULAR_TRANSPORT	7.82E-01	9.67E-01
AMINOACYLTRNA_SYNTHETASES	1.47E-01	9.68E-01
LEUKOCYTE_TRANSENDOTHELIAL_MIGRATION	5.89E-01	9.69E-01
HEPARAN_SULFATE_BIOSYNTHESIS	1.00E+00	9.69E-01
CELL_COMMUNICATION	9.20E-01	9.70E-01
HISTIDINE_METABOLISM	6.86E-01	9.71E-01
RNA_POLYMERASE	7.08E-01	9.72E-01

NITROGEN_METABOLISM	6.77E-01	9.72E-01
BUTANOATE_METABOLISM	9.30E-01	9.73E-01
NEUROACTIVE_LIGANDRECEPTOR_INTERACTION	8.56E-01	9.76E-01
PYRIMIDINE_METABOLISM	2.43E-01	9.76E-01
UREA_CYCLE_AND_METABOLISM_OF_AMINO_GROUPS	6.94E-01	9.77E-01
GAMMAHEXACHLOROCYCLOHEXANE_DEGRADATION	6.74E-01	9.77E-01
BLOOD_GROUP_GLYCOLIPID_BIOSYNTHESISNEOLACTOSERIES	1.00E+00	9.77E-01
TIGHT_JUNCTION	6.30E-01	9.79E-01
PHENYLALANINE_METABOLISM	3.95E-01	9.79E-01
ECMRECEPTOR_INTERACTION	9.16E-01	9.79E-01
REGULATION_OF_ACTIN_CYTOSKELETON	2.11E-01	9.79E-01
BIOSYNTHESIS_OF_STEROIDS	1.00E+00	9.79E-01
FRUCTOSE_AND_MANNOSE_METABOLISM	6.24E-01	9.80E-01
GANGLIOSIDE_BIOSYNTHESIS	1.00E+00	9.81E-01
ADIPOCYTOKINE_SIGNALING_PATHWAY	6.18E-01	9.81E-01
STARCH_AND_SUCROSE_METABOLISM	3.53E-01	9.81E-01
TYPE_I_DIABETES_MELLITUS	1.00E+00	9.83E-01
NGLYCAN_BIOSYNTHESIS	5.81E-01	9.83E-01
BENZOATE_DEGRADATION_VIA_COA_LIGATION	9.05E-01	9.84E-01
PANTOTHENATE_AND_COA_BIOSYNTHESIS	7.03E-01	9.84E-01
CIRCADIAN_RHYTHM	1.00E+00	9.85E-01
GLOBOSIDE_METABOLISM	1.00E+00	9.85E-01
CYTOKINECYTOKINE_RECEPTOR_INTERACTION	7.41E-01	9.85E-01
AMINOSUGARS_METABOLISM	2.06E-01	9.85E-01
GLYCOSPHINGOLIPID_METABOLISM	5.74E-01	9.86E-01
DORSOVENTRAL_AXIS_FORMATION	1.00E+00	9.86E-01
CELL_CYCLE	8.90E-01	9.87E-01
PURINE_METABOLISM	5.72E-01	9.88E-01

BISPHENOL_A_DEGRADATION	1.00E+00	9.89E-01
TRYPTOPHAN_METABOLISM	7.64E-01	9.89E-01
NGLYCAN_DEGRADATION	1.00E+00	9.89E-01
NEURODEGENERATIVE_DISORDERS	2.40E-01	9.90E-01
SULFUR_METABOLISM	1.00E+00	9.90E-01
GLYCOSYLPHOSPHATIDYLINOSITOLGPIANCHOR_BIOSYNTHESIS	1.00E+00	9.90E-01
FATTY_ACID_METABOLISM	3.65E-01	9.90E-01
OGLYCAN_BIOSYNTHESIS	1.00E+00	9.91E-01
NICOTINATE_AND_NICOTINAMIDE_METABOLISM	9.48E-01	9.92E-01
HEDGEHOG_SIGNALING_PATHWAY	1.00E+00	9.92E-01
AXON_GUIDANCE	9.88E-01	9.92E-01
BLOOD_GROUP_GLYCOLIPID_BIOSYNTHESISLACTOSERIES	1.00E+00	9.93E-01
CALCIUM_SIGNALING_PATHWAY	3.13E-01	9.93E-01
ANDROGEN_AND_ESTROGEN_METABOLISM	6.01E-01	9.93E-01
GLYCOSAMINOGLYCAN_DEGRADATION	1.00E+00	9.93E-01
FC_EPSILON_RI_SIGNALING_PATHWAY	8.56E-01	9.94E-01
TGFBETA_SIGNALING_PATHWAY	9.01E-01	9.94E-01
B_CELL_RECEPTOR_SIGNALING_PATHWAY	8.43E-01	9.94E-01
PROTEASOME	1.00E+00	9.94E-01
UBIQUITIN_MEDIATED_PROTEOLYSIS	5.89E-01	9.94E-01
ALANINE_AND_ASPARTATE_METABOLISM	9.95E-02	9.95E-01
LONGTERM_POTENTIATION	8.47E-01	9.95E-01
ADHERENS_JUNCTION	8.73E-01	9.95E-01
MATURITY_ONSET_DIABETES_OF_THE_YOUNG	6.55E-01	9.95E-01
GAP_JUNCTION	9.31E-01	9.96E-01
RIBOSOME	9.82E-01	9.96E-01
NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY	5.73E-01	9.97E-01
COMPLEMENT_AND_COAGULATION_CASCADES	5.50E-01	9.98E-01

INOSITOL_PHOSPHATE_METABOLISM	9.56E-01	9.99E-01
PHOSPHATIDYLINOSITOL_SIGNALING_SYSTEM	8.27E-01	1.00E+00
GLYCINE,_SERINE_AND_THREONINE_METABOLISM	6.12E-01	1.00E+00
KERATAN_SULFATE_BIOSYNTHESIS	5.29E-01	1.00E+00
RIBOFLAVIN_METABOLISM	5.66E-01	1.00E+00
PROPANOATE_METABOLISM	8.36E-01	1.00E+00
ASCORBATE_AND_ALDARATE_METABOLISM	1.00E+00	1.00E+00
PHENYLALANINE,_TYROSINE_AND_TRYPTOPHAN_BIOSYNTHESIS	4.10E-01	1.00E+00
NOTCH_SIGNALING_PATHWAY	1.05E-01	1.00E+00
OXIDATIVE_PHOSPHORYLATION	2.89E-01	1.00E+00
T_CELL_RECEPTOR_SIGNALING_PATHWAY	4.67E-01	1.00E+00
TYPE_II_DIABETES_MELLITUS	3.32E-01	1.00E+00
CAPROLACTAM_DEGRADATION	5.12E-01	1.00E+00
LONGTERM_DEPRESSION	9.81E-01	1.00E+00
PYRUVATE_METABOLISM	1.00E+00	1.00E+00
GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM	1.00E+00	1.00E+00
PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	1.22E-01	1.00E+00
GLUTATHIONE_METABOLISM	5.31E-01	1.00E+00
REDUCTIVE_CARBOXYLATE_CYCLE_CO2_FIXATION	4.30E-01	1.00E+00
STREPTOMYCIN_BIOSYNTHESIS	4.00E-01	1.00E+00
ARACHIDONIC_ACID_METABOLISM	4.62E-01	1.00E+00
1_AND_2METHYLNAPHTHALENE_DEGRADATION	3.26E-01	1.00E+00
AMINOPHOSPHONATE_METABOLISM	2.48E-01	1.00E+00
BETAALANINE_METABOLISM	3.24E-01	1.00E+00
ALKALOID_BIOSYNTHESIS_II	2.75E-01	1.00E+00
SELENOAMINO_ACID_METABOLISM	4.93E-01	1.00E+00
CELL_ADHESION_MOLECULES_CAMS	1.06E-01	1.00E+00
CYSTEINE_METABOLISM	2.91E-01	1.00E+00

ARGININE_AND_PROLINE_METABOLISM	4.90E-01	1.00E+00
ANTIGEN_PROCESSING_AND_PRESENTATION	8.44E-01	1.00E+00
ABC_TRANSPORTERS__GENERAL	5.47E-01	1.00E+00
PENTOSE_AND_GLUCURONATE_INTERCONVERSIONS	1.00E+00	1.00E+00
GALACTOSE_METABOLISM	4.64E-01	1.00E+00
FOLATE_BIOSYNTHESIS	8.97E-02	1.00E+00
MAPK_SIGNALING_PATHWAY	9.26E-01	1.00E+00
ATP_SYNTHESIS	8.45E-01	1.00E+00
FOCAL_ADHESION	4.95E-01	1.00E+00
METHIONINE_METABOLISM	5.40E-01	1.00E+00
DNA_POLYMERASE	9.47E-02	1.00E+00
INSULIN_SIGNALING_PATHWAY	1.04E-01	1.00E+00
FATTY_ACID_ELONGATION_IN_MITOCHONDRIA	1.00E+00	1.00E+00
WNT_SIGNALING_PATHWAY	9.72E-01	1.00E+00
GLYCEROLIPID_METABOLISM	2.28E-01	1.00E+00
C21STEROID_HORMONE_METABOLISM	1.00E+00	1.00E+00
ONE_CARBON_POOL_BY_FOLATE	5.40E-01	1.00E+00
NITROBENZENE_DEGRADATION	5.62E-01	1.00E+00
METHANE_METABOLISM	1.00E+00	1.00E+00
JAKSTAT_SIGNALING_PATHWAY	3.62E-02	1.00E+00
CARBON_FIXATION	8.26E-02	1.00E+00
CUSTOM_CALCIUM_CHANNEL_A1	1.00E+00	1.00E+00
NUCLEOTIDE_SUGARS_METABOLISM	5.90E-01	1.00E+00
GLYCOLYSIS__GLUCONEOGENESIS	7.51E-02	1.00E+00
LYSINE_DEGRADATION	3.33E-01	1.00E+00
VALINE,_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS	1.00E+00	1.00E+00
TOLLLIKE_RECEPTOR_SIGNALING_PATHWAY	5.48E-01	1.00E+00

Table 5.10: Pathway analysis results using Ingenuity pathways (Ingenuity® Systems, www.ingenuity.com) gene-sets as a reference

Gene-set	NOMINAL_GSEA_PVAL_95PERC_CUTOFF	FDR_95PERC_CUTOFF
Aryl.Hydrocarbon.Receptor.Signaling	1.13E-02	6.83E-01
VDR.RXR.Activation	3.52E-01	7.23E-01
Actin.Cytoskeleton.Signaling	4.15E-01	7.36E-01
Phototransduction.Pathway	4.33E-01	7.38E-01
LXR.RXR.Activation	4.78E-01	7.44E-01
Acute.Phase.Response.Signaling	4.57E-01	7.45E-01
SAPK.JNK.Signaling	4.63E-01	7.47E-01
cAMP-mediated.Signaling	4.30E-01	7.53E-01
Neuregulin.Signaling	3.75E-01	7.60E-01
Apoptosis.Signaling	4.60E-01	7.61E-01
PTEN.Signaling	3.40E-01	7.68E-01
PPARa.RXRa.Activation	4.73E-01	7.68E-01
Clutamate.Receptor.Signaling	4.81E-01	7.69E-01
Serotonin.Receptor.Signaling	5.05E-01	7.70E-01
Wnt.beta-catenin.Signaling	3.59E-01	7.71E-01
Dopamine.Receptor.Signaling	5.73E-01	7.77E-01
Complement.System	6.58E-01	7.84E-01
Protein.Ubiquitination.Pathway	6.49E-01	7.85E-01
Fc.Epsilon.RI.Signaling	5.61E-01	7.87E-01
TR.RXR.Activation	3.13E-01	7.87E-01
JAK.Stat.Signaling	3.99E-01	7.89E-01
IL-4.Signaling	6.00E-01	7.90E-01
Chemokine.Signaling	6.89E-01	7.90E-01
PDGF.Signaling	6.84E-01	7.91E-01
Huntingtons.Disease.Signaling	5.48E-01	7.93E-01

NFKB.Signaling	5.99E-01	7.93E-01
IL-10.Signaling	3.41E-01	7.94E-01
Mitochondrial.Dysfunction	5.14E-01	7.96E-01
Cell.Cycle.G2.M.DNA.Damage.Checkpoint.Regulation	5.82E-01	7.98E-01
Hepatic.Fibrosis.Hepatic.Stellate.Cell.Activation	5.63E-01	7.99E-01
p53.Signaling	5.45E-01	8.00E-01
IGF-1.Signaling	6.25E-01	8.13E-01
Coagulation.System	3.37E-01	8.22E-01
BMP.Signaling.pathway	7.20E-01	8.26E-01
14-3-3-mediated.Signaling	3.02E-01	8.32E-01
Natural.Killer.Cell.Signaling	3.28E-01	8.37E-01
Interferon.Signaling	2.89E-01	8.41E-01
Insulin.Receptor.Signaling	2.41E-01	8.42E-01
Role.of.BRCA1.in.DNA.Damage.Response	7.60E-01	8.42E-01
IL-6.Signaling	7.43E-01	8.43E-01
ERK.MAPK.Signaling	7.30E-01	8.47E-01
LPS.IL-1.Mediated.Inhibition.of.RXR.Function	7.64E-01	8.66E-01
Amyotrophic.Lateral.Sclerosis.Signaling	7.78E-01	8.71E-01
Axonal.Guidance.Signaling	8.52E-01	8.76E-01
NRF2-mediated.Oxidative.Stress.Response	2.25E-01	8.78E-01
Toll-like.Receptor.Signaling	8.04E-01	8.78E-01
Integrin.Signaling	8.49E-01	8.81E-01
Nitric.Oxide.Signaling.in.the.Cardiovascular.System	1.00E+00	8.81E-01
Ephrin.Receptor.Signaling	2.66E-01	8.81E-01
G-Protein.Coupled.Receptor.Signaling	2.44E-01	8.83E-01
Sonic.Hedgehog.Signaling	1.00E+00	8.86E-01
Calcium.Signaling	1.00E+00	8.86E-01
Tight.Junction.Signaling	8.40E-01	8.91E-01

B.Cell.Receptor.Signaling	8.33E-01	8.92E-01
Ceramide.Signaling	1.00E+00	8.95E-01
PXR.RXR.Activation	8.98E-01	8.96E-01
Endoplasmic.Reticulum.Stress.Pathway	1.00E+00	9.01E-01
VEGF.Signaling	1.00E+00	9.01E-01
Antigen.Presentation.Pathway	1.00E+00	9.02E-01
Hypoxia.Signaling.in.the.Cardiovascular.System	1.00E+00	9.12E-01
PPAR.Signaling	1.00E+00	9.12E-01
Cell.Cycle.G1.S.Checkpoint.Regulation	1.00E+00	9.13E-01
GM-CSF.Signaling	1.00E+00	9.19E-01
FGF.Signaling	1.00E+00	9.19E-01
EGF.Signaling	1.00E+00	9.22E-01
FXR.RXR.Activation	9.29E-01	9.23E-01
p38.MAPK.Signaling	1.00E+00	9.26E-01
Hepatic.Cholestasis	9.46E-01	9.28E-01
RAR.Activation	1.89E-01	9.37E-01
Leukocyte.Extravasation.Signaling	2.00E-01	9.42E-01
Death.Receptor.Signaling	1.00E+00	9.56E-01
Glucocorticoid.Receptor.Signaling	9.51E-01	9.59E-01
T.Cell.Receptor.Signaling	8.31E-02	1.00E+00
Eicosanoid.Signaling	1.72E-01	1.00E+00
IL-2.Signaling	2.24E-01	1.00E+00
Estrogen.Receptor.Signaling	1.88E-01	1.00E+00
Neurotrophin.TRK.Signaling	1.59E-01	1.00E+00
PI3K.AKT.Signaling	4.95E-02	1.00E+00
TGF-beta.Signaling	1.77E-01	1.00E+00
Xenobiotic.Metabolism.Signaling	1.13E-01	1.00E+00
Erythropoietin.Signaling	1.65E-01	1.00E+00

Table 5.11: Result of GSEA-based pathway analysis after SNP pruning using Chinese and Japanese HapMap3 data

SNP pruning	**Gene-set**	**NOMINAL** **GSEA PVAL 95%**	**FDR** **95% CUTOFF**
Default	GLUTAMATE METABOLISM	1.70E-03	1.51E-01
CHB+JPT prune	GLUTAMATE METABOLISM	2.80E-03	1.76E-01

Table 5.12: Gene list in the glutamate metabolism pathway. (NA=not applicable)

Gene_Symbol	Gene_P-value	No of SNP	Best_SNP_rs	Best_SNP_pval
ABAT	6.04E-01	69	rs11862060	2.77E-02
CAD	9.45E-01	15	rs11691585	4.63E-01
CPS1	6.40E-01	76	rs2287604	4.17E-02
EPRS	8.18E-01	40	rs6683340	1.64E-01
GAD1	5.69E-01	38	rs12996521	4.41E-02
GAD2	8.29E-01	31	rs17667121	2.25E-01
GFPT1	1.50E-02	32	rs11683511	2.31E-03
GCLC	1.20E-02	59	rs513041	2.10E-04
GCLM	6.77E-01	37	rs1413899	6.34E-02
GLS	3.53E-02	41	rs10497709	3.86E-03
GLUD1	4.91E-02	10	rs17343324	1.10E-02
GLUD2	5.40E-02	12	rs7051454	1.65E-02
GLUL	1.95E-01	45	rs1925829	6.68E-03
GOT1	9.93E-03	48	rs3750897	2.68E-04
GOT2	4.08E-01	34	rs154469	3.98E-02
GPT	NA	0	NA	NA
GSR	2.06E-01	20	rs3779647	2.77E-02
GSS	1.23E-01	23	rs2281626	1.96E-02
PPAT	7.53E-01	18	rs17086853	1.85E-01
QARS	9.68E-01	1	rs6785832	8.01E-01
ALDH5A1	4.20E-01	58	rs807530	1.77E-02
ALDH4A1	3.80E-01	20	rs4920550	5.12E-02
GMPS	3.49E-01	19	rs6774542	5.86E-02
GFPT2	NA	0	NA	NA
GLS2	5.48E-01	22	rs2638319	7.87E-02
NADSYN1	5.43E-02	66	rs7940244	1.14E-03
NAGK	4.37E-03	28	rs2058552	4.53E-04
GNPNAT1	2.86E-02	20	rs2884345	4.90E-03
GPT2	9.50E-01	1	rs2034598	6.32E-01
ADC	1.50E-01	21	rs7552729	1.98E-02
EARS2	5.07E-01	20	rs3826263	1.11E-01

Figures

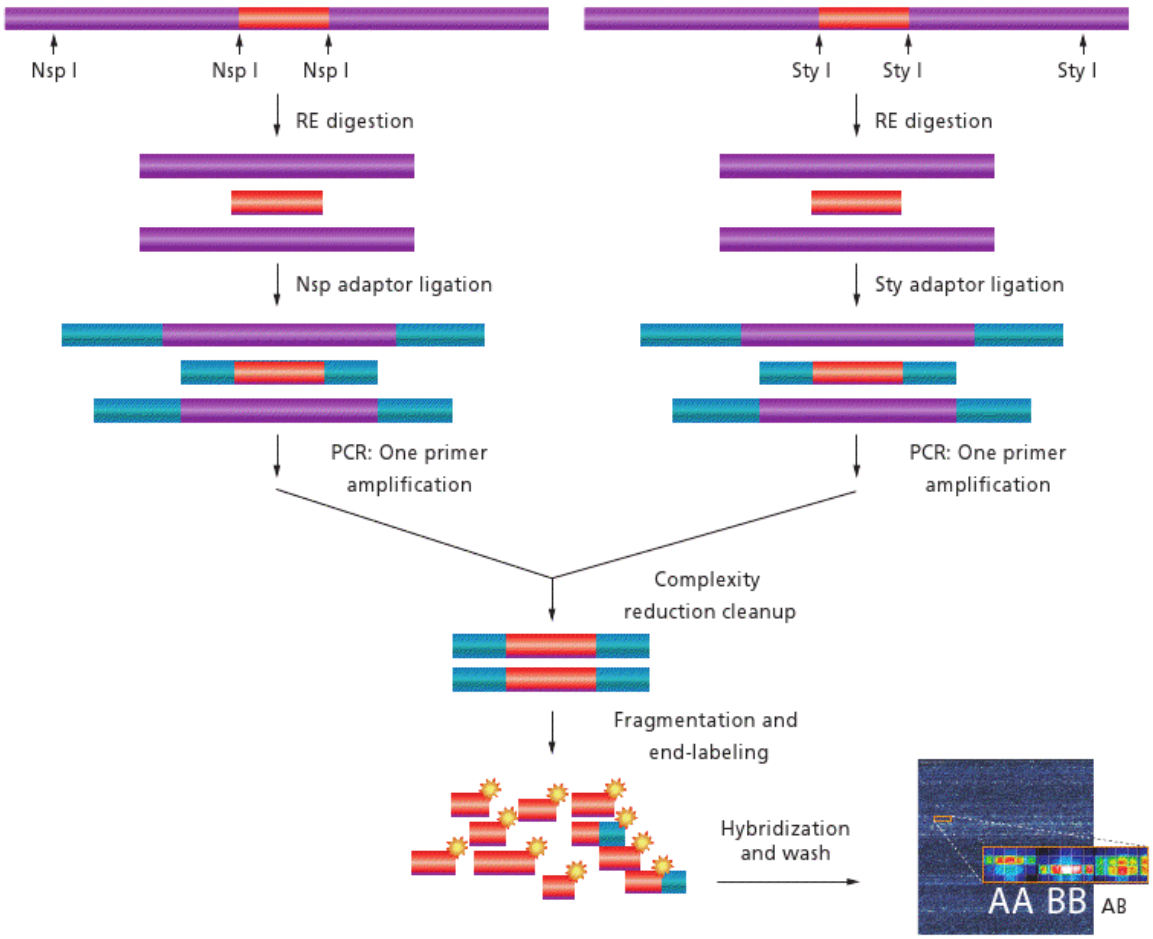
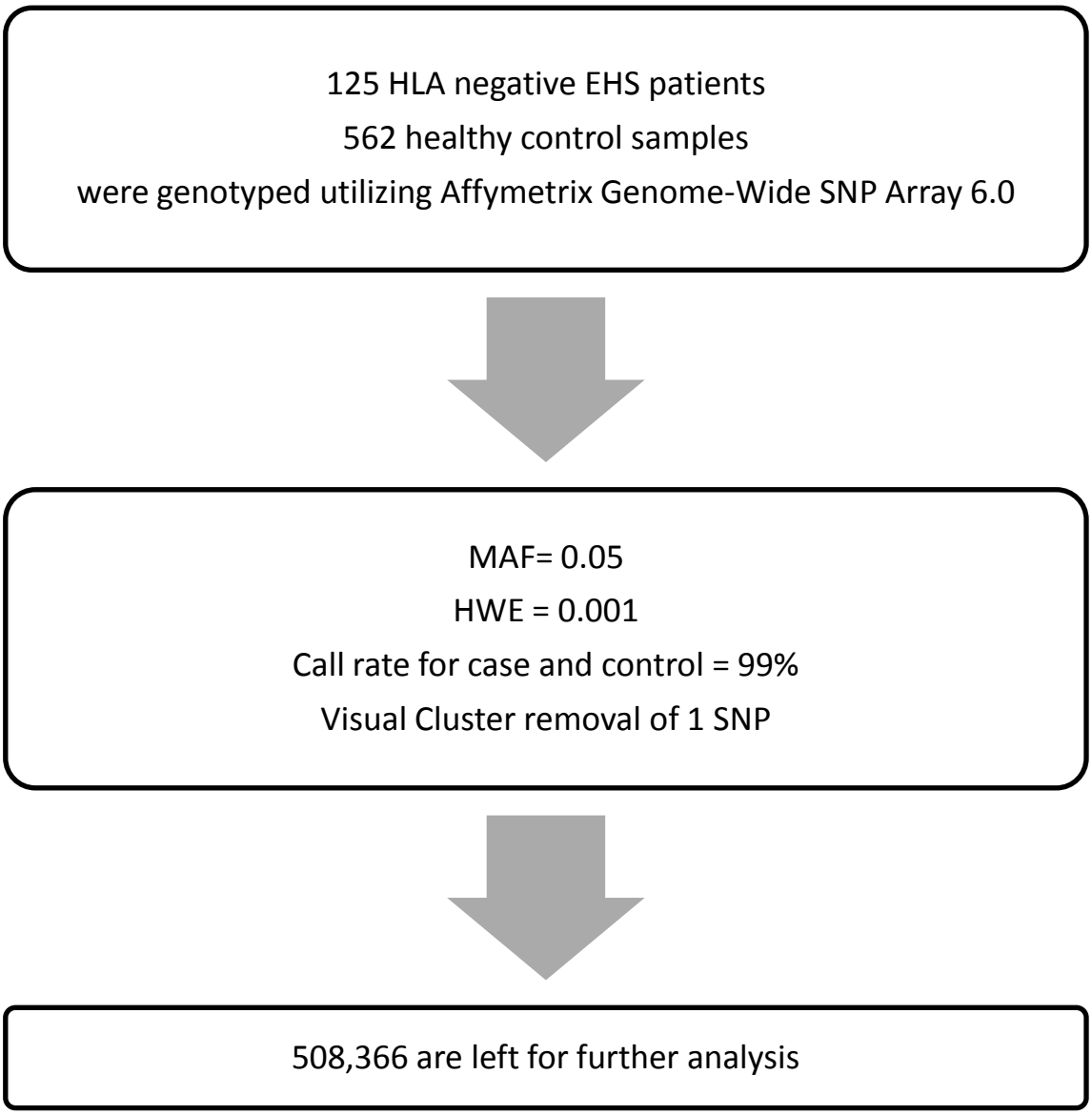


Figure 2.1: Schematic presentation of overview processes in Affymetrix Genome-Wide SNP Array 6.0. Figure modified from www.affymetrix.com



All the statistical analyses were performed using R statistical environment version 2.9.0 or PLINK 1.09.

Figure 2.2: Schematic presentation of overflow of statistical analysis for GWAS of Japanese *HLA-DQB1*06:02* negative EHS

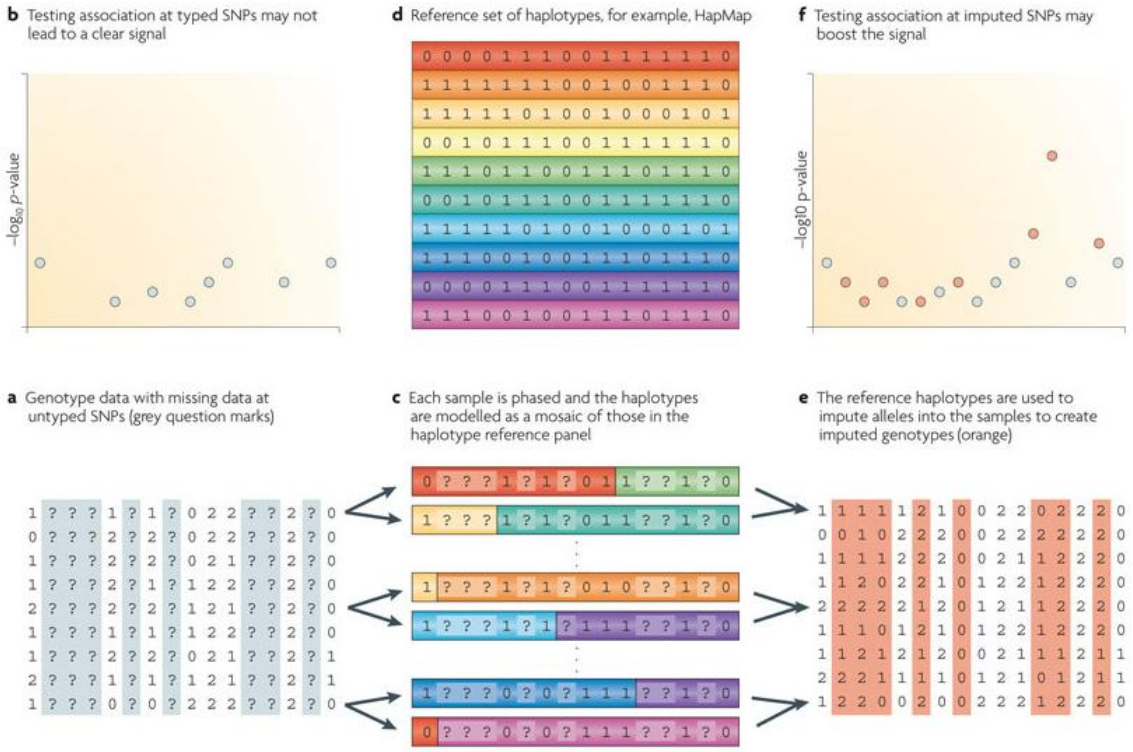


Figure 2.3: Illustration figure for imputation analysis for unrelated individuals. The raw data consist of un-typed SNPs or missing SNPs (part a). Association analysis on these markers alone will significantly decrease the power to detect true associated SNP (part b). The figure highlights three phased individuals (part c). These haplotypes are then compared with those of HapMap (for example: JPT+CHB HapMap II build 36) (part d). The phased study haplotypes have been coloured according to which reference haplotypes they match. Missing genotypes in the study sample are then imputed using those matching haplotypes in the reference set. MACH version 1.0 was used to estimate haplotypes, map crossover and error rates using 50 iterations of the Markov chain Monte Carlo algorithm were generated (part e). For the quality control of the imputed genotypes, I retained imputed genotypes with the estimated $r^2 > 0.3$. Imputed genotypes were re-analyzed by allelic, dominant, recessive models and Cochran-Armitage Trend Test utilizing PLINK 1.7[34]

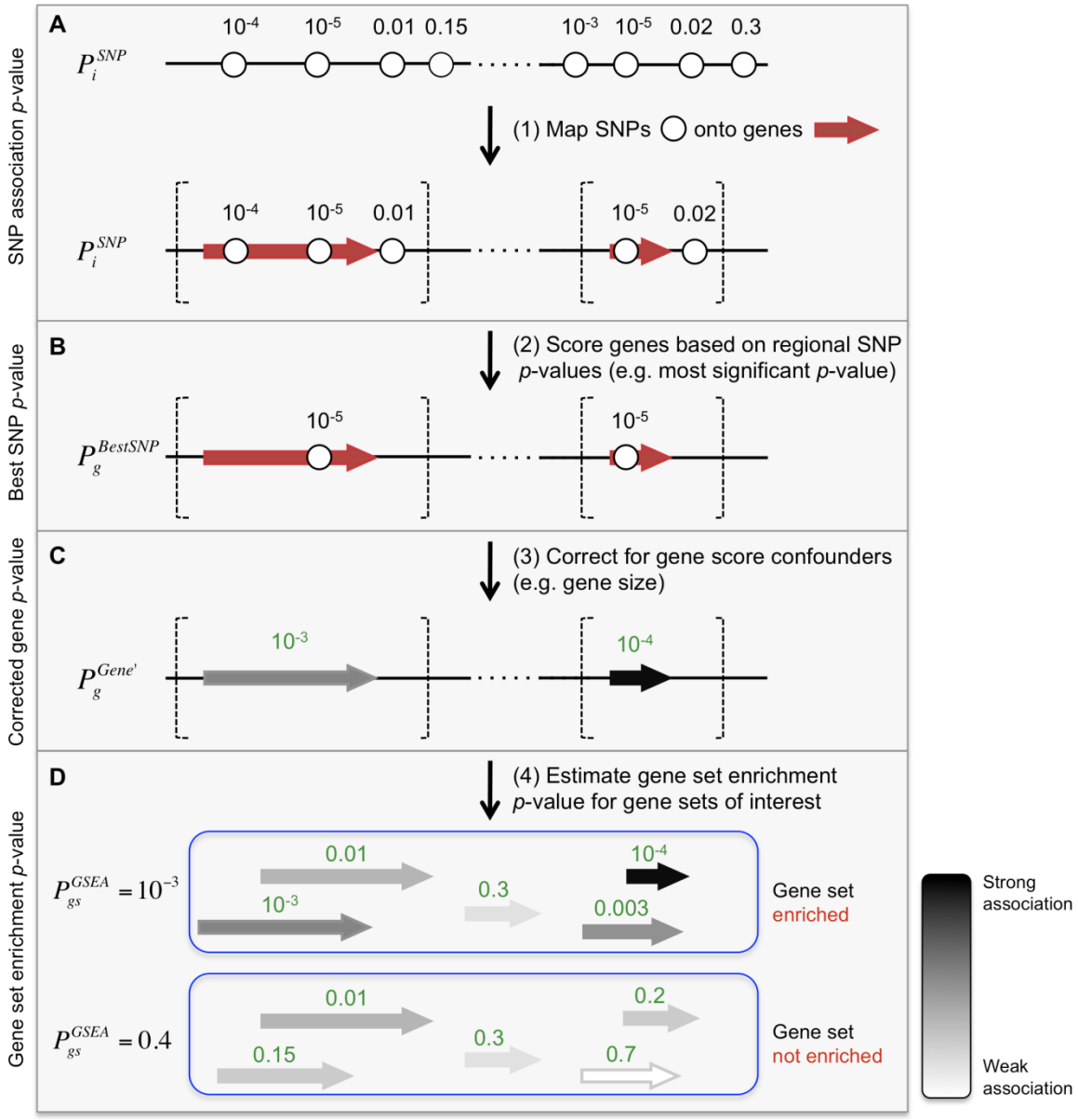
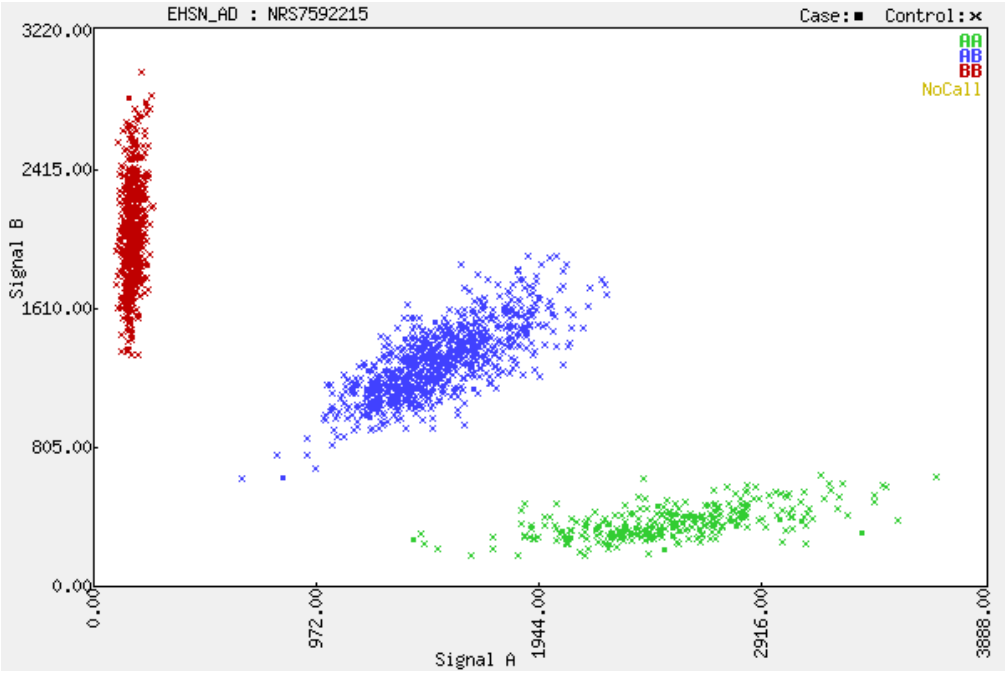


Figure 4.1: Description of Meta-Analysis Gene-set Enrichment of variant Associations (MAGENTA) method[54]. (A) Step 1: Map SNPs from GWAS onto genes. (B) Step 2: Score genes based on their local SNP (most significant P-value) (C) Step 3: Correct for confounding effects on the gene score (step-wise multivariate linear regression analysis) (D) Step 4: Calculate a gene set enrichment P-value for each biological pathway or gene set of interest. Non-parametric statistical test was applied to test whether for all genes in gene set are enriched for highly ranked gene scores more than would be expected by chance, compared to randomly sampled gene sets of identical size from the genome. Figure excerpt from Segre *et al.*,[54]

(a)



(b)

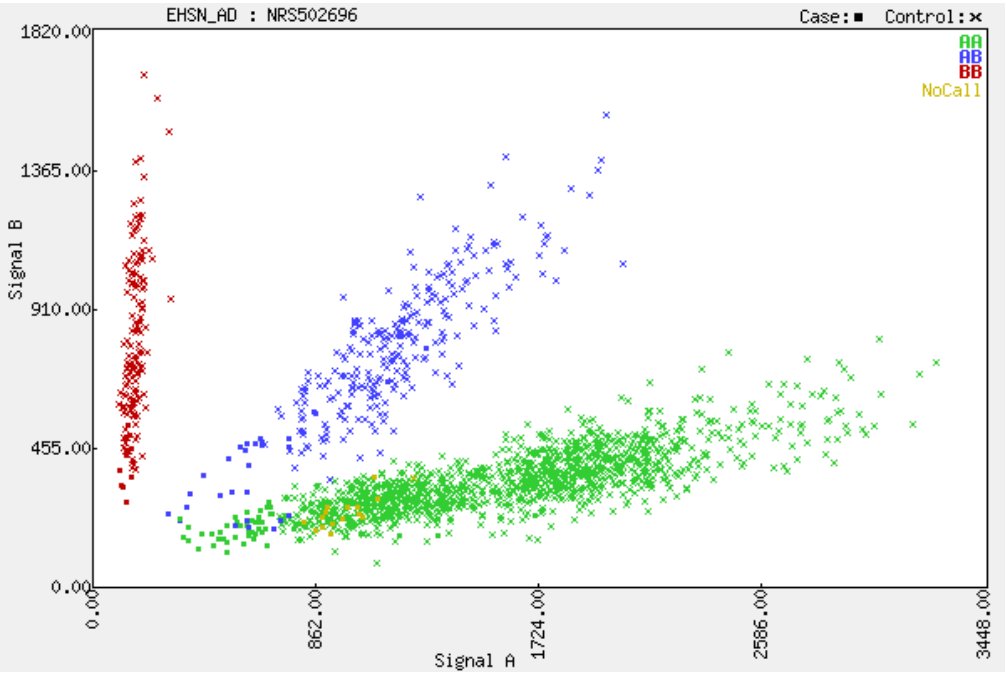


Figure 5.1: Examples of cluster plots generated from Affymetrix Genotyping Console 4.0. (a) Example of a good cluster plot where three distinct clusters were generated and identified. (b) Example of a bad cluster plot where part of the signals overlapped and alleles are often miscalled

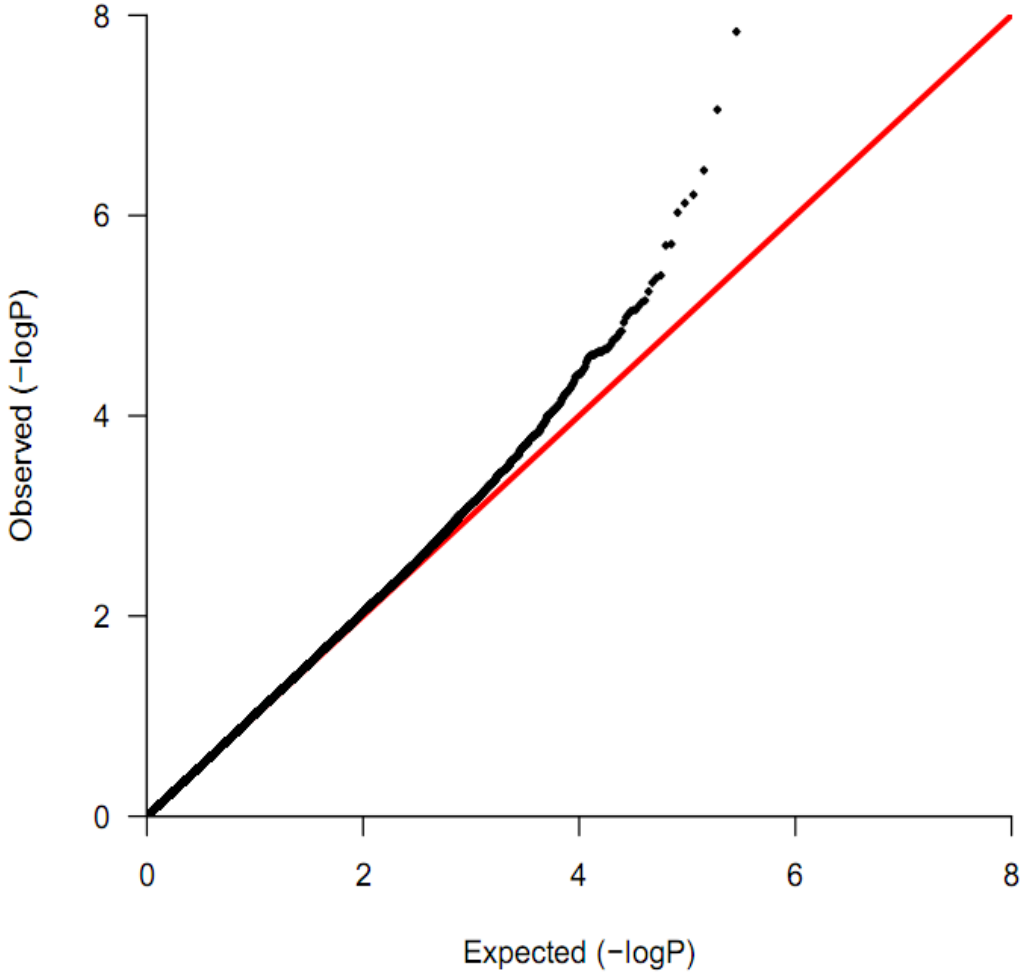


Figure 5.2: Q-Q plot for GWAS of *HLA-DQB1*06:02* negative EHS in Japanese population. The graph was drawn based on Cochran-Armitage P-values after standard quality control .No significant population stratification was observed with genomic inflation factor λ=1.008.

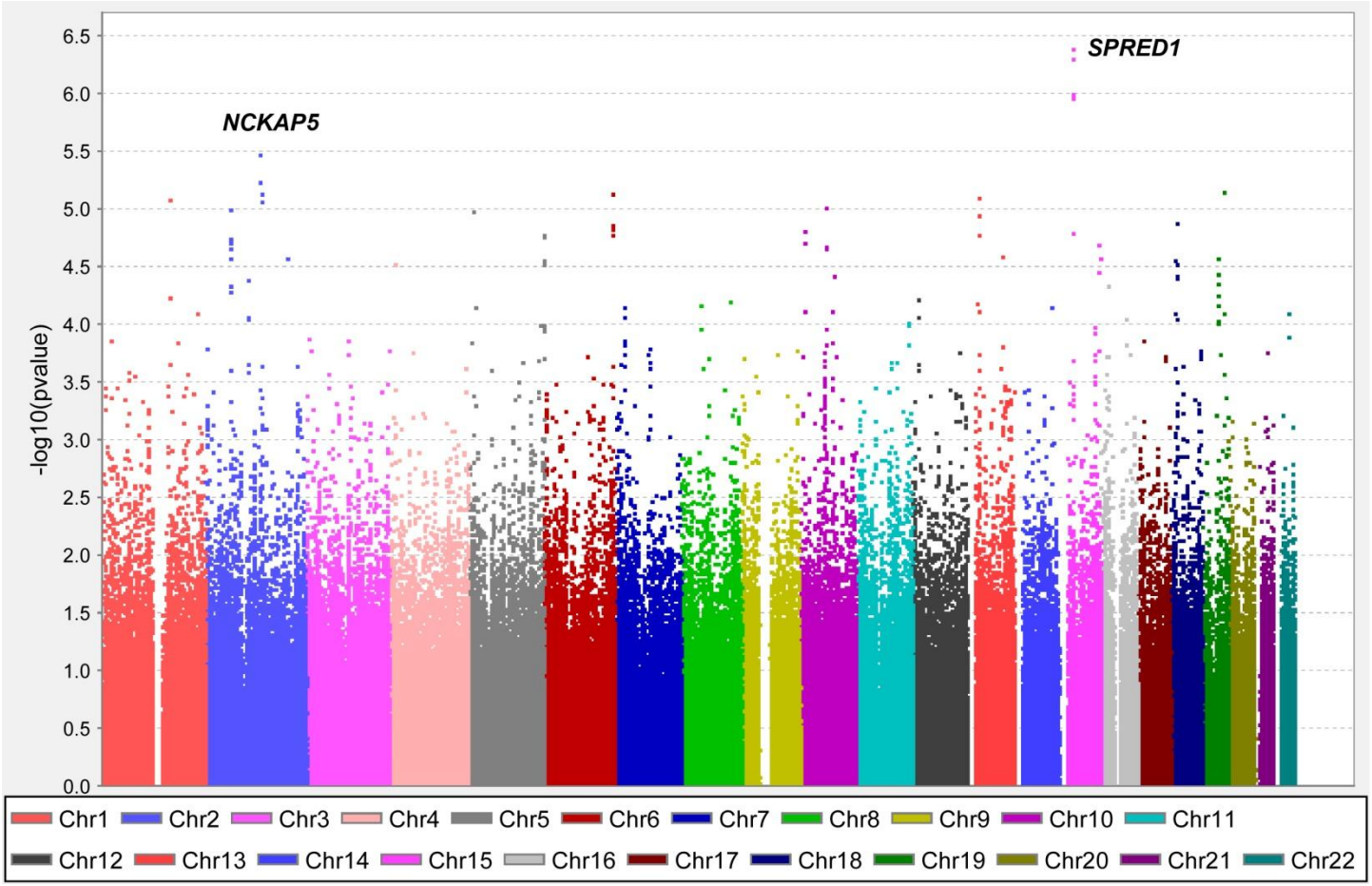


Figure 5.3: Manhattan plot of the GWAS data from the Japanese population. Manhattan plot was plotted based on P-value calculated using the Cochran-Armitage trend test.

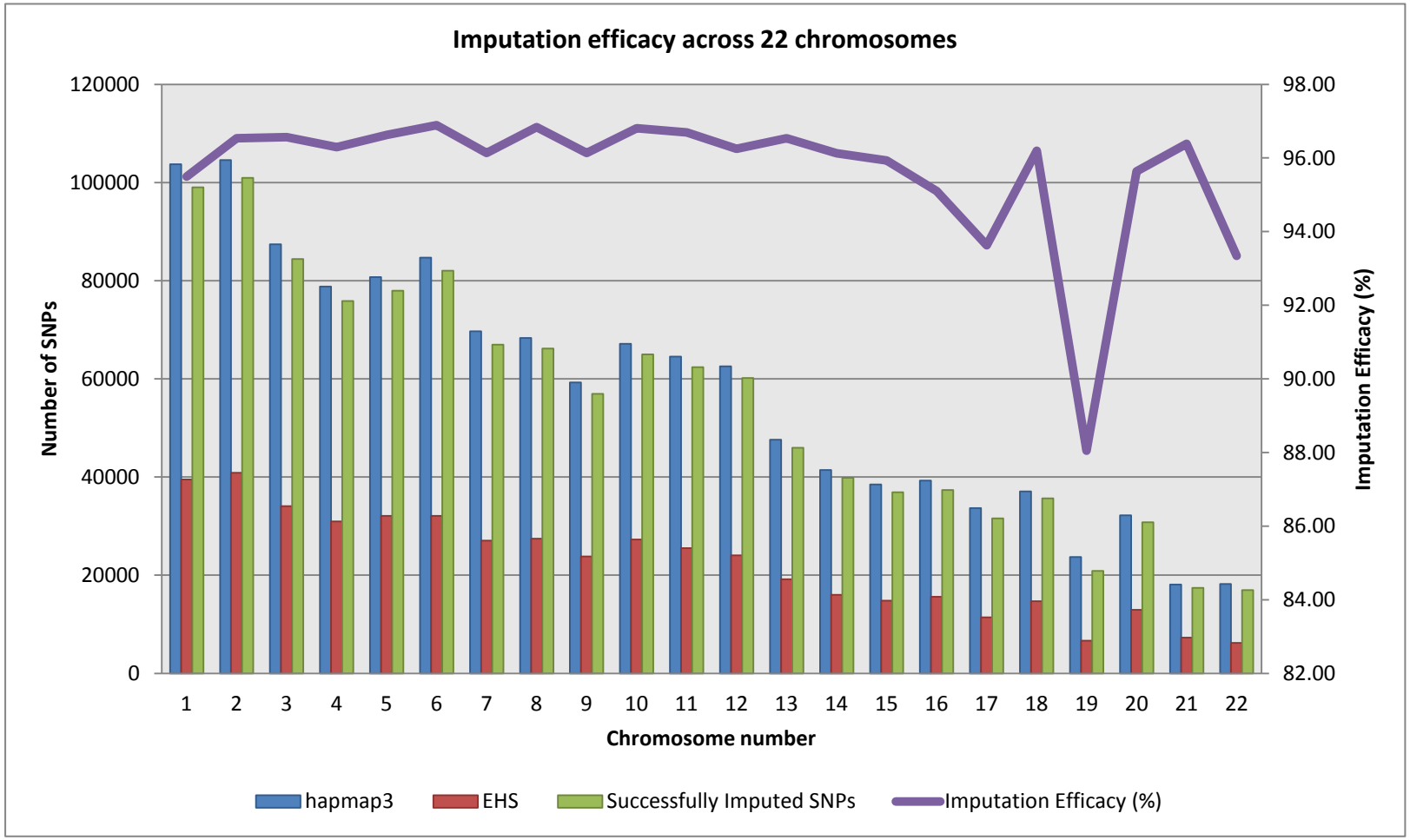


Figure 5.4: Imputation efficacy across 22 autosomes in which whole genome imputation was performed on *HLA-DQB1*06:02* negative EHS GWAS data using HapMap3 revision 2 Build36 data as a reference. Imputation accuracy (purple line) was calculated by the ratio of successfully imputed SNP after quality control with the total available SNPs in the reference

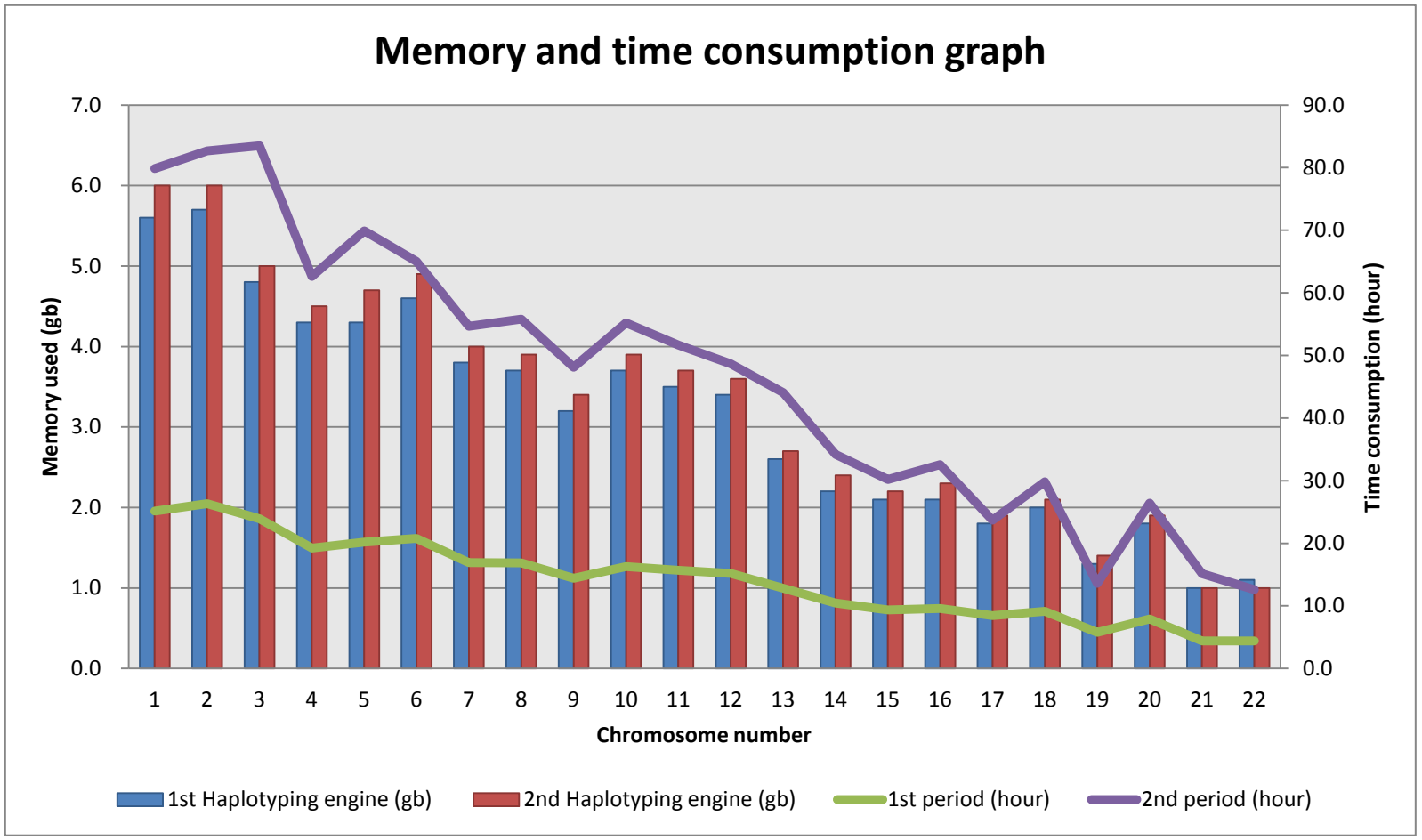


Figure 5.5: Computer memory power needed and time consumption for whole genome imputation. Imputation was performed in two steps: haplotyping and the actual imputation. The graph bar (blue and red bar) indicates the computer memory power (gigabytes) needed for each of the chromosomes. The line graph (purple and green lines) indicates the time needed to complete each steps across 22 chromosomes.

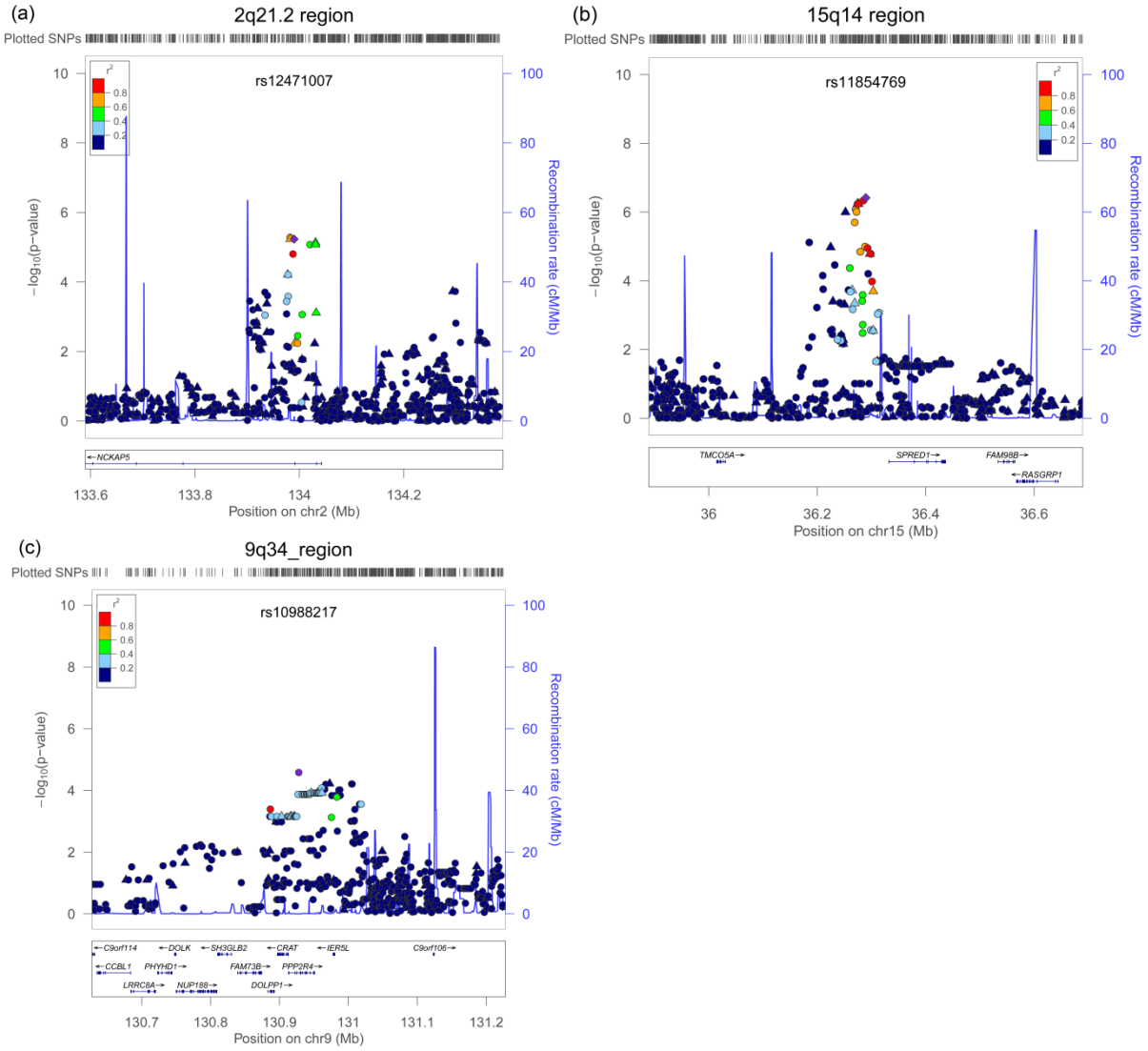


Figure 5.6: Regional association plots for three *HLA-DQB1*06:02* negative EHS risk loci. (a) The region 2q21.2 Cochran-Armitage trend test P-value, the SNP rs16826003 located in an intron of *NCKAP5* gene (b) The region 15q14 Cochran-Armitage trend test , SNP rs11854769 is located 42 kb upstream of the nearest gene, *SPRED1* (c) The 9q34 region based on P-minimum, rs10988217 is located in the an intron of *CRAT* gene. Each of the top markers is indicated by a purple diamonds. SNPs that were genotyped using Affymetrix 6.0 are marked by triangles. Imputed SNPs are plotted as circles. The color intensity represents the extent of the LD with the marker SNP, red ($r2 \geq 0.8$), orange ($0.6 < r2 \leq 0.8$), green ($0.4 < r2 \leq 0.6$), light blue ($0.2 < r2 \leq 0.4$), and dark blue ($r2 \leq 0.2$). Light blue in the background indicates local recombination rate.

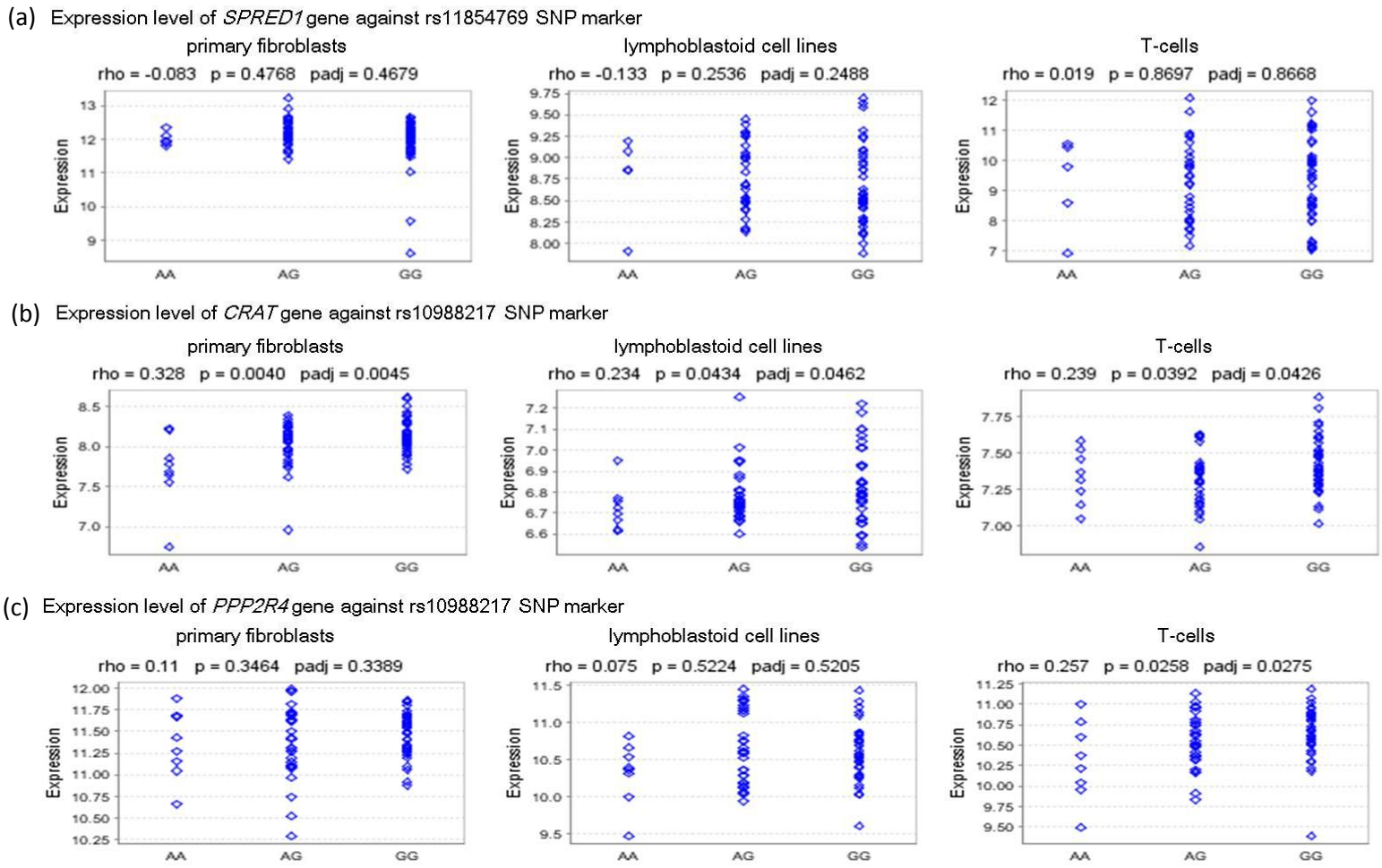


Figure 5.7: These eQTL analyses were performed based on data from the Sanger Institute GENEVAR project[39], this expression data is based on three cell types (fibroblast, lymphoblastoid cell line, and T-cell) from 75 unrelated individuals of Western European origin. (a) The plot displays the relationship between *SPRED1* gene expression and rs11854769. (b) The plot displays the relationship between *CRAT* gene expression and rs10988217. (c) The plot displays the relationship between *PPP2R4* gene expression and rs10988217.

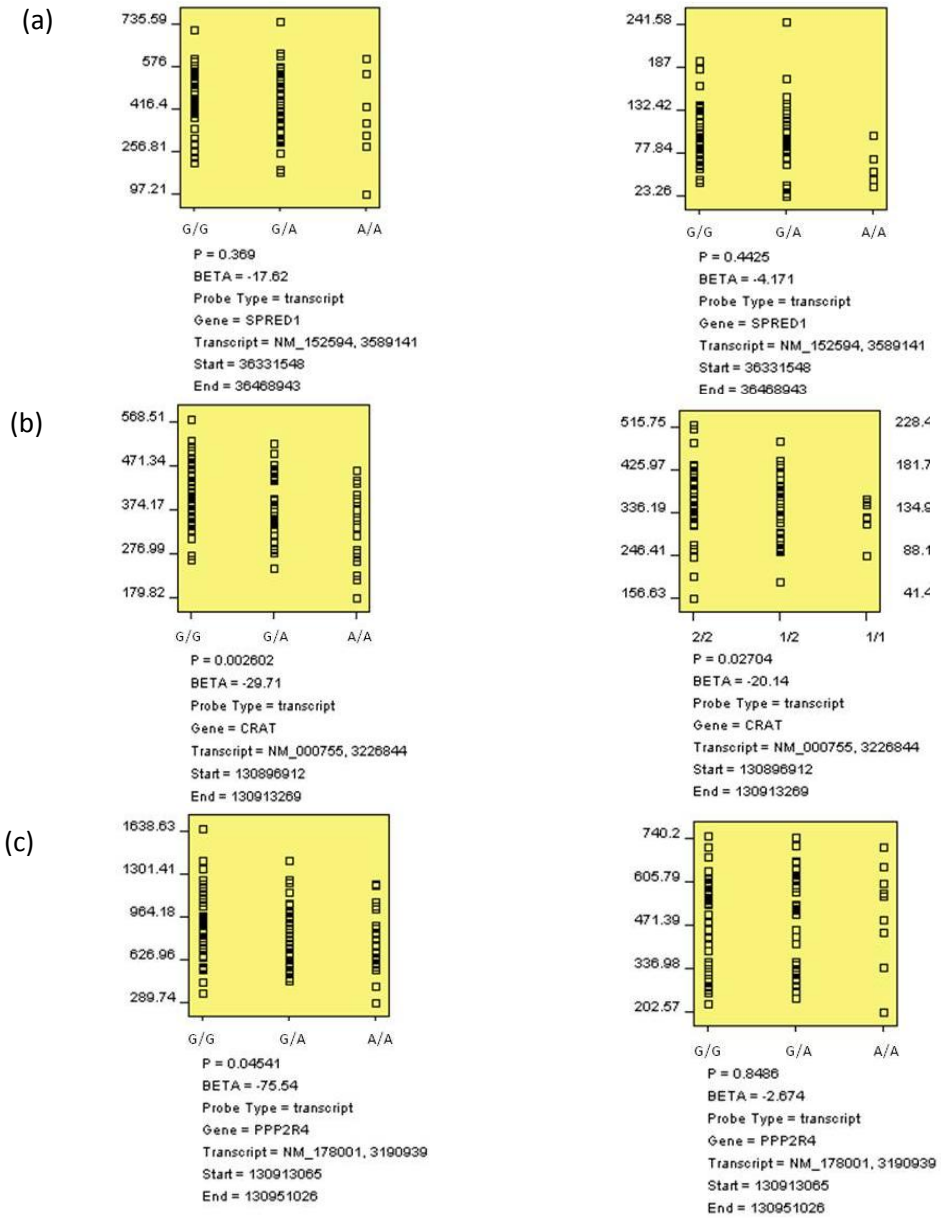


Figure 5.8: These eQTL association analyses were performed based on transcript expression data from the SNPExpress Database [41]; these data were derived from brain samples from 93 individuals of European ancestry (left) and Peripheral Blood Mononuclear Cell (PMBC) samples from 80 individuals of European ancestry (right). (a) The plot displays the relationship between *SPRED1* gene expression and rs11854769. (b) The plot displays the relationship between *CRAT* gene expression and rs10988217. (c) The plot displays the relationship between *PPP2R4* gene expression and rs10988217.

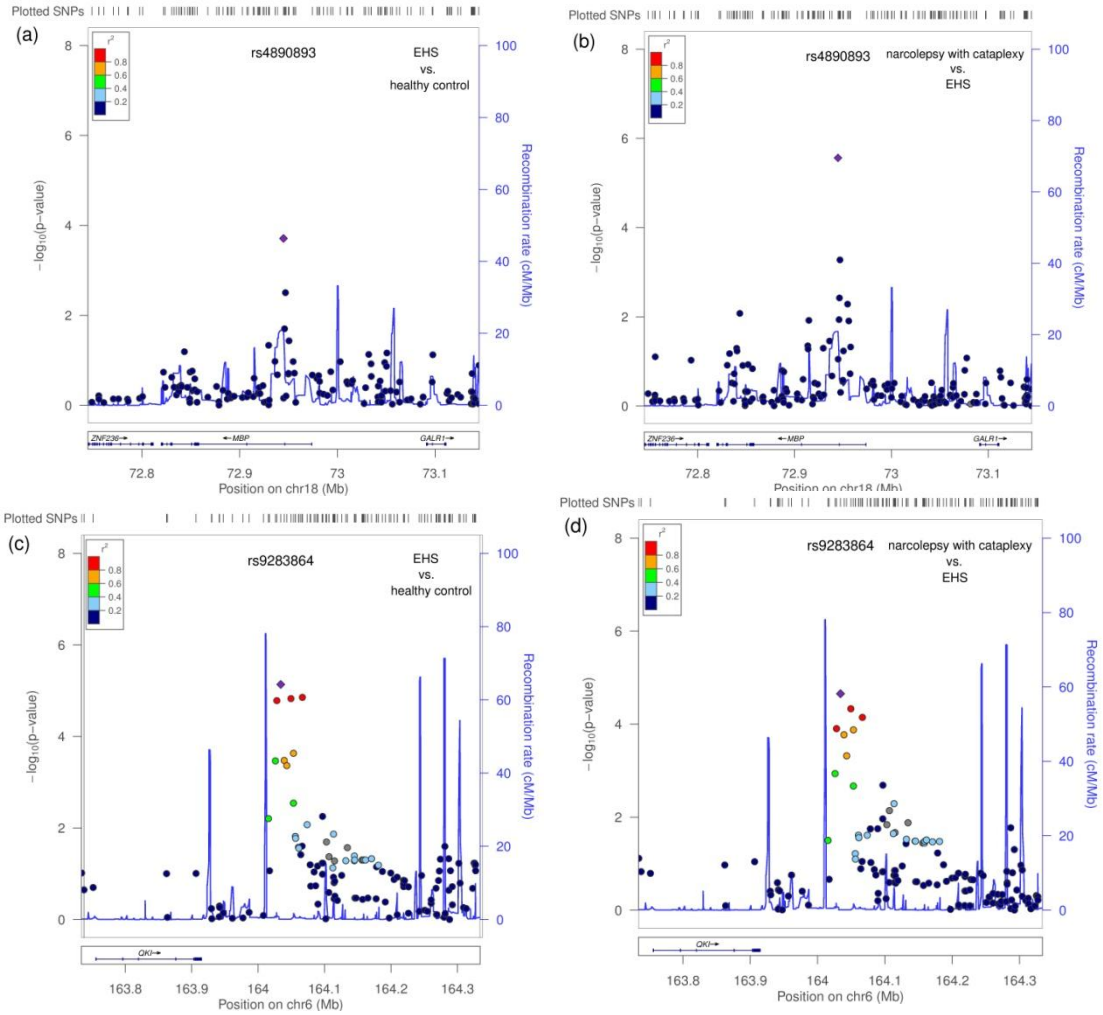


Figure 5.9: Regional association plots for risk loci at *QKI* and *MBP* regions (a) *HLA-DQB1*06:02* negative EHS GWAS showing the region 18q23 Cochran-Armitage trend test *P*-value, the SNP rs4890893 is located in the intron of the *MBP* gene (b) Comparison test between EHS and narcolepsy with cataplexy showing similar region as figure a (c) EHS GWAS showing the region 6q26 Cochran-Armitage trend test *P*-value, the SNP rs9283864 is located 146kb downstream of the *QKI* gene (d) Comparison test between EHS and narcolepsy showing similar region as figure b. Each of the top markers is indicated by a purple diamond. The color intensity represents the extent of the LD with the marker SNP: red ($r2 \geq 0.8$), orange ($0.6 < r2 \leq 0.8$), green ($0.4 < r2 \leq 0.6$), light blue ($0.2 < r2 \leq 0.4$), and dark blue ($r2 \leq 0.2$). Light blue in the background indicates the local recombination rate.

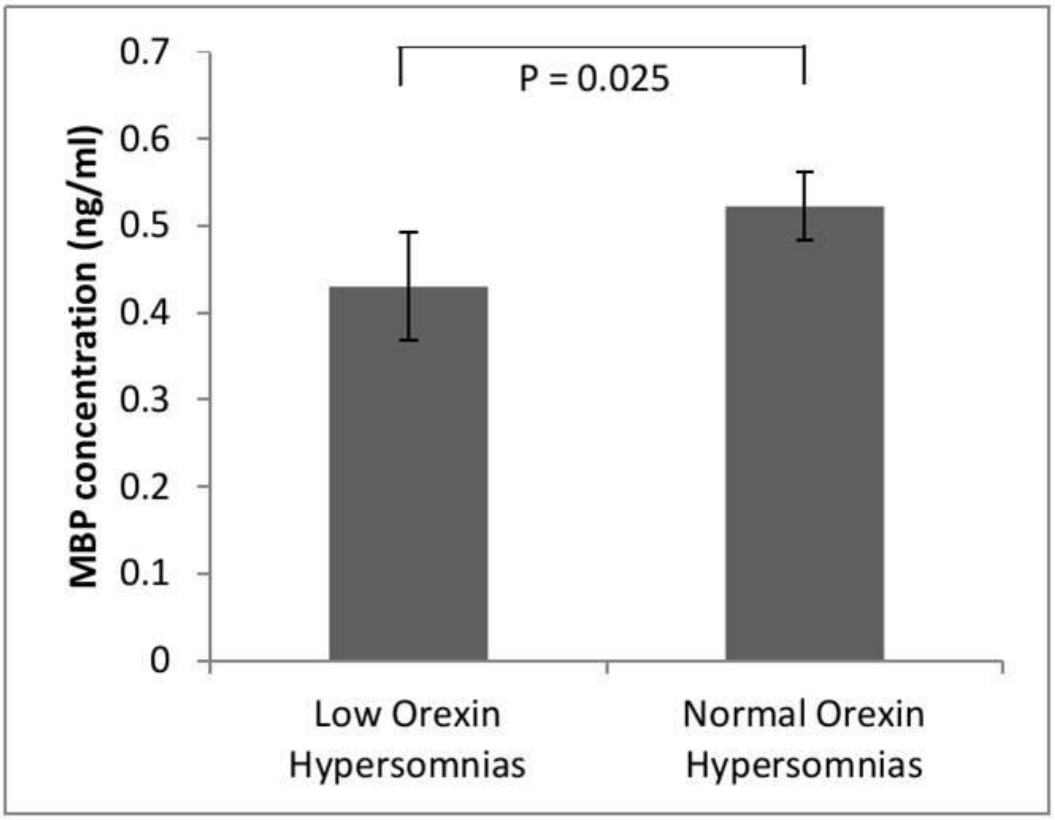


Figure 5.10: CSF MBP concentration (ng/mL ± s.e.m.) for low-orexin hypersomnias and normal orexin-hypersomnias

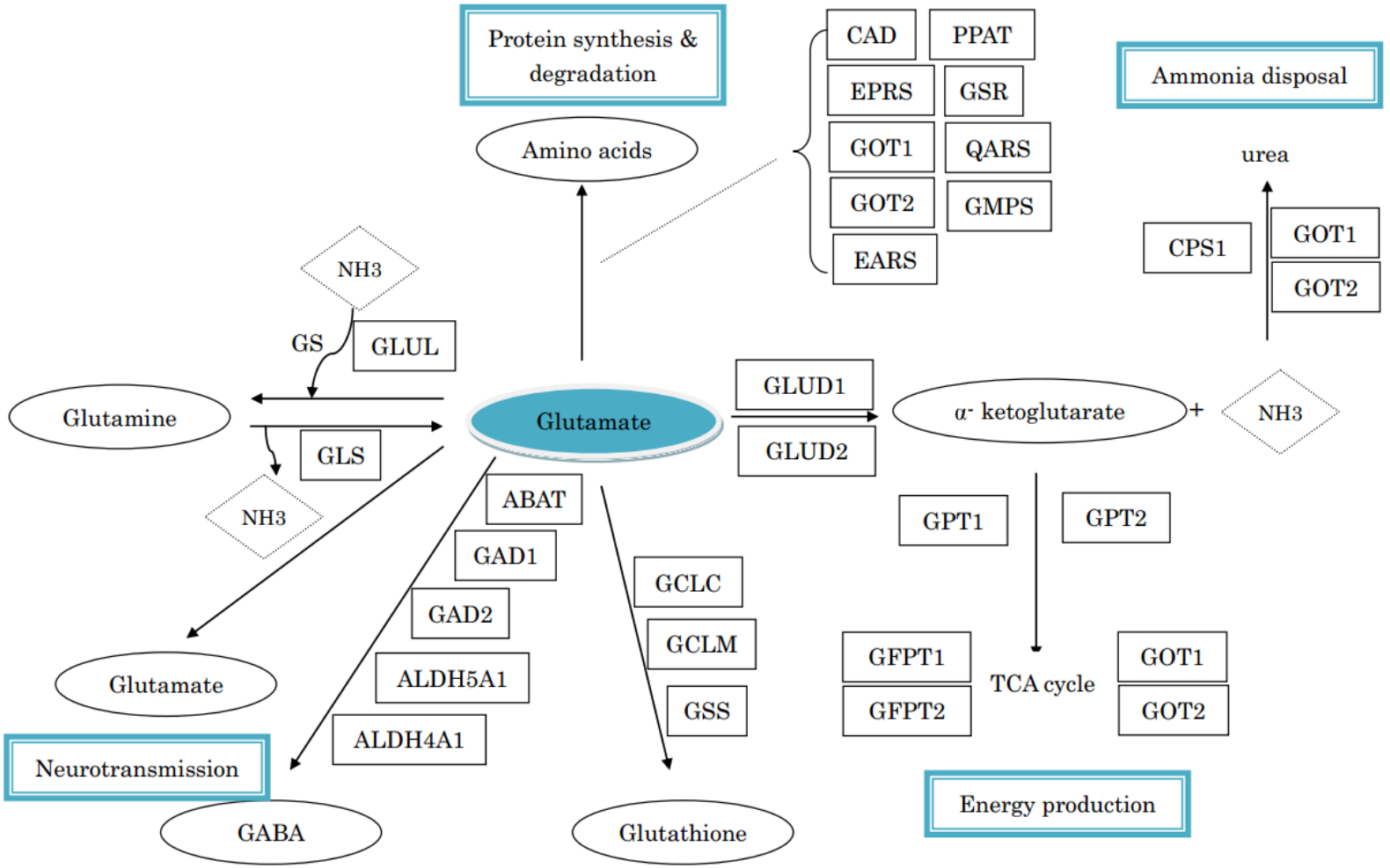


Figure 5.11: The gene components in glutamate metabolism pathway were grouped according to their functions. Blue square boxes indicate the grouping of functions, square boxes indicate the gene component in the pathway and circles indicate the protein/amino acid.

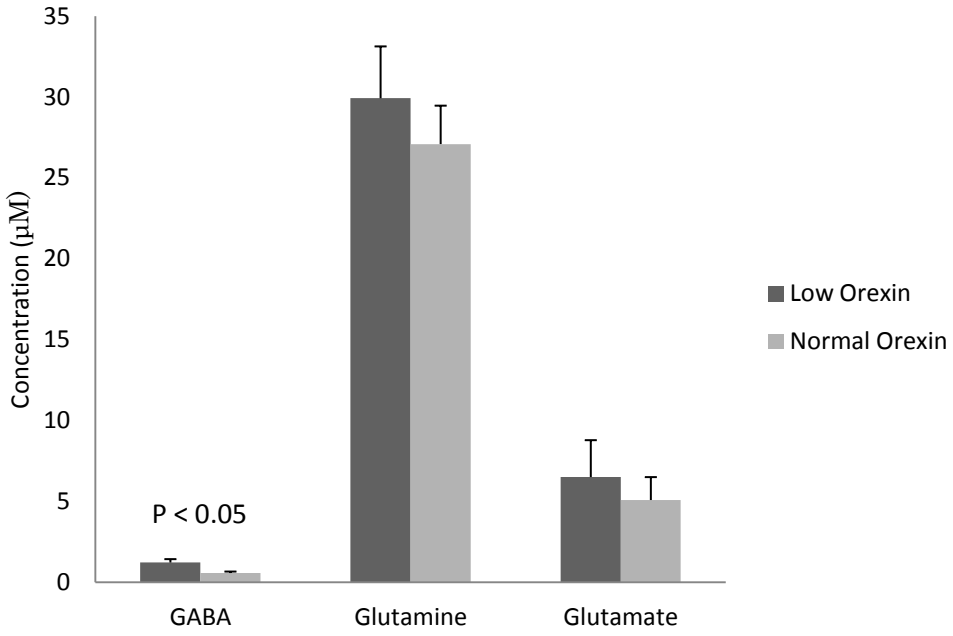
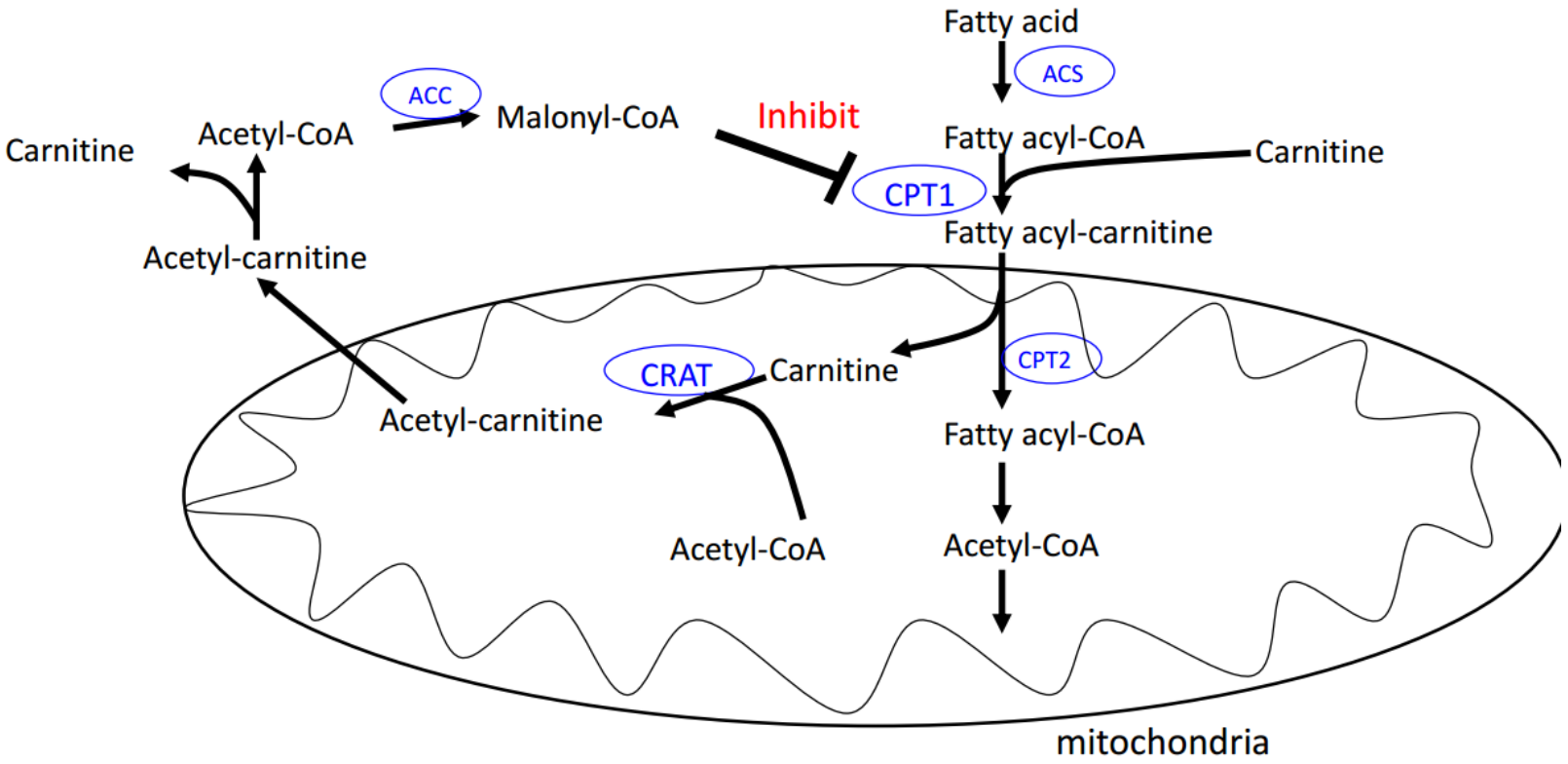


Figure 5.12: GABA, Glutamine, Glutamate concentration (µM ± s.e.m.) for low-orexin hypersomnias and normal orexin-hypersomnias



ACC: acetyl-CoA carboxylase, ACS: acyl-CoA synthase,
CPT1: carnitine palmitoyltransferase I , CPT2: carnitine palmitoyltransferase 2,
CRAT: carnitine acetyltransferase

Figure 6.1 Carnitine shuttle pathways in relation to *CRAT* and *CPT1B*.

Appendix

Khor SS, 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, Tokunaga K. Genome-wide association study of *HLA-DQB1*06:02* negative essential hypersomnia. PeerJ 1:e66, April, 2013

Khor SS, Miyagawa T, Imanishi A, Toyoda H, Honda M, Shimizu T, Kanbayashi T, Tokunaga K. Common variants in QKI and MBP contribute to myelin basic protein increase in low-orexin hypersomnia. The Journal of Neurology, Neurosurgery, and Psychiatry (Under Review, 2013)

Yamasaki M, Miyagawa T, Toyoda H, Khor SS, Koike A, 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. Journal of Human Genetics (In press, 2014)

Khor SS, Miyagawa T, Kodama T, Imanishi A, Toyoda H, Honda M, Shimizu T, Kanbayashi T, Tokunaga K. Pathway analysis on *HLA-DQB1*06:02* negative essential hypersomnia. (manuscript in preparation)

Selected scientific meeting attended:

Khor SS, 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, Tokunaga K. Genome-wide association study of *HLA-DQB1*06:02* negative essential hypersomnia. **Oral and Poster Presentation**, The 38th Annual Meeting of Japanese Society of Sleep Research, June 27-28, 2013, Akita, Japan.

Khor SS, 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, Tokunaga K. Genome-wide association study of *HLA-DQB1*06:02* negative essential hypersomnia. **Poster Presentation**, 10th International Workshop on Advanced Genomics, May 21-23, 2013, Tokyo, Japan.

Khor SS, 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, Tokunaga K. Genome-wide association study of *HLA-DQB1*06:02* negative essential hypersomnia. **Poster Presentation**, 57th Annual meeting of the Japan Society of Human Genetics, October 24-27, 2012, Tokyo, Japan.

Khor SS, 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, Tokunaga K. Genome-wide association study of *HLA-DQB1*06:02* negative essential hypersomnia. **Best poster nomination poster presentation**, 55th Annual meeting of the Japan Society of Human Genetics, October 27-30, 2010, Saitama, Japan.

Khor SS, 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, Tokunaga K. Genome-wide association study of *HLA-DQB1*06:02* negative essential hypersomnia. **Poster Presentation**, The 19th Annual Meeting of the Japanese Society for Histocompatibility and Immunogenetics , September 17-19, 2010, Tokyo, Japan.