Mol Neurobiol (2017) 54:4343–4352
DOI 10.1007/s12035-016-9996-x

mtDNA Heteroplasmy in Monozygotic Twins Discordant for Schizophrenia

Hong Li[1,6] · Rui Bi[2] · Yu Fan[2] · Yong Wu[2,3] · Yanqing Tang[4] · Zongchang Li[1] · Ying He[1] · Jun Zhou[1] · Jinsong Tang[1] · Xiaogang Chen[1] · Yong-Gang Yao[2,3,5]

Received: 19 March 2016 / Accepted: 14 June 2016 / Published online: 24 June 2016

Abstract Although monozygotic (MZ) twins have theoretically identical nuclear DNA sequences, there may be phenotypic differences between them caused by somatic mutations and epigenetic changes affecting each genome. In this study, we collected eight families of MZ twins discordant for schizophrenia with the aim of investigating the potential role of mitochondrial DNA (mtDNA) heteroplasmy in causing the phenotypic differences between the twin pairs. Next-generation sequencing (NGS) technology was used to screen the whole mitochondrial genome of the twin pairs and their parents. The mtDNA heteroplasmy level was found to be nearly identical between the twin pairs but was distinctly different between each mother and their offspring. These results suggest that the discordance of schizophrenia between MZ twins may not be attributable to the difference in mtDNA heteroplasmy, and the high concordance of mtDNA heteroplasmy between MZ twins may indicate the relatively equal distribution of mtDNA during embryo separation of MZ twins and/or the modulation effect from the same nuclear genetic background. Furthermore, we observed an overrepresentation of heteroplasmy in noncoding regions and an elevated ratio of nonsynonymous heteroplasmy, suggesting the possible effects of a purifying selection in shaping the pattern of mtDNA heteroplasmy.

Keywords Monozygotic twins · mtDNA · Heteroplasmy · Schizophrenia

Hong Li and Rui Bi contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s12035-016-9996-x) contains supplementary material, which is available to authorized users.

✉ Jinsong Tang
tangjinsonghn@gmail.com

✉ Xiaogang Chen
chenxghn@163.com

✉ Yong-Gang Yao
yaoyg@mail.kiz.ac.cn

[1] Institute of Mental Health of the Second Xiangya Hospital of Central South University, The China National Clinical Research Center for Mental Health Disorders, National Technology Institute of Psychiatry, Key Laboratory of Psychiatry and Mental Health of Hunan Province, Changsha, Hunan 410011, China

[2] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China

[3] Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650204, China

[4] Department of Psychiatry, the First Affiliated Hospital of China Medical University, Shenyang 110122, China

[5] CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China

[6] Department of Psychiatry, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450000, China

Introduction

Schizophrenia (MIM 181500) is a common psychiatric disorder affected by both genetic and environmental factors [1]. Although the heritability of schizophrenia is as high as 80 % [2], the genetic etiology of schizophrenia is very complex and remains largely unknown.

Due to the important role of mitochondria in energy metabolism, neurological diseases that are linked to tissues with high energy demands, such as brain, have been reported to be associated with mitochondrial defects [3]. Accumulating evidence suggests that mitochondrial dysfunction, either caused

by mitochondrial DNA (mtDNA) mutations or nuclear encoded mitochondrial genes, is involved in the pathogenesis of schizophrenia [4–10]. There are hundred to thousand copies of mtDNA in a single cell, and the coexistence of different mtDNA haplotypes within an individual or within a cell is called mtDNA heteroplasmy [11, 12]. Heteroplasmy of pathogenic mtDNA mutations has been considered to be one of the reasons for the incomplete penetrance of several mitochondrial diseases [12, 13]. The presence of disease symptoms is (partially) dependent on whether the allele frequency of a mtDNA pathogenic mutation exceeds a certain threshold [14]. However, to quantify the mtDNA heteroplasmy in a cell is not easy [11]. Recent studies have made great efforts to improve the sensitivity of heteroplasmy detection, and next-generation sequencing (NGS) has been established as being an effective method for the identification of low levels of mtDNA heteroplasmy [15–21].

In this study, we collected eight sets of monozygotic (MZ) twins and their parents. Each MZ twin pair showed a discordant phenotype in that only one individual of the pair had developed schizophrenia, while the other one was mentally healthy. As the nuclear genomes of MZ twins are theoretically identical, we hypothesized that different levels of mtDNA heteroplasmy might account for the phenotypic discordance between the twins. NGS technology was used to screen for mtDNA variants and heteroplasmy in these eight pairs of MZ twins and their parents. Our study has provided a straightforward way of testing this hypothesis in respect of schizophrenia and revealed several important insights concerning mtDNA heteroplasmy.

Materials and Methods

Sample Collection

Eight families with MZ twins discordant for schizophrenia were recruited in this study. Demographical information of each family member is shown in supplementary Table S1. The schizophrenia patients were diagnosed independently by two psychiatrists following the criterion of Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). Informed consents conforming to the tenets of the Declaration of Helsinki were obtained from each participant or guardian prior to this study. This study was approved by the ethics committees of the Second Xiangya Hospital of Central South University and Kunming Institute of Zoology.

Whole mtDNA Genome Sequencing, Alignment, and Variant Calling

Genomic DNA from the eight MZ twin pairs and their parents was obtained from the peripheral blood by using AxyPrep™ Blood Genomic DNA Miniprep Kit (Axygen, USA) according to the manufacturer's instruction. Whole genome sequencing was performed at the Beijing Genomics Institute (BGI)-Shenzhen using the Illumina HiSeq Platform (500-bp library, 90-bp reads). Each set of paired end read files was run through Trimmomatic [22] to remove low-quality base pairs with the parameters as "LEADING:3; TRAILING:3; SLIDINGWINDOW:4:15; MINLEN:36." Quality-filtered reads were mapped against the human genome (UCSC [http://genome.ucsc.edu], assembly hg19) by BWA version 0.79a [23]. The mapped read sets were sorted by genomic position by SortSam and then merged by MergeSam Files in picard-tools-1.107 (https://github.com/broadinstitute/picard). Possible PCR duplicates in Illumina mate-pair reads were removed by MarkDuplicates in picard-tools-1.107. The reads uniquely mapped the mitochondrial genome sequences of hg19 were extracted and were further realigned relative to the Revised Cambridge Reference Sequence (rCRS) (GenBank Accession No. NC_012920) [24] to avoid the contamination of nuclear sequences homologous to mtDNA genome (NUMT) [25]. We used the unified genotyper (UG) in GATK 2.8 for SNV discovery and calculation of genotype likelihoods on each family by the parameters as recommended [26]. We considered all SNVs called by the GATK UG with a Phred-quality score >Q10 as a starting point before filtering in order to maximize sensitivity. To avoid potential errors caused by the flaw in the algorithm of UG, we also programmed the same pipeline individually for each sample and excluded the inconsistent sites.

Phylogenetic Analysis of the Mitochondrial Genome

A phylogenetic tree for the 32 mtDNA sequences from the 8 MZ twin families was constructed following the same approach as described in our recent studies [27, 28]. The haplogroup status of each mtDNA was determined according to the updated East Asian mtDNA tree and Phylotree (http://phylotree.org/tree/main.htm; mtDNA tree Build 17, 18 Feb 2016) [29] and was validated by MitoTool (www.mitotool.org) [30]. mtDNA variants relative to the rCRS were displayed in an mtDNA phylogenetic tree. Potentially functional private variants (variants that are nonsynonymous or in the tRNA/rRNA genes) located in the terminal branches of the tree were further analyzed for their uniqueness, conservation, and pathogenicity as described in our previous studies [27, 28].

Assessing the Pattern of the Observed mtDNA Differences

The average coverage for each mtDNA genome was around 250×. We set the detection threshold for mtDNA heteroplasmy at 4 % so that only heteroplasmy with minor allele frequency more than 4 %, or with at least 10 high quality reads were considered. As heteroplasmy observed in long C

stretch regions may be caused by sequencing errors, the mtDNA poly-C regions (regions 302–316, 566–573, and 16181–16194 in the rCRS) were excluded from the analysis.

mtDNA variants across the mitochondrial genome were scored relative to the rCRS [24]. The distribution pattern of the mtDNA mutations were counted in respect of the gene-coding and noncoding regions. The variants (heteroplasmic or homoplasmic) shared by the mothers and their offspring in the same family were only counted once. In order to exclude the sampling bias in our study, 118 complete mtDNA sequences (GenBank Accession No. AY255133–AY255180, DQ272107–DQ272126, HM030499–HM030548) from across China were assessed to obtain an overall distribution pattern of the variants [31–33]. As the reported 118 individuals were unrelated, the count of each variant in these mtDNA sequences was equal to the number of individuals sharing this variant. The differences in the variant distribution between two groups were compared by using the Fisher's exact test with MitoTool [30]. The correlation of mtDNA heteroplasmy with age was estimated by GraphPad Prism software (GraphPad Software, La Jolla, CA, USA).

The absolute value of the heteroplasmy difference was analyzed among four groups: group "M vs. T1," the heteroplasmy difference between mother (M) and offspring 1 (T1: healthy offspring of each family; supplementary Table S1); group "M vs. T2," the heteroplasmy difference between mother and offspring 2 (T2: proband of each family); group "M vs. T," the average heteroplasmy difference between the mothers and the twin pairs (T); group "T1 vs. T2," the heteroplasmy difference between the MZ twin pairs. The statistical difference between pairs of these four groups was estimated with Student's *t* test by using GraphPad Prism software (GraphPad Software, La Jolla, CA, USA).

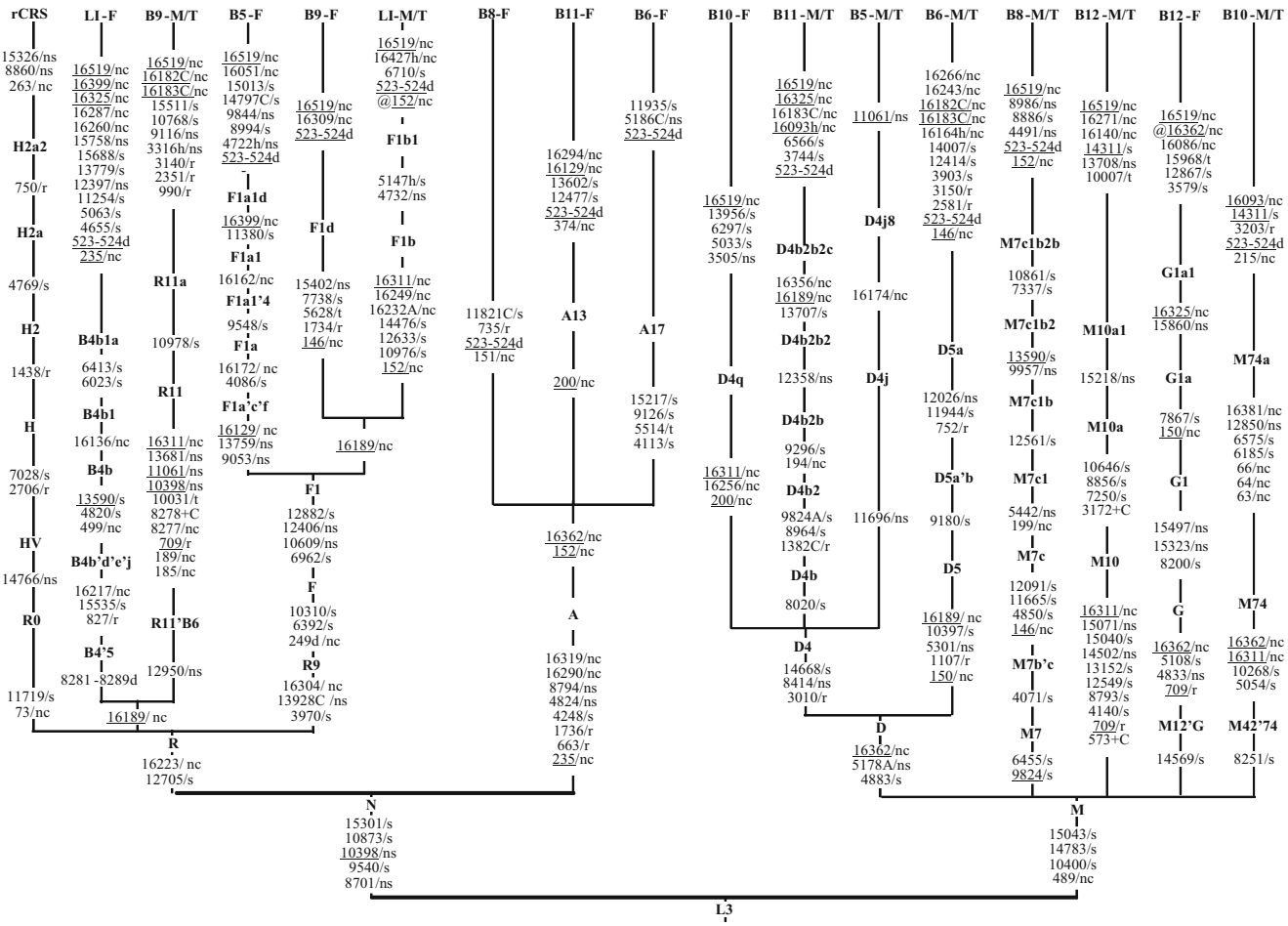

Fig. 1 Haplogroup tree of 32 complete mtDNA sequences from the eight families with monozygotic (MZ) twins. The revised Cambridge Reference Sequence (rCRS) [24] was included in the tree to show the phylogenetic position of each lineage. Fathers, mothers, and twins are indicated by *F*, *M*, and *T*, respectively. Synonymous and nonsynonymous variants were marked as *s* and *ns*, respectively. Variants in the rRNA or tRNA genes are indicated by *r* or *t*, respectively. Variants in the noncoding region are labeled by *nc*. Suffixes *A* and *C* mean transversions. Recurrent mutations are *underlined*. Back mutations are *underlined* and *marked* @. Variants which had a heteroplasmic level over 10 % or with at least 20 high quality reads are marked by *h*. Variants 309+C(C) and 315+C were not considered

Results

Absence of Potentially Pathogenic mtDNA Germline Mutations

The monozygosity of the twins was confirmed by the high concordance of their nuclear genome (authors' unpublished data). The entire mtDNA genome sequences of the 32 individuals from the eight families have been deposited in GenBank under accession numbers KX228162–KX228193. mtDNA variants (either homoplasmic or with over 10 % heteroplasmy) relative to rCRS were listed in supplementary Table S2 and were displayed in the tree (Fig. 1). All the mtDNAs could be classified as belonging to recognized East Asian haplogroups (Fig. 1 and supplementary Table S2). Considering that potentially pathogenic mutations are most likely to be private mutations that are located in the terminal branches of the tree [8, 27], we analyzed the private variants which were nonsynonymous or located in the tRNA/rRNA genes in the mtDNA of the mothers and the twins. Except for m.3150T>C and m.8986A>G, the other private variants are haplogroup-defining variants and are common in the general population (Table 1), and are unlikely to be pathogenic albeit they may be functional [34]. The potential for variants m.3150T>C and m.8986A>G to cause deleterious effects was considered and excluded, because these variants are not conserved and have also been identified in mtDNA sequences in the general populations according to the web-based search [35]. In conclusion, there were no potentially pathogenic mtDNA mutations at the germline level identified in any of the MZ twin pair families.

The Distribution Pattern of mtDNA Heteroplasmy

Our cutoff criteria were the presence of a mutation in more than 10 reads or having an allele frequency higher than 4 % in all reads (approximately 250×). A total of 16 heteroplasmic mutations were identified (Table 2). The distribution pattern of the mutations across the mitochondrial genome was similar among the heteroplasmic mutations, homoplasmic variants, and the germline variants in 118 previously reported Chinese mtDNA sequences [31–33] (Fig. 2a–c and supplementary Table S3). Eight of the 16 heteroplasmic mutations (50 %) were in the noncoding control region. As the noncoding region only covers ~1180 bp (~7 %) of the mtDNA genome, the frequency of heteroplasmic mutations in this region is

Table 1 The nonsynonymous and tRNA/rRNA private germline variants in monozygotic twins and their mothers

Family[a]	Private variant (amino acid change)	Gene	Reported (population context)[b]	Haplogroup-specific variant[c]	Conservation Index (CI)[d]	Pathogenic score[e]
B5	m.11061C>T (p.S101F)	*MT-ND4*	yes	yes	0.54	0.479
B6	m.2581A>G	*MT-RNR2*	yes	yes	0.25	–
	m.3150T>C	*MT-RNR2*	yes	no	0.25	–
B8	m.4491G>A (p.V8I)	*MT-ND2*	yes	yes	0.17	0.439
	m.8986A>G (p.M154V)	*MT-ATP6*	yes	no	0.50	0.497
B9	m.990T>C	*MT-RNR1*	yes	yes	0.71	–
	m.2351T>C	*MT-RNR2*	yes	yes	0.33	–
	m.3140A>G	*MT-RNR2*	yes	yes	1.00	–
	m.3316G>A (p.A4T)	*MT-ND1*	yes	yes	0.38	0.463
	m.9116T>C (p.I197T)	*MT-ATP6*	yes	yes	0.63	0.671
B10	m.3203A>G	*MT-RNR2*	yes	yes	0.69	–
B12	m.10007T>C	*MT-TG*	yes	yes	0.44	–
	m.13708G>A (p.A458T)	*MT-ND5*	yes	yes	0.37	0.409

[a] The complete mtDNA genomes of mothers and twins in family B11 and family LI contained no private nonsynonymous and mt-tRNA/rRNA variants and were not included in the table

[b] The uniqueness of each variant was assessed following the described strategy [35] on 3 December 2015 (e.g., both "C11061T mtDNA" and "11061C>T mtDNA" were queried)

[c] The haplogroup-specific variant was determined according to the available global mtDNA phylogenetic tree (Phylotree, http://phylotree.org/tree/main.htm; mtDNA tree Build 17, 18 Feb 2016)

[d] The conservation index (CI) [57] was calculated by using MitoTool (http://www.mitotool.org/) based on 43 primate species. A CI value of 0.54 means that 54 % of 43 primate species share the same wide-type allele with human sequence (GenBank Accession number NC_012920)

[e] The pathogenicity score was consulted from Pereira et al.'s study [58]. The score ranges from 0 to 1. Higher pathogenicity score is associated with greater possibility that the variant is pathogenic

Table 2 Heteroplasmic variants identified in the monozygotic twin families

Family	Variant[a]	Location	F[b]	M[b]	T1[b]	T2[b]	M vs. T[c]	T1 vs. T2[d]
B5	m.4722A>G (ns)	ND2	**0.548**	–	–	–	–	–
B5	m.5178C>A (ns)	ND2	–	0.992	**0.936**	**0.936**	0.056	0.000
B5	m.8414C>T (ns)	ATP8	–	0.976	**0.931**	**0.956**	0.032	0.025
B5	m.14668C>T (s)	ND6	–	0.988	0.972	**0.956**	0.024	0.016
B5	m.16174C>T	D-Loop	–	**0.952**	**0.952**	**0.947**	0.002	0.005
B5	m.16362T>C	D-Loop	–	0.964	0.964	**0.912**	0.026	0.052
B6	m.16164A>G	D-Loop	–	**0.864**	**0.896**	**0.872**	0.020	0.024
B6	m.16223C>T	D-Loop	0.996	**0.96**	0.968	**0.96**	0.004	0.008
B9	m.3316G>A (ns)	ND1	–	**0.225**	0.032	0.016	0.201	0.016
B11	m.10400C>T (s)	ND3	–	**0.956**	1.000	0.996	0.042	0.004
B11	m.16093T>C	D-Loop	–	0.968	**0.948**	**0.928**	0.030	0.020
B11	m.16294C>T	D-Loop	**0.944**	–	–	–	–	–
B12	m.16086T>C	D-Loop	**0.908**	–	–	–	–	–
LI	m.5147G>A (s)	ND2	–	**0.772**	1.000	0.996	0.226	0.004
LI	m.6392T>C (s)	CO1	–	**0.908**	0.984	0.992	0.080	0.008
LI	m.16427C>T	D-Loop	–	**0.104**	**0.088**	**0.072**	0.024	0.016
Average difference[e]							0.059	0.015

[a] Nonsynonymous and synonymous variants are marked by "ns" and "s" in brackets, respectively

[b] Variant frequency in father (F), mother (M), offspring 1 (T1), and offspring 2 (T2). Variants with minor allele frequency higher than 4 % are marked in bold

[c] The absolute value of the heteroplasmy difference between mother and the average value of twins

[d] The absolute value of the heteroplasmy difference between the twin pairs

[e] The average value of the difference in "M vs. T" group and "T1 vs. T2" group

significantly higher than expected ($P = 0.02$, Fig. 2d). A similar pattern of overrepresentation of variants in the noncoding region was also observed for the homoplasmic variants in these eight families ($P < 0.001$, Fig. 2d), as well as in the 118 Chinese complete mtDNA sequences ($P < 0.001$, Fig. 2d). Interestingly, the nonrandom distribution of variants in the noncoding region was more evident for heteroplasmic mutations when compared with homoplasmic variants and variants in the general Chinese populations (supplementary Table S3; Fig. 2d). Further analysis of variants in the coding region revealed a higher ratio of heteroplasmic nonsynonymous (ns) variants among the total heteroplasmic variants (ns/all = 4/8; 50 %) than the homoplasmic ns ratio in the homoplasmic variants (ns/all = 116/324; 35.8 %), and the ns ratio of 118 Chinese mtDNA sequences (ns/all = 885/2676; 33.1 %) (supplementary Table S3). All of these findings fit with the suggestion of a purifying acting on the selection of mtDNA mutations [15, 36, 37].

The correlation between the number of heteroplasmic mutations and age was further analyzed using the data obtained from the fathers and mothers of the families. We did not include the twin individuals in the analysis because mtDNA heteroplasmy could pass from the mothers to their offspring and be modulated by nuclear genetic background [11, 38, 39]. The number of heteroplasmic mutations showed no apparent trend toward increasing with age (Fig. 2e). This result should be received with caution, as the pattern may have been caused by our relatively small sample size and the oversampling of individuals with an age of 40–50 years old.

Difference of mtDNA Heteroplasmy Among the MZ Twins and Their Mother

Using the homoplasmic or heteroplasmic mtDNA variants identified in our study, we investigated whether there were any differences relating to these mtDNA variants between the MZ twin pairs. We compared the differences between the MZ twin pairs according to the spectrum of mtDNA variants and the observed heteroplasmic levels. The average heteroplasmy difference between each twin pair was around 1.5 % (Table 2), indicating a high concordance of mtDNA heteroplasmy between the MZ twin pairs. Furthermore, we observed a substantially elevated heteroplasmy difference between the mothers and one of her twin pair (average difference 5.6–6.1 %; Fig. 3a) or between the mothers and the average of her twin pair (average difference 5.9 %; Fig. 3b). This pattern indicated that less

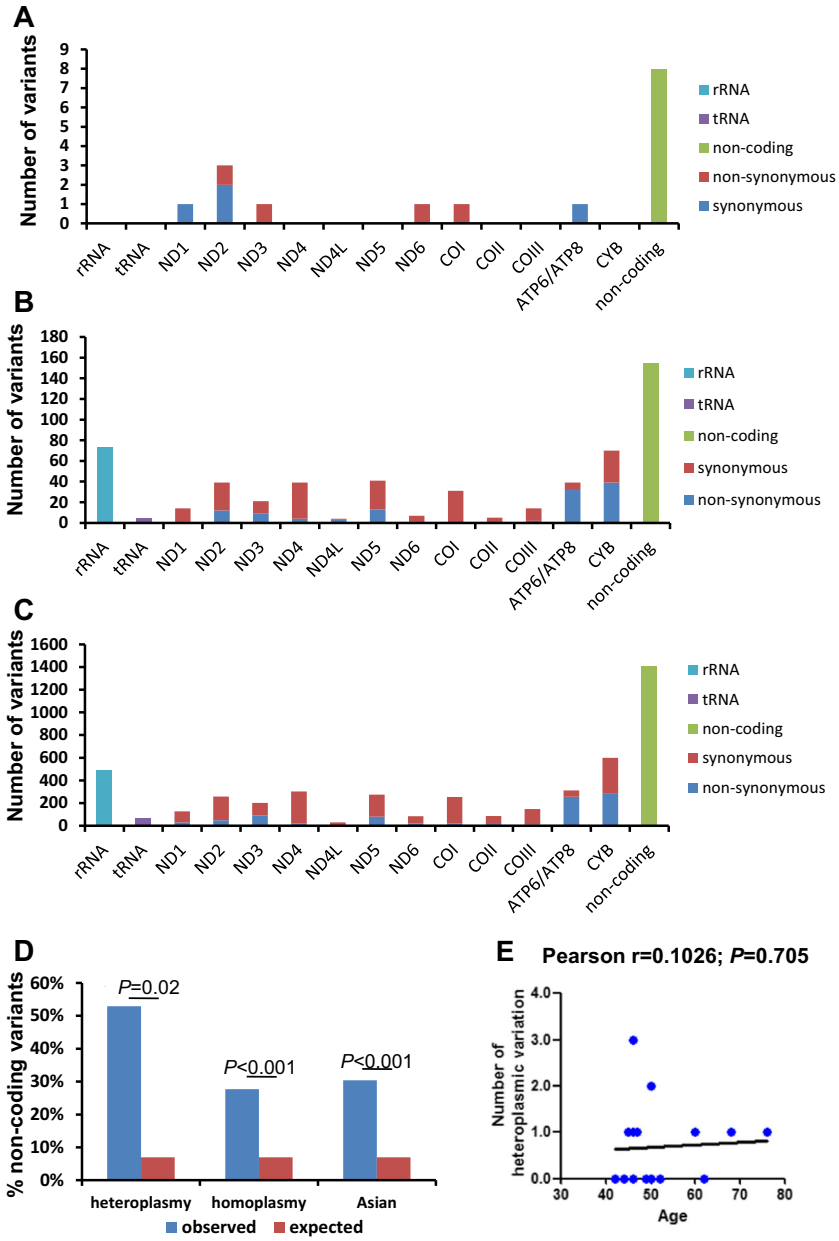

Fig. 2 The distribution pattern of mtDNA variants across the mitochondrial genome. The distribution pattern of **a** heteroplasmic mutations, **b** homoplasmic variants, and **c** variants in 118 reported Chinese mtDNA complete sequences [31–33]. **d** The ratio of the noncoding variants was compared by the Fisher's exact test with MitoTool [30]. **e** The correlation of mtDNA heterogeneity with age was estimated by GraphPad Prism software (GraphPad Software, La Jolla, CA, USA)

variability exists between twin individuals relative to that of between a mother and her offspring.

The role of aging in shaping heteroplasmy difference between the twins and their mothers or between the twin pairs was not evident, as demonstrated by the lack of correlation of age with the heteroplasmy difference, both for the heteroplasmy difference between the mothers and twins (Fig. 3c) or between the twin pairs (Fig. 3d). Again, this result was based on the limited number of samples for consideration. Taken together, we found that the mtDNA heteroplasmic level was similar between MZ twin pairs but was different between twins and their mothers.

Discussion

The pathogenesis of schizophrenia is complex and the clinical phenotype may differ between MZ twins [40, 41]. As MZ twins have theoretically identical nuclear genomic DNA sequences, it is possible that some other genetics factors inherited in a non-Mendelian fashion could potentially modify the phenotype of schizophrenia. Moreover, a study of identical twins and their mothers offers us a good opportunity to trace the occurrence of mutations, either in an inherited or somatic way.

In this study, we used NGS to investigate the differences of the mtDNA variants and levels of heteroplasmy

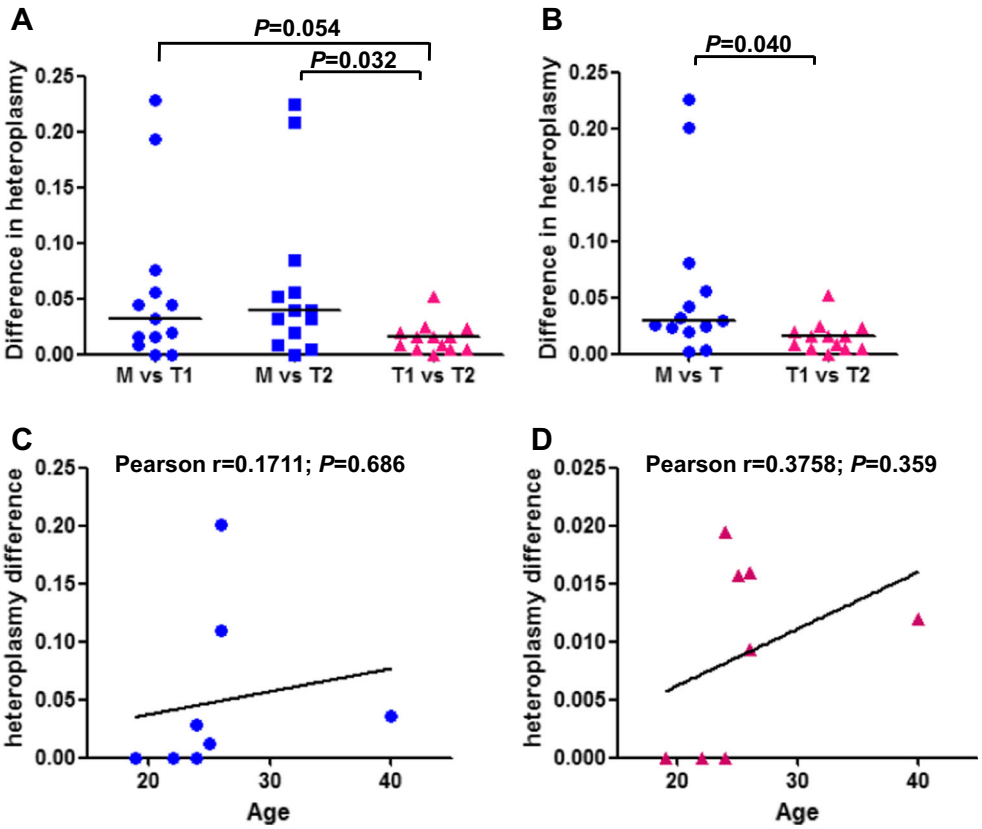

Fig. 3 The difference of mtDNA heteroplasmy among mother, offspring 1 and offspring 2. **a**–**b** The absolute values of difference of heteroplasmy among mother (*M*), offspring 1 (*T1*), offspring 2 (*T2*), and the average values of offspring 1 and offspring 2 (*T*). The difference of heteroplasmy was compared by the Student's *t* test using GraphPad Prism software (GraphPad Software, La Jolla, CA, USA). **c** The correlation of heteroplasmy difference between the mothers and twins with age of twins. **d** The correlation of heteroplasmy difference between each twin pair with their age. Correlation test was performed by using GraphPad Prism software (GraphPad Software, La Jolla, CA, USA)

in MZ twin pairs who were discordant for schizophrenia. However, we found no clear evidence for an mtDNA contribution to the phenotypic difference in these MZ twins, indicating that heteroplasmy of common mtDNA variants might have very limited effects on the risk of schizophrenia. Indeed, most of the positive results in the literature concerning common mtDNA variants/haplogroups and schizophrenia are probably false positives [42]. Our result is in agreement with previous studies concerning the role of mtDNA heteroplasmy in discordant MZ twins with other human diseases [36, 43], indicating that factors, such as epigenetic changes and environmental factors, may play a more important role in shaping the discordance between MZ twins with disease [1, 41, 44]. However, we could not exclude a possible combinational effect of multiplex natures of mtDNA heteroplasmy in schizophrenia, especially considering the fact that mtDNA somatic mutations and deletions in neurons, albeit at a minimum level of heteroplasmy, might have a deleterious effect [45].

The direct quantification of the difference of mtDNA mutations and heteroplasmic levels between each twin pair and their parents revealed several important insights concerning mtDNA heteroplasmy. There is an overrepresentation of the noncoding variants, and this tendency was more evident for mtDNA heteroplasmic mutations than for homoplasmic germline variants. In the coding region, an elevated ratio of nonsynonymous heteroplasmic variations was identified. All these findings were consistent with previous studies [15, 36, 37], suggesting potentially negative selection acting on the mitochondrial genome, especially on the generation of low levels of mtDNA heteroplasmy [37]. Furthermore, we identified no significant correlation between age and the number of heteroplasmic mutations, which was different from the previous reports that mtDNA heterogeneity accumulates during aging [20, 39, 46–48]. The exact reason for this discrepancy may be due to insufficient number of samples analyzed in this study.

Unlike the nuclear genome, mtDNA is located in mitochondria and is maternally inherited through the cytoplasm [49]. Previous studies have suggested an uneven distribution of mitochondria at the time of embryo separation, which starts from the two-cell stage and leads to the differential pattern of mtDNA heterogeneity between daughter cells [50–52]. These findings indicated that the variability of mtDNA heterogeneity may exist between identical twins. However, inconsistent with this notion, the NGS of the mitochondrial genomes of the MZ twins in this study unexpectedly revealed similar levels of mtDNA heterogeneity in the MZ twin pairs. The average heteroplasmy difference in the MZ twin

pairs was around 1.5 %, which was significantly lower than the difference between the mothers and their twin offspring (5.9 %), suggesting that the approximately equal mutation rate and mitochondria distribution occurred during early embryo separation and the subsequent lifetime of the MZ twins. The identification of mtDNA mutations at a lower level of heteroplasmy between mother and offspring further confirmed previous observations for a pedigree-specific occurrence of certain mtDNA variants and potential modulation effect from the family genetic background [38, 39]. However, how these mutations are generated and inherited in these families remains unknown.

In contrast with our results, a recent study by Wang et al. [53] identified several sites with distinct mtDNA heteroplasmic level between the MZ twins by using NGS technology. However, an audit of the mtDNA data reported by Wang et al. [53] using a phylogenetic approach [54] showed that this data contained many potential errors. For instance, heteroplasmy at site 15301 (an ancient variant which distinguishes macrohaplogroups M and N) was identified in 8 of 10 twin pairs in their study, such a high frequency had never been observed in the available reported mtDNA data in the general populations (>24,275 mtDNA complete sequences, Phylotree: http://phylotree.org/tree/main.htm; MitoTool: http://www.mitotool.org/) and was most likely caused by contamination or sequencing errors; there were also a lot of haplogroup-defining variants missed in their study (Supplementary Table S4), which significantly weakened the overall conclusion in the original study [53]. As the NGS approach is very sensitive at identifying low levels of mutations, there is correspondingly a high chance of showing erroneous results from any traces of contamination in the sample. Therefore, the mtDNA data obtained from NGS should also be checked for data quality by using a phylogenetic method [35, 54–56] to ensure reliable conclusions can be made.

In conclusion, based on our cutoff standard to score mtDNA mutations using NGS technology, we found that there was no significant difference in mtDNA heterogeneity between MZ twin pairs discordant for schizophrenia. We also observed shared mutations between the mothers and their offspring at a lower level of heteroplasmy, with a similar level of heteroplasmy between MZ twin pairs, but a different pattern between the twins and their mothers. The pattern of mtDNA heteroplasmy bears the signature of purifying selection. Future studies investigating mtDNA heteroplasmy in a large set of MZ twin families will be promising to reveal more clues about how mtDNA heteroplasmy is passed from mothers to their offspring and how mtDNA heteroplasmy changes during the embryo separation and the subsequent lifetime of the MZ twin pair.

Acknowledgments We are grateful to the subjects who donated DNA samples. We thank Ian Logan for helpful comments and language editing. This study was supported by the Strategic Priority Research Program (B) of the Chinese Academy of Sciences (XDB02020003), the National Natural Science Foundation of China (31171225, 81271484, 81471361, and 81371480), and the National Key Basic Research and Development Program (973) (Grant No. 2012CB517904).

Compliance with Ethical Standards

Conflict of Interest None

References

1. Brown AS (2011) The environment and susceptibility to schizophrenia. Prog Neurobiol 93(1):23–58
2. Cannon TD, Kaprio J, Lönnqvist J, Huttunen M, Koskenvuo M (1998) The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study. Arch Gen Psychiatry 55(1):67–74
3. Mattson MP, Gleichmann M, Cheng A (2008) Mitochondria in neuroplasticity and neurological disorders. Neuron 60(5):748–766
4. Bamne MN, Talkowski ME, Moraes CT, Manuck SB, Ferrell RE, Chowdari KV, Nimgaonkar VL (2008) Systematic association studies of mitochondrial DNA variations in schizophrenia: focus on the ND5 gene. Schizophr Bull 34(3):458–465
5. Bandelt H-J, Yao YG, Kivisild T (2005) Mitochondrial genes and schizophrenia. Schizophr Res 72(2–3):267–269
6. Rollins B, Martin MV, Sequeira PA, Moon EA, Morgan LZ, Watson SJ, Schatzberg A, Akil H et al (2009) Mitochondrial variants in schizophrenia, bipolar disorder, and major depressive disorder. PLoS One 4(3):e4913
7. Washizuka S, Kametani M, Sasaki T, Tochigi M, Umekage T, Kohda K, Kato T (2006) Association of mitochondrial complex I subunit gene NDUFV2 at 18p11 with schizophrenia in the Japanese population. Am J Med Genet B Neuropsychiatr Genet 141B(3):301–304
8. Bi R, Tang J, Zhang W, Li X, Chen SY, Yu D, Chen X, Yao YG (2016) Mitochondrial genome variations and functional characterization in Han Chinese families with schizophrenia. Schizophr Res 171(1–3):200–206
9. Zhang W, Tang J, Zhang AM, Peng MS, Xie HB, Tan L, Xu L, Zhang YP et al (2014) A matrilineal genetic legacy from the last glacial maximum confers susceptibility to schizophrenia in Han Chinese. J Genet Genomics 41(7):397–407
10. Li X, Zhang W, Tang J, Tan L, Luo XJ, Chen X, Yao YG (2015) Do nuclear-encoded core subunits of mitochondrial complex I confer genetic susceptibility to schizophrenia in Han Chinese populations? Sci Rep 5:11076
11. Yao YG, Kajigaya S, Young NS (2015) Mitochondrial DNA mutations in single human blood cells. Mutat Res 779:68–77
12. Stewart JB, Chinnery PF (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet 16(9):530–542
13. Taylor RW, Turnbull DM (2005) Mitochondrial DNA mutations in human disease. Nat Rev Genet 6(5):389–402
14. DiMauro S, Schon EA (2003) Mitochondrial respiratory-chain diseases. N Engl J Med 348(26):2656–2668
15. Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M (2010) Detecting heteroplasmy from high-throughput

sequencing of complete human mitochondrial DNA genomes. Am J Hum Genet 87(2):237–249
16. Li M, Stoneking M (2012) A new approach for detecting low-level mutations in next-generation sequence data. Genome Biol 13(5): R34
17. Sosa MX, Sivakumar IK, Maragh S, Veeramachaneni V, Hariharan R, Parulekar M, Fredrikson KM, Harkins TT et al (2012) Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. PLoS Comput Biol 8(10): e1002737
18. Greaves LC, Nooteboom M, Elson JL, Tuppen HA, Taylor GA, Commane DM, Arasaradnam RP, Khrapko K et al (2014) Clonal expansion of early to mid-life mitochondrial DNA point mutations drives mitochondrial dysfunction during human ageing. PLoS Genet 10(9):e1004620
19. Williams SL, Mash DC, Zuchner S, Moraes CT (2013) Somatic mtDNA mutation spectra in the aging human putamen. PLoS Genet 9(12):e1003990
20. Kennedy SR, Salk JJ, Schmitt MW, Loeb LA (2013) Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. PLoS Genet 9(9):e1003794
21. Kloss-Brandstatter A, Weissensteiner H, Erhart G, Schafer G, Forer L, Schonherr S, Pacher D, Seifarth C et al (2015) Validation of next-generation sequencing of entire mitochondrial genomes and the diversity of mitochondrial DNA mutations in oral squamous cell carcinoma. PLoS One 10(8):e0135643
22. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120
23. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760
24. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23(2):147
25. Yao YG, Kong QP, Salas A, Bandelt HJ (2008) Pseudomitochondrial genome haunts disease studies. J Med Genet 45(12):769–772
26. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491–498
27. Bi R, Li W-L, Chen M-Q, Zhu Z, Yao Y-G (2011) Rapid identification of mtDNA somatic mutations in gastric cancer tissues based on the mtDNA phylogeny. Mutat Res 709–710:15–20
28. Bi R, Zhang AM, Jia X, Zhang Q, Yao YG (2012) Complete mitochondrial DNA genome sequence variation of Chinese families with mutation m.3635G>A and Leber hereditary optic neuropathy. Mol Vis 18:3087–3094
29. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386–E394
30. Fan L, Yao YG (2013) An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. Mitochondrion 13(4):360–363
31. Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003) Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73(3):671–676
32. Kong QP, Sun C, Wang HW, Zhao M, Wang WZ, Zhong L, Hao XD, Pan H et al (2011) Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. Mol Biol Evol 28(1):513–522
33. Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L et al (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet 15(13):2076–2086
34. Bi R, Zhang W, Yu D, Li X, Wang HZ, Hu QX, Zhang C, Lu W et al (2015) Mitochondrial DNA haplogroup B5 confers genetic susceptibility to Alzheimer's disease in Han Chinese. Neurobiol Aging 36(3):1604.e7–16
35. Bandelt H-J, Salas A, Taylor RW, Yao Y-G (2009) Exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches. Hum Mutat 30(2):191–196
36. Avital G, Buchshtav M, Zhidkov I, Tuval Feder J, Dadon S, Rubin E, Glass D, Spector TD et al (2012) Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. Hum Mol Genet 21(19): 4214–4224
37. Ye K, Lu J, Ma F, Keinan A, Gu Z (2014) Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. Proc Natl Acad Sci U S A 111(29):10654–10659
38. Guo Y, Li CI, Sheng Q, Winther JF, Cai Q, Boice JD, Shyr Y (2013) Very low-level heteroplasmy mtDNA variations are inherited in humans. J Genet Genomics 40(12):607–615
39. Yao YG, Kajigaya S, Feng X, Samsel L, McCoy JP Jr, Torelli G, Young NS (2013) Accumulation of mtDNA variations in human single CD34+ cells from maternally related individuals: effects of aging and family genetic background. Stem Cell Res 10(3):361–370
40. Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 60(12):1187–1192
41. Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, Kalidindi S, Picchioni M et al (2011) Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet 20(24):4786–4796
42. Mosquera-Miguel A, Torrell H, Abasolo N, Arrojo M, Paz E, Ramos-Rios R, Agra S, Paramo M et al (2012) No evidence that major mtDNA European haplogroups confer risk to schizophrenia. Am J Med Genet B Neuropsychiatr Genet 159B(4):414–421
43. Detjen AK, Tinschert S, Kaufmann D, Algermissen B, Nurnberg P, Schuelke M (2007) Analysis of mitochondrial DNA in discordant monozygotic twins with neurofibromatosis type 1. Twin Res Hum Genet 10(3):486–495
44. Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 102(30):10604–10609
45. Grunewald A, Rygiel KA, Hepplewhite PD, Morris CM, Picard M, Turnbull DM (2016) Mitochondrial DNA depletion in respiratory chain-deficient parkinson disease neurons. Ann Neurol 79(3):366–378
46. Li M, Schroder R, Ni S, Madea B, Stoneking M (2015) Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proc Natl Acad Sci U S A 112(8):2491–2496
47. Ross JM, Stewart JB, Hagstrom E, Brene S, Mourier A, Coppotelli G, Freyer C, Lagouge M et al (2013) Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. Nature 501(7467):412–415
48. Sondheimer N, Glatz CE, Tirone JE, Deardorff MA, Krieger AM, Hakonarson H (2011) Neutral mitochondrial heteroplasmy and the influence of aging. Hum Mol Genet 20(8):1653–1659
49. Falkenberg M, Larsson NG, Gustafsson CM (2007) DNA replication and transcription in mammalian mitochondria. Annu Rev Biochem 76:679–699
50. Van Blerkom J, Davis P, Alexander S (2000) Differential mitochondrial distribution in human pronuclear embryos leads to disproportionate inheritance between blastomeres: relationship to

microtubular organization, ATP content and competence. Hum Reprod 15(12):2621–2633
51. Kustova ME, Sokolova VA, Bass MG, Zakharova FM, Sorokin AV, Vasil'ev VB (2008) Distribution of foreign mitochondrial DNA during the first splittings of the transmitochondrial mouse embryos. Tsitologiia 50(11):983–987
52. Lee HS, Ma H, Juanes RC, Tachibana M, Sparman M, Woodward J, Ramsey C, Xu J et al (2012) Rapid mitochondrial DNA segregation in primate preimplantation embryos precedes somatic and germline bottleneck. Cell Rep 1(5):506–515
53. Wang Z, Zhu R, Zhang S, Bian Y, Lu D, Li C (2015) Differentiating between monozygotic twins through next generation mtGenome sequencing. Anal Biochem 490:1–6
54. Yao YG, Macauley V, Kivisild T, Zhang YP, Bandelt HJ (2003) To trust or not to trust an idiosyncratic mitochondrial data set. Am J Hum Genet 72(5):1341–1346
55. Yao YG, Salas A, Logan I, Bandelt HJ (2009) mtDNA data mining in GenBank needs surveying. Am J Hum Genet 85(6):929–933
56. Yao YG, Bandelt HJ, Young NS (2007) External contamination in single cell mtDNA analysis. PLoS One 2(8):e681
57. Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303(5655):223–226
58. Pereira L, Soares P, Radivojac P, Li B, Samuels DC (2011) Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. Am J Hum Genet 88(4):433–439

Supplementary materials

Supplementary Table S1. Demographical information for members of 8 families with monozygotic twins

Sample [a]	Sex [b]	Age	Kinship	Onset age (year old)	Duration of psychosis (years)
B5-F	M	45	father	-	-
B5-M	F	47	mother	-	-
B5-T1	M	24	son	-	-
B5-T2	M	24	son	23	1
B6-F	M	52	father	-	-
B6-M	F	50	mother	-	-
B6-T1	F	25	daughter	-	-
B6-T2	F	25	daughter	21	4
B8-F	M	44	father	-	-
B8-M	F	42	mother	-	-
B8-T1	M	22	son	-	-
B8-T2	M	22	son	12	10
B9-F	M	62	father	-	-
B9-M	F	60	mother	-	-
B9-T1	F	26	daughter	-	-
B9-T2	F	26	daughter	25	1
B10-F	M	50	father	-	-
B10-M	F	50	mother	-	-
B10-T1	F	24	daughter	-	-
B10-T2	F	24	daughter	21	3
B11-F	M	76	father	-	-
B11-M	F	68	mother	-	-
B11-T1	M	40	son	-	-
B11-T2	M	40	son	32	8
B12-F	M	46	father	-	-
B12-M	F	46	mother	-	-
B12-T1	M	19	son	-	-
B12-T2	M	19	son	17.5	1.5
LI-F	M	49	father	-	-
LI-M	F	46	mother	-	-
LI-T1	M	26	son	-	-
LI-T2	M	26	son	24	2

[a] F: father; M: mother; T1: offspring 1; T2: offspring 2, and this individual has schizophrenia.
[b] M-male; F-female.

Supplementary Table S2. Sequence variants in mtDNA complete sequences of 32 individuals from 8 families

Sample [a]	Haplogroup	mtDNA variation [b]
B5-F	F1a1d	73, 249d, 263, 315+C, 523-524d, 750, 1438, 2706, 3970, 4086, 4722R, 4769, 6392, 6962, 7028, 8860, 8994, 9053, 9548, 9844, 10310, 10609, 11380, 11719, 12406, 12882, 13759, 13928C, 14766, 14797C, 15013, 15326, 16051, 16129, 16162, 16172, 16304, 16399, 16519
B5-M	D4j8	73, 263, 315+C, 489, 750, 1438, 2706, 3010, 4769, 4883, 5178A, 7028, 8414, 8701, 8860, 9540, 10398, 10400, 10873, 11061, 11696, 11719, 12705, 14668, 14766, 14783, 15043, 15301, 15326, 16174, 16223, 16362
B5-T1	D4j8	73, 263, 315+C, 489, 750, 1438, 2706, 3010, 4769, 4883, 5178A, 7028, 8414, 8701, 8860, 9540, 10398, 10400, 10873, 11061, 11696, 11719, 12705, 14668, 14766, 14783, 15043, 15301, 15326, 16174, 16223, 16362
B5-T2	D4j8	73, 263, 315+C, 489, 750, 1438, 2706, 3010, 4769, 4883, 5178A, 7028, 8414, 8701, 8860, 9540, 10398, 10400, 10873, 11061, 11696, 11719, 12705, 14668, 14766, 14783, 15043, 15301, 15326, 16174, 16223, 16362
B6-F	A17	73, 152, 235, 263, 315+C, 523-524d, 663, 750, 1438, 1736, 2706, 4113, 4248, 4769, 4824, 5186C, 5514, 7028, 8794, 8860, 9126, 11719, 11935, 12705, 14766, 15217, 15326, 16223, 16290, 16319, 16362
B6-M	D5a	73, 146, 150, 263, 315+C, 489, 523-524d, 750, 752, 1107, 1438, 2581, 2706, 3150, 3903, 4769, 4883, 5178A, 5301, 7028, 8701, 8860, 9180, 9540, 10397, 10398, 10400, 10873, 11719, 11944, 12026, 12414, 12705, 14007, 14766, 14783, 15043, 15301, 15326, 16164R, 16182C, 16183C, 16189, 16223, 16243, 16266, 16362
B6-T1	D5a	73, 146, 150, 263, 315+C, 489, 523-524d, 750, 752, 1107, 1438, 2581, 2706, 3150, 3903, 4769, 4883, 5178A, 5301, 7028, 8701, 8860, 9180, 9540, 10397, 10398, 10400, 10873, 11719, 11944, 12026, 12414, 12705, 14007, 14766, 14783, 15043, 15301, 15326, 16164R, 16182C, 16183C, 16189, 16223, 16243, 16266, 16362
B6-T2	D5a	73, 146, 150, 263, 315+C, 489, 523-524d, 750, 752, 1107, 1438, 2581, 2706, 3150, 3903, 4769, 4883, 5178A, 5301, 7028, 8701, 8860, 9180, 9540, 10397, 10398, 10400, 10873, 11719, 11944, 12026, 12414, 12705, 14007, 14766, 14783, 15043, 15301, 15326, 16164R, 16182C, 16183C, 16189, 16223, 16243, 16266, 16362
B8-F	A	73, 151, 152, 235, 263, 315+C, 523-524d, 663, 735, 750, 1438, 1736, 2706, 4248, 4769, 4824, 7028, 8794, 8860, 11719, 11821C, 12705, 14766, 15326, 16223, 16290, 16319, 16362

Sample [a]	Haplogroup	mtDNA variation [b]
B8-M	M7c1b2b	73, 146, 152, 199, 263, 315+C, 489, 523-524d, 750, 1438, 2706, 4071, 4491, 4769, 4850, 5442, 6455, 7028, 7337, 8701, 8860, 8886, 8986, 9540, 9824, 9957, 10398, 10400, 10861, 10873, 11665, 11719, 12091, 12561, 12705, 13590, 14766, 14783, 15043, 15301, 15326, 16223, 16519
B8-T1	M7c1b2b	73, 146, 152, 199, 263, 315+C, 489, 523-524d, 750, 1438, 2706, 4071, 4491, 4769, 4850, 5442, 6455, 7028, 7337, 8701, 8860, 8886, 8986, 9540, 9824, 9957, 10398, 10400, 10861, 10873, 11665, 11719, 12091, 12561, 12705, 13590, 14766, 14783, 15043, 15301, 15326, 16223, 16519
B8-T2	M7c1b2b	73, 146, 152, 199, 263, 315+C, 489, 523-524d, 750, 1438, 2706, 4071, 4491, 4769, 4850, 5442, 6455, 7028, 7337, 8701, 8860, 8886, 8986, 9540, 9824, 9957, 10398, 10400, 10861, 10873, 11665, 11719, 12091, 12561, 12705, 13590, 14766, 14783, 15043, 15301, 15326, 16223, 16519
B9-F	F1d	73, 146, 249d, 263, 315+C, 523-524d, 750, 1438, 1734, 2706, 3970, 4769, 5628, 6392, 6962, 7028, 7738, 8860, 10310, 10609, 11719, 12406, 12882, 13928C, 14766, 15326, 15402, 16189, 16304, 16309, 16519
B9-M	R11a	73, 185, 189, 263, 315+C, 709, 750, 990, 1438, 2351, 2706, 3140, 3316R, 4769, 7028, 8277, 8278+C, 8860, 9116, 10031, 10398, 10768, 10978, 11061, 11719, 12950, 13681, 14766, 15326, 15511, 16182C, 16183C, 16189, 16311, 16519
B9-T1	R11a	73, 185, 189, 263, 315+C, 709, 750, 990, 1438, 2351, 2706, 3140, 4769, 7028, 8277, 8278+C, 8860, 9116, 10031, 10398, 10768, 10978, 11061, 11719, 12950, 13681, 14766, 15326, 15511, 16182C, 16183C, 16189, 16311, 16519
B9-T2	R11a	73, 185, 189, 263, 315+C, 709, 750, 990, 1438, 2351, 2706, 3140, 4769, 7028, 8277, 8278+C, 8860, 9116, 10031, 10398, 10768, 10978, 11061, 11719, 12950, 13681, 14766, 15326, 15511, 16182C, 16183C, 16189, 16311, 16519
B10-F	D4q	73, 200, 263, 315+C, 489, 750, 1438, 2706, 3010, 3505, 4769, 4883, 5033, 5178A, 6297, 7028, 8414, 8701, 8860, 9540, 10398, 10400, 10873, 11719, 12705, 13956, 14668, 14766, 14783, 15043, 15301, 15326, 16223, 16256, 16311, 16362, 16519
B10-M	M74a	63, 64, 66, 73, 215, 263, 315+C, 489, 523-524d, 750, 1438, 2706, 3203, 4769, 5054, 6185, 6575, 7028, 8251, 8701, 8860, 9540, 10268, 10398, 10400, 10873, 11719, 12705, 12850, 14311, 14766, 14783, 15043, 15301, 15326, 16093, 16223, 16311, 16362, 16381
B10-T1	M74a	63, 64, 66, 73, 215, 263, 315+C, 489, 523-524d, 750, 1438, 2706, 3203, 4769, 5054, 6185, 6575, 7028, 8251, 8701, 8860, 9540, 10268, 10398, 10400, 10873, 11719, 12705, 12850, 14311, 14766, 14783, 15043, 15301, 15326, 16093, 16223, 16311, 16362, 16381

Sample [a]	Haplogroup	mtDNA variation [b]
B10-T2	M74a	63, 64, 66, 73, 215, 263, 315+C, 489, 523-524d, 750, 1438, 2706, 3203, 4769, 5054, 6185, 6575, 7028, 8251, 8701, 8860, 9540, 10268, 10398, 10400, 10873, 11719, 12705, 12850, 14311, 14766, 14783, 15043, 15301, 15326, 16093, 16223, 16311, 16362, 16381
B11-F	A13	73, 152, 200, 235, 263, 315+C, 374, 523-524d, 663, 750, 1438, 1736, 2706, 4248, 4769, 4824, 7028, 8794, 8860, 11719, 12477, 12705, 13602, 14766, 15326, 16129, 16223, 16290, 16294, 16319, 16362
B11-M	D4b2b2c	73, 194, 263, 315+C, 489, 523-524d, 750, 1382C, 1438, 2706, 3010, 3744, 4769, 4883, 5178A, 6566, 7028, 8020, 8414, 8701, 8860, 8964, 9296, 9540, 9824A, 10398, 10400, 10873, 11719, 12358, 12705, 13707, 14668, 14766, 14783, 15043, 15301, 15326, 16093, 16183C, 16189, 16223, 16325, 16356, 16362, 16519
B11-T1	D4b2b2c	73, 194, 263, 315+C, 489, 523-524d, 750, 1382C, 1438, 2706, 3010, 3744, 4769, 4883, 5178A, 6566, 7028, 8020, 8414, 8701, 8860, 8964, 9296, 9540, 9824A, 10398, 10400, 10873, 11719, 12358, 12705, 13707, 14668, 14766, 14783, 15043, 15301, 15326, 16093Y, 16183C, 16189, 16223, 16325, 16356, 16362, 16519
B11-T2	D4b2b2c	73, 194, 263, 315+C, 489, 523-524d, 750, 1382C, 1438, 2706, 3010, 3744, 4769, 4883, 5178A, 6566, 7028, 8020, 8414, 8701, 8860, 8964, 9296, 9540, 9824A, 10398, 10400, 10873, 11719, 12358, 12705, 13707, 14668, 14766, 14783, 15043, 15301, 15326, 16093Y, 16183C, 16189, 16223, 16325, 16356, 16362, 16519
B12-F	G1a1	73, 150, 263, 315+C, 489, 709, 750, 1438, 2706, 3579, 4769, 4833, 5108, 7028, 7867, 8200, 8701, 8860, 9540, 10398, 10400, 10873, 11719, 12705, 12867, 14569, 14766, 14783, 15043, 15301, 15323, 15326, 15497, 15860, 15968, 16086, 16223, 16325, 16519
B12-M	M10a1	73, 263, 315+C, 489, 573+C, 709, 750, 1438, 2706, 3172+C, 4140, 4769, 7028, 7250, 8701, 8793, 8856, 8860, 9540, 10007, 10398, 10400, 10646, 10873, 11719, 12549, 12705, 13152, 13708, 14311, 14502, 14766, 14783, 15040, 15043, 15071, 15218, 15301, 15326, 16140, 16223, 16271, 16311, 16519
B12-T1	M10a1	73, 263, 315+C, 489, 573+C, 709, 750, 1438, 2706, 3172+C, 4140, 4769, 7028, 7250, 8701, 8793, 8856, 8860, 9540, 10007, 10398, 10400, 10646, 10873, 11719, 12549, 12705, 13152, 13708, 14311, 14502, 14766, 14783, 15040, 15043, 15071, 15218, 15301, 15326, 16140, 16223, 16271, 16311, 16519
B12-T2	M10a1	73, 263, 315+C, 489, 573+C, 709, 750, 1438, 2706, 3172+C, 4140, 4769, 7028, 7250, 8701, 8793, 8856, 8860, 9540, 10007, 10398, 10400, 10646, 10873, 11719, 12549, 12705, 13152, 13708, 14311, 14502, 14766, 14783, 15040, 15043, 15071, 15218, 15301, 15326, 16140, 16223, 16271, 16311, 16519
LI-F	B4b1a	73, 235, 263, 315+C, 499, 523-524d, 750, 827, 1438, 2706, 4655, 4769, 4820, 5063, 6023, 6413, 7028, 8281-8289d, 8860, 11254, 11719, 12397, 13590, 13779, 14766, 15326, 15535, 15688, 15758, 16136, 16189, 16217, 16260, 16287, 16325, 16399, 16519

Sample [a]	Haplogroup	mtDNA variation [b]
LI-M	F1b1	73, 249d, 263, 315+C, 523-524d, 750, 1438, 2706, 3970, 4732, 4769, 5147R, 6392, 6710, 6962, 7028, 8860, 10310, 10609, 10976, 11719, 12406, 12633, 12882, 13928C, 14476, 14766, 15326, 16189, 16232A, 16249, 16304, 16311, 16427Y, 16519
LI-T1	F1b1	73, 249d, 263, 315+C, 523-524d, 750, 1438, 2706, 3970, 4732, 4769, 5147, 6392, 6710, 6962, 7028, 8860, 10310, 10609, 10976, 11719, 12406, 12633, 12882, 13928C, 14476, 14766, 15326, 16189, 16232A, 16249, 16304, 16311, 16427Y, 16519
LI-T2	F1b1	73, 249d, 263, 315+C, 523-524d, 750, 1438, 2706, 3970, 4732, 4769, 5147, 6392, 6710, 6962, 7028, 8860, 10310, 10609, 10976, 11719, 12406, 12633, 12882, 13928C, 14476, 14766, 15326, 16189, 16232A, 16249, 16304, 16311, 16427Y, 16519

[a] F: father; M: mother; T1: offspring 1; T2: offspring 2, and this individual has schizophrenia.
[b] Suffixes A and C mean transversions. Suffixes Y (heteroplasmy for both “C” and “T”) and R (heteroplasmy for both “A” and “G”) mean variants had a heteroplasmic level over 10% or with at least 20 high-quality reads.

Supplementary Table S3. The distribution of mtDNA variants across the mitochondrial genome

Variants [a]	NC[b]	rRNA	tRNA	ND1	ND2	ND3	ND4	ND4L	ND5	ND6	COI	COII	COIII	ATP[c]	CYB
Homoplasmy	154	73	4	14	39	21	39	4	41	7	31	5	14	39	70
ns				1	12	9	4	3	13	1	0	0	2	32	39
s				13	27	12	35	1	28	6	31	5	12	7	31
Heteroplasmy	8	0	0	1	3	1	0	0	0	1	1	0	0	1	0
ns				1	2	0	0	0	0	0	0	0	0	1	0
s				0	1	1	0	0	0	1	1	0	0	0	0
Chinese [d]	1410	488	69	127	258	202	302	30	275	84	254	86	147	306	599
ns				30	48	92	18	5	80	21	19	24	11	256	281
s				97	210	110	284	25	195	63	235	62	136	56	318

[a]Non-synonymous and synonymous variants are marked by “ns” and “s”, respectively.
[b] NC: variants in non-coding region.
[c] Including variants in the *MT-ATP6* and *MT-ATP8* genes.
[d] Variants in 118 Chinese complete mtDNA sequences [1-3].

Supplementary Table S4. Variants and potential errors in the previously reported mtDNA sequences of 10 pairs of identical twins [4].

Sample	Variants [a]	Haplogroup	Missing variants [b]
MZ-1	73, 150, 152, 185,263, 456,489, 681,750, 1107,1438,2706, 4769, 4883, 5178A, 5301, 6253, 7028,8860, 9180,9540, 10397, 10398, 10400, 10873, 11719, 12705, 14766, 14783, 15043,*15301*, 15440, 15470, 16148,*16183C*, *16189*, 16223, 16362, 16519	D5b1c1	1048, 5153, 5899+C, 8701, 12666C, 15326, 15724
MZ-2	73,150, 263, 489, 750, 752, 1107, 1812, 2706, 3918, 4769, 4883, 5178A,5301,7028,8701, 8860, 9540, *10397*, 10398, *10400*, 10873, 11719,11944, 12026, *12705*, 14766,14783, 15043,*15301*, 16129, 16172, 16223, 16362	D5a2	9180, 15326, 16189
MZ-3	73, 146,150, *240*, 263, 709,750, 769, 1119, 1438,2706, 3497, 3571, 4769, 7028, 8603, 8772, 8860, 9545, 11719, 12882, 14766,*15346*, 16140, 16217, 16265, 16274, 16298, 16311, 16519	B4c1b2a	8281-8289d, 15326, 16189, 16335,
MZ-4	73,*207*, 263, 489, 750,1438, 1503, 2706, 3552A, 4062, 4715, 4769, 5821, 6338, 7028, 7196A, 7853, 8584, 8701, 8860, 9540, 9545, 10398, 10400, 10873, 11719, 11914, 12705, 13263, 14318, 14766, 14783, 15043,*15301*, 15487T, *16189*, 16223, 16298, 16327, 16519	C7a	249d, 15326
MZ-5	73, 152, 204, 263, 489, 709, 750, 1438, 1888, 2706, 3732, 4769, 4833, 5108, 7028, 7621, 8701, 8854, 8860, 9540, 10283, 10398, 10400, 10873, 11719, 12705, 14180, 14569, 14766, 14783, 15043,*15301*, 15746, 16223, 16274, 16362	G3a2	143, 15326
MZ-6	73, 153,263, 489,750, 1438, 2706, 3394, 4491, 4769, 4944, 6164, 6366, 7028, 7674, 8701, 8860, 9540, 10398, 10400,10873, 11719, 12705, 14308, 14766, 14783, 15043,*15301*, *16129*, 16223, 16234,16271, 16344, 16362	M9a4a	15326
MZ-7	70, 73, 263, 316C, 489, 709,750, 1438, *2302*, 2706,4140, 4769, 5426, 7028, 7250, 7888, 9540, 10398, 10400,10646, 10724, 10873, 11647, 11719, 11864, 12549,12705, 13152, 14502,14766, 14783, 15040, 15043, 15071, 15218, 16066,16223, *16289*, 16311, 16519	M10a1b	573+C, 3172+C, 8701, 8793, 8856, 8860, 15301, 15326,
MZ-8	73, 152,263, 456, 489, 750, 1438, 2706, 3010,3206, 4769, 4883, 5178A, 5582, *6116*, 7028, 8414, 8473, 8701, 8860, 9540, 10398, 10400, 10873, 11719, 12705, 14668,14766, 14783, 14979, 15043, *15301*, 15596, 16129, 16290, 16223, 16362	D4a	15326
MZ-9	73, 146, 150, 152, 263, 489, 750, 1438, 2706, 4371, 4769, 7028, 8701, 8860, 9182C, 9540, 10398, 10400,10873, 11719, 11984, 12279,12705, 14020, 14560, 14766, 14783, 15043, *15301*, 16129, 16223, 16274, 16354, 16398	M23'75	15326

MZ-10	64, 65A, 73, 195, 237, 263, 489, 501, 750, 1187, 1438, 2706, 3010, 4769, 4883, 5178A, 7028, 8414, 8701, 8860, *9540*, *10398*, 10400, 10873,*11719*, 12705, 13962, 14668,14766, 14783, 15043,*15301*, 15469, 15884C, 16223, 16261, 16294, 16362	D4i	15326

[a] Haplogroup-defining variants were marked in blue. The reported heteroplasmic variants were indicated in italic. Suffixes "A", "C" and "T" referred to transversions.

[b] Missing variants were predicted by a phylogenetic approach as described in our previous studies [5,6].

Supplementary Reference

1. Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003) Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73 (3):671-676.
2. Kong QP, Sun C, Wang HW, Zhao M, Wang WZ, Zhong L, Hao XD, Pan H, Wang SY, Cheng YT, Zhu CL, Wu SF, Liu LN, Jin JQ, Yao YG, Zhang YP (2011) Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. Mol Biol Evol 28 (1):513-522.
3. Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang YP (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet 15 (13):2076-2086.
4. Wang Z, Zhu R, Zhang S, Bian Y, Lu D, Li C (2015) Differentiating between monozygotic twins through next generation mtGenome sequencing. Anal Biochem 490:1-6.
5. Yao YG, Macauley V, Kivisild T, Zhang YP, Bandelt HJ (2003) To trust or not to trust an idiosyncratic mitochondrial data set. Am J Hum Genet 72 (5):1341-1346.
6. Yao YG, Salas A, Logan I, Bandelt HJ (2009) mtDNA data mining in GenBank needs surveying. Am J Hum Genet 85 (6):929-933.