HGG Advances

ARTICLE

The mitochondrial genome as a modifier of autism versus cancer phenotypes in *PTEN* hamartoma tumor syndrome

Ruipeng Wei,[1,2] Lamis Yehia,[1] Ying Ni,[3] and Charis Eng[1,4,5,6,7,*]

Summary

Cancer and autism spectrum disorder/developmental delay (ASD/DD) are two common clinical phenotypes in individuals with germline *PTEN* variants (*PTEN* hamartoma tumor syndrome, PHTS). Burgeoning studies have shown that genomic and metabolomic factors may act as modifiers of ASD/DD versus cancer in PHTS. Recently, we showed copy number variations to be associated with ASD/DD versus cancer in these PHTS individuals. We also found that mitochondrial complex II variants occurring in 10% of PHTS individuals modify breast cancer risk and thyroid cancer histology. These studies suggest that mitochondrial pathways could act as important factors in PHTS phenotype development. However, the mitochondrial genome (mtDNA) has never been systematically studied in PHTS. We therefore investigated the mtDNA landscape extracted from whole-genome sequencing data from 498 PHTS individuals, including 164 with ASD/DD (PHTS-onlyASD/DD), 184 with cancer (PHTS-onlyCancer), 132 with neither ASD/DD nor cancer (PHTS-neither), and 18 with both ASD/DD and cancer (PHTS-ASDCancer). We demonstrate that PHTS-onlyASD/DD has significantly higher mtDNA copy number than PHTS-onlyCancer group ($p = 9.2 \times 10^{-3}$ in all samples; $p = 4.2 \times 10^{-3}$ in the H haplogroup). PHTS-neither group has significantly higher mtDNA variant burden than PHTS-ASDCancer group ($p = 4.6 \times 10^{-2}$); the PHTS-noCancer group (PHTS-onlyASD/DD and PHTS-neither groups) also shows higher variant burden than the PHTS-Cancer group (PHTS-onlyCancer and PHTS-ASD/Cancer groups; $p = 3.3 \times 10^{-2}$). Our study implicates the mtDNA as a modifier of ASD/DD versus cancer phenotype development in PHTS.

Introduction

PTEN (MIM: 601728), a classical tumor suppressor gene, is a shared risk gene for cancer and autism spectrum disorder (ASD, MIM: 605309).[1–4] PTEN dysfunction can cause a wide range of phenotypes including macrocephaly, benign overgrowths and malignant neoplasia, metabolic alterations, and neurodevelopmental disorders (NDD), such as ASD and developmental delay (DD).[5] Two common, but seemingly disparate, clinical phenotypes of individuals with *PTEN* hamartoma tumor syndrome (PHTS, MIM: 158350), defined as having germline *PTEN* variants, are cancer and ASD/DD. Indeed, *PTEN* is one of the most prevalent and penetrant risk genes for ASD, with 2%–5% ASD individuals carrying germline *PTEN* variants.[1,4,6–10] However, it is still impossible to accurately predict development of cancer and/or ASD/DD for any single PHTS individual (versus group), including in those having an identical *PTEN* genotype.

Burgeoning studies have shown that genomic and metabolic factors may act as modifiers of ASD/DD versus cancer in PHTS.[11–13] We showed that variants in subunits of succinate dehydrogenase, also known as mitochondrial complex II (collectively referred to as SDHx), modify breast cancer risk and thyroid cancer histology in PHTS individuals though these occur only in 10%.[14,15] Relatedly, we recently illustrated that metabolic profiles could serve as a biomarker and enable more accurate prediction of PHTS-ASD/DD versus PHTS-Cancer at the individual level.[11,12] More recently, we showed that copy number variations in genomic DNA are associated with ASD/DD compared with cancer development in individuals with PHTS.[13] These data support the hypothesis that genetic or genomic modifying factors exist in PHTS.

The crosstalk between SDHx and metabolism suggests that mitochondrial pathways may act as an important factor in PHTS phenotype development. Importantly, despite identifying several possible modifiers of phenotypic outcomes in PHTS,[11–13] the vast majority of individuals with PHTS remain without known modifying factors. Mitochondria are not only the powerhouses of cells, but they also play a key role in regulating cellular metabolism.[16] This semi-independent cellular organelle contains its own genome housed as mitochondrial DNA (mtDNA), which encodes both RNA and protein. The 13 mitochondrial gene-coding proteins are key components of enzyme complexes of the oxidative phosphorylation

[1]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA; [2]Department of Nutrition, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; [3]Center for Immunotherapy & Precision Immuno-Oncology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA; [4]Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH 44195, USA; [5]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; [6]Germline High Risk Cancer Focus Group, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA
[7]Lead contact
*Correspondence: engc@ccf.org
https://doi.org/10.1016/j.xhgg.2023.100199.

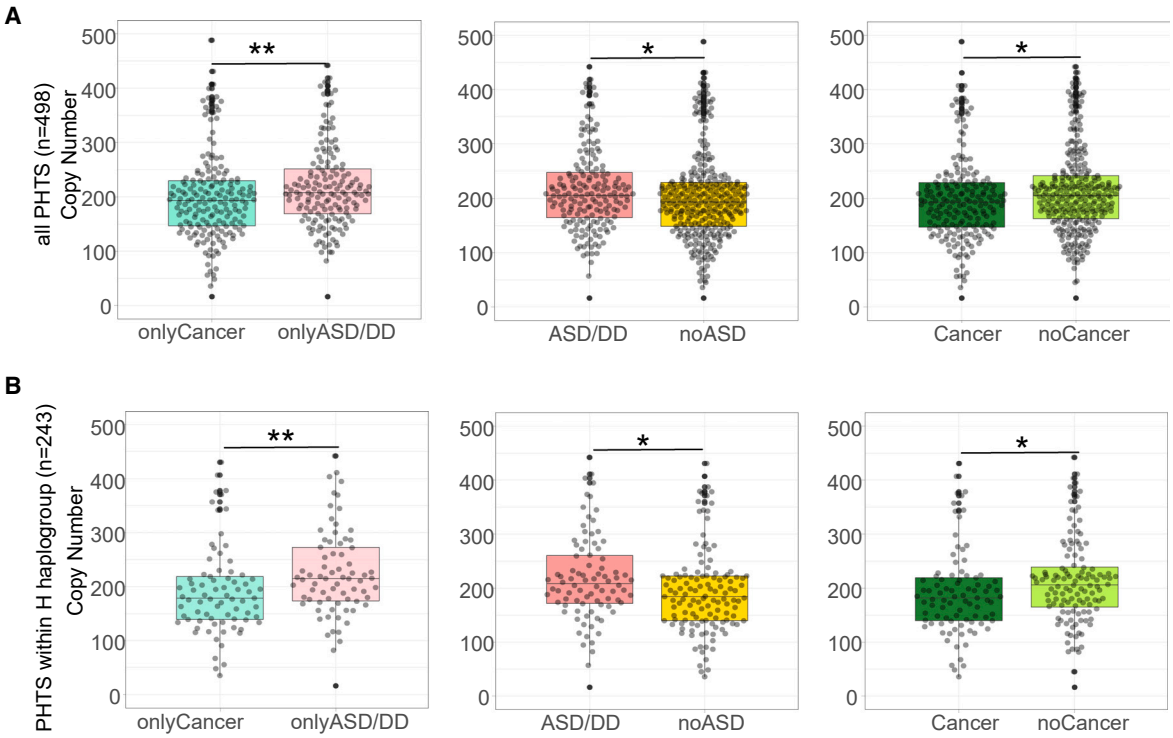


Figure 1. Mitochondrial DNA copy number differences among different PHTS phenotype groups

(A) Beeswarm plots of copy number differences between different cancer/ASD phenotypes within all samples (n = 498). p values are calculated by the Mann-Whitney test.

(B) Beeswarm plots of copy number differences between different cancer/ASD phenotypes within the H haplogroup (n = 213). p values are calculated by the Mann-Whitney test.

The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). Each dot represents one sample's mtDNA CN value. The upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points and are plotted individually.

**0.001 < p value ≤ 0.01. *p value <0.05.

system/electron transport chain.[16] Based on these observations and premises, we hypothesized that the mitochondrial genome (mtDNA) content and variation could also be modifiers of ASD/DD versus cancer phenotype development in PHTS.

Material and methods

Study participants, sample collection, processing, and sequencing

We performed whole-genome sequencing from germline genomic DNA derived from 599 peripheral blood samples from individuals with PHTS, followed at the Cleveland Clinic's PTEN Multidisciplinary Clinic and Center of Excellence (Cleveland, Ohio, USA) (Table S3). Each individual had only one sequencing sample. We were interested in the dichotomy of cancer and ASD/DD phenotypes. We categorized samples into four groups by phenotype, including onlyASD/DD, onlyCancer, ASDCancer, and neither (no ASD/DD and no cancer). The ASD/DD group includes individuals diagnosed with ASD/DD regardless of cancer status; the Cancer group includes individuals diagnosed with cancer regardless of ASD/DD status. This study was approved by the Cleveland Clinic Institutional Review Board, and all participants gave written informed consent. To ensure uniform phenotyping, after informed consent was obtained, checklists to document the presence or absence of specific features were completed by specialist genetic counselors or physicians (who have served in the PTEN Multidisciplinary Clinic for years and led by C.E.) concurrently with withdrawal of blood specimen. DNA sequencing of germline genomic DNA was performed at the Broad Institute Genomic Services (Cambridge, Massachusetts, USA). Whole-genome sequencing was performed with Illumina HiSeq XTM Ten platform with over 30x coverage in 150-bp paired-end reads.

Alignment and extraction of chrM and chr22 reads

All sequencing reads were aligned to genome reference build hg38. Mapping of reads to chromosome 22 (chr22), positions 46,000,000 to 46,100,000, and to the mitochondrial chromosome (chrM) were extracted from the BAM files using SAMtools version 1.16.1.[17] Mitochondrial reads were realigned to the revised Cambridge Reference Sequence using the Burrows-Wheeler Aligner (BWA) tool version 0.7.17.[18]

Copy number estimation of the mitochondrial genome

To assess mtDNA copy number (CN) for each sample, nuclear chromosome 22 positions 46,000,000 to 46,100,000 (genome

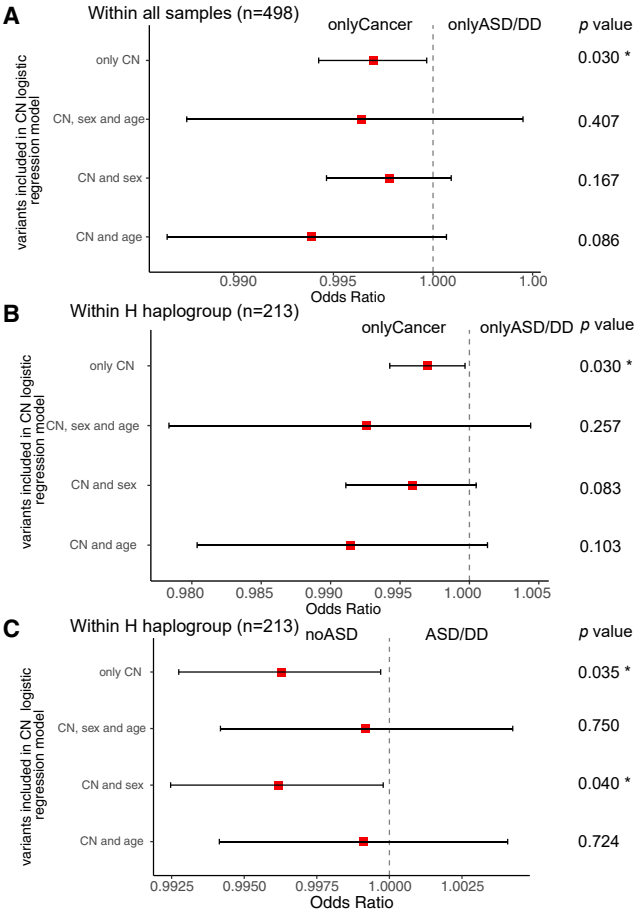


Figure 2. Forest plots of mtDNA CN analysis among different PHTS clinical phenotype groups

(A) Logistic regression results of mtDNA CN between onlyCancer and onlyASD/DD group (all participants, n = 498) with unadjusted variants, adjusted sex and age at consent, sex only, and age at consent only, respectively.

(B) Logistic regression results of mtDNA CN between onlyCancer and onlyASD/DD group (limited to the H haplogroup, n = 213) with unadjusted variants, adjusted sex and age at consent, sex only, and age at consent only, respectively.

(C) Logistic regression results of mtDNA CN between noASD and ASD group (limited to the H haplogroup, n = 213) with unadjusted variants, adjusted sex and age at consent, sex only, and age at consent only, respectively.

The red squares indicate the odds ratios, and error bars indicate the 95% confidence intervals of the odds ratios. *p value < 0.05.

build hg38) were chosen as a proxy representing two copies per cell due to its low level of copy number variants (CNVs) in the population.[19] Using read depth as a proxy of DNA CN, the mtDNA CN was estimated as $\frac{D_m}{D_n} \times 2$, where D_m denotes the average read depth across the mitochondrial chromosome, and D_n denotes the average read depth of the two-copy region, chr22: 46,000,000–46,100,000.

Variant calling and downstream quality control

We used the Genome Analysis Toolkit (GATK) version 4.2.0.0 and its mitochondrial mode to call mitochondrial DNA single-nucleotide variants (SNVs).[20] We omitted all variants in the regions from mitochondrial base positions 302 to 316, 513 to 526, 566 to 573, 8,860, and 16,181 to 16,194, which are known to yield false positive variant calls.[21] Because heteroplasmic variants generally impact disease phenotypes only at levels above 80%,[22] all variants with heteroplasmy levels less than this threshold were excluded from further consideration. Population frequency of each variant was obtained from GenBank (data last accessed on Dec. 10, 2021).[23] We used HaploGrep 2.1.21 to define the haplogroup for each sample.[24]

Statistical procedures

All statistical computations were performed using R version 4.2.0. Associations between mtDNA CN and PHTS clinical phenotypes were tested using the logistic regression models:

$$\text{logit}(p) = \alpha + \beta_0 \times \log(\text{CN}) + \beta_1 \times \text{age at consent} + \beta_2 \times \text{sex} \quad \text{(Equation 1)}$$

$$\text{logit}(p) = \alpha + \beta_0 \times \log(\text{CN}) + \beta_1 \times \text{age at consent} \quad \text{(Equation 2)}$$

$$\text{logit}(p) = \alpha + \beta_0 \times \log(\text{CN}) + \beta_1 \times \text{sex} \quad \text{(Equation 3)}$$

$$\text{logit}(p) = \alpha + \beta_0 \times \log(\text{CN}) \quad \text{(Equation 4)}$$

Where, α, β_0, β_1, and β_2 are coefficients of the corresponding regression, and p = probability(binary clinical feature). p values were extracted from the model fit with the anova() function, using the chi-squared test argument.

Rare SNV association analyses were performed using similar models:

$$\text{logit}(p) = \alpha + \beta_0 \times (\text{rare})\ \text{SNV} + \beta_1 \times \text{age} + \beta_2 \times \text{gender} \quad \text{(Equation 5)}$$

$$\text{logit}(p) = \alpha + \beta_0 \times (\text{rare})\ \text{SNV} + \beta_1 \times \text{age} \quad \text{(Equation 6)}$$

$$\text{logit}(p) = \alpha + \beta_0 \times (\text{rare})\ \text{SNV burden} + \beta_2 \times \text{gender} \quad \text{(Equation 7)}$$

$$\text{logit}(p) = \alpha + \beta_0 \times (\text{rare})\ \text{SNV burden} \quad \text{(Equation 8)}$$

Here, the "(rare) SNV" term represents either the numerical variable log(total number of mtDNA SNVs) (or) log(rare mtDNA SNVs). p values were computed as above.

Phenotype terms for all the above logistic regression models include PHTS-onlyASD/DD, PHTS-onlyCancer, PHTS-neither, PHTS-ASD, PHTS-Cancer, PHTS-ASD/Cancer, PHTS-noCancer, and PHTS-noASD.

We used the nonparametric Mann-Whitney test, Fisher's exact test, and chi-squared test, which are the unadjusted tests of the above-described logistic regression models, to identify differences between the features of the groups clustered by PHTS clinical phenotypes. p values <0.05 were considered as statistically significant. Our participants belong to over 15 mitochondrial haplogroups, which makes it challenging to perform variant-level association analyses, knowing that mitochondrial variants are core determinants of the different haplogroups and may therefore lead to spurious associations. Since the H haplogroup is the largest one among all the haplogroups, we decided to perform targeted analyses limited to this haplogroup (n = 213).

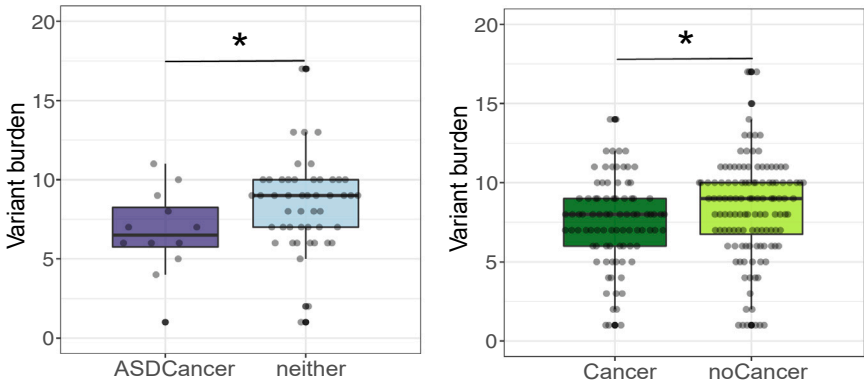


Figure 3. Beeswarm plots of mitochondrial DNA single-variant burden differences among different PHTS phenotypes within the H haplogroup (n = 213)

The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). Each dot represents one sample's mtDNA CN value. The upper whisker extends from the hinge to the largest value no further than 1.5 * interquartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points and are plotted individually.

p values are calculated by the Mann-Whitney test. *p value < 0.05.

Results

Mitochondrial haplogroup distribution and participant phenotypes

Our analytical series consisted of 498 samples after excluding samples from individuals who have maternal relationships, performing standard quality control, haplogroup calling, and excluding *PTEN* wild-type family members. Of the 498 samples, 213 (42.8%) belong to the H haplogroup, while the others mainly belong to the U, J, K T, L, and W haplogroups. In this series, 164 (32.93%) individuals have ASD/DD but not cancer (designated as onlyASD/DD), 184 (36.94%) have cancer but not ASD/DD (designated as onlyCancer), and 18 (3.61%) have both ASD/DD and cancer (designated as ASDCancer). The remaining 132 (26.51%) participants have neither ASD/DD nor cancer (designated as neither). When the samples are narrowed down within the H haplogroup, the estimated distribution of individuals within each of the phenotype groups parallels the distribution within the entire series.

Mitochondrial DNA copy number distribution in individuals with PHTS

We previously showed that a higher genomic DNA autosomal CNV burden is associated with the PHTS-ASD/DD phenotype compared with PHTS individuals without ASD/DD.[13] We therefore hypothesized that PHTS-ASD/DD individuals may also have a higher mtDNA CN. To address this hypothesis, we tested mtDNA CN differences among different PHTS-related phenotype groups (Figure S1). In addition to the onlyASD/DD, onlyCancer, ASDCancer, and neither groups, we also examined the ASD/DD group, which includes individuals diagnosed with ASD/DD regardless of cancer status, and the Cancer group, which includes individuals diagnosed with Cancer regardless of ASD/DD status. We observed that the onlyASD/DD group has a higher median mtDNA CN compared with the onlyCancer group ($p = 9.2 \times 10^{-3}$); the ASD/DD group has a higher mtDNA CN burden than the noASD group ($p = 2.4 \times 10^{-2}$); and the noCancer group has a higher mtDNA CN burden than the Cancer group ($p = 3.2 \times 10^{-2}$) (Figures 1A and S1). Since idiopathic ASD is more frequently diagnosed in males than in females, we performed statistical comparisons while controlling for biological sex. Additionally, because individuals with PHTS-ASD/DD are younger than PHTS individuals with cancer at the time of consent, we also performed statistical comparisons controlling for age alone and controlling for both age and sex. After controlling for age at consent, the onlyASD/DD group still showed a trend toward higher burden of mtDNA CN compared with the onlyCancer group ($p = 8.6 \times 10^{-2}$) (Figure 2A and Tables S1 and S2).

Mitochondrial DNA copy number distribution in PHTS individuals belonging to the H haplogroup

To test whether the mtDNA CN differences among the different phenotypes still held within the H haplogroup, we performed similar comparisons utilizing a more focused set of individuals. Within the H haplogroup, we similarly found that the onlyASD/DD group has a higher mtDNA CN compared with the onlyCancer group ($p = 4.2 \times 10^{-3}$) and with the neither group ($p = 3.4 \times 10^{-3}$ controlling for sex); the ASD/DD group has higher mtDNA CN than the noASD group ($p = 1.4 \times 10^{-2}$, $p = 4.0 \times 10^{-2}$ when controlling for age at consent); and the Cancer group has lower mtDNA CN than the noCancer group ($p = 1.5 \times 10^{-2}$) (Figures 1B, 2B, and S2).

Mitochondrial DNA single-nucleotide variant burden in PHTS individuals with the H haplogroup

Next, we tested the association between mtDNA SNV burden and PHTS clinical phenotypes in individuals belonging to the H haplogroup. Haplogroups are defined by specific sets of SNVs.[24] Therefore, we should either control for the haplogroups or analyze SNV burden within each haplogroup. However, with the limited sample size (n = 498), it became prudent to only focus on the largest haplogroup, the H haplogroup. Overall, the entire series included 8,156

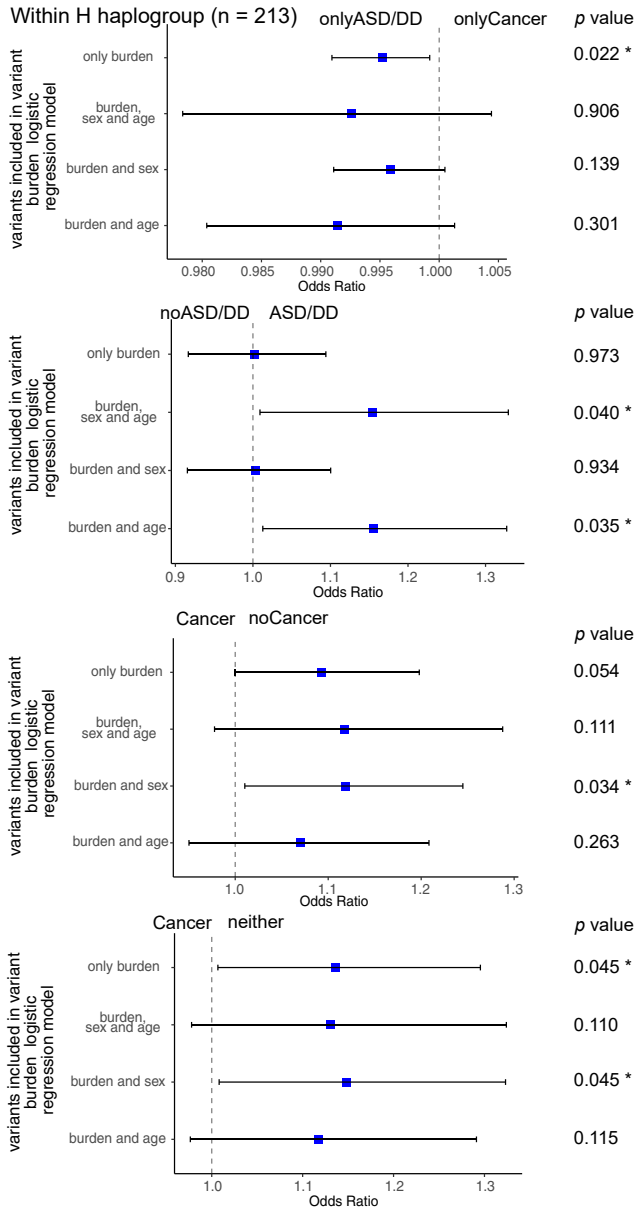


Figure 4. Forest plots of mtDNA variant burden analysis among different PHTS clinical phenotype groups within the H haplogroup (n = 213)

The blue squares indicate the odds ratios, and error bars indicate the 95% of confidence interval of the odds ratios. Clinical phenotype groups have been indicated at the top of each forest plot. *p value < 0.05.

non-reference SNVs (1,017 unique variants, Table S4). Within the H haplogroup, we identified 1,691 SNVs (369 unique variants). After performing all pairwise comparisons, the neither group showed a higher SNV burden compared with the ASDCancer group ($p = 4.6 \times 10^{-2}$) and with the Cancer group ($p = 0.045$ controlling for age at consent); the noASD group had a higher SNV burden than the ASD/DD group ($p = 4.0 \times 10^{-2}$, controlling for age and sex); and the noCancer group had a higher SNV burden than the Cancer group ($p = 3.3 \times 10^{-2}$, $p = 3.4 \times 10^{-2}$ controlling for sex) (Figures 3 and 4 and Tables S1 and S2).

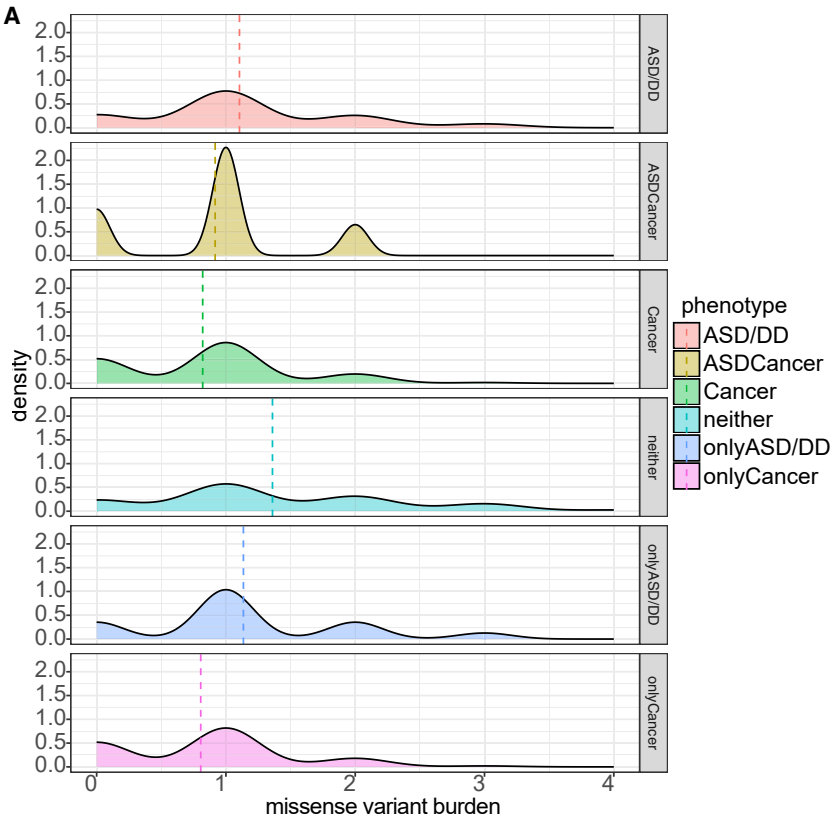
Mitochondrial DNA single-nucleotide variant effects in PHTS individuals with the H haplogroup

Within the H haplogroup, we identified 251 SNVs with predicted functional consequences (48 unique variants) based on the NCBI database.[25] Among the 251 SNVs, 225 SNVs (35 unique variant) are missense variants, 25 non-coding transcript exon variants, and only 1 synonymous variant. Missense variants, non-coding transcript exon variants, and synonymous variants have moderate, modifier, and low impact, respectively. The missense variant histogram indicates that the neither group has more samples with higher missense variant burden than the onlyCancer group (Figure 5A). After performing all pairwise comparisons, the neither group showed a higher missense variant burden than the Cancer and onlyCancer groups (both $p = 1.0 \times 10^{-2}$); the noASD/DD group showed a higher missense variant burden than the ASD/DD group ($p = 8.0 \times 10^{-4}$); and the onlyASD/DD group showed a higher missense variant burden than the onlyCancer group (Figure 5B). We did not find any statistically significant comparisons after examining non-coding transcript exon variants and the synonymous variant.

Discussion

Here, we present results investigating the relationship between PHTS clinical phenotypes and mtDNA content and variation. These observations begin to provide important insights into the clinical challenge of predicting ASD/DD and/or cancer development in PHTS.

ASD/DD has strong genetic underpinnings and metabolic comorbidities.[26–28] Although previous studies have explored mtDNA within individuals with sporadic ASD/DD,[29–32] the mtDNA landscape in PHTS-ASD/DD has never been investigated. On this PHTS background, we found that PHTS-ASD/DD has higher mtDNA CN than PHTS-Cancer or PHTS-neither. Notably, our observations remain consistent before and after controlling for age, which is highly associated with PHTS-ASD/DD and PHTS-Cancer. Intriguingly, the PHTS-ASD/DD group has lower mtDNA SNV (versus CN) burden than that of PHTS-neither. We attempted to annotate all the SNVs we identified from the whole-genome sequencing; however, only a small number has known annotations.[33,34] This implies that most SNVs we found are novel, non-functional, or neutral. We interpret those SNVs as polymorphism-like and protective rather than pathogenic or likely pathogenic. Additionally and notably, we found that individuals with PHTS-ASD/DD have a higher burden of mitochondrial missense variants than individuals with PHTS-Cancer. Importantly, sporadic ASD/DD has been previously correlated with mitochondrial disorders.[28] The mtDNA variant n.3397A>G (p. ND1:M31V) has been found in 25 individuals with sporadic non-PHTS-ASD/DD, but in our series, only one participant with ASD/DD harbored this variant.[32] These observations may be



B

Comparisons/Test	Mann-Whitney test
onlyASD/DD vs. onlyCancer	0.010*
onlyASD/DD vs. ASDCancer	0.43
onlyASD/DD vs. neither	0.23
onlyCancer vs. ASDCancer	0.55
onlyCancer vs. neither	0.0011**
ASDCancer vs. neither	0.15
ASD/DD vs. noASD	0.00080***
ASD/DD vs. neither	0.17
Cancer vs. noCancer	0.64
Cancer vs. neither	0.0011**

* 0.05 > p > 0.01
** 0.01 ≥ p > 0.001
*** 0.001 ≥ p

Figure 5. Missense variant burden among different PHTS phenotypes within the H haplogroup (n = 213)

(A) Density plot of missense variants among different phenotype groups. Dashed lines represent the mean missense variant burden.

(B) Mann-Whitney test results of the missense variant burden comparisons between different phenotype groups. *** $p \leq 0.001$. **$0.001 < p$ value ≤ 0.01. *$0.01 < p$ value < 0.05.

explained by the fact that our series of individuals with PHTS represents syndromic ASD/DD rather than the sporadic form and the mtDNA variation likely being modifying instead of etiologic in this context.[35]

Our study does have certain limitations. Interrogating mtDNA in peripheral blood is, at best, an indirect measure of the mtDNA content in cancer tissues or in the brain, the primary tissues of interest based on phenotype. However, for such studies, it is impossible (perhaps unethical) to obtain brain tissues from living individuals with ASD/DD. Importantly, mtDNA from blood can serve as a representative tissue source for mtDNA.[29,30] Although this may be a limitation when analyzing CN, our aim was to analyze germline mtDNA sequence content, which should largely be the same in all tissues. Owing to the rarity of PHTS, another limitation is sample size (n = 498), which left us underpowered to detect specific phenotype-associated SNVs, rare SNVs, and SNVs within specific genes or regions. The mtDNA gene-coding variants, n.8251G>A (p. COII:G222G), n.8269G>A (p. COII:*228*), n.8271A>G (non-coding variant in *mt-NC7*), and others, found by previous ASD studies, have not been found in our participant series.[31,32] Because these SNVs were identified in a much limited sample size in previous studies, it is plausible that these SNVs are associated with sporadic versus *PTEN*-related syndromic ASD and are more etiologic versus modifying. A natural extension of our work would be to query the mitochondrial whole-genome sequence for large deletions. Together with the nuclear mitochondrial genome results, these data will help inform the association between ASD/DD and mtDNA within the PHTS background. Another caveat in our study is that although we follow these research participants prospectively for a median of 4 years (range = 1 month to 16 years; SD = 4 years), we cannot ascertain cancer status in individuals who did not have cancer at the time of consent nor during follow-up. However, we are more certain that individuals without ASD/DD indeed do not have these NDDs given the younger (typically pediatric) ages at diagnosis of NDD. Similarly, we cannot predict which ASD/DD individuals will develop cancer decades from consent. This is part of the motivation of our studies. It is also important to point out that the PHTS component cancers (except for thyroid cancer) do not arise until 30–35 years old and beyond, in the context that the median age of our ASD/DD persons is 11.35 years old at the time of consent.

Despite these limitations, our participant series comprises one of the largest PHTS cohorts worldwide, including comprehensive and uniform phenotyping by PHTS experts. The use of whole-genome sequencing yields extremely high-depth coverage of the mitochondrial genome, which enables comprehensive and sensitive detection of all mtDNA substitutions and avoids the

error-prone imputation process. Because all samples were processed at the same center with the same sequencing protocols, batch effects are minimized, and meta-analysis is not necessary. This overall uniformity enabled us to uncover significant associations despite a limited sample size for each phenotype group.

In summary, we analyzed mtDNA content and variation in the peripheral blood of individuals with PHTS and found that individuals with ASD/DD have higher mtDNA CN on a PHTS background. We also found that the PHTS-ASD/DD group has lower mtDNA SNV (versus CN) burden than PHTS-neither, and PHTS-ASD/DD individuals have more mitochondrial missense variants than PHTS-Cancer, an observation that also held true within the H haplogroup. Thus, our data suggest that specific mtDNA content/variation could act as modifiers in the background of germline *PTEN* variants. However, future studies are needed to validate these findings in other PHTS cohorts and within the other mitochondrial haplogroups. Characterizing the mtDNA landscape is an important step for more precisely predicting ASD/DD versus cancer phenotypes in individuals with PHTS, a clinically heterogeneous syndrome.

Data and code availability

Data will be available on reasonable request to the corresponding author. All the bioinformatics and statistical tools are described in the material and methods. Scripts and code for generation of association statistics and additional statistical analyses are available at https://github.com/weiruipeng/EngLab_mtDNAproject.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2023.100199.

Acknowledgments

We thank all our research participants who contributed to this study. We are grateful to the clinical research team at the PTEN Multidisciplinary Clinic and Center of Excellence at the Cleveland Clinic for administrative support. We also thank Dr. Thomas LaFramboise (Case Western Reserve University) for scientific advice and helpful discussions. This study was funded, in part, by the PTEN Research Foundation (CCF19-001), the Ambrose Monell Foundation (Switch grant), the Breast Cancer Research Foundation, and the NICHD/NIH (R01HD105049) (all to C.E.). L.Y. is an Ambrose Monell Cancer Genomic Medicine Fellow at the Cleveland Clinic Genomic Medicine Institute. C.E. is the Sondra J. and Stephen R. Hardis Endowed Chair of Cancer Genomic Medicine at the Cleveland Clinic and an ACS Clinical Research Professor.

Declaration of interests

The authors have no relevant conflicts of interest.

Received: January 20, 2023
Accepted: April 13, 2023

Web resources

OMIM: Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

Genome Analysis Toolkit (GATK), https://gatk.broadinstitute.org/hc/en-us.

MITOMAP, https://www.mitomap.org/MITOMAP.

Haplogrep2, https://haplogrep.i-med.ac.at/category/haplogrep2/

HmtDB, https://www.hmtdb.uniba.it/

References

1. Butler, M.G., Dasouki, M.J., Zhou, X.-P., Talebizadeh, Z., Brown, M., Takahashi, T.N., Miles, J.H., Wang, C.H., Stratton, R., Pilarski, R., and Eng, C. (2005). Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. J. Med. Genet. *42*, 318–321. https://doi.org/10.1136/jmg.2004.024646.
2. Crawley, J.N., Heyer, W.-D., and LaSalle, J.M. (2016). Autism and cancer share risk genes, pathways and drug targets. Trends Genet. *32*, 139–146. https://doi.org/10.1016/j.tig.2016.01.001.
3. Forés-Martos, J., Catalá-López, F., Sánchez-Valle, J., Ibáñez, K., Tejero, H., Palma-Gudiel, H., Climent, J., Pancaldi, V., Fañanás, L., Arango, C., et al. (2019). Transcriptomic metaanalyses of autistic brains reveals shared gene expression and biological pathway abnormalities with cancer. Mol. Autism. *10*, 17. https://doi.org/10.1186/s13229-019-0262-8.
4. Trost, B., Thiruvahindrapuram, B., Chan, A.J.S., Engchuan, W., Higginbotham, E.J., Howe, J.L., Loureiro, L.O., Reuter, M.S., Roshandel, D., Whitney, J., et al. (2022). Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell *185*, 4409–4427.e18. https://doi.org/10.1016/j.cell.2022.10.009.
5. Yehia, L., Keel, E., and Eng, C. (2020). The clinical spectrum of PTEN mutations. Annu. Rev. Med. *71*, 103–116. https://doi.org/10.1146/annurev-med-052218-125823.
6. Fernandez, B.A., and Scherer, S.W. (2017). Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. Dialogues Clin. Neurosci. *19*, 353–371.
7. Klein, S., Sharifi-Hannauer, P., and Martinez-Agosto, J.A. (2013). Macrocephaly as a clinical indicator of genetic subtypes in autism. Autism Res. *6*, 51–56. https://doi.org/10.1002/aur.1266.
8. Orrico, A., Galli, L., Buoni, S., Orsi, A., Vonella, G., and Sorrentino, V. (2009). Novel PTEN mutations in neurodevelopmental disorders and macrocephaly. Clin. Genet. *75*, 195–198. https://doi.org/10.1111/j.1399-0004.2008.01074.x.
9. Tilot, A.K., Frazier, T.W., and Eng, C. (2015). Balancing proliferation and connectivity in PTEN-associated autism spectrum disorder. Neurotherapeutics *12*, 609–619. https://doi.org/10.1007/s13311-015-0356-8.
10. Varga, E.A., Pastore, M., Prior, T., Herman, G.E., and McBride, K.L. (2009). The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental

delay, and macrocephaly. Genet. Med. *11*, 111–117. https://doi.org/10.1097/GIM.0b013e31818fd762.
11. Yehia, L., Ni, Y., Feng, F., Seyfi, M., Sadler, T., Frazier, T.W., and Eng, C. (2019). Distinct alterations in tricarboxylic acid cycle metabolites associate with cancer and autism phenotypes in cowden syndrome and bannayan-riley-ruvalcaba syndrome. Am. J. Hum. Genet. *105*, 813–821. https://doi.org/10.1016/j.ajhg.2019.09.004.
12. Yehia, L., Ni, Y., Sadler, T., Frazier, T.W., and Eng, C. (2022). Distinct metabolic profiles associated with autism spectrum disorder versus cancer in individuals with germline PTEN mutations. NPJ Genom. Med. *7*, 16–19. https://doi.org/10.1038/s41525-022-00289-x.
13. Yehia, L., Seyfi, M., Niestroj, L.-M., Padmanabhan, R., Ni, Y., Frazier, T.W., Lal, D., and Eng, C. (2020). Copy number variation and clinical outcomes in patients with germline PTEN mutations. JAMA Netw. Open *3*, e1920415. https://doi.org/10.1001/jamanetworkopen.2019.20415.
14. Ni, Y., Zbuk, K.M., Sadler, T., Patocs, A., Lobo, G., Edelman, E., Platzer, P., Orloff, M.S., Waite, K.A., and Eng, C. (2008). Germline mutations and variants in the succinate dehydrogenase genes in cowden and cowden-like syndromes. Am. J. Hum. Genet. *83*, 261–268. https://doi.org/10.1016/j.ajhg.2008.07.011.
15. Ni, Y., He, X., Chen, J., Moline, J., Mester, J., Orloff, M.S., Ringel, M.D., and Eng, C. (2012). Germline SDHx variants modify breast and thyroid cancer risks in Cowden and Cowden-like syndrome via FAD/NAD-dependant destabilization of p53. Hum. Mol. Genet. *21*, 300–310. https://doi.org/10.1093/hmg/ddr459.
16. Alberts, B. (2017). Molecular Biology of the Cell, 6th ed. (W.W. Norton & Company). https://doi.org/10.1201/9781315735368.
17. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). Genome project data processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
18. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
19. McMahon, S., and LaFramboise, T. (2014). Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates. Carcinogenesis *35*, 1046–1054. https://doi.org/10.1093/carcin/bgu012.
20. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
21. Li, M., Schröder, R., Ni, S., Madea, B., and Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proc. Natl. Acad. Sci. USA *112*, 2491–2496. https://doi.org/10.1073/pnas.1419651112.
22. Stewart, J.B., and Chinnery, P.F. (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat. Rev. Genet. *16*, 530–542. https://doi.org/10.1038/nrg3966.
23. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013). Nucleic Acids Res. *41*, D36–D42. https://doi.org/10.1093/nar/gks1195.
24. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44*, W58–W63. https://doi.org/10.1093/nar/gkw233.
25. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. *50*, D20–D26. https://doi.org/10.1093/nar/gkab1112.
26. Geschwind, D.H. (2011). Genetics of autism spectrum disorders. Trends Cognit. Sci. *15*, 409–416. https://doi.org/10.1016/j.tics.2011.07.003.
27. Geschwind, D.H., and State, M.W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. *14*, 1109–1120. https://doi.org/10.1016/S1474-4422(15)00044-7.
28. Kim, Y., Vadodaria, K.C., Lenkei, Z., Kato, T., Gage, F.H., Marchetto, M.C., and Santos, R. (2019). Mitochondria, metabolism, and redox mechanisms in psychiatric disorders. Antioxidants Redox Signal. *31*, 275–317. https://doi.org/10.1089/ars.2018.7606.
29. Patowary, A., Nesbitt, R., Archer, M., Bernier, R., and Brkanac, Z. (2017). Next generation sequencing mitochondrial DNA analysis in autism spectrum disorder. Autism Res. *10*, 1338–1343. https://doi.org/10.1002/aur.1792.
30. Chalkia, D., Singh, L.N., Leipzig, J., Lvova, M., Derbeneva, O., Lakatos, A., Hadley, D., Hakonarson, H., and Wallace, D.C. (2017). Association between mitochondrial DNA haplogroup variation and autism spectrum disorders. JAMA Psychiatr. *74*, 1161–1168. https://doi.org/10.1001/jamapsychiatry.2017.2604.
31. Piryaei, F., Houshmand, M., Aryani, O., Dadgar, S., and Soheili, Z.-S. (2012). Investigation of the mitochondrial ATPase 6/8 and tRNA(lys) genes mutations in autism. Cell J. *14*, 98–101.
32. Weissman, J.R., Kelley, R.I., Bauman, M.L., Cohen, B.H., Murray, K.F., Mitchell, R.L., Kern, R.L., and Natowicz, M.R. (2008). Mitochondrial disease in autism spectrum disorder patients: a cohort analysis. PLoS One *3*, e3815. https://doi.org/10.1371/journal.pone.0003815.
33. Clima, R., Preste, R., Calabrese, C., Diroma, M.A., Santorsola, M., Scioscia, G., Simone, D., Shen, L., Gasparre, G., and Attimonelli, M. (2017). HmtDB 2016: data update, a better performing query system and human mitochondrial DNA haplogroup predictor. Nucleic Acids Res. *45*, D698–D706. https://doi.org/10.1093/nar/gkw1066.
34. Ruiz-Pesini, E., Lott, M.T., Procaccio, V., Poole, J.C., Brandon, M.C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., and Wallace, D.C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res. *35*, D823–D828. https://doi.org/10.1093/nar/gkl927.
35. Grandhi, S., Bosworth, C., Maddox, W., Sensiba, C., Akhavanfard, S., Ni, Y., and LaFramboise, T. (2017). Heteroplasmic shifts in tumor mitochondrial genomes reveal tissue-specific signals of relaxed and positive selection. Hum. Mol. Genet. *26*, 2912–2922. https://doi.org/10.1093/hmg/ddx172.

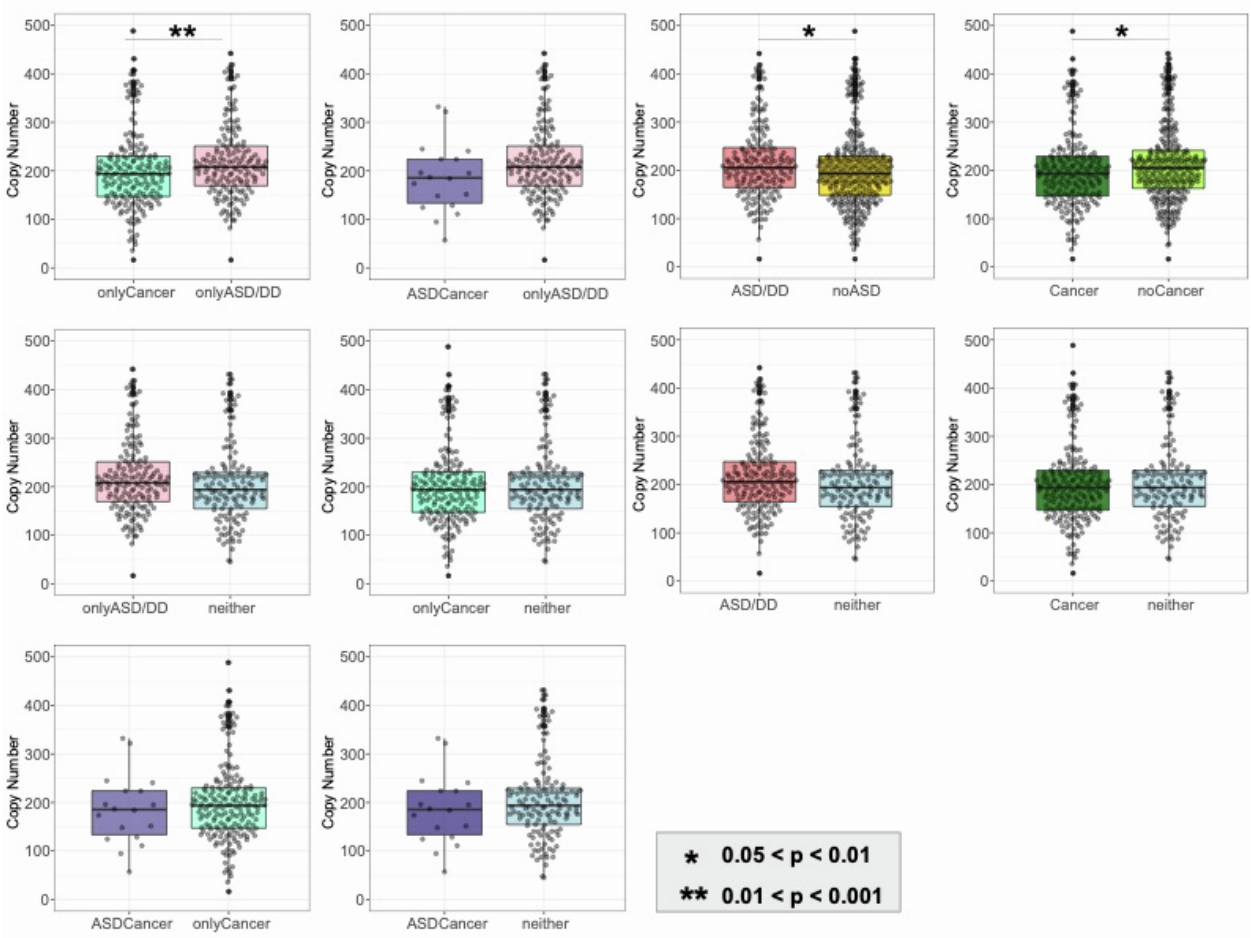
HGGA, Volume *4*

Supplemental information

The mitochondrial genome as a modifier of autism versus cancer phenotypes in *PTEN* hamartoma tumor syndrome

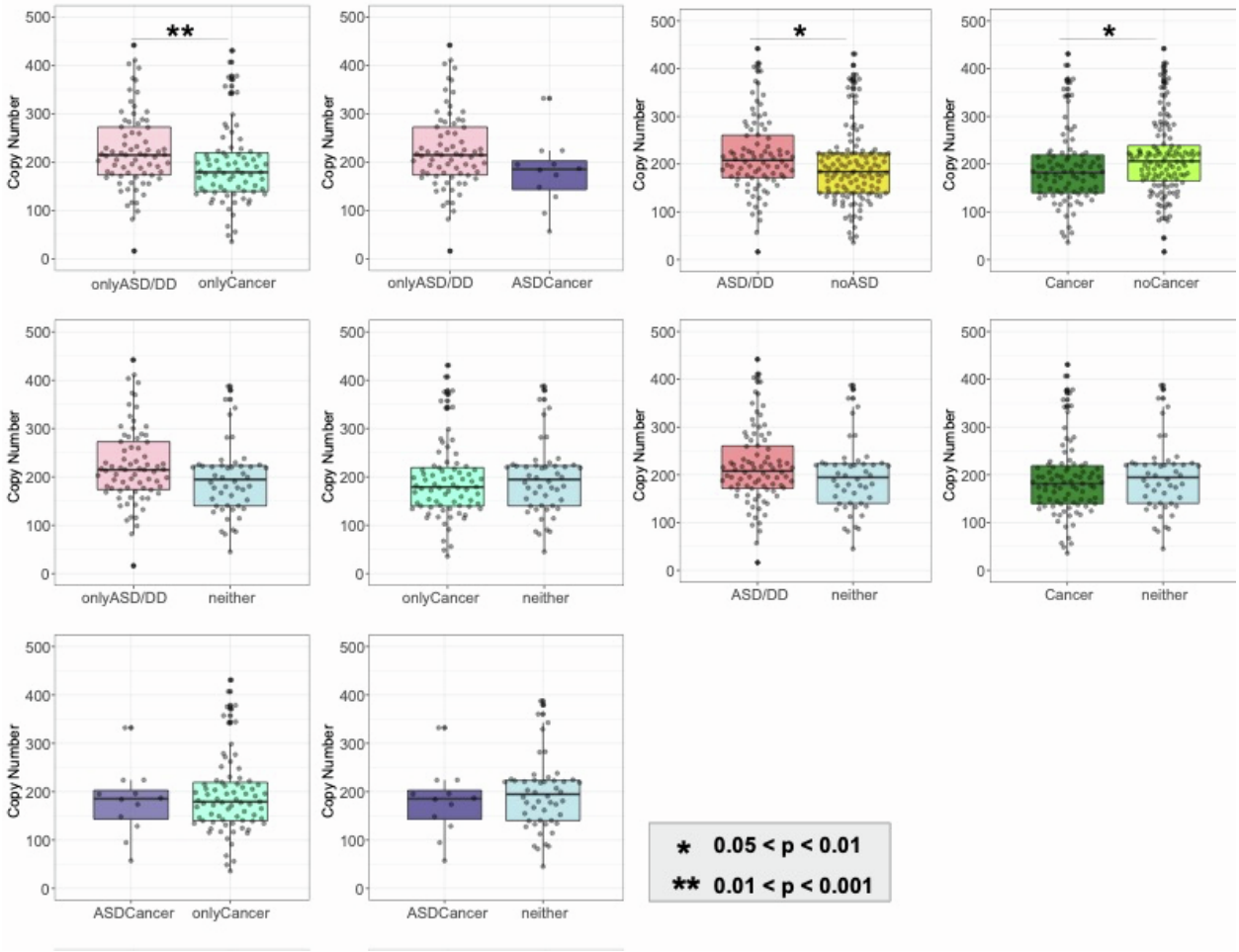
Ruipeng Wei, Lamis Yehia, Ying Ni, and Charis Eng

Figure S1. Beeswarm plots of mtDNA CN differences between different PHTS phenotypes within all research participants (n=498)



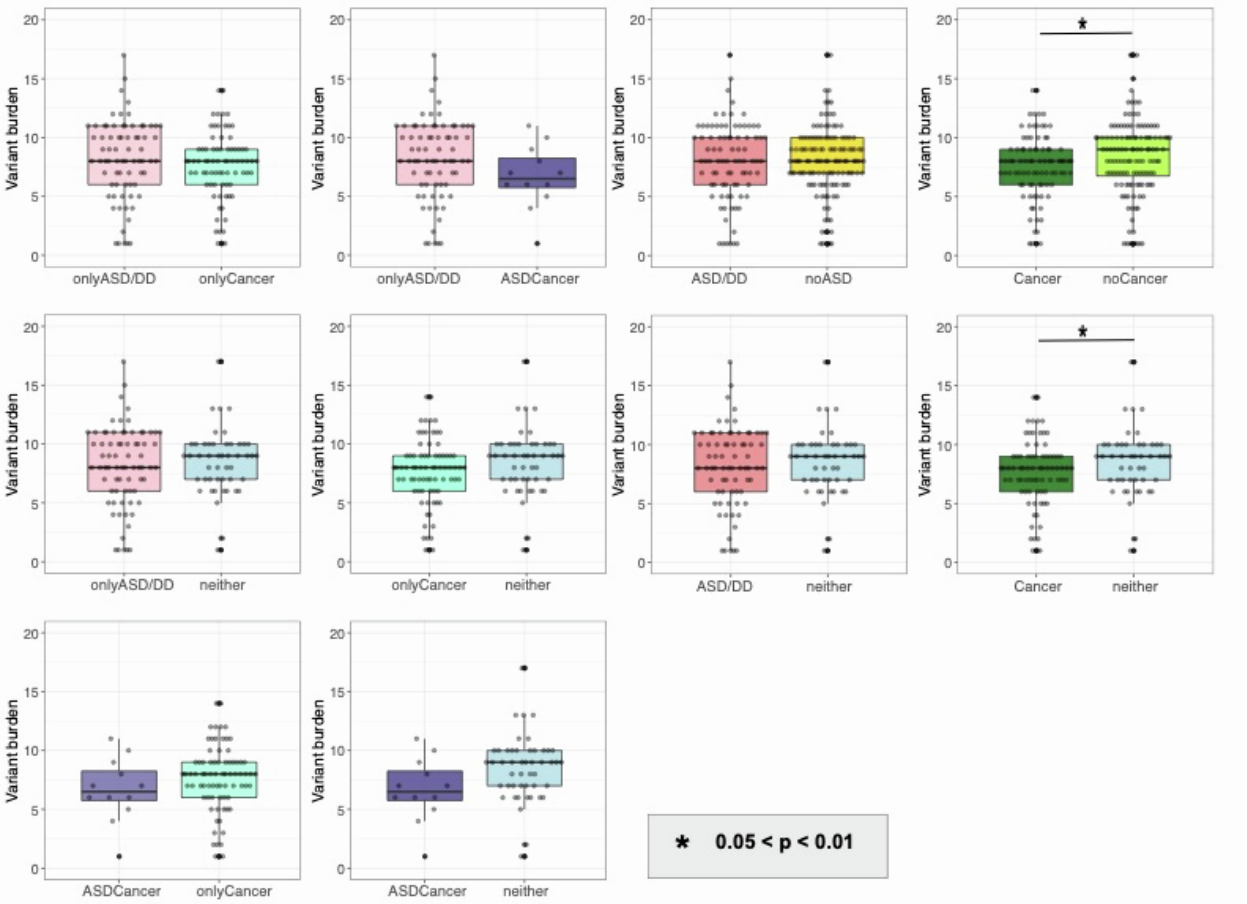
The beeswarm plots represent 10 mtDNA CN comparisons between different PHTS phenotypes within all PHTS samples. Each dot represents one sample's mtDNA CN value. The upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points and are plotted individually. Different colors indicate different PHTS phenotypes. PHTS-ASD/DD group has higher mtDNA CN than PHTS individuals without ASD/DD.

Figure S2. Beeswarm plots of mtDNA CN differences between different PHTS phenotypes within the H haplogroup (n=243)



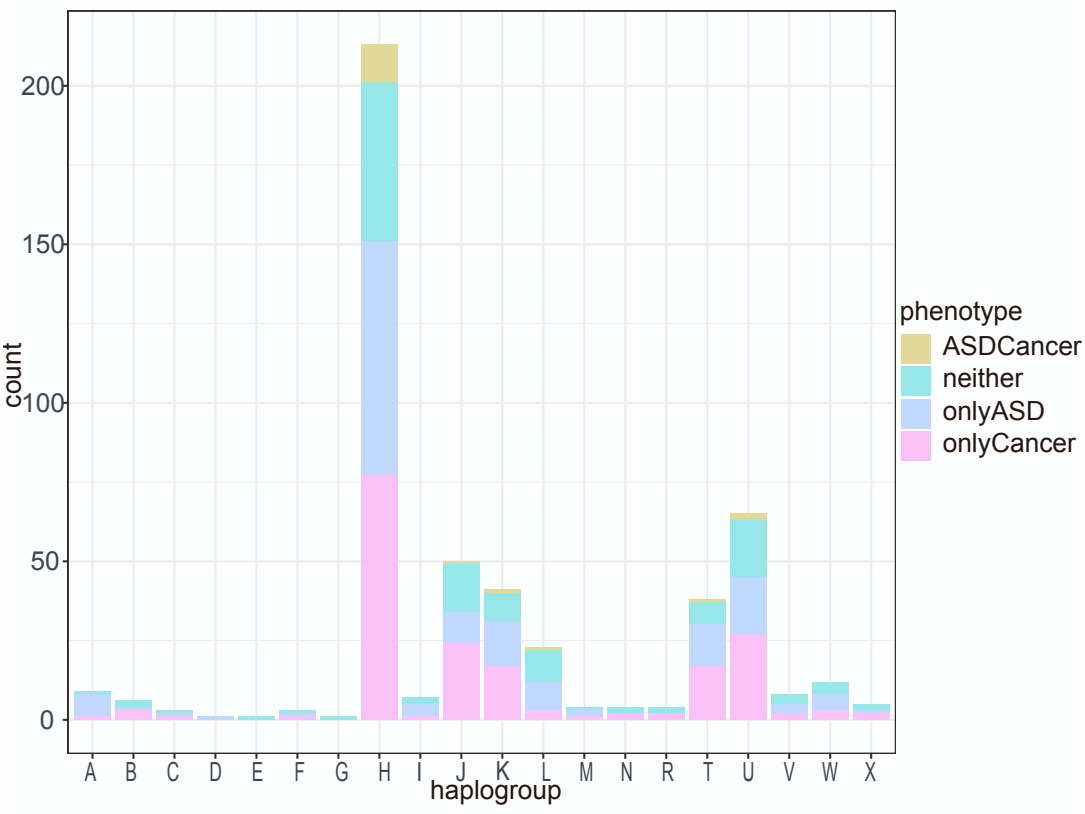
The beeswarm plots represent 10 mtDNA CN comparisons between different PHTS phenotypes within the H haplogroup. Each dot represents one sample's mtDNA CN value. The upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points and are plotted individually. Different colors indicate different PHTS phenotypes. PHTS-ASD/DD group has higher mtDNA CN than PHTS individuals without ASD/DD.

Figure S3. Beeswarm plots of SNV burden differences between different PHTS phenotypes within the H haplogroup (n=243)



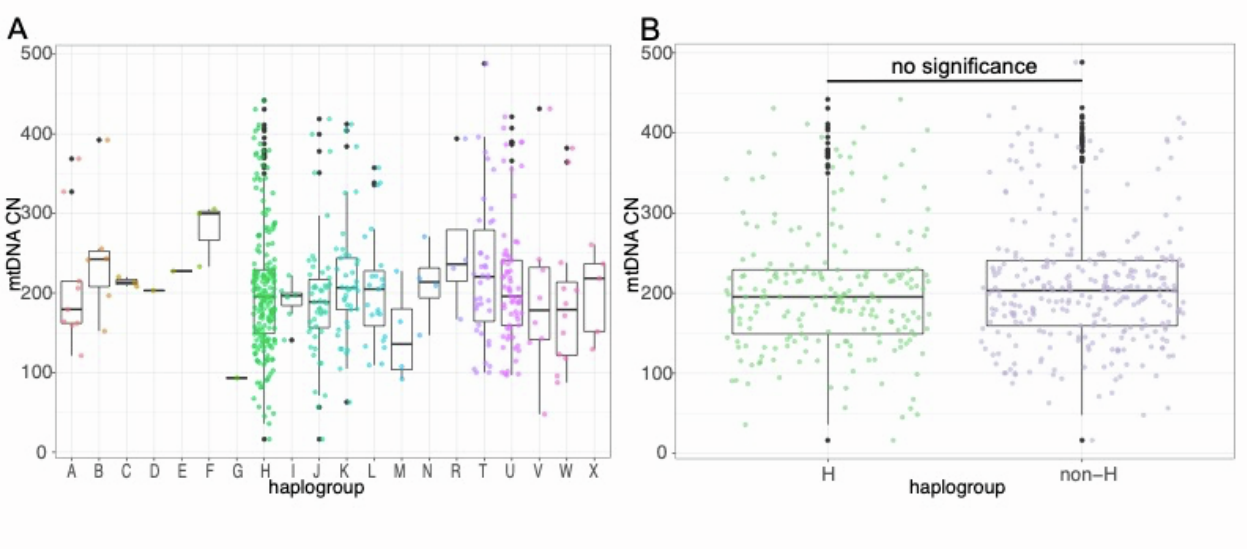
The beeswarm plots represent 10 mtDNA SNV burden comparisons between different PHTS phenotypes within the H haplogroup. Each dot represents one sample's mtDNA CN value. The upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points and are plotted individually. Different colors indicate different PHTS phenotypes. PHTS-noCancer group has higher SNV burden than those PHTS individuals who have cancer or who have both ASD/DD and cancer.

Figure S4. mtDNA phenotype distribution among each haplogroup



The bar plot represents the phenotype distribution across the different PHTS phenotypes within all samples (n = 498).

Figure S5. mtDNA CN distribution among the haplogroups



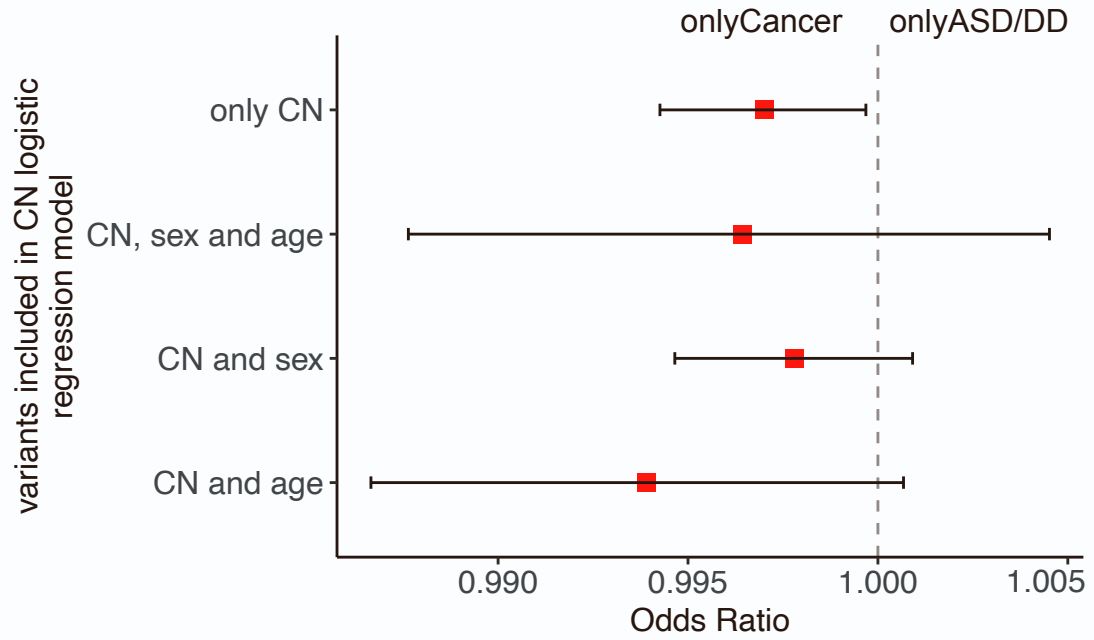
A. Beeswarm plots of copy number among all haplogroups in our sample set (n=498).

B. Beeswarm plots of copy number among H and non-H haplogroups in our sample set (n=498).

There are no statistically significant differences in mtDNA CN between the H haplogroup and non-H haplogroup.

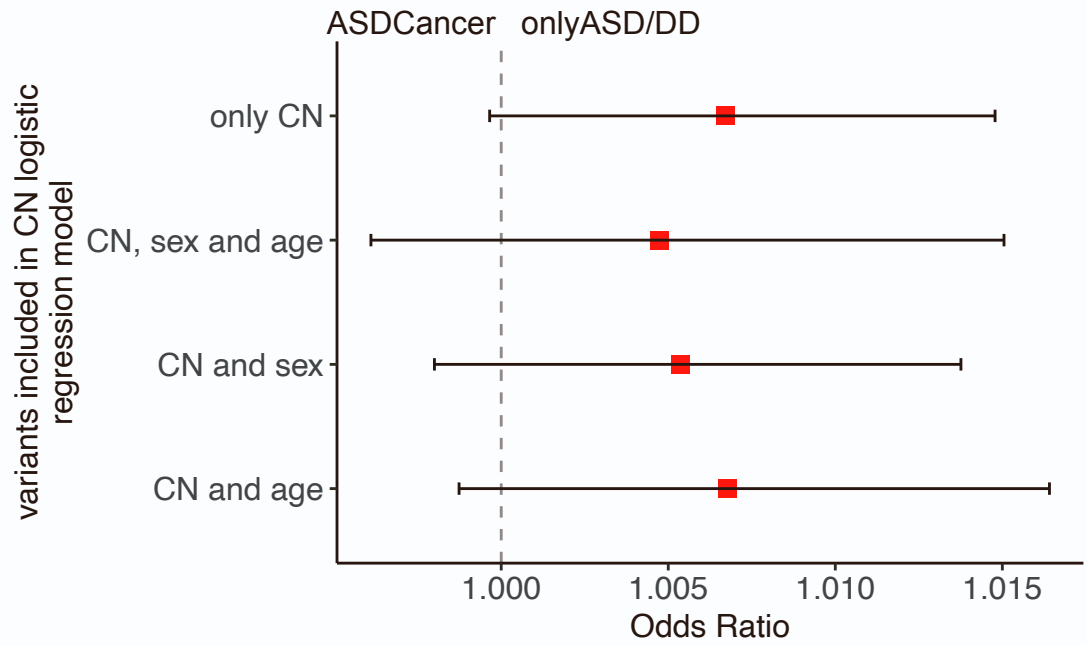
Each dot represents one sample's mtDNA CN value. The upper whisker extends from the hinge to the largest value no further than 1.5 * inter-quartile range (IQR) from the hinge. The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are "outlying" points and are plotted individually.

Figure S6. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-onlyCancer groups within all haplogroups (n = 498).



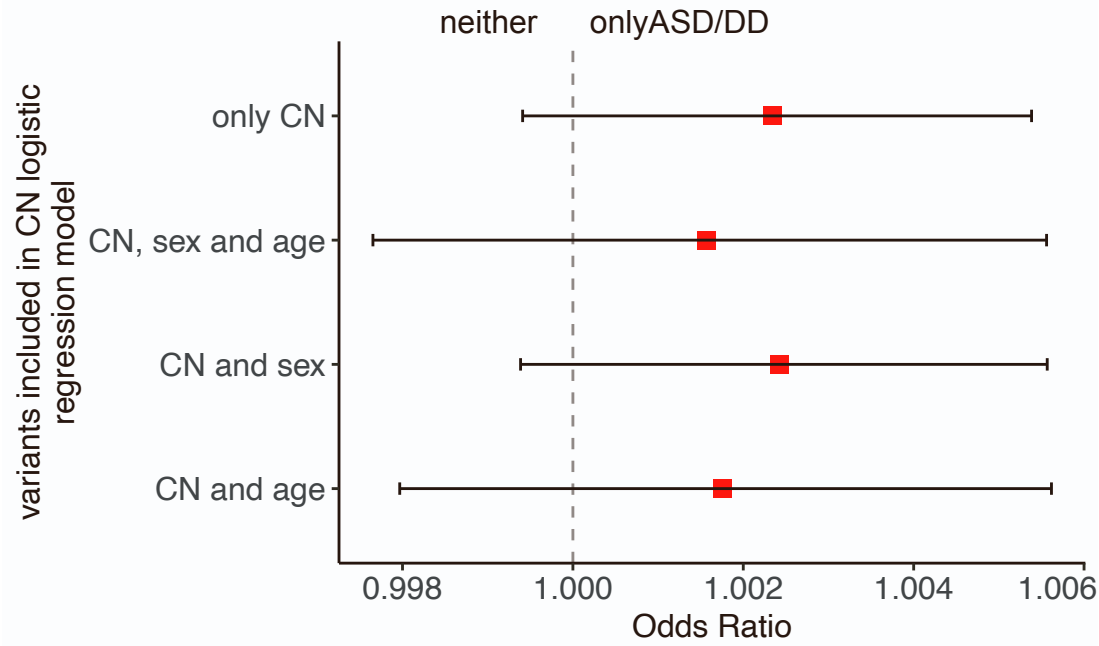
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S7. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-ASDCancer groups within all haplogroups (n = 498).



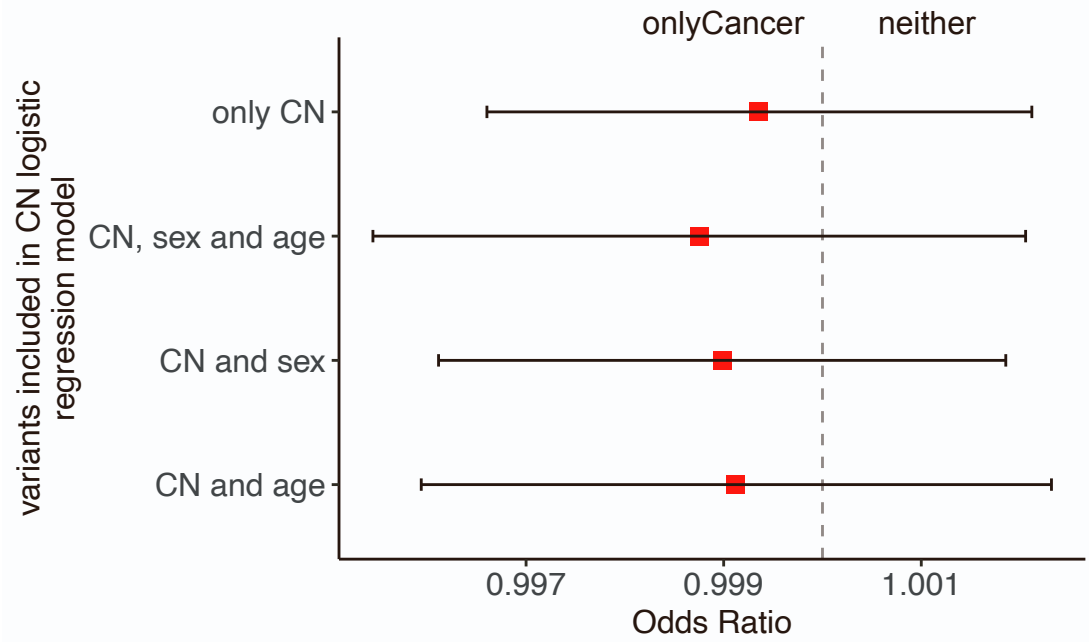
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S8. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-neither groups within all haplogroups (n = 498).



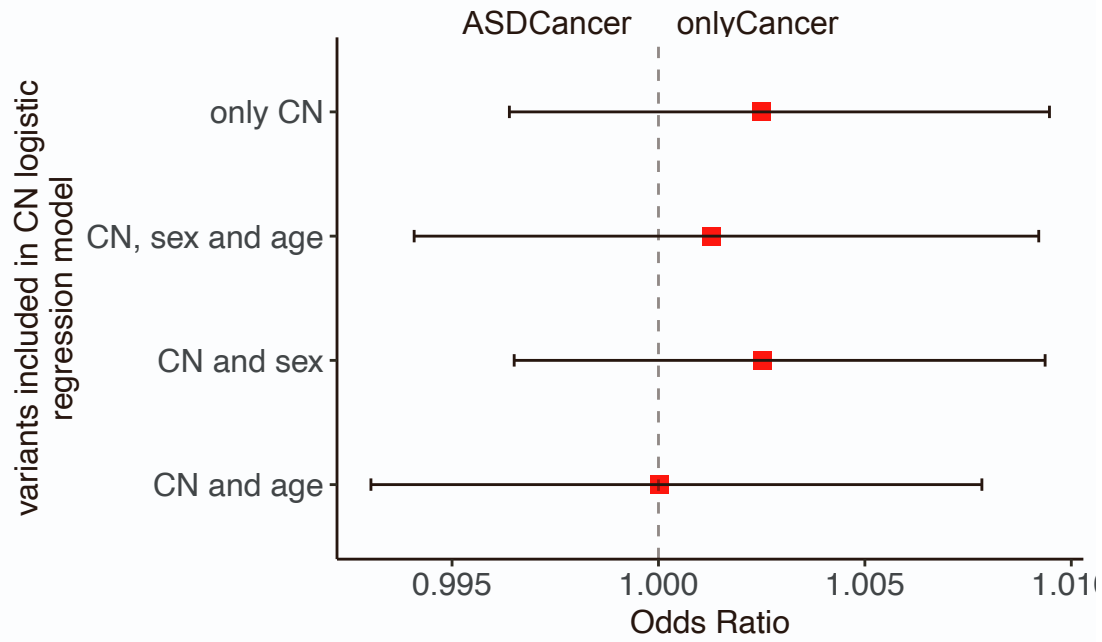
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S9. Forest plot of mtDNA CN analysis among PHTS-neither and PHTS-onlyCancer groups within all haplogroups (n = 498).



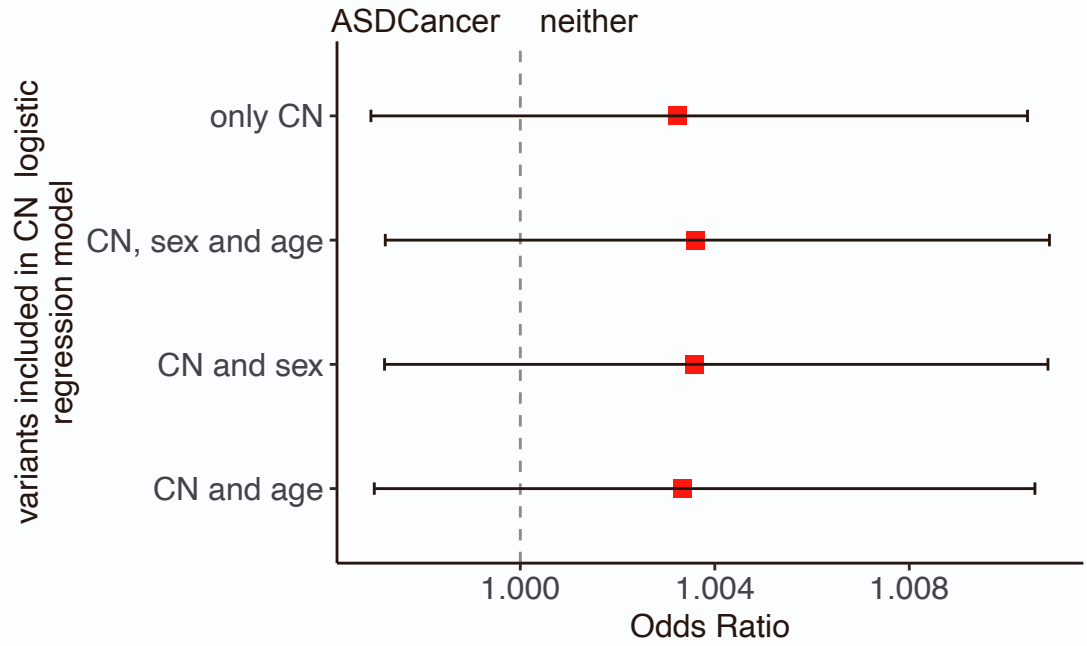
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S10. Forest plot of mtDNA CN analysis among PHTS-ASDCancer and PHTS-onlyCancer groups within all haplogroups (n = 498).



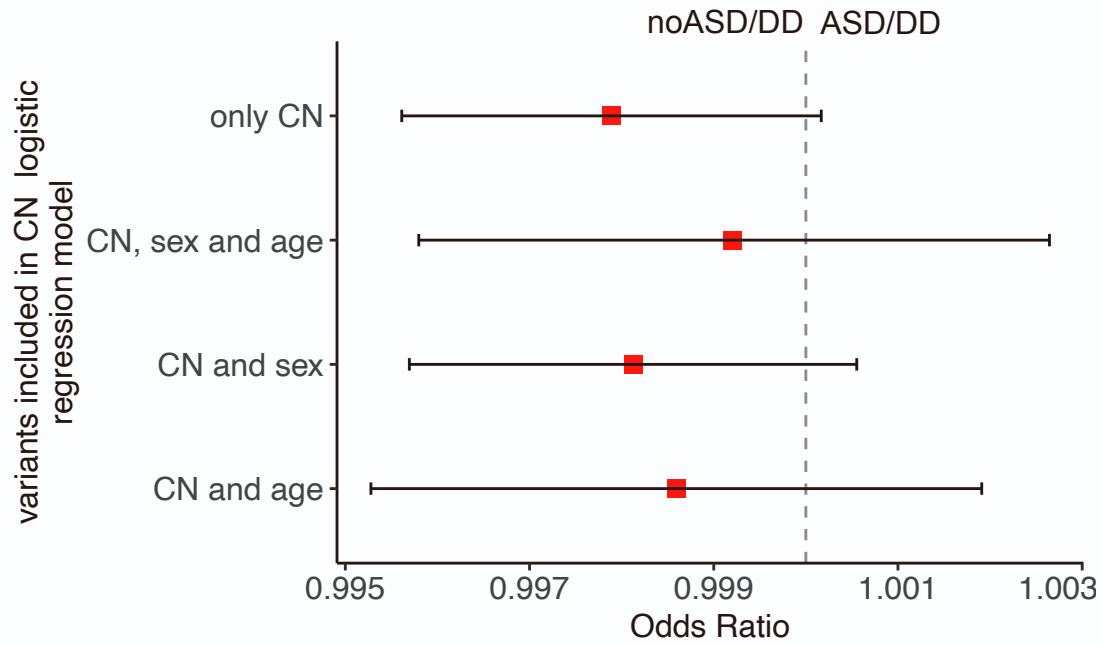
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S11. Forest plot of mtDNA CN analysis among PHTS-ASDCancer and PHTS-neither groups within all haplogroups (n = 498).



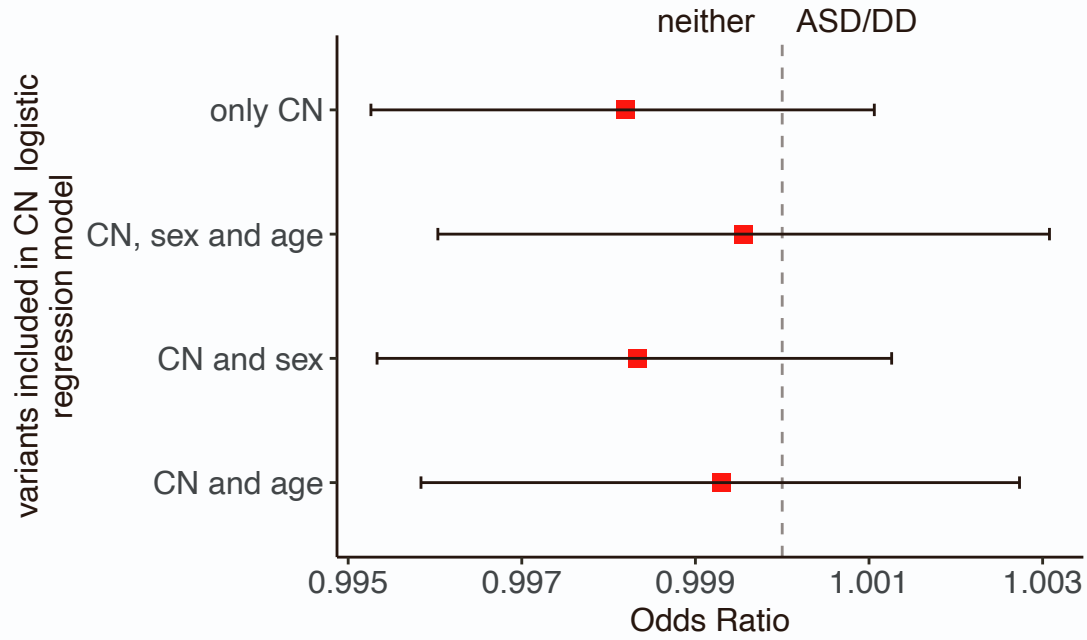
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S12. Forest plot of mtDNA CN analysis among PHTS-ASD/DD and PHTS-noASr groups within all haplogroups (n = 498).



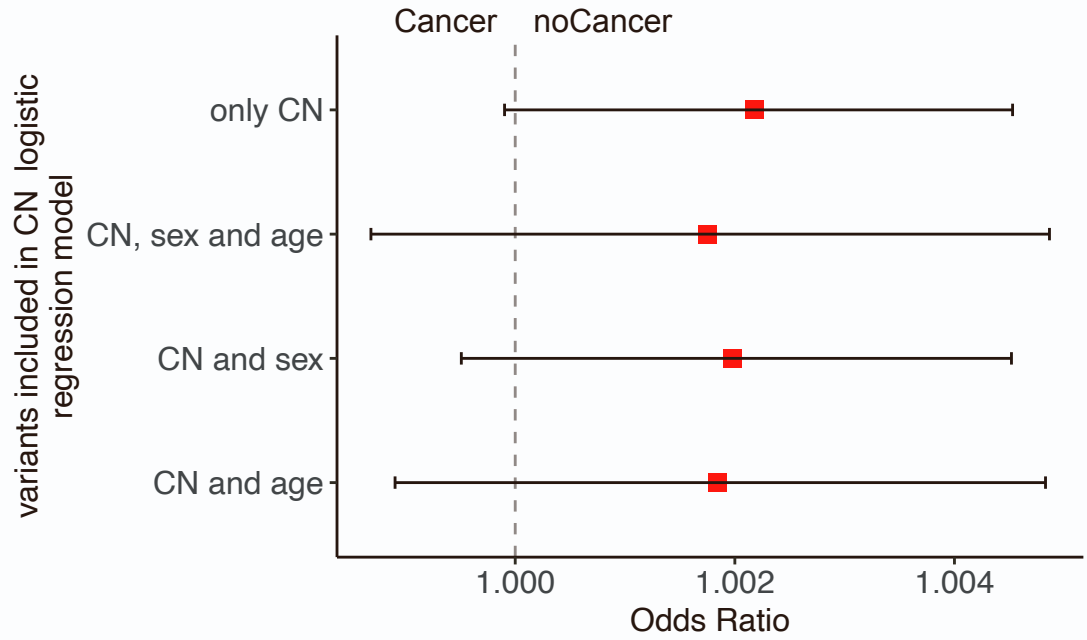
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S13. Forest plot of mtDNA CN analysis among PHTS-ASD/DD and PHTS-neither groups within all haplogroups (n = 498).



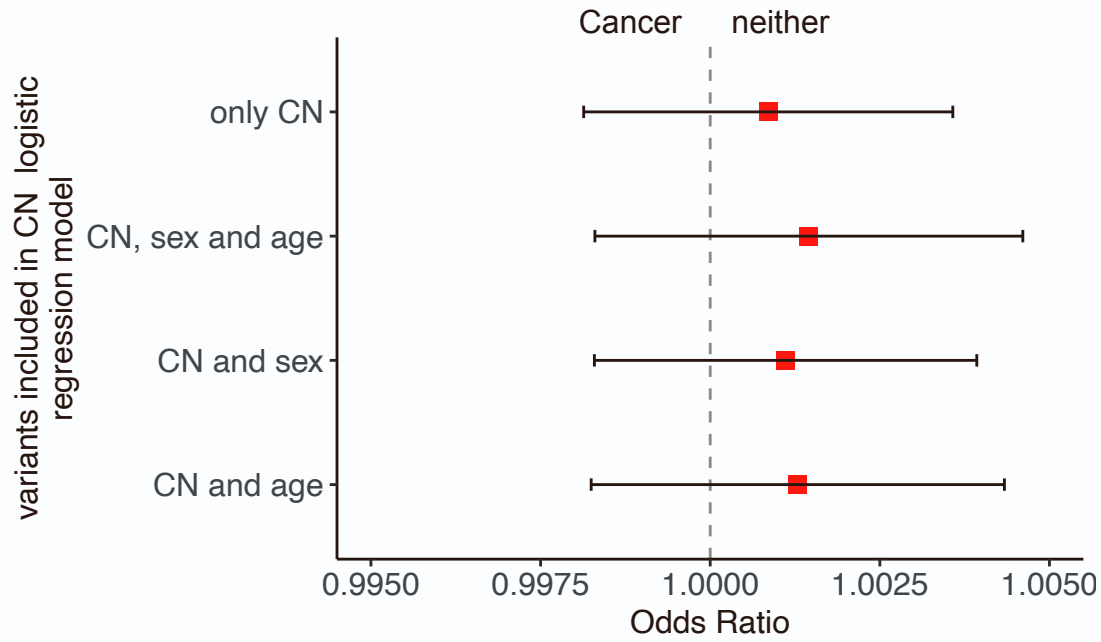
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S14. Forest plot of mtDNA CN analysis among PHTS-Cancer and PHTS-noCancer groups within all haplogroups (n = 498).



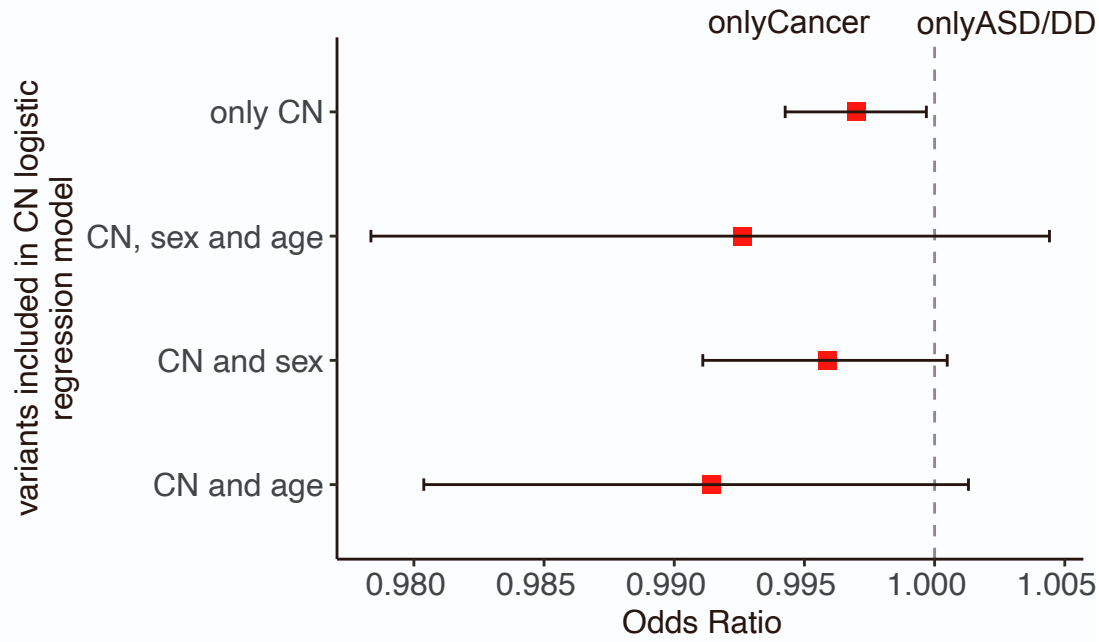
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S15. Forest plot of mtDNA CN analysis among PHTS-neither and PHTS-noCancer groups within all haplogroups (n = 498).



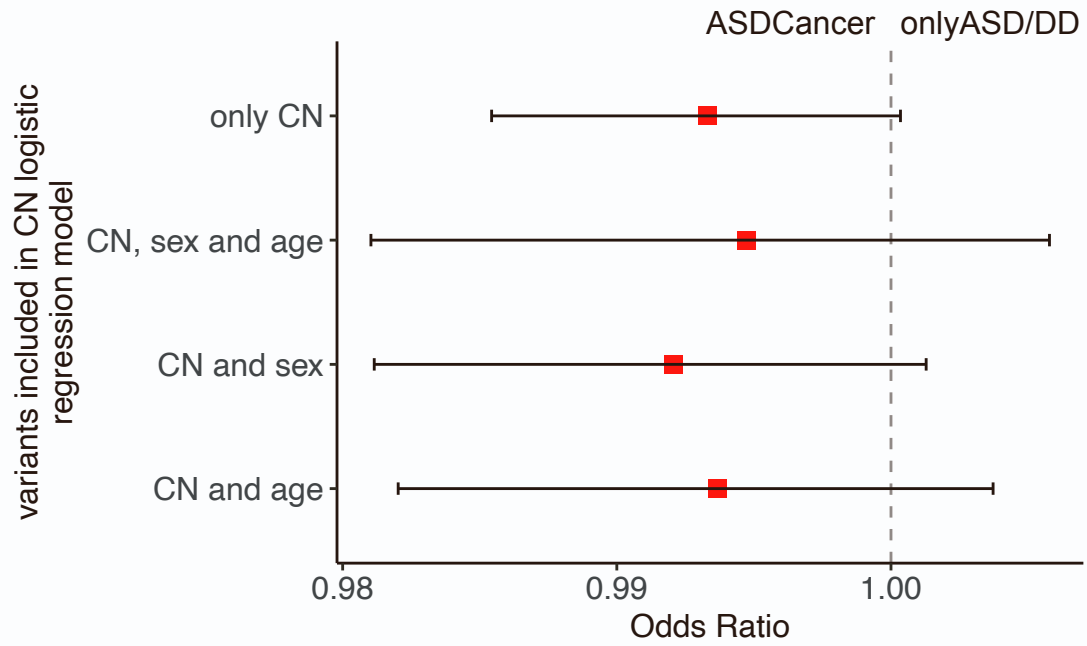
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S16. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-onlyCancer groups within the H haplogroup (n = 213).



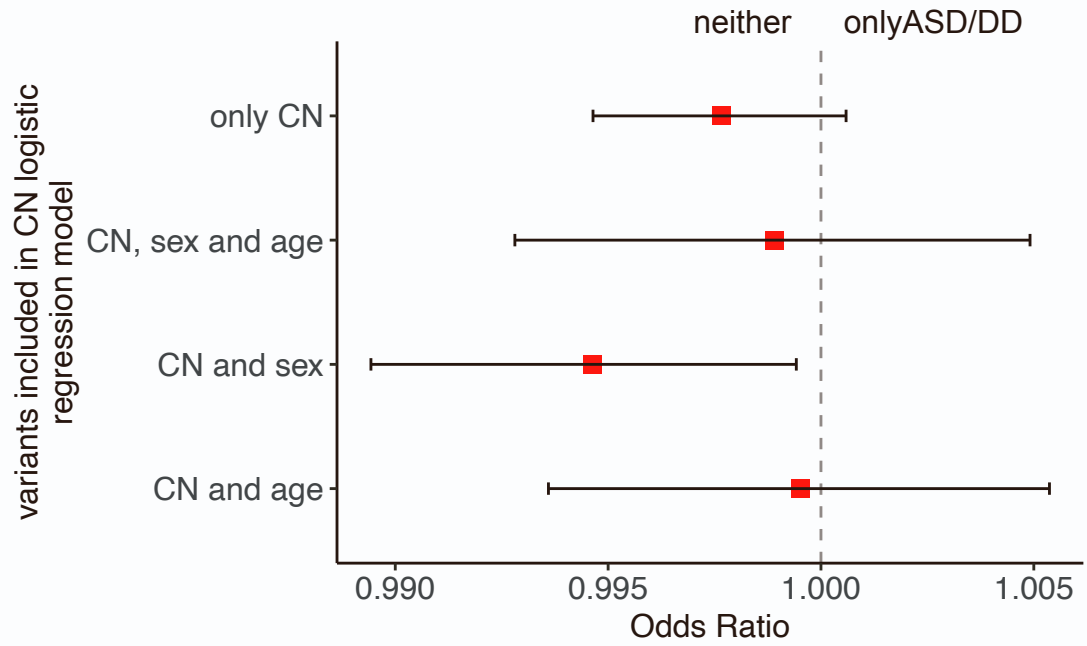
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S17. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-ASDCancer groups within the H haplogroup (n = 213).



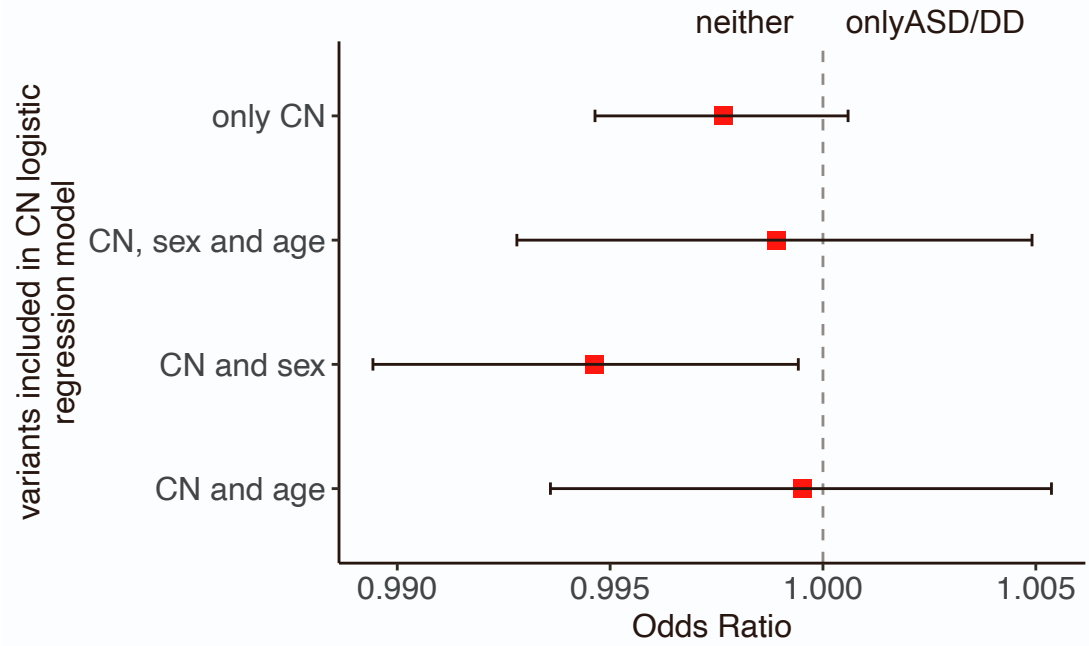
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S18. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-neither groups within the H haplogroup (n = 213).



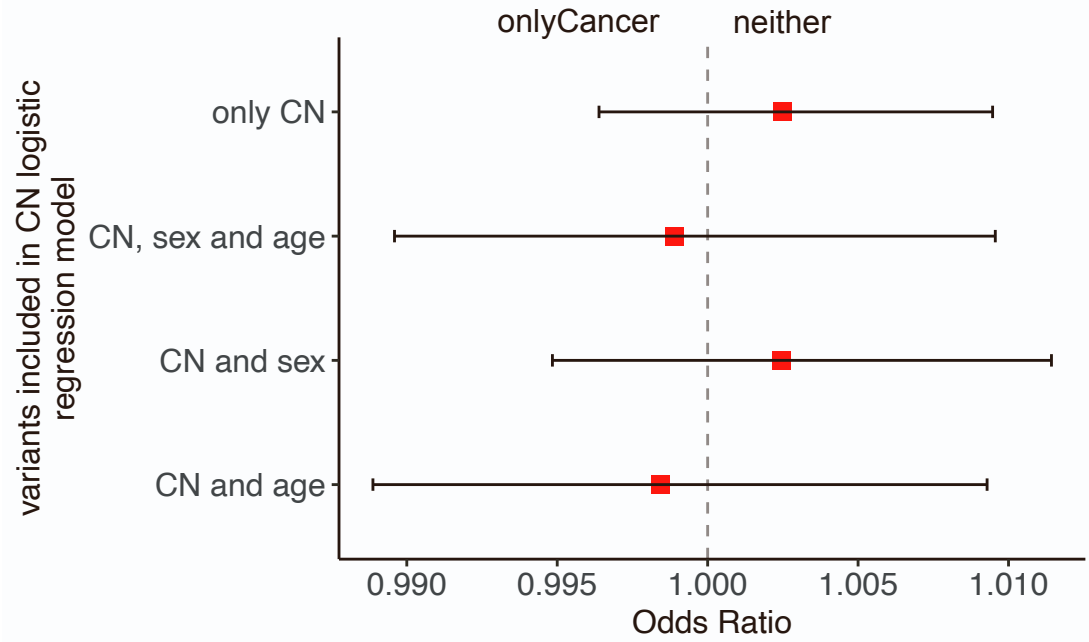
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S19. Forest plot of mtDNA CN analysis among PHTS-onlyASD/DD and PHTS-neither groups within the H haplogroup (n = 213).



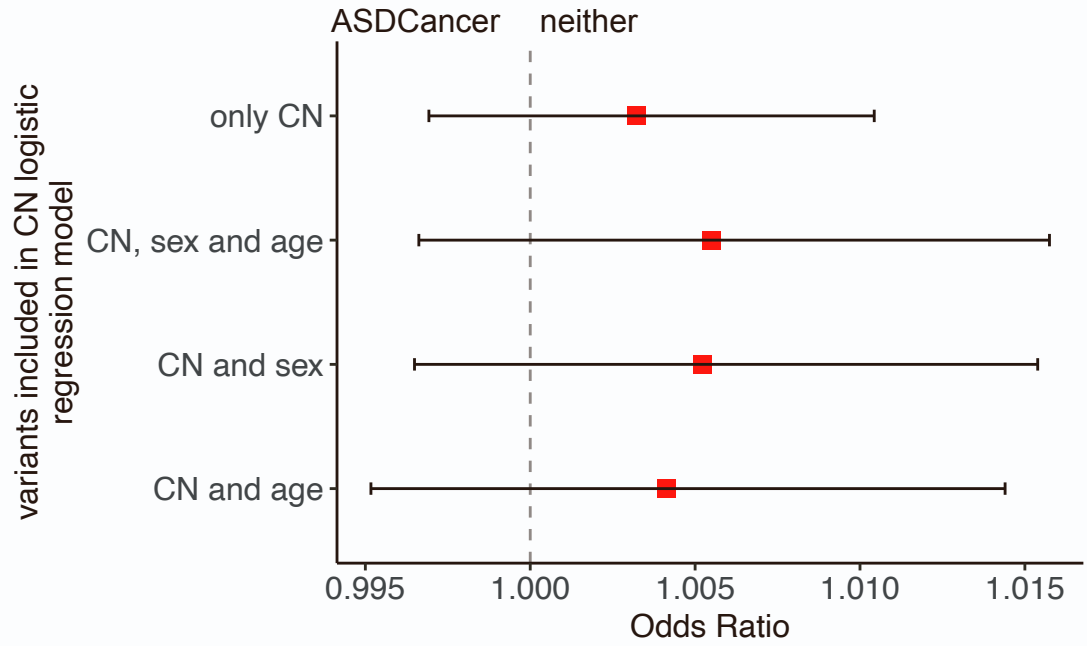
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S20. Forest plot of mtDNA CN analysis among PHTS-neither and PHTS-onlyCancer groups within the H haplogroup (n = 213).



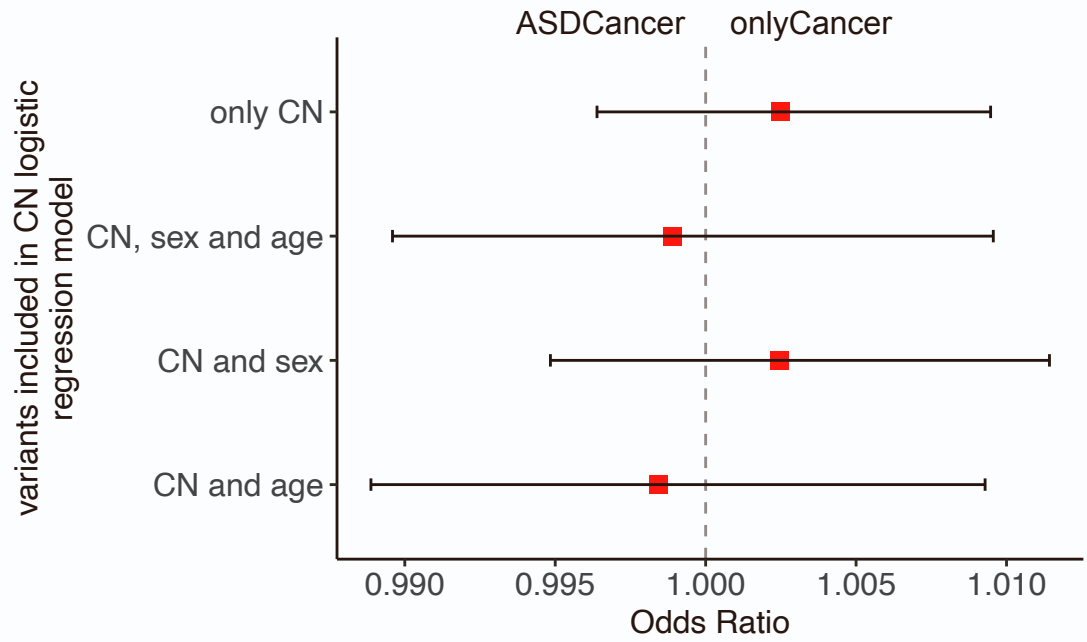
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S21. Forest plot of mtDNA CN analysis among PHTS-ASDCancer and PHTS-neither groups within the H haplogroup (n = 213).



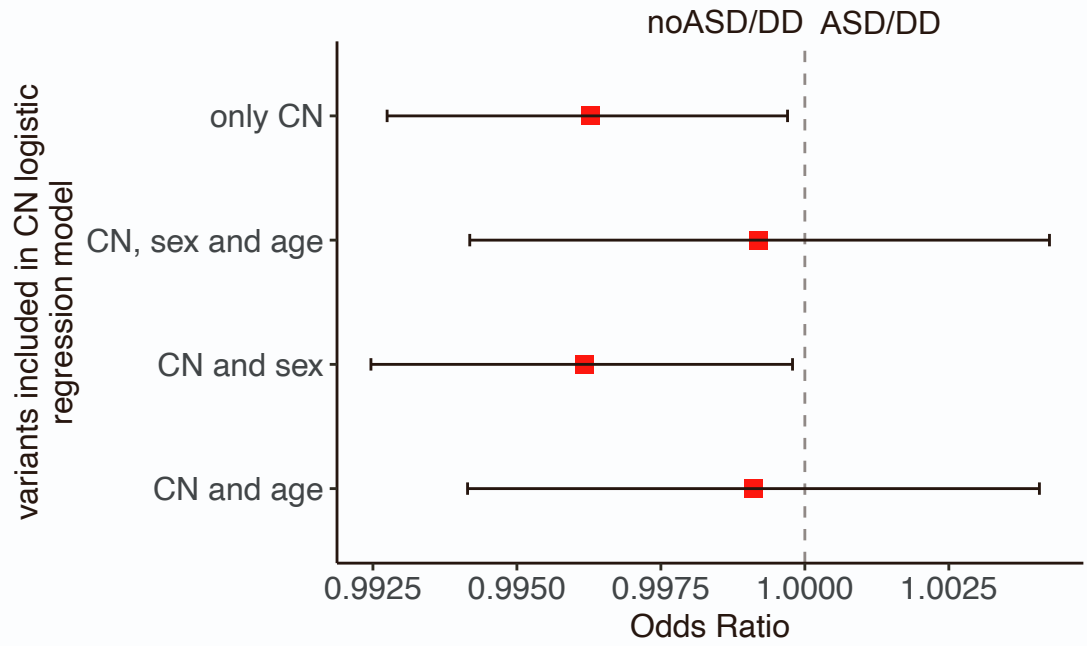
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S22. Forest plot of mtDNA CN analysis among PHTS-ASDCancer and PHTS-onlyCancer groups within the H haplogroup (n = 213).



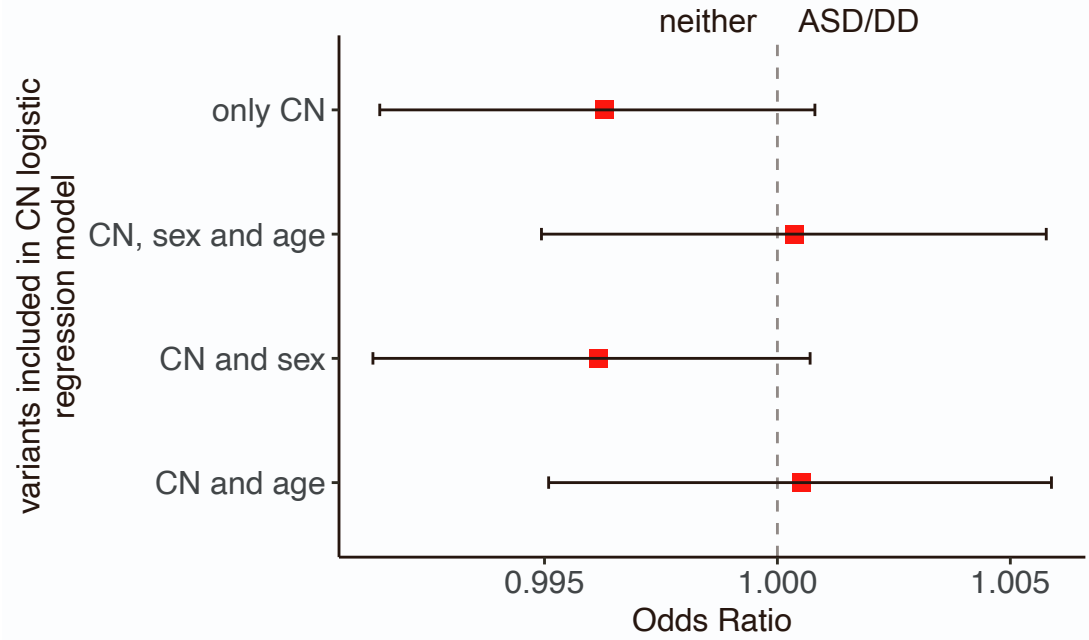
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S23. Forest plot of mtDNA CN analysis among PHTS-ASD/DD and PHTS-noASD/DD groups within the H haplogroup (n = 213).



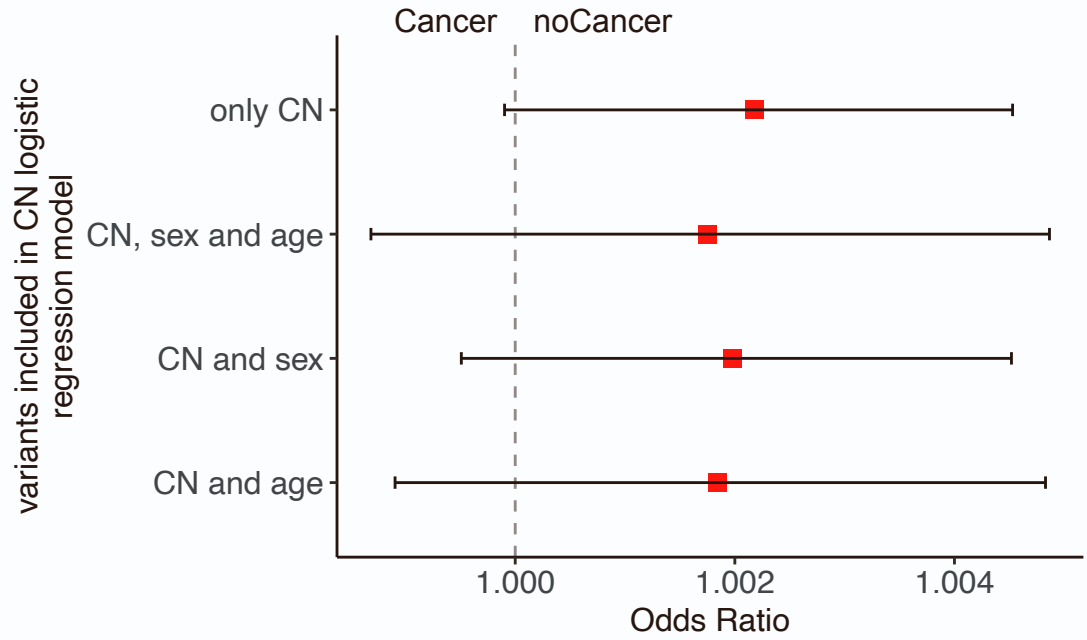
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S24. Forest plot of mtDNA CN analysis among PHTS-ASD/DD and PHTS-neither groups within the H haplogroup (n = 213).



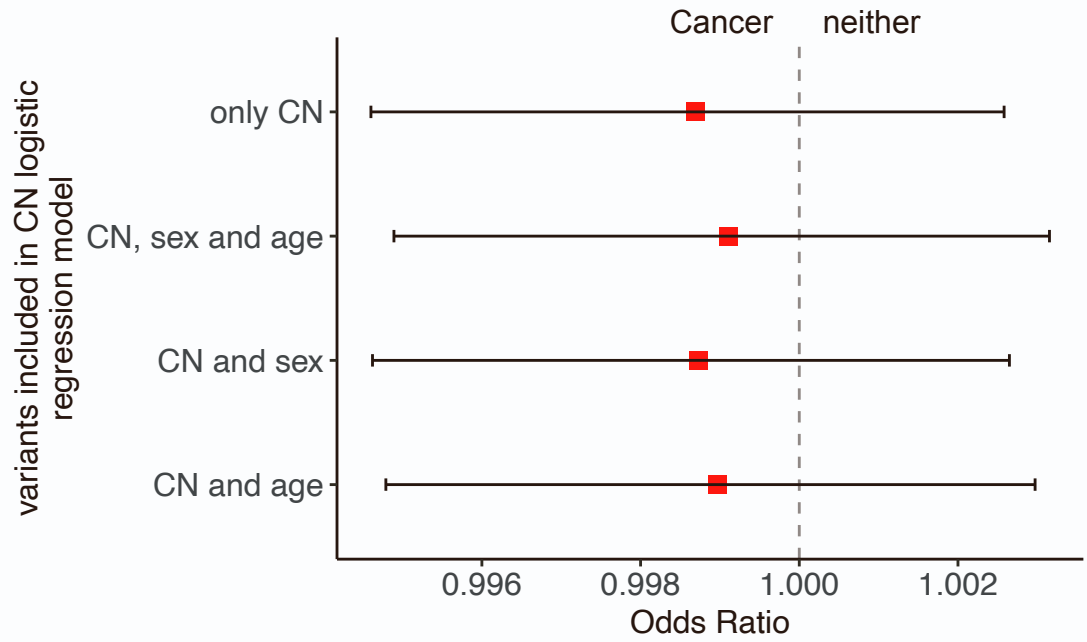
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S25. Forest plot of mtDNA CN analysis among PHTS-Cancer and PHTS-noCancer groups within the H haplogroup (n = 213).



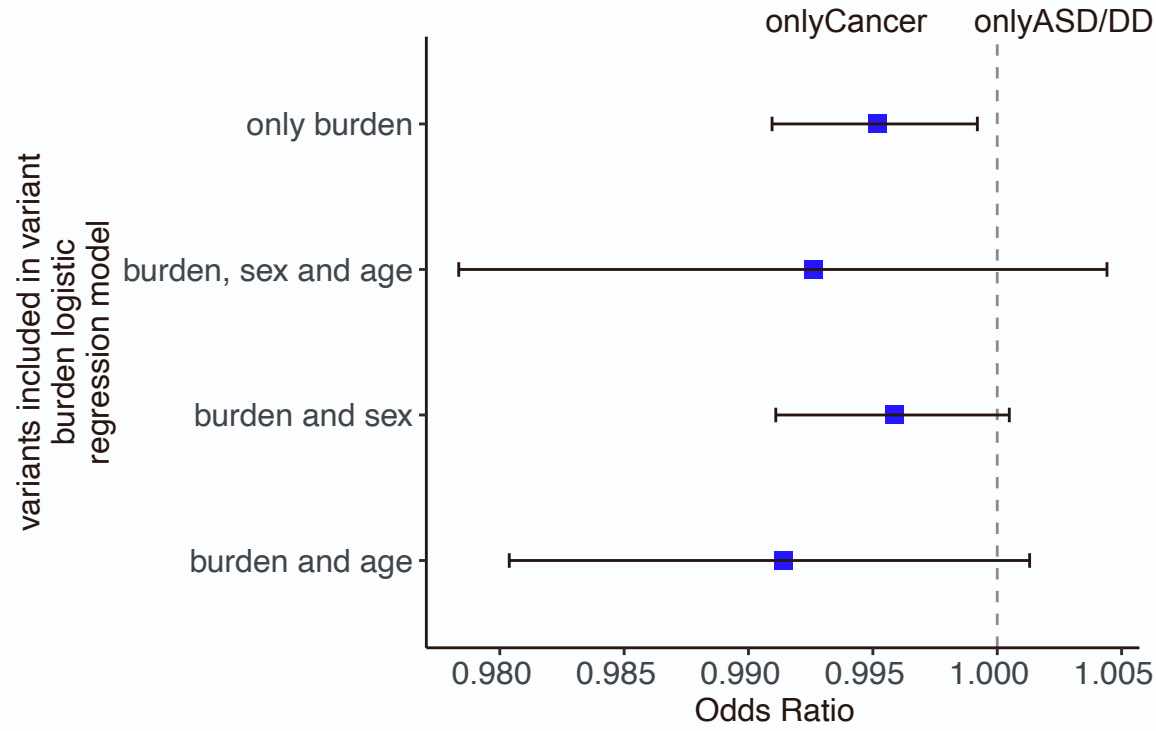
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S26. Forest plot of mtDNA CN analysis among PHTS-Cancer and PHTS-neither groups within the H haplogroup (n = 213).



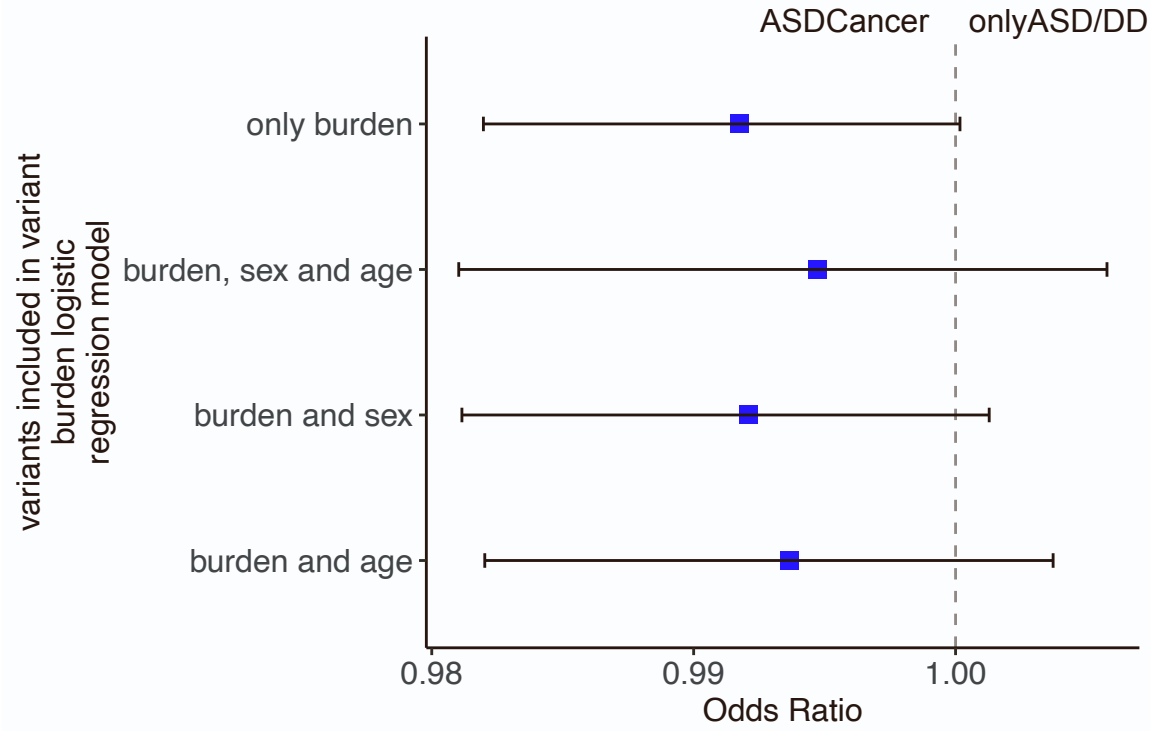
The red squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S27. Forest plot of mtDNA variant burden analysis among PHTS-onlyASD/DD and PHTS-onlyCancer groups within the H haplogroup (n = 213).



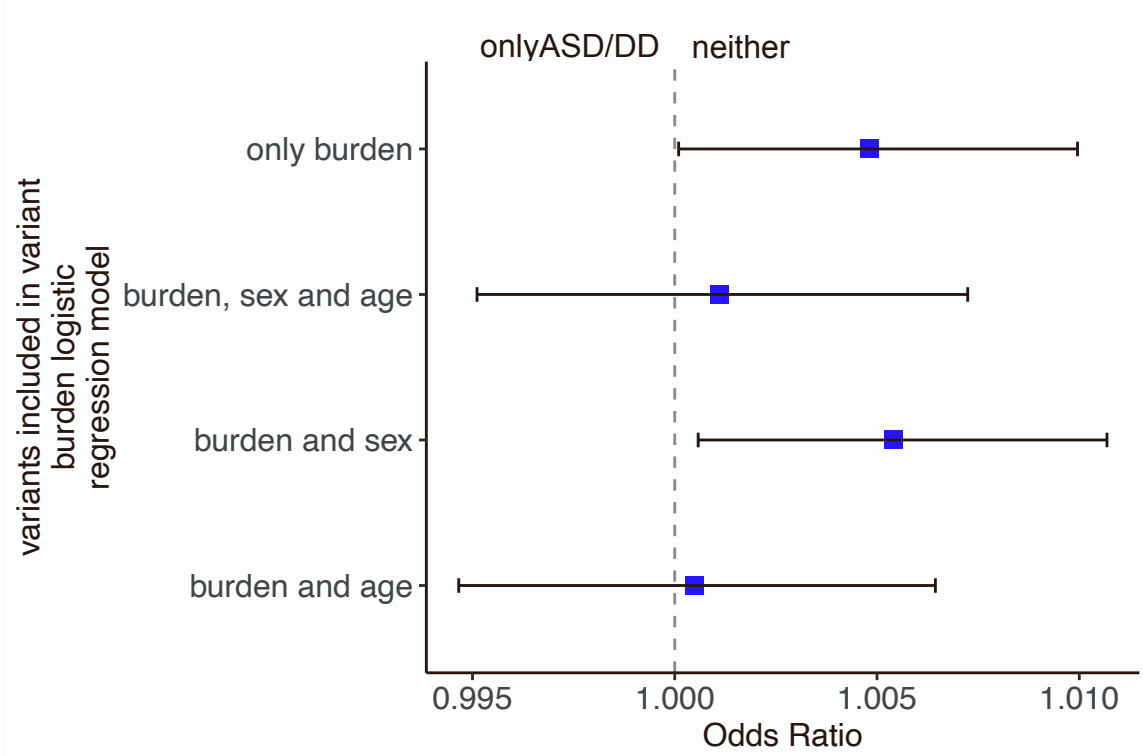
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S28. Forest plot of mtDNA variant burden analysis among PHTS-ASD/DD and PHTS-ASDCancer groups within the H haplogroup (n = 213).



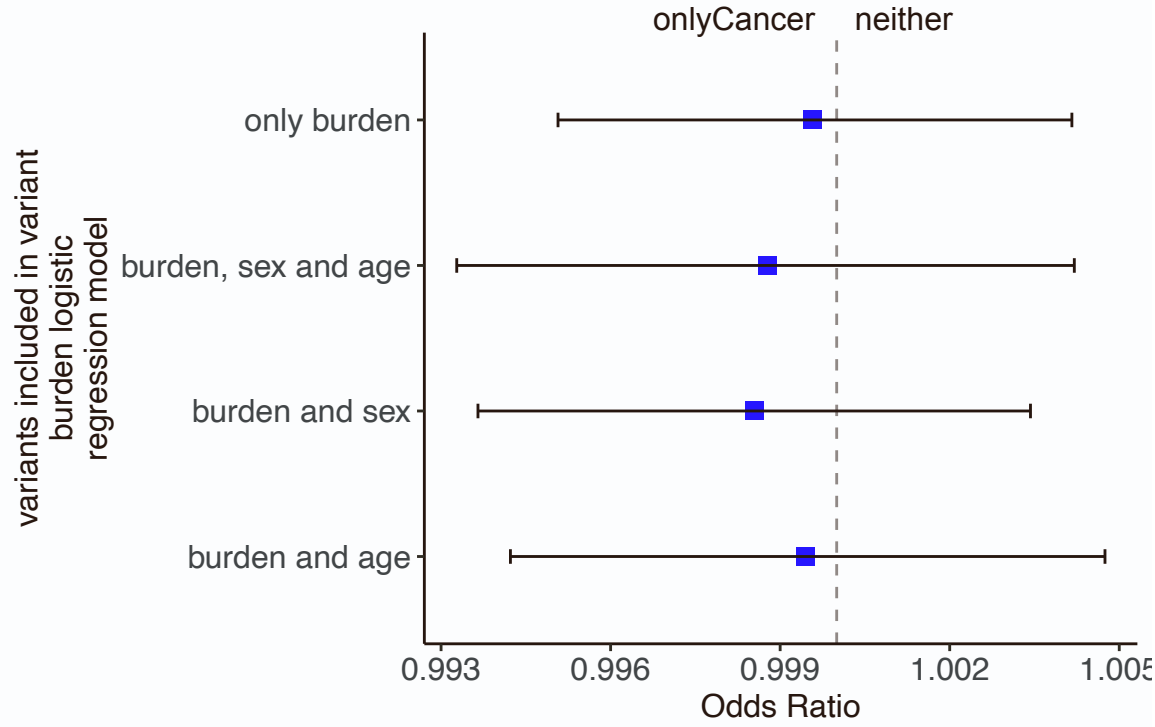
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S29. Forest plot of mtDNA variant burden analysis among PHTS-onlyASD/DD and PHTS-neither groups within the H haplogroup (n = 213).



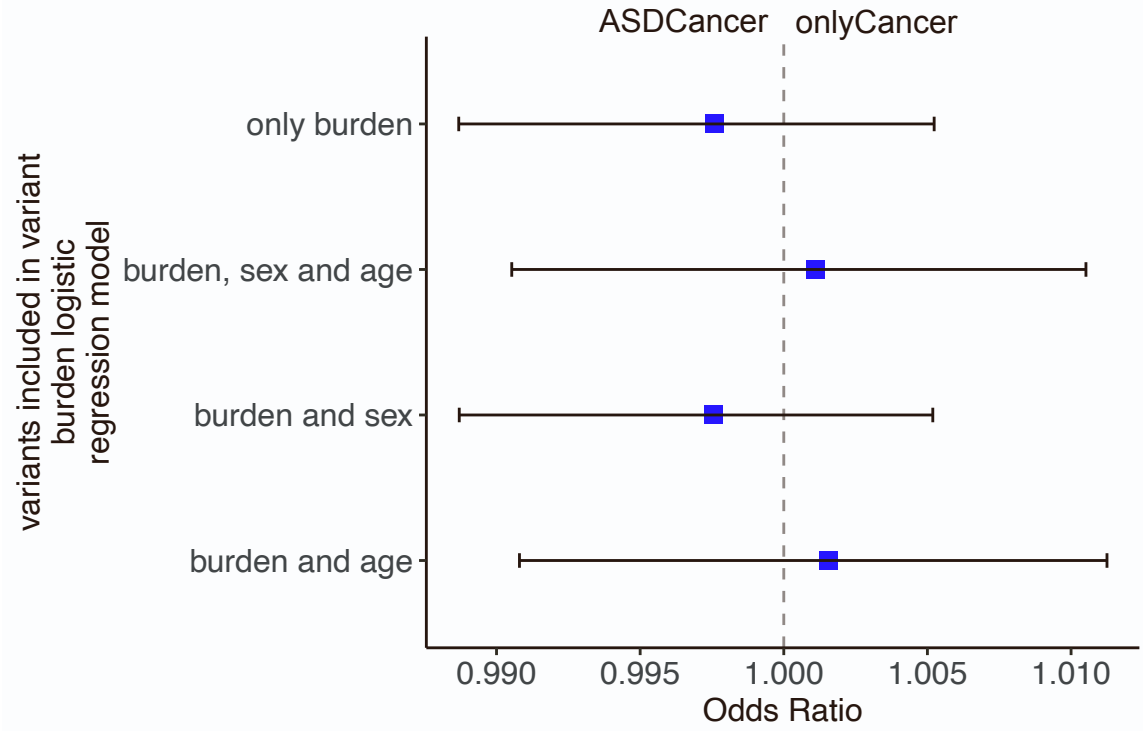
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S30. Forest plot of mtDNA variant burden analysis among PHTS-onlyCancer and PHTS-neither groups within the H haplogroup (n = 213).



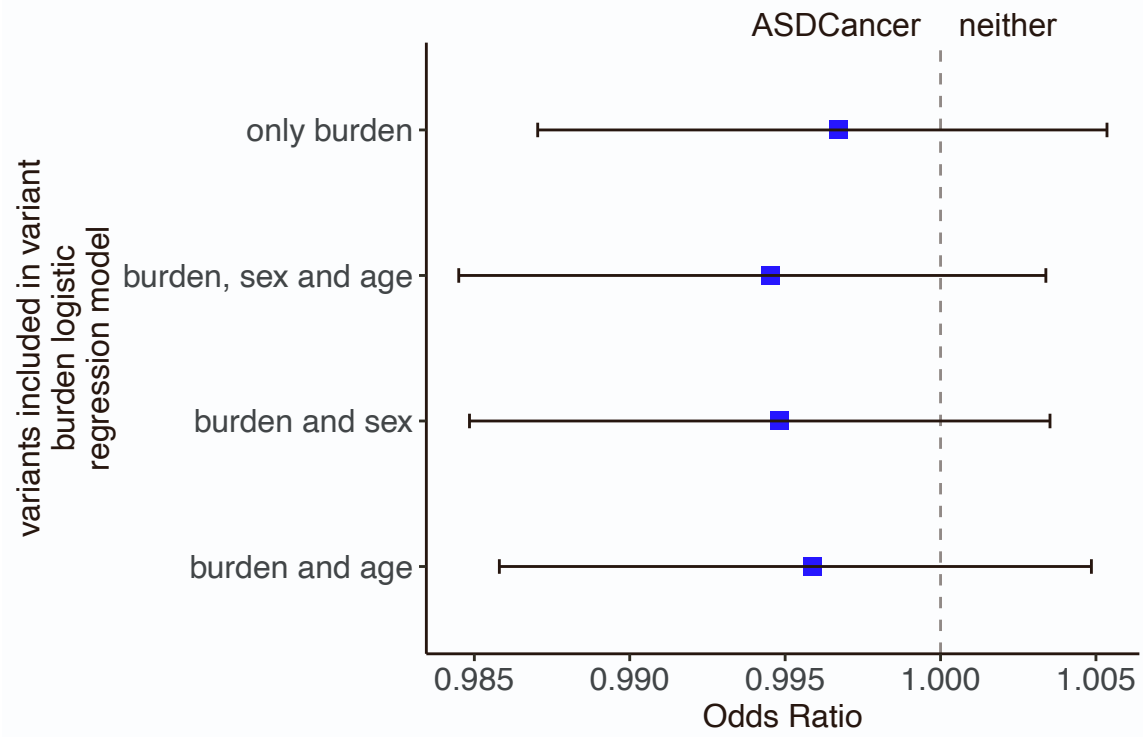
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S31. Forest plot of mtDNA variant burden analysis among PHTS-onlyCancer and PHTS-ASDCancer groups within the H haplogroup (n = 213).



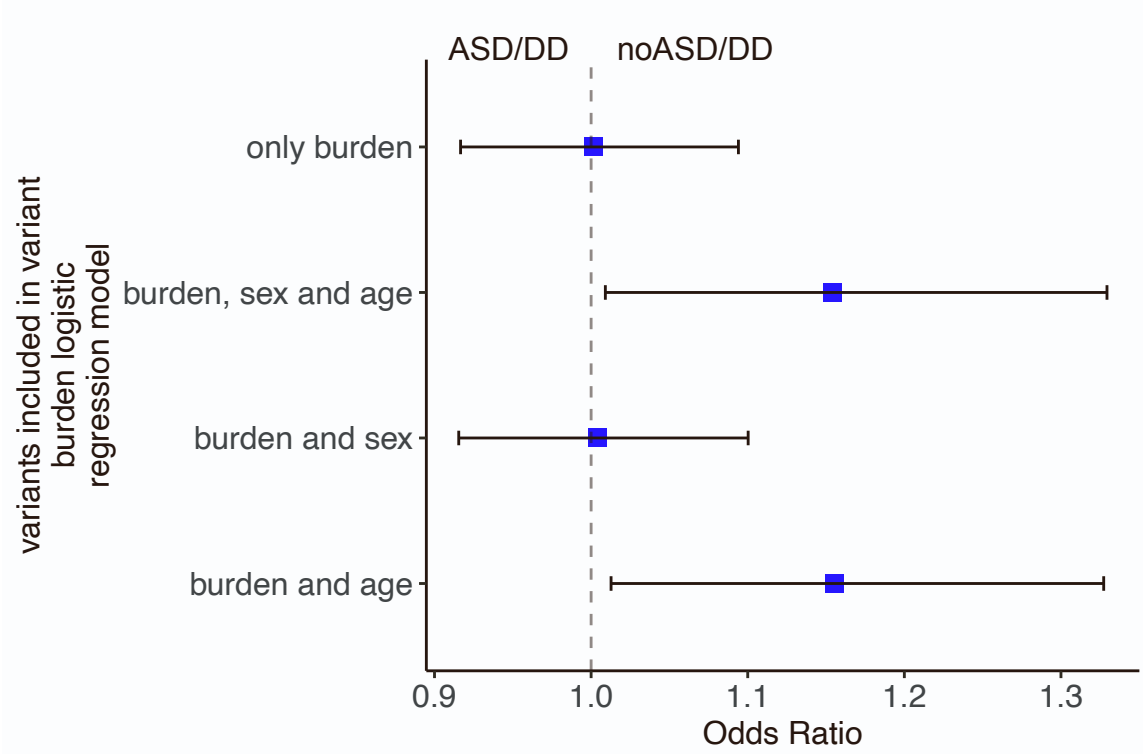
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S32. Forest plot of mtDNA variant burden analysis among PHTS-ASDCancer and PHTS-neither groups within the H haplogroup (n = 213).



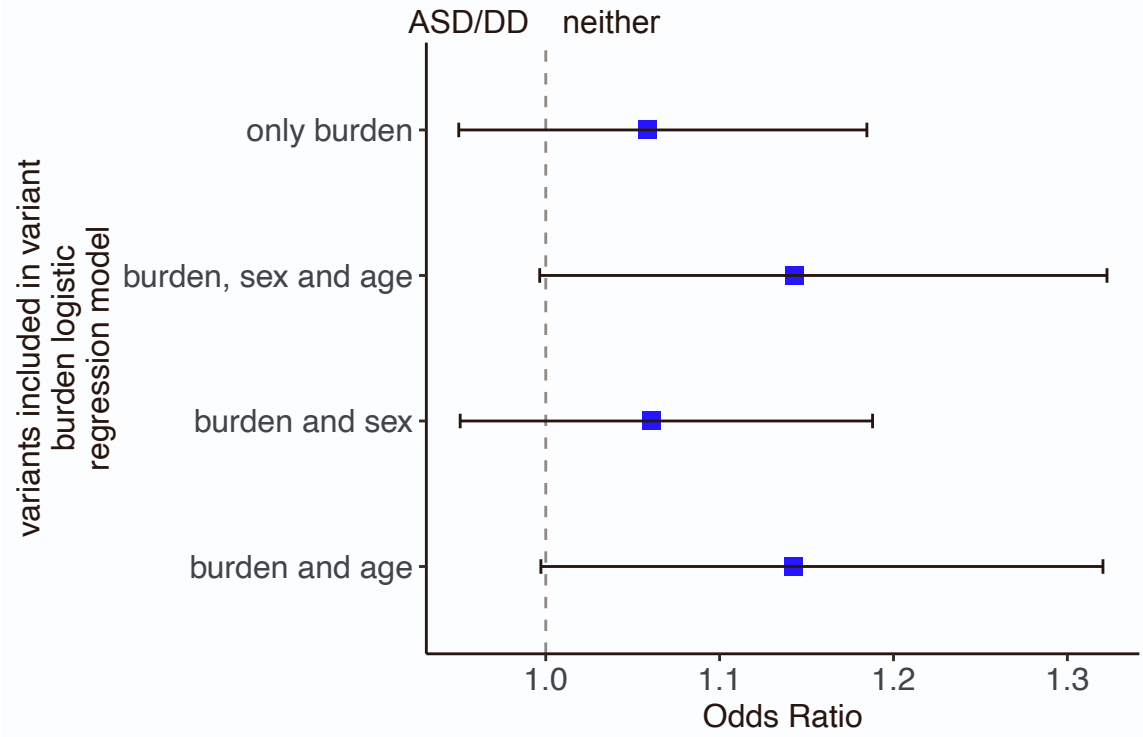
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S33. Forest plot of mtDNA variant burden analysis among PHTS-ASD/DD and PHTS-noASD/DD groups within the H haplogroup (n = 213).



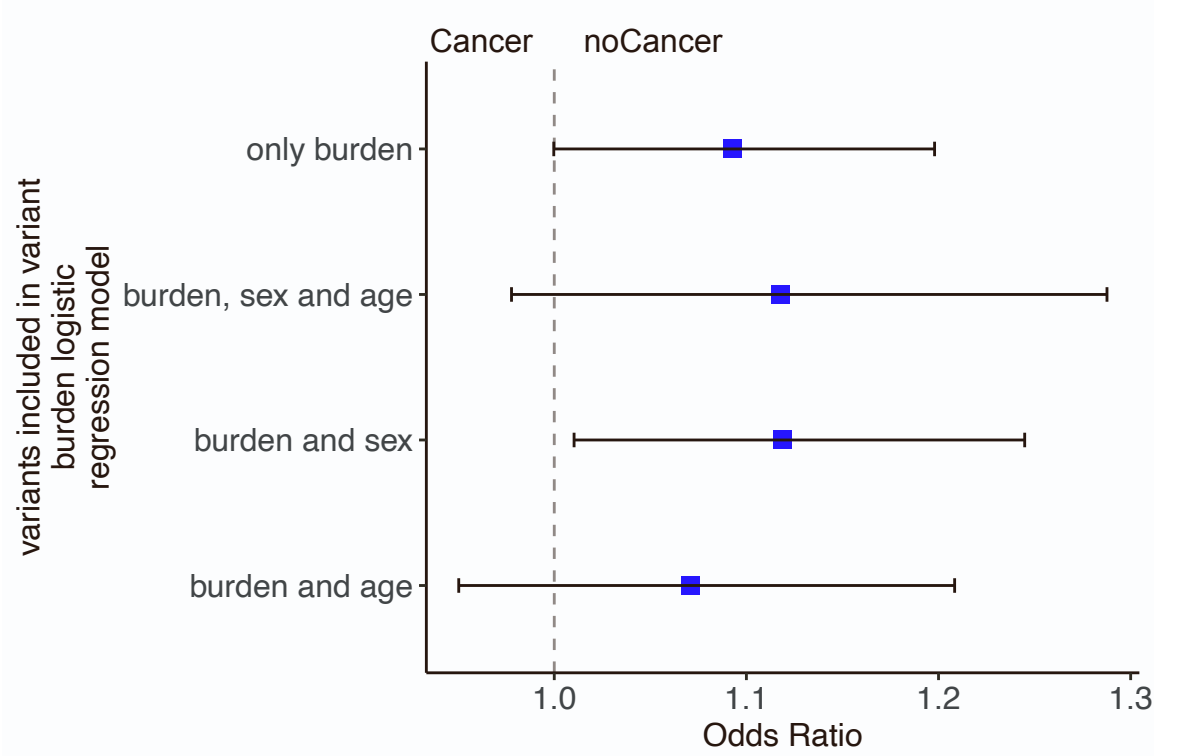
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S34. Forest plot of mtDNA variant burden analysis among PHTS-ASD/DD and PHTS-neither groups within the H haplogroup (n = 213).



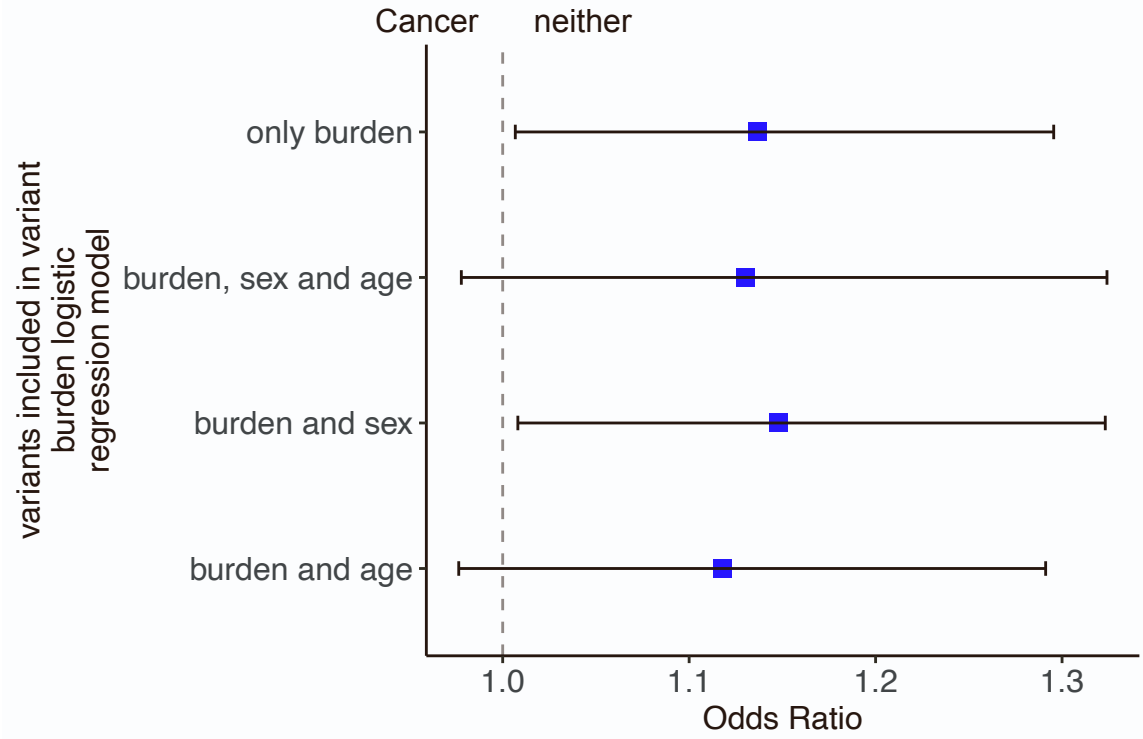
The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S35. Forest plot of mtDNA variant burden analysis among PHTS-Cancer and PHTS-noCancer groups within the H haplogroup (n = 213).



The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Figure S36. Forest plot of mtDNA variant burden analysis among PHTS-Cancer and PHTS-neither groups within the H haplogroup (n = 213).



The blue squares indicate the odds ratios, error bars indicate the 95% of confidential intervals of the odds ratios. Clinical phenotype groups are indicated at the top of each forest plot. Corresponding p values and odds ratios can be located in Table S1 and Table S2, respectively.

Table S1. Statistical results of the comparisons of mtDNA copy number and SNV burden across all PHTS phenotype groups

Comparisons/Test	Mann-Whitney test	Controlling for neither age nor sex	Controlling for age[a] and sex)	Controlling for sex	Controlling for age
mtDNA copy number across all PHTS phenotype groups (n = 498)					
onlyASD/DD vs. onlyCancer (n = 348)	0.00918[b]**	0.03*	0.407	0.167	0.0862
onlyASD/DD vs. ASDCancer (n = 182)	0.0841	0.079	0.319	0.179	0.128
onlyASD/DD vs. neither (n = 296)	0.0666	0.12	0.433	0.122	0.368
onlyCancer vs. ASDCancer (n = 202)	0.492	0.45	0.739	0.440	0.994
onlyCancer vs. neither (n = 316)	0.591	0.64	0.458	0.488	0.586
ASDCancer vs. neither (n = 150)	0.433	0.433	0.296	0.299	0.332
ASD/DD vs. noASD (n = 498)	0.0242*[c]	0.069	0.650	0.130	0.405
ASD/DD vs. neither (n = 498)	0.129	0.221	0.851	0.268	0.689
Cancer vs. noCancer (n = 498)	0.0323*	0.064	0.265	0.120	0.220
Cancer vs. neither (n = 498)	0.511	0.54	0.367	0.441	0.405
mtDNA copy number among the phenotype groups with the H haplogroup (n=213)					
onlyASD/DD vs. onlyCancer (n = 151)	0.00418**	0.03*	0.257	0.0833	0.103
onlyASD/DD vs. ASDCancer (n = 86)	0.0670	0.08	0.397	0.121	0.249
onlyASD/DD vs. neither (n = 124)	0.0545	0.122	0.720	0.0340*	0.872
onlyCancer vs. ASDCancer (n = 89)	0.805	0.645	0.825	0.556	0.759
onlyCancer vs. neither (n = 127)	0.520	0.451	0.658	0.556	0.836
ASDCancer vs. neither (n = 62)	0.527	0.344	0.249	0.267	0.390
ASD/DD vs. noASD (n = 213)	0.0135*	0.035*	0.750	0.0400*	0.724
ASD/DD vs. neither (n = 213)	0.123	0.117	0.891	0.852	0.107
Cancer vs. noCancer (n = 213)	0.0146*	0.0635	0.265	0.120	0.220
Cancer vs. neither (n = 213)	0.465	0.519	0.672	0.535	0.622
mtDNA variant burden within the H haplogroup (n = 213)					

onlyASD/DD vs. onlyCancer (n = 151)	0.190	0.0221*	0.906	0.139	0.301
onlyASD/DD vs. ASDCancer (n = 86)	0.100	0.0732	0.253	0.599	0.941
onlyASD/DD vs. neither (n = 124)	0.882	0.0531	0.153	0.504	0.158
onlyCancer vs. ASDCancer (n = 89)	0.226	0.853	0.109	0.343	0.0655
onlyCancer vs. neither (n = 127)	0.0843	0.563	0.338	0.133	0.266
ASDCancer vs. neither (n = 62)	0.0457*	0.476	0.0964	0.0984	0.0656
ASD/DD vs. noASD (n = 213)	0.805	0.973	0.0399*	0.934	0.0348*
ASD/DD vs. neither (n = 213)	0.524	0.308	0.0619	0.296	0.0607
Cancer vs. noCancer (n = 213)	0.0329*	0.054	0.111	0.0342*	0.263
Cancer vs. neither (n = 213)	0.0843	0.0448*	0.110	0.0452*	0.115

[a]Age represents age at consent

[b]** $0.001 < p \text{ value} \leqslant 0.01$

[c]* $p \text{ value} < 0.05$

Table S2. Odds ratios of the comparisons of mtDNA copy number and SNV burden across all PHTS phenotype groups

Comparisons/Test	Controlling for neither age nor sex (Odd Ratio, 95% confidence interval)	Controlling for age[a] and sex (Odd Ratio, 95% confidence interval)	Controlling for sex (Odd Ratio, 95% confidence interval)	Controlling for age (Odd Ratio, 95% confidence interval)
mtDNA copy number across all PHTS phenotype groups (n = 498)				
onlyASD/DD vs. onlyCancer (n = 348)	0.997 (0.994-1.000)	0.996 (0.988-1.005)	0.998 (0.995-1.001)	0.998 (0.995-1.001)
onlyASD/DD vs. ASDCancer (n = 182)	1.007 (1.000-1.015)	1.005 (0.996-1.015)	0.998 (0.995-1.001)	1.007 (0.999-1.016)
onlyASD/DD vs. neither (n = 296)	1.002 (0.999-1.005)	1.002 (0.998-1.006)	0.998 (0.995-1.001)	1.002 (0.998-1.006)
onlyCancer vs. ASDCancer (n = 202)	0.999 (0.997-1.002)	1.001 (0.994-1.009)	0.998 (0.995 1.001)	1.000 (0.993-1.008)
onlyCancer vs. neither (n = 316)	0.999 (0.997-1.002)	0.999 (0.995-1.002)	0.998 (0.995-1.002)	0.999 (0.996-1.002)
ASDCancer vs. neither (n = 150)	1.003 (0.997-1.010)	1.003 (0.997-1.011)	0.998 (0.995-1.001)	1.003 (0.997-1.011)
ASD/DD vs. noASD (n = 498)	0.998 (0.996-1.000)	0.999 (0.996-1.003)	0.998 (0.996-1.001)	0.999 (0.995-1.002)
ASD/DD vs. neither (n = 498)	0.998 (0.995-1.001)	1.000 (0.996-1.003)	0.998 (0.995-1.001)	0.999 (0.996-1.003)
Cancer vs. noCancer (n = 498)	1.002 (1.000-1.005)	1.002 (0.999,1.005)	1.002 (1.000-1.005)	1.002 (0.999-1.005)
Cancer vs. neither (n = 498)	1.001 (0.998-1.004)	1.001 (0.998-1.005)	1.001 (0.998-1.004)	1.001 (0.998-1.004)
mtDNA copy number among the phenotype groups with the H haplogroup (n=213)				
onlyASD/DD vs. onlyCancer (n = 151)	0.997 (0.994-1.000)	0.9938 (0.978-1.004)	0.996 (0.991-1.000)	0.991 (0.980-1.001)
onlyASD/DD vs. ASDCancer (n = 86)	0.993 (0.985-1.000)	0.995 (0.981-1.006)	0.992 (0.981-1.001)	0.993 (0.982-1.003)
onlyASD/DD vs. neither (n = 124)	0.998 (0.995-1.001)	0.999 (0.993-1.005)	0.995 (0.989-0.999)	1.000 (0.994-1.005)
onlyCancer vs. ASDCancer (n = 89)	1.001 (0.998-1.003)	0.999 (0.990- 1.010)	1.002 (0.995-1.011)	1.001 (0.995-1.006)
onlyCancer vs. neither (n = 127)	1.003 (0.996-1.009)	1.001 (0.996-1.007)	1.001 (0.997-1.006)	1.002 (0.995-1.006)
ASDCancer vs. neither (n = 62)	1.003 0.997-1.010)	0.995 (0.981-1.006)	1.005 (0.996-1.015)	1.004 (0.995-1.014)
ASD/DD vs. noASD (n = 213)	0.996 0.993-1.000)	0.999 (0.994-1.004)	0.996 (0.993,0.999)	0.999 (0.995-1.003)
ASD/DD vs. neither (n = 213)	0.996 (0.991-1.001)	1.000 (0.995-1.006)	0.996 (0.991-1.001)	1.001 (0.995-1.006)
Cancer vs. noCancer (n = 213)	1.002 (1.000-1.005)	1.002 (0.999-1.005)	1.002 (1.000-1.005)	1.002 (0.999-1.005)

Cancer vs. neither (n = 213)	0.999 (0.995-1.003)	0.999 (0.995-1.003)	0.999 (0.995-1.003)	0.999 (0.995-1.003)
mtDNA variant burden within the H haplogroup (n = 213)				
onlyASD/DD vs. onlyCancer (n = 151)	0.995 (0.991-0.999)	1.017 (0.767-1.390)	0.912 (0.804-1.029)	1.140 (0.897-1.491)
onlyASD/DD vs. ASDCancer (n = 86)	0.992 (0.982-1.000)	0.829 (0.581-1.133)	0.795 (0.613-0.999)	0.991 (0.780 1.268)
onlyASD/DD vs. neither (n = 124)	1.005 (1.000-1.010)	1.119 (0.961-1.312)	1.039 (0.929-1.166)	1.114 (0.961-1.300)
onlyCancer vs. ASDCancer (n = 89)	1.000 (0.995-1.004)	1.292 (0.965-1.834)	1.112 (0.894-1.392)	1.35 (0.997-1.912)
onlyCancer vs. neither (n = 127)	0.998 (0.989-1.005)	1.080 (0.926-1.273)	1.110 (0.973-1.282)	1.086 (0.940-1.263)
ASDCancer vs. neither (n = 62)	0.997 (0.987-1.005)	1.223 (0.975-1.574)	1.215 (0.973-1.558)	1.248 (0.998-1.611)
ASD/DD vs. noASD (n = 213)	1.002 (0.917-1.094)	1.154 (1.009-1.329)	1.004 (0.915-1.100)	1.156 (1.013-1.327)
ASD/DD vs. neither (n = 213)	1.059 (0.950-1.185)	1.143 (0.997-1.323)	1.061 (0.951-1.188)	1.143 (0.997-1.320)
Cancer vs. noCancer (n = 213)	1.093 (1.000-1.198)	1.118 (0.978-1.288)	1.119 (1.010-1.245)	1.071 (0.950-1.208)
Cancer vs. neither (n = 213)	1.136 (1.007-1.296)	1.13 (0.978-1.324)	1.148 (1.008-1.323)	1.118 (0.976-1.291)

[a]Age represents age at consent

Table S3. Clinical phenotypic characteristics of the 498 individuals with PHTS

Clinical phenotypic characteristics		Data
Sex, No. (%)		
	Female	266 (53.4%)
	Male	232 (46.6%)
Age at consent, mean (SD)[range],years		31.19 (21.9) [0.5-85]
Autism spectrum disorder, No. (%)		182 (36.5%)
Cancer, No. (%)		202 (40.6%)
PHTS component malignant neoplasms, No. (%)		128 (25.7%)
	Breast cancer	113 (22.7)
	Thyroid cancer	62 (12.5)
	Renal cell cancer	28 (5.6%)
	Endometrial cancer	33 (6.6%)
	Colon cancer	17 (3.4%)
	Melanoma	10 (2.0%)
Other malignant neoplasms[a], No. (%)		74 (14.9%)

[a] These are non-component cancers.