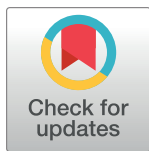# Schizophrenia-associated mt-DNA SNPs exhibit highly variable haplogroup affiliation and nuclear ancestry: Bi-genomic dependence raises major concerns for link to disease

Christian M. Hagen[1], Vanessa F. Gonçalves[2], Paula L. Hedley[1], Jonas Bybjerg-Grauholm[1], Marie Bækvad-Hansen[1], Christine S. Hansen[1], Jørgen K. Kanters[3], Jimmi Nielsen[4], Ole Mors[5], Alfonso B. Demur[6], Thomas D. Als[7], Merete Nordentoft[8], Anders Børglum[7], Preben B. Mortensen[9], James Kennedy[2], Thomas M. Werge[6], David M. Hougaard[1], Michael Christiansen[1,3]*

1 Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark, 2 Centre for Addiction and Mental Health, University of Toronto, Toronto, Canada, 3 Department of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark, 4 Aalborg Psychiatric Hospital, Aalborg University Hospital, Aalborg, Denmark, 5 Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, 6 Mental Health Centre, Sct Hans, Capital Region of Denmark, Denmark, 7 Institute of Medical Genetics, Aarhus University, Aarhus, Denmark, 8 Mental Health Centre, Capital Region of Denmark, Denmark, 9 Center for Register Research, Institute of Economics, Aarhus University, Aarhus, Denmark

* mic@ssi.dk

Abstract

Mitochondria play a significant role in human diseases. However, disease associations with mitochondrial DNA (mtDNA) SNPs have proven difficult to replicate. An analysis of eight schizophrenia-associated mtDNA SNPs, in 23,743 Danes without a psychiatric diagnosis and 2,538 schizophrenia patients, revealed marked inter-allelic differences in mitochondrial haplogroup affiliation and nuclear ancestry. This bi-genomic dependence could entail population stratification. Only two mitochondrial SNPs, m.15043A and m.15218G, were significantly associated with schizophrenia. However, these associations disappeared when corrected for haplogroup affiliation and nuclear ancestry. The extensive bi-genomic dependence documented here is a major concern when interpreting historic, as well as designing future, mtDNA association studies.

Introduction

Genetic variants in mitochondrial DNA (mtDNA)—and in nuclear genes coding for mitochondrial function—have been associated with disease [1–3]. More than 300 variants [4, 5] in mtDNA and genes involved in mitochondrial function [6] have been reported to cause mitochondrial disease, which is clinically characterized by complex metabolic, neurological, muscular and psychiatric symptoms [7, 8]. SNPs in mtDNA and mitochondrial haplogroups

(mtDNA hgs), which are evolutionarily fixed SNP sets with a characteristic geographical distribution, have been proposed as potential disease modifiers[8]. This has been reported in neurological degenerative diseases such as Alzheimer's disease [9–12] and Parkinson's disease [12–14], metabolic diseases and cancers [15], as well as psychiatric diseases, notably schizophrenia (SZ) and bipolar disease [16–18].

Association studies of mtDNA variants and disease have been difficult to replicate[8]. However, the definition of a methodological paradigm for association studies with mtDNA variants [19] implicitly assumes that mtDNA variants are independent of the nuclear genome (gDNA). In a recent Danish study on mtDNA hgs and their nuclear ancestry, we demonstrated a marked difference in nuclear ancestry between individual mtDNA hgs [20]. This means that mtDNA hgs entail population stratification also at the level of gDNA. The effect of such a stratification on disease association will depend on the admixture structure of the particular population, the population history, epidemiology and genetic epidemiology of the disease, as well as the number of persons included in the study. The extensive fine-scale heterogeneity of gDNA and significant admixture documented in the UK [21] and Europe [22] further increase the risk of spurious false positive associations, if the mtDNA/gDNA interaction is not corrected in association studies.

Using DNA-array data from the Danish iPSYCH study on 2,538 SZ patients and 23,743 population controls, we show that eight mtDNA SNPs, previously associated with SZ [16–18, 23], exhibit considerable inter-allelic differences both with respect to mtDNA hg affiliation and nuclear ancestry. This phenomenon, which we name bi-genomic dependence, affects the association between an mtDNA SNP and mtDNA, as well as gDNA, and can lead to both false negative and positive associations with disease. We demonstrate that, in this cohort, it is only possible to replicate the association results for two of the original eight SNPs, when correcting for population stratification. Both mtDNA hg affiliation and nuclear ancestry affects the strength of association. Finally, we show that none of the SNPs are associated with SZ when examined on a particular mtDNA hg background, with correction for bi-genomic dependence. As none of previous studies of mtDNA SNPs have been performed with correction for population stratification, let alone bi-genomic dependence, our results indicate that all such published associations should be considered preliminary. In principle, this conclusion should not be limited to associations with SZ.

Results

From a literature search of mtDNA SNPs previously associated with SZ, we identified eight that were also typed by the PsychChip, Table 1. PsychChip data from 23,743 normal Danes and 2,538 SZ patients (Detailed in S1 Table) showed that the SNPs were present in the population with frequencies varying from 0.2%– 20.6%, Table 1. There was no appreciable difference in mtDNA hg or nuclear ancestry distribution between controls and SZ patients, S1 and S2 Figs.

Haplogroup distribution of mtDNA SNPs

The potential affiliation, based on PhyloTree, of SNPs to different mtDNA hgs is shown in Table 1, and the actual mtDNA hg distribution in the controls (not different from that of the SZ patients, S1 Fig) is shown in Fig 1. For all SNPs, there is a marked difference in the actual mtDNA hg distribution between the two alleles at the same position. Thus, when comparing persons with either of two alleles at the same mtDNA position, the comparison is between groups with widely differing mtDNA distributions. A PCA analysis of the mtDNA sequences in persons with either the A or G allele at position 15,043 is shown in Fig 2A. This analysis

Table 1. Schizophrenia associated mtDNA SNPs, their frequency, mtDNA hg association, and link with disease.

Variant	Locus	Molecular effect	Effect on risk of SZ	% DK iPSYCH controls (SZ)	Theoretical mtDNA hg associations[*]
m.1438A [17]	MT-RNR1	n.a.	SZ↑	4.6 (4.8)	L0d, L1, C1b4, C4c2, M32′56, D5a2, I5c1, H1b1g, H2, H13a2b5, H14a2c, J1b1b1c, P2, P5
m.3197C [18]	MT-RNR2	n.a.	SZ↑	7.3 (8.2)	U5a′b, L3f1a1, H14b, U2e1a1a
m.3666A [16]	MT-ND1	p.G120G	SZ↑	0.3 (0.4)	L0d1c1a1a, L0d1b2b2c2, L1, M13c, H1ak, R9b2, F1a1c, U5a2b3a1
m.4769A [17]	MT-ND2	p.M100M	SZ↑	4.4 (4.6)	R2, B4a1a1c, B4a1a4, L0d2b1, H2a
m.9377G [16]	MT-CO3	p.W57W	SZ↓	0.5 (0.6)	G2a, Q2a1, D5b2, A2ac, H1c9a, U6a1b4, K1a4a1a2a, L2e, L3e2b1
m.10398G [23]	MT-ND3	p.T114A	SZ↑	20.6 (19.1)	K1, J, R11, B4m, B4c1c, R12′21, P4, U6a5c, K2a11, U8c, N1a1, W1e1a, W3a1d, N8, Y, S3, X2f1, R0a2k1, B5
m.15043A [18]	MT-CYB	p.G99G	SZ↑	5.7 (4.6)	M, N1a1, L2c2a, J1c2a3, T2f1a, U6a7, U2c1a
m.15218G [18]	MT-CYB	p.T158A	SZ↑	4.3 (5.2)	U5a1, M7a1a2, M10a1, HV1a′b′c, H13a2c

[*]Based on PhyloTree.

https://doi.org/10.1371/journal.pone.0208828.t001

shows that the difference in mtDNA sequence, while subtle, indicates independent mtDNA distribution between the alleles.

Nuclear ancestry of mtDNA SNPs

The distribution of nuclear ancestries in control persons as a function of each mtDNA allele revealed major differences, both between different positions (inter-SNP) in the mtDNA
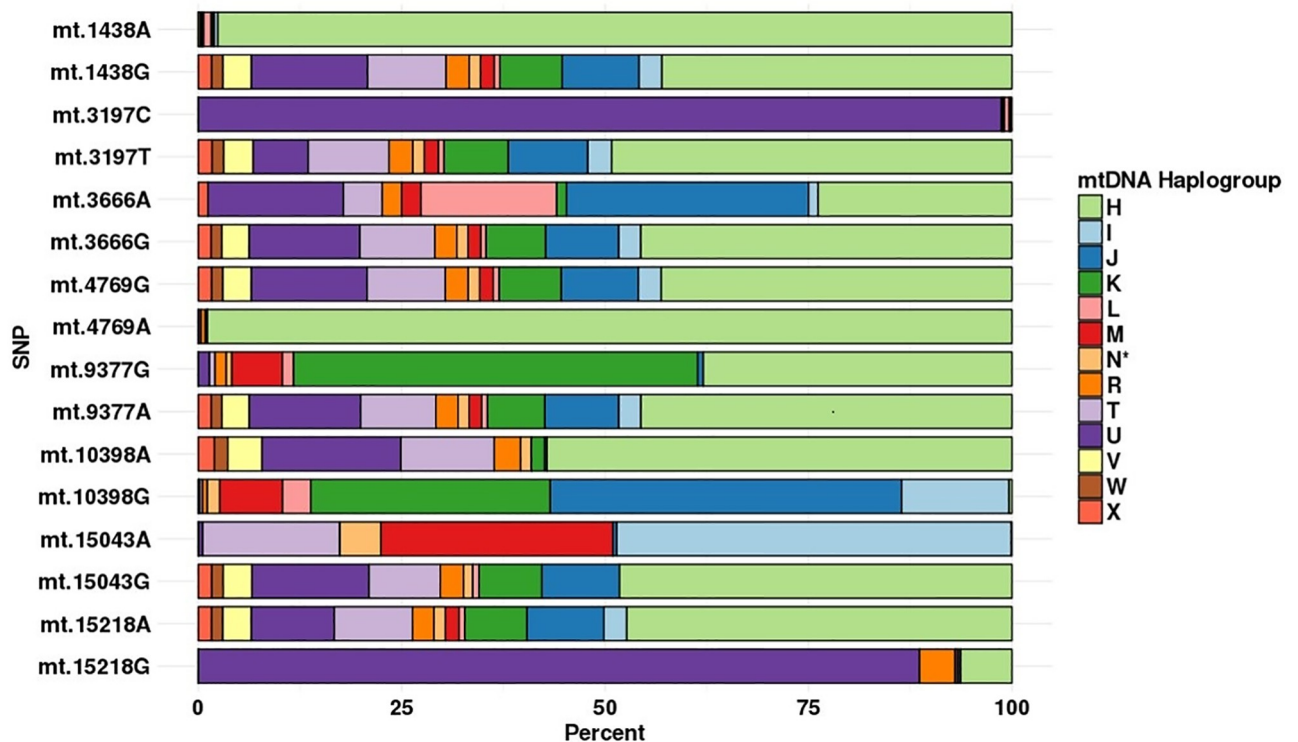
Fig 1. The percentage distribution of mtDNA hgs, per allele, among control individuals.

https://doi.org/10.1371/journal.pone.0208828.g001
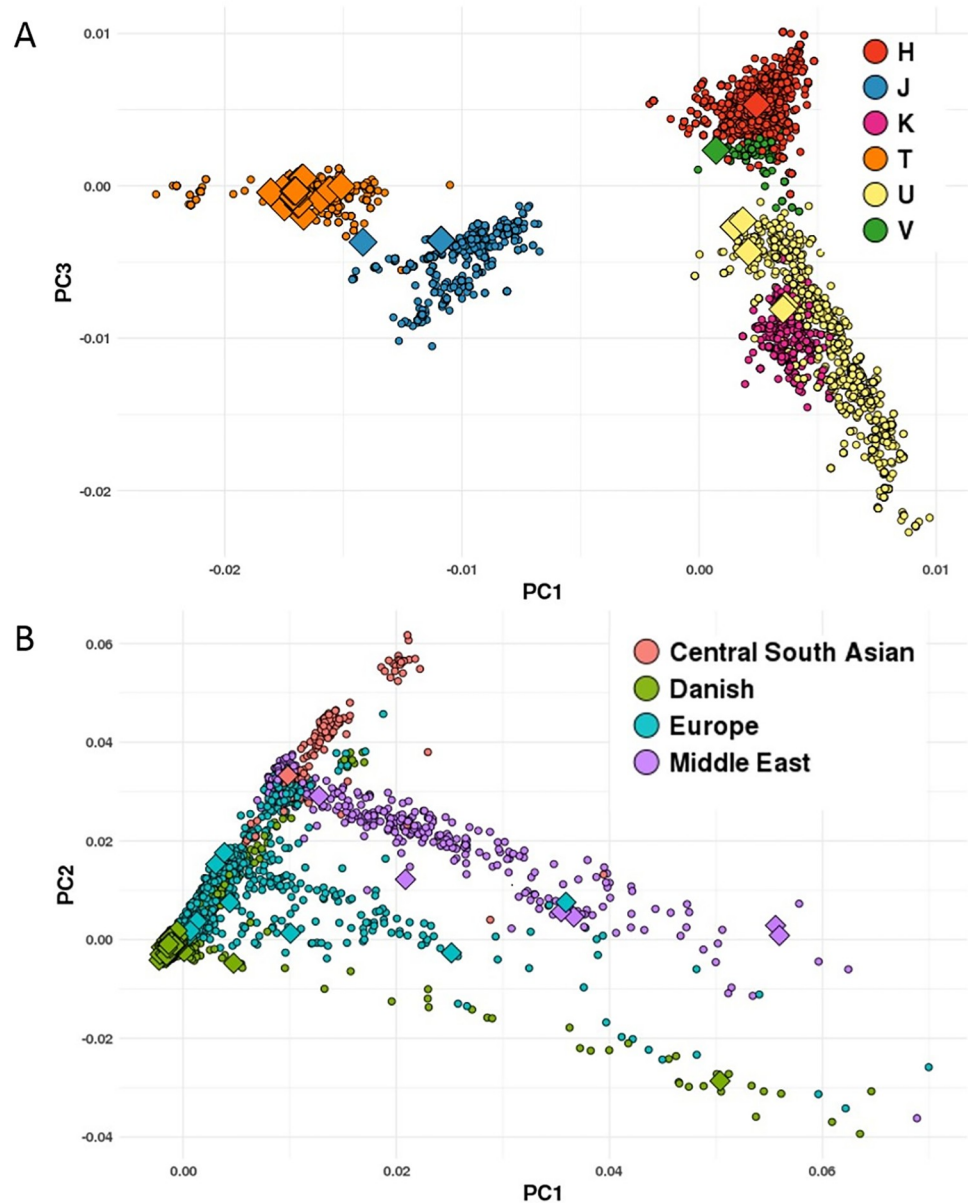
Fig 2. PCA of **A**. mtDNA (PC1 versus PC3) and **B**. gDNA (PC1 versus PC2), with the m.15043A (filled diamonds) and m.15043G (filled circles) alleles. Only persons with "European" mtDNA haplogroup (H, HV, V, U, K, J & T) were included. The PCA was performed using PCs defined by control samples.

molecule and between different alleles (inter-allelic) at the same position, Fig 3. Apart from subtle differences in the distributions of Greenlandic and Asian ancestral affiliations, the distribution is similar in SZ patients, S2 Fig. The nuclear ancestry was statistically different between the alleles of seven of the eight SNPs assessed by chi-squared test; m.1438, m.3197, m.4769, m.10398 and m.15043 were statistically highly significant (P-value = $2.6e^{-7}$, $1.34e^{-6}$, $2.45e^{-7}$, $2.2e^{-16}$ $2.2e^{-16}$ respectively) and m.3666 and m15218 were statistically significant (P-value = 0.002 and 0.008 respectively whereas m.9377 was not statistically different. Most marked is the difference in Danish nuclear ancestry for m.15043 A/G, with an inter-allelic
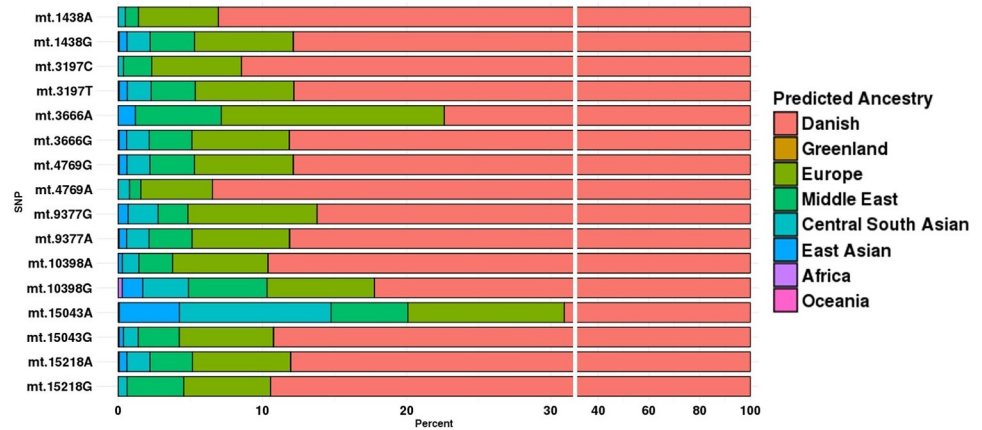
Fig 3. The percentage distribution of nuclear ancestries, per allele, among control individuals.

difference ~ 20 percent points in Danish ancestry, and the A-allele has a prominent Central South Asian ancestry not seen in the G-allele, Fig 3 and S2 Fig. PCA analyses of the gDNA, Fig 2B, of the two alleles, reveal subtle differences in the genomic structure independent of the nuclear ancestry between the two alleles. Thus, when comparing persons with either of two alleles at the same mtDNA position, there is a risk for comparison between groups with widely different ancestries, thus increased risk of confounding by population structure.

Association between schizophrenia and mtDNA SNPs

The association of each mtDNA SNP with SZ was assessed (Fig 4). In consequence of the inter-allelic differences in mtDNA hg affiliations and nuclear ancestry demonstrated above, several association analyses were performed. Five SNPs, m.1438A, m.3197C, m.3666A, m.4769A, and m.9377G showed no association with SZ, both when all persons were included and where selection was made to reduce effects of varying mtDNA and nuclear ancestry affiliations. The m.10398G SNP was marginally significantly, while m.15043A was significantly associated with reduced risk of SZ in All and All-Danish mtDNA hgs. The m.15218G was associated significantly, or borderline significantly, with a reduced risk for SZ irrespective of the grouping, Fig 4. The high level of diversity in the PCA analysis of m.15043A (and the remaining SNPs—data not shown), Fig 2A, prompted us to examine whether a fixation of the mtDNA hg, i.e. limiting the analysis to persons with a specific hg, in cases where it was reasonably frequent, would result in significant associations. The result, Table 2, was the contrary—on a fixed mtDNA hg background, none of the SNPs exhibited a significant association with SZ.

Discussion

Here we show the dependence of mtDNA SNPs with both mtDNA hgs and gDNA clusters due to population structure and shared demographic history of mtDNA and gDNA, we have called this relationship "bi-genomic dependence". This has the consequence that an association between a particular allele, in a specific SNP, and disease is not exclusively the result of the presence of that particular allele; since each mtDNA allele is associated with the unique distributions of both mtDNA hgs and gDNA clusters. Furthermore, bi-genomic dependence can be accounted for by including mtDNA and gDNA principle components from PCAs in association analyses—thus incorporating a bi-genomic measure of population stratification and
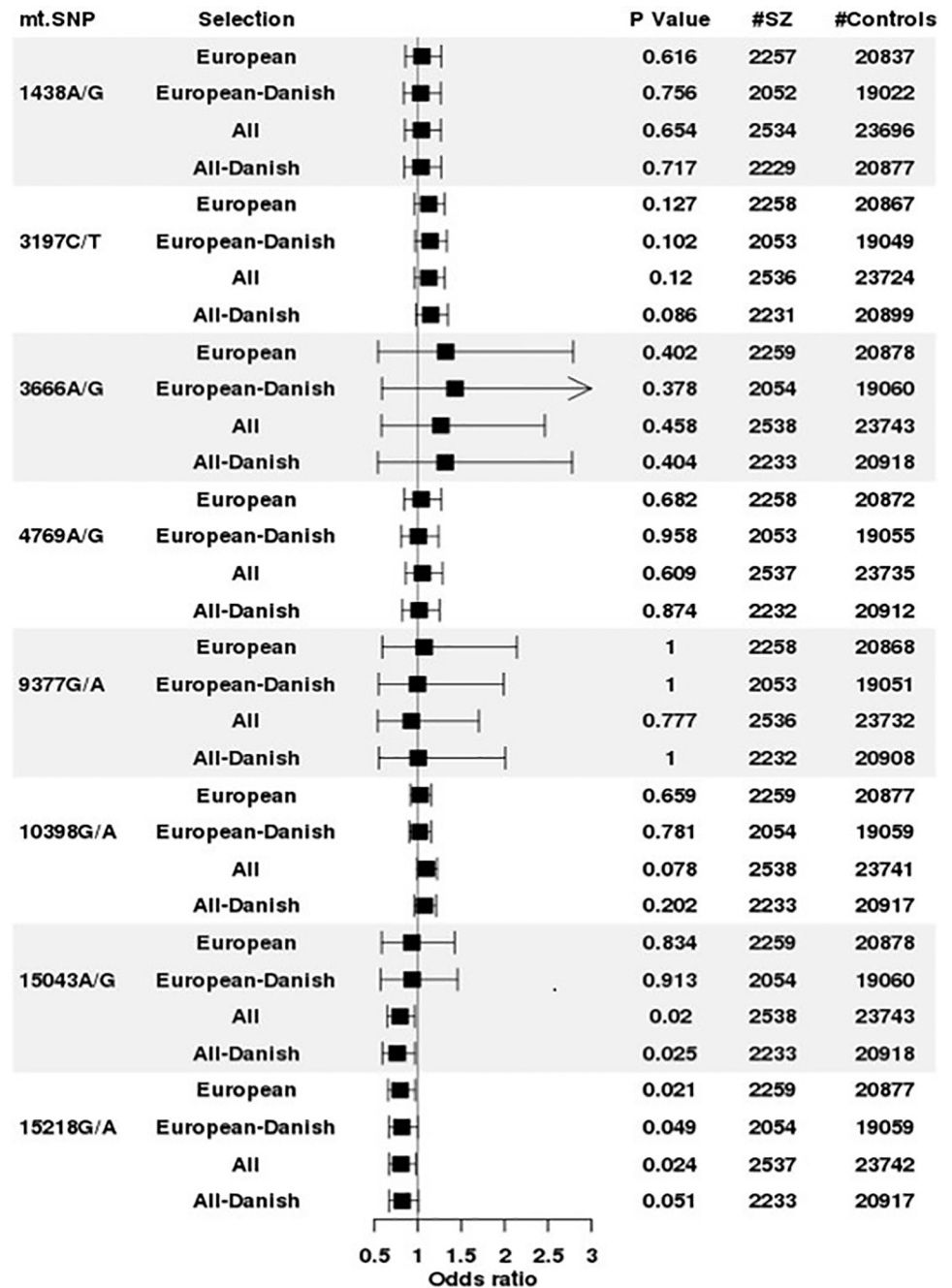
Fig 4. Association between individual mtDNA SNPs and schizophrenia as a function of the selection of the involved persons. (**All hgs**: All persons, **EU hgs**: persons with a "classical European" mtDNA hg; **EU_hgs_DANES**: persons with both a European mtDNA hg and a Danish nuclear ancestry and **ALL_hgs_DANES**: persons with a Danish nuclear ancestry).

https://doi.org/10.1371/journal.pone.0208828.g004

admixture into mtDNA association analyses. Currently, such a measure of bi-genomic dependence is not incorporated into studies of mtDNA-disease-association which consider an association as evidence for a specific effect of an allele on the function of a protein or RNA coded for by the mtDNA and, consequently, as a cause of pathophysiological changes.

Table 2. Association of SNPs with SZ in Danish nuclear ancestry and a specific mtDNA haplogroup.

mtDNA hg	Allele	Odds-ratio (95% cfi)	p	SZ (N)	Controls (N)
H	m.1438A	1.05 (0.84–1.29)	0.7	110	1,000
	m.3666A	1.12 (0.13–4.71)	0.7	2	17
	m.4769A	1.03 (0.83–1.28)	0.7	105	966
	m.9377A	0.95 (0.38–3.07)	0.8	1,046	9,917
	m.10398A	0.79 (0.18–7.14)	0.7	1,049	9,947
	m.15218A	1.09 (0.47–3.09)	1	1,045	9,900
J	m.3666A	0.74 (0.08–3.06)	1	2	22
K	m.9377A	0.90 (0.38–2.58)	0.8	133	1,509
	m.10398A	1.40 (0.90–2.12)	0.1	33	106
U	m.3197C	1.08 (0.85–1.38)	0.6	189	1,572
	m.3666A	2.59 (0.46–10.13)	0.2	3	10
	m.15218A	0.81 (0.63–1.04)	0.1	211	1,939
T	m.15043A	0.99 (0.59–1.58)	1	22	208

The linkage disequilibrium between different alleles of a SNP and mtDNA hgs and sub-hgs is not surprising as hgs are defined by series of evolutionarily conserved SNPs. The particular distribution of subsets of mtDNA hgs, sub-hgs and individual SNPs that are associated with a particular allele at a specific SNP will depend on the population history and the extent and source of admixture. In most countries, and in particular in Europe, such history is very complicated and incompletely clarified. m.10398G was found associated with SZ in Han Chinese, however, when the cohort was broken down with respect to mtDNA hgs, the association disappeared [24]. This illustrates that a specific allele's mtDNA hg distribution may induce spurious association with SZ. Spurious associations between mtDNA SNPs and a particular phenotype, when restricted to persons of a specific mtDNA hg [25], may however be due to population stratification at the sub-hg level. In the iPSYCH cohort there was no association between SZ and any of the European mtDNA hgs (Data not shown).s mtDNA replicates independently of the cell cycle and without recombination, mtDNA hgs and SNPs should be independent of the gDNA, but only if the population is infinitely large and in the absence of population substructure. This is often not the case, due to geographical population substructure, recent admixture, socially and culturally defined restrictions in choice of spouse. A recent study showed that most Danish grandparents to present-day high school students chose spouses within a short distance of their birthplace [26]. This practice will, with time, lead to regionalization, and a southwestern to northeastern gradient was found [26]. Furthermore, immigrants may seek a partner from within their ethnic community. Such effects have been eliminated in some studies by restricting the participants in mtDNA association studies to persons with a three-generation presence in the population. However, it has not been documented that this is sufficient to obviate association or linkage disequilibrium between mtDNA SNPs and specific gDNA clusters. Extensive gDNA micro-scale heterogeneity has been documented in the UK [21] and Western France [27] and admixture has been an important factor in the accretion of the present-day genomic variation of Europe [22, 28]. The UK study [21] showed that this is not just a result of recent demic changes; however, recent migrations may lead to widespread bi-genomic dependence.

Schizophrenia is a complex syndromic disease [29] with geographically varying prevalence [30] and characterized by a markedly elevated prevalence among first and second generation immigrants [31, 32], particularly among persons with dark skin moving to Nordic latitudes [33]. These epidemiological characteristics of SZ obviously increase the risk of spurious

associations caused by subtle admixture and bi-genomic dependence. However, it does not *per se* refute the mitochondrial pathogenic paradigm [34] where variation in mitochondrial function, believed to interfere with ATP production [35, 36], inflammation and signaling [37, 38] as well as Ca^{2+}-homeostasis [39], and apoptosis [38], is considered to be of paramount importance for development of disease. Several neuroanatomical post-mortem findings in SZ brains indicate perturbed mitochondrial function [40], but such findings are difficult to distinguish from changes caused by drug treatment.

The iPSYCH data are prospective and signs of immigration are apparent [20], but they also showed that the variation in ancestry differed greatly between mtDNA hgs—even within traditional European hgs, i.e. mtDNA hg U, where ancient European sub-hgs occurred together with U-sub-hgs of recent Near Eastern and Central Asian origin [20]. Thus, bi-genomic dependence is likely to be a confounder and may lead both to false positive as well as false negative associations with disease. The method of correction for bi-genomic dependence in association studies will depend on the specific mtDNA SNP examined, the population structure and history, as well as the size of the study population.

If population stratification involving gDNA is inherent when performing association studies with mtDNA SNPs, it should be expected that diseases with geographically varying prevalence would be likely to find associated with specific mtDNA SNPs. The largest mtDNA association study to date [18] found mtDNA SNPs associated with ulcerative colitis, exhibiting a European North-South and East-West gradient [41], and with multiple sclerosis exhibiting a longitudinal prevalence gradient [42] and effect of immigration [43]. The same study found that the prevalence of mtDNA SNPs associated with Parkinson's disease was lower in African and Asian people [44]. Furthermore, the incidence of primary biliary cirrhosis is very high in North East England, 50% lower in the rest of England and Scandinavia, and 90% lower in the Middle East and Asia [45].

A major problem with the interpretation of mtDNA SNP variants is the difficulties associated with performing a meaningful and reproducible assessment of mitochondrial function. In vitro studies of mitochondrial function, e.g. enzymatic activity measurements of OXPHOS components in cells, tissues [46] or cybrids [47] as well as allotopic expression [15], are difficult to interpret as they also interfere with the inherent cellular control of mitochondrial function [38]. Furthermore, it should be noted, that mtDNA hgs and sub-hgs are cladistics groups and not functional units. Thus, in the Danish population, the U-hg is composed of a range of sub-hgs, e.g. U5a, U5b, U6, U7, and U8, with widely differing nuclear ancestries, reflecting migrations rather than selection [20]. It is thus meaningless to ascribe a specific functional effect to a particular mtDNA hg—without having carefully examined both mtDNA and nuclear genetic variation and corrected for stratifications in both.

Previous conflicting studies of disease associations with mtDNA have been suggested to be the result of insufficient power [48], insufficient stratification respect to sex, age, geographical background [49] or population admixture [50], or the use of small areas of recruitment risking "occult" founder effects [51]. The fact that careful control, as here, of these factors and the bi-genomic dependence, results in none of eight previously SZ associated mtDNA SNPs being associated with SZ in the very large Danish iPSYCH cohort, suggests that previously reported associations could indeed be spurious findings due to cryptic population stratification. Meta-analyses pooling studies from different populations [15] does not necessarily solve this problem—it may aggravate it by introducing further sub-stratification of the total population analyzed. The extensive bi-genomic dependence demonstrated in the Danish population makes this phenomenon the most parsimonious explanation of non-replicable associations with mtDNA variants, not only associations with SZ, but obviously bi-genomic dependence can interfere with associations between mtDNA and all types of diseases and traits.

Materials and methods

Ethics statement

The iPSYCH cohort study (www.ipsych.au.dk) is register-based using data from Danish national health registries. The study was approved by the Scientific Ethics Committees of the Central Denmark Region (www.komite.rm.dk) (J.nr.: 1-10-72-287-12) and executed according to guidelines from the Danish Data Protection Agency (www.datatilsynet.dk) (J.nr.: 2012-41-0110). Passive, but not informed, consent was obtained, in accordance with Danish Law nr. 593 of June 14, 2011, para 10, on the scientific ethics administration of projects within health research.

SZ patients and controls

As part of the iPSYCH recruitment protocol, 23,743 controls, born between May 1st 1981 and Dec 31st 2005 were selected at random from the Danish Central Person Registry. Among persons born within the same time span 2,538 persons assigned an ICD-10 F20 were identified in the Danish National Patient Registry. All were singletons, were alive one year after their birth, and had a mother registered in the Danish Central Person Registry. DNA was available from DBS cards obtained from the Danish Neonatal Screening Biobank at Statens Serum Institut [52] Demographics of patients and controls are given in S1 Table.

Genetic analysis and mtDNA SNPing

From each DBS card two 3.2-mm disks were excised from which DNA extracted using Extract-N-Amp Blood PCR Kit (Sigma-Aldrich, St Louis, MO, USA)(extraction volume: 200 μL). The extracted DNA samples were whole genome amplified (WGA) in triplicate using the REPLIg kit (Qiagen, Hilden, Germany), then pooled into a single aliquot. Finally, WGA DNA concentrations were estimated using the Quant-IT Picogreen dsDNA kit (Invitrogen, Carlsbad, CA, USA). The amplified samples were genotyped at the Broad Institute (MA, USA) using the PsychChip (Illumina, CA, USA) typing 588,454 variants, developed by the Psychiatric Genetic Consortia. We then isolated the 418 mitochondrial loci and reviewed the genotype calls, before exporting into the PED/MAP format using GenomeStudio (Illumina, CA, USA). Haplo-grouping of mtDNA was performed using the defining SNPs reported in www.phylotree.org [53].

Nuclear ancestry

Nuclear ancestry estimation was done using ADMIXTURE 1.3.050 in the supervised approach. Briefly, reference populations consisting of Human Genome Diversity Project (HGDP) (http://www.hagsc.org/hgdp/), a Danish (716 individuals) and a Greenlandian (592 individuals) genotyping SNP data set were used. The final reference data set consisted of 103,268 autosomal SNPs and 2,248 individuals assigned to one of nine population groups: Africa, America, Central South Asia, Denmark, East Asia, non-Danish Europe, Greenland, Middle East and Oceania. The number of clusters, K was set to eight, based on principal component analysis clustering (data not shown). The subpopulations were merged with the reference population data set and analyzed using ADMIXTURE. For prediction of the ancestry of individuals within the mtDNA hgs we created a random forest model [54] based on the reference data set, with the clusters Q1-8 as predictors and population groups as outcome. Thus, the ancestry analysis of the individual person was the result of a supervised prediction. Prediction was done using R3 version 3.2.2, using the Caret package.

Statistics

The statistical significance of differences in mtDNA SNP proportions between controls and SZ patients was assessed using a permutation version of Fisher's exact test. Samples with missing sequence data were excluded. To assess differences in allele distribution within the predicted ancestries we used Pearson's Chi-squared test. Ancestries with allele counts below six were not included. Calculations were performed using R (version 3.1.3). Principal component analysis (PCA), was prepared using PLINK (v.1.90b3.31). For the PCA the reference population variants were extracted from the iPSYCH control sample, LD pruned (indep-pairwise 50 5 0.5) and allowing only SNPs with 99% genotyping rate. Prior to PCA of mtDNA data, samples were loaded into GenomeStudio (version 2011.a), a custom cluster was created using Gentrain (version 2), following automatic clustering, all positions with heterozygotes were manually curated. The data was exported relative to the forward strand using PLINK Input Report Plug-in (version 2.1.3). Eigenvectors were calculated using PLINK (v1.90b3.31). PCA plots were created using the package ggplot2 (version 1.0.1) in R (version 3.1.3).

Supporting information

S1 Table. Demographics and nuclear ancestry and mtDNA haplogroups of the cohort. (DOCX)

S1 Fig. Distribution of mtDNA haplogroups of schizophrenic patients with specific SNPs. (TIF)

S2 Fig. Distribution of nuclear ancestries of schizophrenic persons with specific SNPs. (TIF)

Acknowledgments

Author Contributions

Conceptualization: Vanessa F. Gonçalves, Jørgen K. Kanters, James Kennedy, Thomas M. Werge, Michael Christiansen.

Data curation: Christian M. Hagen, Paula L. Hedley, Jonas Bybjerg-Grauholm, Marie Bækvad-Hansen, Christine S. Hansen, Ole Mors, Merete Nordentoft, Preben B. Mortensen, Michael Christiansen.

Formal analysis: Christian M. Hagen, Christine S. Hansen, Alfonso B. Demur, Thomas M. Werge.

Funding acquisition: Ole Mors, Merete Nordentoft, Anders Børglum, Preben B. Mortensen, Thomas M. Werge, David M. Hougaard.

Investigation: Christian M. Hagen, Vanessa F. Gonçalves, Paula L. Hedley, Marie Bækvad-Hansen, Jørgen K. Kanters, Jimmi Nielsen, Alfonso B. Demur, Thomas D. Als, James Kennedy, Michael Christiansen.

References

1. Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, et al. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science. 1988; 242(4884):1427–30. PMID: 3201231.

2. Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. Nature. 1988; 331(6158):717–9. https://doi.org/10.1038/331717a0 PMID: 2830540.

3. Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgeois M, Viegas-Pequignot E, et al. Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. Nat Genet. 1995; 11(2):144–9. https://doi.org/10.1038/ng1095-144 PMID: 7550341.

4. Koopman WJ, Willems PH, Smeitink JA. Monogenic mitochondrial disorders. N Engl J Med. 2012; 366(12):1132–41. https://doi.org/10.1056/NEJMra1012478 PMID: 22435372.

5. Calvo SE, Mootha VK. The mitochondrial proteome and human disease. Annu Rev Genomics Hum Genet. 2010; 11:25–44. https://doi.org/10.1146/annurev-genom-082509-141720 PMID: 20690818.

6. Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, et al. A mitochondrial protein compendium elucidates complex I disease biology. Cell. 2008; 134(1):112–23. https://doi.org/10.1016/j.cell.2008.06.016 PMID: 18614015.

7. Vafai SB, Mootha VK. Mitochondrial disorders as windows into an ancient organelle. Nature. 2012; 491(7424):374–83. https://doi.org/10.1038/nature11707 PMID: 23151580.

8. Chinnery PF, Hudson G. Mitochondrial genetics. Br Med Bull. 2013; 106:135–59. https://doi.org/10.1093/bmb/ldt017 PMID: 23704099.

9. Chinnery PF, Taylor GA, Howell N, Andrews RM, Morris CM, Taylor RW, et al. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology. 2000; 55(2):302–4. PMID: 10908912.

10. Carrieri G, Bonafe M, De Luca M, Rose G, Varcasia O, Bruni A, et al. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. Hum Genet. 2001; 108(3):194–8. PMID: 11354629.

11. van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, et al. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci Lett. 2004; 365(1):28–32. https://doi.org/10.1016/j.neulet.2004.04.051 PMID: 15234467.

12. Coskun P, Wyrembak J, Schriner SE, Chen HW, Marciniack C, Laferla F, et al. A mitochondrial etiology of Alzheimer and Parkinson disease. Biochim Biophys Acta. 2012; 1820(5):553–64. https://doi.org/10.1016/j.bbagen.2011.08.008 PMID: 21871538.

13. Hudson G, Nalls M, Evans JR, Breen DP, Winder-Rhodes S, Morrison KE, et al. Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. Neurology. 2013; 80(22): 2042–8. https://doi.org/10.1212/WNL.0b013e318294b434 PMID: 23645593.

14. Pyle A, Foltynie T, Tiangyou W, Lambert C, Keers SM, Allcock LM, et al. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann Neurol. 2005; 57(4):564–7. https://doi.org/10.1002/ana.20417 PMID: 15786469.

15. Marom S, Friger M, Mishmar D. MtDNA meta-analysis reveals both phenotype specificity and allele heterogeneity: a model for differential association. Sci Rep. 2017; 7:43449. https://doi.org/10.1038/srep43449 PMID: 28230165.

16. Sequeira A, Martin MV, Rollins B, Moon EA, Bunney WE, Macciardi F, et al. Mitochondrial mutations and polymorphisms in psychiatric disorders. Front Genet. 2012; 3:103. https://doi.org/10.3389/fgene.2012.00103 PMID: 22723804.

17. Rollins B, Martin MV, Sequeira PA, Moon EA, Morgan LZ, Watson SJ, et al. Mitochondrial variants in schizophrenia, bipolar disorder, and major depressive disorder. PLoS One. 2009; 4(3):e4913. https://doi.org/10.1371/journal.pone.0004913 PMID: 19290059.

18. Hudson G, Gomez-Duran A, Wilson IJ, Chinnery PF. Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases. Plos Genet. 2014; 10(5):e1004369. https://doi.org/10.1371/journal.pgen.1004369 PMID: 24852434.

19. Saxena R, de Bakker PI, Singer K, Mootha V, Burtt N, Hirschhorn JN, et al. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. Am J Hum Genet. 2006; 79(1):54–61. https://doi.org/10.1086/504926 PMID: 16773565.

20. Bybjerg-Grauholm J, Hagen, C. M. Goncalves, V. F., Bækvad-Hansen, M., Hansen, C.S., Hedley,P.L., Kanters, J.K., Nielsen, J., Theisen, M., Mors, O., Kennedy, J., Als, T.D., Demur, A.B., Nordentoft, M., Børglum,A.D., Mortensen, P.B., Werge, T.M., Hougaard, D.M., Christiansen, M. Complex spatio-temporal distribution and genogeographical affinity of mitochondrial DNA haplogroups in 24,216 danes. biorxiv/2017/148494. 2017.

21. Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, et al. The fine-scale genetic structure of the British population. Nature. 2015; 519(7543):309–14. https://doi.org/10.1038/nature14230 PMID: 25788095.

22. Busby GB, Hellenthal G, Montinaro F, Tofanelli S, Bulayeva K, Rudan I, et al. The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. Curr Biol. 2015; 25(19):2518–26. https://doi.org/10.1016/j.cub.2015.08.007 PMID: 26387712.

23. Zhang W, Tang J, Zhang AM, Peng MS, Xie HB, Tan L, et al. A matrilineal genetic legacy from the last glacial maximum confers susceptibility to schizophrenia in Han Chinese. J Genet Genomics. 2014; 41(7):397–407. https://doi.org/10.1016/j.jgg.2014.05.004 PMID: 25064678.

24. Du C, El Harchi A, Zhang H, Hancox JC. Modification by KCNE1 variants of the hERG potassium channel response to premature stimulation and to pharmacological inhibition. Physiol Rep. 2013; 1(6):e00175. https://doi.org/10.1002/phy2.175 PMID: 24400172.

25. Ji Y, Zhang AM, Jia X, Zhang YP, Xiao X, Li S, et al. Mitochondrial DNA haplogroups M7b1'2 and M8a affect clinical expression of leber hereditary optic neuropathy in Chinese families with the m.11778G—>a mutation. Am J Hum Genet. 2008; 83(6):760–8. https://doi.org/10.1016/j.ajhg.2008.11.002 PMID: 19026397.

26. Athanasiadis G, Cheng JY, Vilhjalmsson BJ, Jorgensen FG, Als TD, Le Hellard S, et al. Nationwide Genomic Study in Denmark Reveals Remarkable Population Homogeneity. Genetics. 2016; 204(2):711–22. https://doi.org/10.1534/genetics.116.189241 PMID: 27535931.

27. Karakachoff M, Duforet-Frebourg N, Simonet F, Le Scouarnec S, Pellen N, Lecointe S, et al. Fine-scale human genetic structure in Western France. Eur J Hum Genet. 2015; 23(6):831–6. https://doi.org/10.1038/ejhg.2014.175 PMID: 25182131.

28. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes mirror geography within Europe. Nature. 2008; 456(7218):98–101. https://doi.org/10.1038/nature07331 PMID: 18758442.

29. Owen MJ, Sawa A, Mortensen PB. Schizophrenia. Lancet. 2016; 388(10039):86–97. https://doi.org/10.1016/S0140-6736(15)01121-6 PMID: 26777917.

30. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol Rev. 2008; 30:67–76. https://doi.org/10.1093/epirev/mxn001 PMID: 18480098.

31. Cantor-Graae E, Pedersen CB. Full spectrum of psychiatric disorders related to foreign migration: a Danish population-based cohort study. JAMA Psychiatry. 2013; 70(4):427–35. https://doi.org/10.1001/jamapsychiatry.2013.441 PMID: 23446644.

32. Cantor-Graae E, Selten JP. Schizophrenia and migration: a meta-analysis and review. Am J Psychiatry. 2005; 162(1):12–24. https://doi.org/10.1176/appi.ajp.162.1.12 PMID: 15625195.

33. Dealberto MJ. Ethnic origin and increased risk for schizophrenia in immigrants to countries of recent and longstanding immigration. Acta Psychiatr Scand. 2010; 121(5):325–39. https://doi.org/10.1111/j.1600-0447.2009.01535.x PMID: 20105146.

34. Wallace DC. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu Rev Genet. 2005; 39:359–407. https://doi.org/10.1146/annurev.genet.39.110304.095751 PMID: 16285865.

35. Kenney MC, Chwa M, Atilano SR, Falatoonzadeh P, Ramirez C, Malik D, et al. Molecular and bioenergetic differences between cells with African versus European inherited mitochondrial DNA haplogroups: implications for population susceptibility to diseases. Biochim Biophys Acta. 2014; 1842(2):208–19. https://doi.org/10.1016/j.bbadis.2013.10.016 PMID: 24200652.

36. Larsen S, Diez-Sanchez C, Rabol R, Ara I, Dela F, Helge JW. Increased intrinsic mitochondrial function in humans with mitochondrial haplogroup H. Biochim Biophys Acta. 2014; 1837(2):226–31. https://doi.org/10.1016/j.bbabio.2013.10.009 PMID: 24184346.

37. Atilano SR, Malik D, Chwa M, Caceres-Del-Carpio J, Nesburn AB, Boyer DS, et al. Mitochondrial DNA variants can mediate methylation status of inflammation, angiogenesis and signaling genes. Hum Mol Genet. 2015; 24(16):4491–503. https://doi.org/10.1093/hmg/ddv173 PMID: 25964427.

38. Kenney MC, Chwa M, Atilano SR, Falatoonzadeh P, Ramirez C, Malik D, et al. Inherited mitochondrial DNA variants can affect complement, inflammation and apoptosis pathways: insights into mitochondrial-nuclear interactions. Hum Mol Genet. 2014; 23(13):3537–51. https://doi.org/10.1093/hmg/ddu065 PMID: 24584571.

39. Raffaello A, Mammucari C, Gherardi G, Rizzuto R. Calcium at the Center of Cell Signaling: Interplay between Endoplasmic Reticulum, Mitochondria, and Lysosomes. Trends Biochem Sci. 2016; 41(12):1035–49. https://doi.org/10.1016/j.tibs.2016.09.001 PMID: 27692849.

40. Verge B, Alonso Y, Valero J, Miralles C, Vilella E, Martorell L. Mitochondrial DNA (mtDNA) and schizophrenia. Eur Psychiatry. 2011; 26(1):45–56. https://doi.org/10.1016/j.eurpsy.2010.08.008 PMID: 20980130.

41. Burisch J. Crohn's disease and ulcerative colitis. Occurrence, course and prognosis during the first year of disease in a European population-based inception cohort. Dan Med J. 2014; 61(1):B4778. PMID: 24393595.

42. Kurtzke JF. Geographic distribution of multiple sclerosis: An update with special reference to Europe and the Mediterranean region. Acta Neurol Scand. 1980; 62(2):65–80. PMID: 6971038.

43. McLeod JG, Hammond SR, Kurtzke JF. Migration and multiple sclerosis in immigrants to Australia from United Kingdom and Ireland: a reassessment. I. Risk of MS by age at immigration. J Neurol. 2011; 258(6):1140–9. https://doi.org/10.1007/s00415-010-5898-4 PMID: 21264474.

44. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. Lancet Neurol. 2006; 5(6):525–35. https://doi.org/10.1016/S1474-4422(06)70471-9 PMID: 16713924.

45. Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: a systematic review. J Hepatol. 2012; 56(5):1181–8. https://doi.org/10.1016/j.jhep.2011.10.025 PMID: 22245904.

46. Aidt FH, Nielsen SM, Kanters J, Pesta D, Nielsen TT, Norremolle A, et al. Dysfunctional mitochondrial respiration in the striatum of the Huntington's disease transgenic R6/2 mouse model. PLoS Curr. 2013; 5. https://doi.org/10.1371/currents.hd.d8917b4862929772c5a2f2a34ef1c201 PMID: 23568011.

47. Arduino DM, Esteves AR, Swerdlow RH, Cardoso SM. A cybrid cell model for the assessment of the link between mitochondrial deficits and sporadic Parkinson's disease. Methods Mol Biol. 2015; 1265:415–24. https://doi.org/10.1007/978-1-4939-2288-8_31 PMID: 25634293.

48. Samuels DC, Carothers AD, Horton R, Chinnery PF. The power to detect disease associations with mitochondrial DNA haplogroups. Am J Hum Genet. 2006; 78(4):713–20. https://doi.org/10.1086/502682 PMID: 16532401.

49. Raule N, Sevini F, Santoro A, Altilia S, Franceschi C. Association studies on human mitochondrial DNA: methodological aspects and results in the most common age-related diseases. Mitochondrion. 2007; 7(1–2):29–38. https://doi.org/10.1016/j.mito.2006.11.013 PMID: 17306632.

50. Biffi A, Anderson CD, Nalls MA, Rahman R, Sonni A, Cortellini L, et al. Principal-component analysis for assessment of population stratification in mitochondrial medical genetics. Am J Hum Genet. 2010; 86(6):904–17. https://doi.org/10.1016/j.ajhg.2010.05.005 PMID: 20537299.

51. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, et al. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet. 2004; 75(5):910–8. https://doi.org/10.1086/425590 PMID: 15382008.

52. Agerbo E, Mortensen PB, Wiuf C, Pedersen MS, McGrath J, Hollegaard MV, et al. Modelling the contribution of family history and variation in single nucleotide polymorphisms to risk of schizophrenia: a Danish national birth cohort-based study. Schizophr Res. 2012; 134(2–3):246–52. https://doi.org/10.1016/j.schres.2011.10.025 PMID: 22108675.

53. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 2009; 30(2):E386–94. https://doi.org/10.1002/humu.20921 PMID: 18853457.

54. Breiman L. Random forests. Machine Learning. 2001; 45(1):5–32. https://doi.org/10.1023/A:1010933404324