The Novel Mutations of Unexplained Abortion Were Analyzed at the Nucleotide Level and Classified According to Gene Function

Ji Yao
  Second Affiliated Hospital of Shantou University Medical College

Peishan Chen
  Second Affiliated Hospital of Shantou University Medical College

Xialan Zhang
  Second Affiliated Hospital of Shantou University Medical College

Wei Liu
  Second Affiliated Hospital of Shantou University Medical College

Manjuan Xu
  Shantou University Medical College

Chuanchun Yang
  CheerLand Biological Technology Co., Ltd

Shanmei Tang
  CheerLand Biological Technology Co., Ltd

Xiaoli Cui
  CheerLand Biological Technology Co., Ltd

Chuan Chen
  CheerLand Biological Technology Co., Ltd

Jian Ou
  Center of Reproduction and Genetics, The Affiliated Suzhou Hospital of Nanjing Medical University,
Suzhou Municipal Hospital

Huiru Tang
  Peking University Shenzhen Hospital

Chengkang Xu (✉ xck202002@163.com)
  Sun Yat-sen University First Affiliated Hospital   https://orcid.org/0000-0003-1237-4141

Research

Abstract

Background Abortion is a major problem affecting women's normal reproductive life. Non-whole chromosomes account for about 50% of abortions. In around 50% of cases, the etiology remains unknown. The present analysis was aimed to screen the mutations at the nucleotide level and detected new genetic origins of human diseases.

Results 72 abortion tissues (with abnormal chromosome reason excluded) were recruited and whole-genome low coverage mate-pair sequencing (WGL-MPS) was performed on each tissue. And used 100 healthy human data as a background control to find specific relevant genes and mutations with stringent bioinformatics analysis. We merged 72 abortion samples after sequencing. 28 sequence mutations in 25 genes potentially associated with unexplained abortion were identified in abortion. These 25 genes were divided into six classifications, including fetal development, cell cycle, genital correlation, kidneys, nerve, and other know functions. The mutation frequency of these sites was greater than 20% compared with the reference sequence in merged abortion data. The frequency of mutations in the samples of the control group at these sites was less than 1%. Furthermore, we verified by 10 unexplained abortion tissues, seven of which had above mutations sites on these 25 genes.

Conclusions We identified 28 mutations in 25 genes mutation that may be associated with unexplained abortion, and some were associated with embryonic development. In conclusion, the results of the present study provide a good research direction for unexplained abortions.

Introduction

Miscarriage is a common complex disease that affects about 10–25% of clinically confirmed pregnancies [1, 2]. The increased risk of miscarriage is associated with age[1], and has been related to a range of causes; obesity, endocrine, immunological dysregulation, embryo, and oocyte aneuploidy, parental chromosomal abnormalities, oocyte aneuploidy and embryo, although causal underlying factors remain largely unknown[3].

Normal pregnancy requires a series of vascular, metabolic, immunological, and endocrine regulating processes. Disorder in these processes may lead to miscarriage. The majority (50–60%) of early pregnancy losses are caused by chromosomal aberrations, which are either de novo abnormalities from parents with normal karyotypes or inherited from the parents[4, 5]. Embryonic aneuploidy, which increases significantly with advanced maternal age, accounts for a large portion of spontaneous abortus[4]. Previous studies have suggested that abnormal embryonic karyotype may contribute to abortion[6–9]. De novo numerical abnormalities, particularly autosomal trisomy, may explain a proportion of abortion[9]. However, up to 50% of recurrent pregnancy loss (RPL) cases fail to find an etiology and are therefore referred to as unexplained recurrent pregnancy losses [10]. Polymorphic variants in different genes have been the target of investigations into genetic susceptibility to idiopathic RPL [11].

In addition to this, are there any common genetic causes for abortions? The development and application of NGS, including whole-genome or whole-exome sequencing (WES) or whole-genome sequencing, makes it possible to screen entire genomes and detect new genetic origins of human disease at the nucleotide level [12]. However, the use of NGS for finding causes of embryonic or fetal deaths and associated developmental abnormalities in pregnancy loss has been less extensive. Evica summarized the next-generation sequencing in recurrent pregnancy loss-approaches and outcomes, including specific mutations or affected genes in different studies[13]. Additional examples of whole family WES approach were provided, with compound heterozygous mutations detected in a variety of genes considered to be contributing to the lethal or abnormal phenotype in the embryo[14–16]. Some coding variants in genes potentially related to the unexplained abortion were identified by using whole-exome sequencing[17–21]. Chen et al identified 275 potential developmental genes integral to 154 CNVs detected by WGS and CMA[22]. Filges et al used family-based whole-exome sequencing to identify causal variants for a recurrent pattern of an undescribed lethal fetal congenital anomaly syndrome and identify mutations in KIF14 as a novel cause of an autosomal recessive lethal fetal ciliopathy phenotype[23].

In this study, we presented a new effective method to detect mutations with NGS and bioinformatics approaches that have potential functional effects associated with unexplained abortion. This method is also more effective in large numbers of populations. This study merged 72 abortions (with chromosome aneuploidy excluded), and used 100 normal human data as a background control to find specific relevant genes and variants. It was found that the 28 variants in 25 genes potentially related to the phenotype. The gene functions of these loci are related to embryo development, immunity, neurodevelopment, genitalia, and so on. We have reported the characteristics. We consider that some of these genes/variants may be potential future biomarkers for unexplained abortion.

Methods

Subjects

All samples in this study are Chinese people from The Affiliated Suzhou Hospital of Nanjing Medical University, Peking University Shenzhen hospital, and the second affiliated hospital of Shantou University. The sampling time is from May 2018 to October 2019. We counted the clinical and demographic data of the 72 abortion tissues and 100 control samples (Table 1).

A total of 72 abortion tissues were evaluated in this case-control study. These abortion tissues came from unexplained pregnancy losses that occurred before 24 weeks of pregnancy. The woman may have miscarriage for the first time or have had a miscarriage before. Samples of abortion tissue were excluded by 1. pregnant women diagnosed with abnormal chromosome abortion or other known causes; 2. pregnant women who had given birth normally ever.

There was usually no change in the human genome from birth to adult. So we collected peripheral blood from 100 individuals' adults with no significant medical history and no reproductive system diseases as the control group to compare genetic variants with abortion tissues. Demographic information, medical

and obstetric histories were also evaluated for both groups. All procedures performed in the study involving abortion tissue samples were by the ethical standards of The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital. The tissue or blood was frozen by liquid nitrogen.

Data extraction

After thawed, washed the aborted tissue with phosphate-buffered saline (PBS), and extracted the Genomic DNA by QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany). Genomic DNA was extracted by QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany) from blood in the same way. Quantity and purity of gDNA were assessed by Qubit® 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and NanoDrop-One (Thermo Scientific, Wilmington, DE, USA). The isolated DNA was stored in the − 80°C refrigerator immediately.

Library preparation and Illumina sequencing

1 ug of gDNA was fragmented to the size range of 200–500 bp by M220 Focused-ultrasonicator (Covaris, UK), and next end repair, tailing and adaptor ligation with VAHTS Universal DNA Library Prep Kit for illumining (VAZYME), 7 cycles were performed in the amplification. Quality controlling the generate sequencing library by Agilent 2100 bioanalyzer (Agilent) and Qubit 3.0 (Invitrogen). Massively parallel sequencing of aborted tissue libraries was performed on the BGI500 platform (BGI, China). The sequencing data was 70M-200M PE150 raw reads per sample.

Sequence data analysis

We used WGL-MPS (whole-genome low-coverage mate-pair sequencing) technology to detect the CCRs (complex chromosomal rearrangements) in maternal chromosomes[24] and further checked for aneuploidy by conventional PGT-A[25]. Therefore, 72 abortion tissue samples without abnormal chromosome structure and number were selected as the experimental group. The number of original reads for each sample was more than 70 million. To eliminate the difference in sequencing depth, we randomly selected 70 million reads for each sample and merged the raw data. This newly merged data had a total of 5.04 billion reads number.

First, we performed a filtering p whole genome rocess by SOAPnuke (V1.5.0) on the merged data. After filtering, the merged data were mapped to the reference genome sequence(hg19)using BWA (V0.7.12). We used a self-developed program to analyze the bam file and counted the number of the four bases of ATCG at each coverage site. We retained high quality (depth was greater than 4 and non-reference frequency is greater than 10%) mutations and annotated the mutation sites with genomic, population frequency, and functional annotations. In the control group, we selected 100 whole genome sequencing samples. We performed the filtering, alignment, bam file statistics, mutation filtering, and annotation steps of the above experimental group to obtain the result file of each sample (**Fig. 1a**).

Selected all the genes related to abortion on hg19, therapeutic abortion (HP:0030449), recurrent spontaneous abortion (HP:0200067), spontaneous abortion (HP:0005268), use the Exomiser(v12.1.0).

Then, found the genes and the mutation sites of these genes in the case sample. Next, retained meaningful mutation sites related to abortion (**Fig. 1b**). Afterward, selected the sites with a non-reference frequency was greater than 10% from the 72 samples. Subsequently, filtered out the variants with allele frequencies above 1% in the ExAC database or unknown population frequencies. The distribution of the population frequency in CHR is in **Fig. 2**. Variants fulfilling the following criteria were retained in exon regions: frameshift deletion, stopgain, nonframeshift insertion, nonsynonymous SNV and splicing. After that found out the corresponding mutation site in controls behind the above screening. Finally, found the non-reference mutation frequency (Non-reference mutation frequency, means to the sum of the frequency of occurrence of all bases different from the reference, regardless of the direction of mutation) of these mutation sites differed between merged sample, and controls significantly.

Results

172 samples with unexplained abortion were recruited in this study, including 72 abortion tissues and 100 controls. Their general information and data were shown in Table 1. We observed more mutations in the case merged sample merged from 72 abortion tissues. We speculated that genetic factors may play an important role in patients with unexplained abortion. As shown in Fig. 3, the non-reference mutations frequency were greater than 10% in the merged abortion tissues sample shown in the red line. In the 100 control samples, remain the proportion of samples with non-reference mutations greater than 10% was < 0.05 shown in the green line. The horizontal axis was the mutation site, with a total of 525; the vertical axis was the proportion of non-reference mutation frequencies. There were significantly different between the integrated sample and controls in the non-ref mutation frequency of 525 mutation sites (Wilcoxon rank-sum test, P-value < 0.01, Fig. 3).

There were many mutations different from the reference sequence that occurred in the merged sample, and in control samples without reproductive problems, almost no samples were mutated. The black threshold line kept only non-reference frequencies greater than 20%, leaving 81 mutations. Next, we selected 28 mutation sites that were closely related to reproductive reproduction from 81 mutation sites remaining filtered by the black threshold line in **Fig. 3**. The mutations with unknown gene functions were not considered. There were 28 loci with heterozygous mutations, shown in Supplementary Table 1.

A total of 28 sequence variants in 25 genes potentially associated with unexplained abortion were identified. According to the different annotation information of genes, these 25 genes were divided into 6 classifications, including fetal development (5 SNP), cell cycle (2 SNP), genital correlation (3 SNP), kidneys (4 SNP), nerve (4 SNP) and other known functions (10 SNP), shows in **Fig. 4a**. Next, we selected ten new abortion tissue samples for analysis. These samples were from pregnant women who had never had normal births, and abnormal chromosomal abortions were excluded. The results showed that three of these samples had no mutations at these 28 sites. For the other seven samples at these 28 sites, the mutation ratio of each classification was shown in Table 2 and **Fig. 4b**. The ratio was the number of mutated sites in the sample in the classification divided by the total number of mutated sites in this classification. It can be obtained from the radar chart (Fig. 4b) that the mutation sites of S2 and S4

samples were mainly concentrated on the sites related to fetal development, and the percentage was up to 80%. S1, S5, and S6 were the same as the previous, and the percentage was 60%. The main mutation sites of S3 related to cell cycle, which percent was 50%. S7 has no obvious tendency of variation sites.

There are four genes associated with embryonic development of the 25 genes that we found, including five mutation sites. A missense mutation was detected in *MST1* (c.469A > C:p.K157Q). It accounted for 49.35% of the merged abortion data. But no such mutation occurred in the controls. Both the RS numbers and population frequency are unknown in the variation database. *MST1* gene was associated with embryogenesis [26]. It was also found a heterozygous mutation in the *RBPJ* (c.548C > G:p.T183R) in chromosome 4. *RBPJ* related to recombination signal-binding protein for the immunoglobulin kappa j region, its replacement symbol is *RBPSUH*. *RBPSUH* and *DLL4* (605185) are both involved in Notch signaling. Krebs et al. (2004) showed that *Dll4* haploinsufficiency or *RBPSUH* knockout in mice resulted in severe vascular defects leading to embryonic lethality [27].

 TRIM71 (c.1166T > C:p.I389T) was found in chromosome 3, which related to embryo development and was known to cause embryonic lethality in animals [28, 29]. In 2005, Mitschka et al. confirmed that loss of *TRIM71* resulted in 100% embryonic lethality in mice. And *TRIM71* -/- mice showed general growth retardation of the trunk, all embryos were dead by E14. 5 at later stages [29]. There were two variations on *FOXD4L4* (c.506T > C:p.I169T), which regulate embryonal development and tissue differentiation. Mutation sites significantly different from control have also been found in these two genes [30].

Other than three genes were associated with genitals *PAK2* (c.982G > A:p.G328R) in chr3, *TEKT4* (c. 239C > A:p.A80D) in chr2 and *NASP* (c.854C > T:p.T285I) in chr1 [31, 32]. Two mutation sites on two cell cycle-related genes, namely *CDC27* (c.1702A > T:p.M568L) in chromosome 17 and *MAD2L1* (c.494A > G:p.E165G) [33, 34]. Three genes related to kidney glands *ABCD1* (c.2033G > A:p.G678D) in chrX[35], *APOL1*(c.71G > A:p.R24K; c.77C > T:p.S26F) in chr22[36], *HSD3B2*(c.1097C > A:p.T366N) in chr1 [37]. Three genes related to neurodevelopment. *NBEA*(c.1909G > A:p.V637I) in chr13 [38], *VPS37A*(c.442T > C: p.S148P; c.174A > G:p.I58M) in chr8 [39]. *NOTCH3*(c.4258A > G:p.S1420G) in chr19 linked to nerves and blood vessels[40, 41]. Besides there are ten genes related to autosomal diseases or proteins, shows in Supplementary Table 1.

Discussion

In the present study, we merged 72 abortions (with abnormal chromosome excluded), and used 100 normal human data as a control to find specific relevant genes and variation. It aimed to screen the entire genome and detect the genetic origin of human diseases at the nucleotide level. 28 rare sequence variants in 25 genes potentially associated with unexplained abortion were identified by using a strict filtering strategy.

As a result, these 25 genes are divided into six classifications, including fetal development, cell cycle, genital correlation, kidneys, nerve, and other know functions. These sites had more than 20% mutations frequency different from the reference sequence in the case. Almost no samples had these mutations in

the control. And we verified by 10 abortion tissues without chromosome problems, most of which had mutation sites on these genes.

We found that four mutation site such as *MST1*(c.469A > C:p.K157Q) et al in euchromosome was associated with an increased incidence of abortion. The functions of these genes have been reported in previous studies and verified in animal models. But these loci have never been reported. Besides, as mentioned below, we also found several other genes related to embryo development. Based on our data, we suggest that the single-gene mutation site has a very large inevitable link to abortion of unknown cause. However, we still need to conduct a larger population analysis, and the sites and genes were then validated in animal models in the follow-up.

It is noteworthy that there is a non-reference mutation on *MUC4* (c.C12119C > T;p.S4040L) that reaches 46.75%, and only 3% in the control samples. Moehle et al. (2006) found that *MUC4* was highly expressed in adult trachea and colon, and fetal lung. Expression was moderate in adult testis, prostate, mammary gland, lung, and stomach [42]. The occurrence of such a high mutation frequency in the streaming product is certainly not an accident. Next, we should continue to expand the research object for further verification.

Conclusion

we found specific relevant genes and variations in abortion and control. It was found that the variation of 28 loci may be related to abortion. The gene functions of these loci are related to Fetal development, cell cycle, neurodevelopment, genital correlation, kidneys, and so on. Although further research is needed, these results help to provide a good research direction for abortions of unknown cause, and subsequent experiments can be performed to verify the sites found.

Declarations

Author's contributions

Chengkang Xu and Huiru Tang designed and directed the whole project. Xialan Zhang, Wei Liu, and Manjuan Xu collected the abortion tissues from pregnancy loss patients and blood samples from control subjects. Peishan Chen and Chuan chen performed the experiments and collected the results. Chuanchun Yang, Shanmei Tang, Xiaoli Cui, and Ji Yao analyzed the data. Ji Yao, Peishan Chen, Xialan Zhang, and Shanmei Tang wrote the first draft of the manuscript. Chengkang Xu, Huiru Tang, and Jian Ou revised the paper. Ji Yao, Peishan Chen, Xialan Zhang, Wei Liu, Chuanchun Yang, Shanmei Tang, and Xiaoli Cui participated in the discussion of the project. Chengkang Xu made substantial contributions to conception

and design; been involved in revising the manuscript critically for important intellectual content and given final approval of the version to be published. All authors contributed to and approved the final manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in the article.

Competing interests

The authors declare that they have no conflict of interest.

Ethics approval and consent to participate

This study was permitted by the ethics committee of The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, and informed consent was gained from all participants before enrollment in the project. This study conformed to the approved guidelines.

Consent for publication

All patients in this study provided their consent for publication.

Reference

1. Nybo Andersen, A.M., et al., *Maternal age and fetal loss: population based register linkage study.* BMJ (Clinical research ed.), 2000. **320**(7251): p. 1708-1712.
2. Wang, X., et al., *Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study.* Fertility and Sterility, 2003. **79**(3): p. 577-584.
3. Larsen, E.C., et al., *New insights into mechanisms behind miscarriage.* BMC Medicine, 2013. **11**(1): p. 154.
4. Sugiura-Ogasawara, M., et al., *Abnormal embryonic karyotype is the most frequent cause of recurrent miscarriage.* Human Reproduction, 2012. **27**(8): p. 2297-2303.
5. Werner, M., et al., *Characteristics of chromosomal abnormalities diagnosed after spontaneous abortions in an infertile population.* Journal of Assisted Reproduction and Genetics, 2012. **29**(8): p. 817-820.
6. Israel, R., et al., *Frequency of abnormal karyotypes among abortuses from women with and without a history of recurrent spontaneous abortion.* Fertility and sterility, 1996. **65**(2): p. 250-253.

7. Carp, H., et al., *Karyotype of the abortus in recurrent miscarriage.* Fertility and Sterility, 2001. **75**(4): p. 678-682.

8. Stephenson, M.D., K.A. Awartani, and W.P. Robinson, *Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case–control study.* Human Reproduction, 2002. **17**(2): p. 446-451.

9. Sullivan, A.E., et al., *Recurrent Fetal Aneuploidy and Recurrent Miscarriage.* Obstetrics & Gynecology, 2004. **104**(4): p. 784-788.

10. Ford, H.B. and D.J. Schust, *Recurrent Pregnancy Loss: Etiology, Diagnosis, and Therapy.* Reviews in Obstetrics Gynecology, 2009. **2**(2): p. 76-83.

11. Pereza, N., et al., *Systematic review and meta-analysis of genetic association studies in idiopathic recurrent spontaneous abortion.* Fertility and Sterility, 2017. **107**(1): p. 150-159.e2.

12. Qiao, Y., et al., *Whole exome sequencing in recurrent early pregnancy loss.* Molecular Human Reproduction, 2016. **22**(5): p. 364-372.

13. Rajcan-Separovic, E., *Next generation sequencing in recurrent pregnancy loss-approaches and outcomes.* Eur J Med Genet, 2020. **63**(2): p. 103644.

14. Bondeson, M.L., et al., *A nonsense mutation in CEP55 defines a new locus for a Meckel-like syndrome, an autosomal recessive lethal fetal ciliopathy.* Clin Genet, 2017. **92**(5): p. 510-516.

15. Dohrn, N., et al., *ECEL1 mutation causes fetal arthrogryposis multiplex congenita.* American Journal of Medical Genetics Part A, 2015. **167**(4): p. 731-743.

16. Wilbe, M., et al., *MuSK: a new target for lethal fetal akinesia deformation sequence (FADS).* J Med Genet, 2015. **52**(3): p. 195-202.

17. Shamseldin, H.E., et al., *Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families.* Genome Biology, 2015. **16**(1): p. 116.

18. Quintero-Ronderos, P., et al., *Novel genes and mutations in patients affected by recurrent pregnancy loss.* PLoS One, 2017. **12**(10): p. e0186149.

19. Fu, M., et al., *Whole-exome sequencing analysis of products of conception identifies novel mutations associated with missed abortion.* Molecular medicine reports, 2018. **18**(2): p. 2027-2032.

20. Shamseldin, H.E., et al., *Molecular autopsy in maternal–fetal medicine.* Genetics in Medicine, 2018. **20**(4): p. 420-427.

21. Stals, K.L., et al., *Diagnosis of lethal or prenatal-onset autosomal recessive disorders by parental exome sequencing.* Prenat Diagn, 2018. **38**(1): p. 33-43.

22. Chen, Y., et al., *Characterization of chromosomal abnormalities in pregnancy losses reveals critical genes and loci for human early development.* Hum Mutat, 2017. **38**(6): p. 669-677.

23. Filges, I., et al., *Exome sequencing identifies mutations in KIF14 as a novel cause of an autosomal recessive lethal fetal ciliopathy phenotype.* Clin Genet, 2014. **86**(3): p. 220-8.

24. Dong, Z., et al., *A robust approach for blind detection of balanced chromosomal rearrangements with whole-genome low-coverage sequencing.* Hum Mutat, 2014. **35**(5): p. 625-36.

25. Practice Committees of the American Society for Reproductive, M., et al., *The use of preimplantation genetic testing for aneuploidy (PGT-A): a committee opinion.* Fertil Steril, 2018. **109**(3): p. 429-436.

26. Bezerra, J., et al., *Biological effects of targeted inactivation of hepatocyte growth factor- like protein in mice.* Journal of Clinical Investigation, 1998. **101**(5): p. 1175-1183.

27. Krebs, L.T., et al., *Haploinsufficient lethality and formation of arteriovenous malformations in Notch pathway mutants.* Genes & development, 2004. **18**(20): p. 2469-2473.

28. R. Maller Schulman, B., et al., *The let-7 microRNA target gene, Mlin41/Trim71 is required for mouse embryonic survival and neural tube closure.* Cell Cycle, 2008. **7**(24): p. 3935-3942.

29. Mitschka, S., et al., *Co-existence of intact stemness and priming of neural differentiation programs in mES cells lacking Trim71.* Scientific Reports, 2015. **5**(1): p. 11126.

30. Freyaldenhoven, B.S., C. Fried, and K. Wielckens, *FOXD4a and FOXD4b, two new winged helix transcription factors, are expressed in human leukemia cell lines.* Gene, 2002. **294**(1): p. 131-140.

31. ZHAO, Z.-s. and E. MANSER, *PAK and other Rho-associated kinases – effectors with surprisingly diverse mechanisms of regulation.* Biochemical Journal, 2005. **386**(2): p. 201-214.

32. Zheng, Z., et al., *TEKT4 Promotes Papillary Thyroid Cancer Cell Proliferation, Colony Formation, and Metastasis through Activating PI3K/Akt Pathway.* Endocrine Pathology, 2018. **29**(4): p. 310-316.

33. Jörgensen, P.M., et al., *Characterisation of the human APC1, the largest subunit of the anaphase-promoting complex.* Gene, 2001. **262**(1): p. 51-59.

34. Li, Y. and R.J.S. Benezra, *Identification of a Human Mitotic Checkpoint Gene: hsMAD2.* Science, 1996. **274**(5285): p. 246-248.

35. Dean, M., A. Rzhetsky, and R. Allikmets, *The Human ATP-Binding Cassette (ABC) Transporter Superfamily.* Journal of Lipid Research, 2001. **11**(7): p. 1156.

36. Genovese, G., et al., *Association of Trypanolytic ApoL1 Variants with Kidney Disease in African Americans.* science, 2010. **329**(5993): p. 841-845.

37. Doi, M., et al., *Salt-sensitive hypertension in circadian clock–deficient Cry-null mice involves dysregulated adrenal Hsd3b6.* Nature Medicine, 2010. **16**(1): p. 67-74.

38. Savelyeva, L., et al., *The neurobeachin gene spans the common fragile site FRA13A.* Human Genetics, 2006. **118**(5): p. 551-558.

39. Zivony-Elboum, Y., et al., *A founder mutation in Vps37A causes autosomal recessive complex hereditary spastic paraparesis.* Journal of Medical Genetics, 2012. **49**(7): p. 462-472.

40. Arboleda-Velasquez, J.F., et al., *Linking Notch signaling to ischemic stroke.* PNAS, 2008. **105**(12): p. 4856-4861.

41. Eikermann-Haerter, K., et al., *Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy syndrome mutations increase susceptibility to spreading depression.* Annals of neurology, 2011. **69**(2): p. 413-418.

42. Moehle, C., et al., *Aberrant intestinal expression and allelic variants of mucin genes associated with inflammatory bowel disease.* Journal of Molecular Medicine, 2006. **84**(12): p. 1055-1066.

Tables

Table 1.
Overview of the cohorts studied.

Clinical characteristics	Case	Control
n	72	100
Sample type	Abortion	blood
Characteristics	The mother had one or more miscarriages, and never had a normal birth.	no disease of infertility
Downsized	Yes	No
Merged	Yes	No

Table 2.
The mutation ratio of each classification at the seven samples in total of 28 mutation sites.

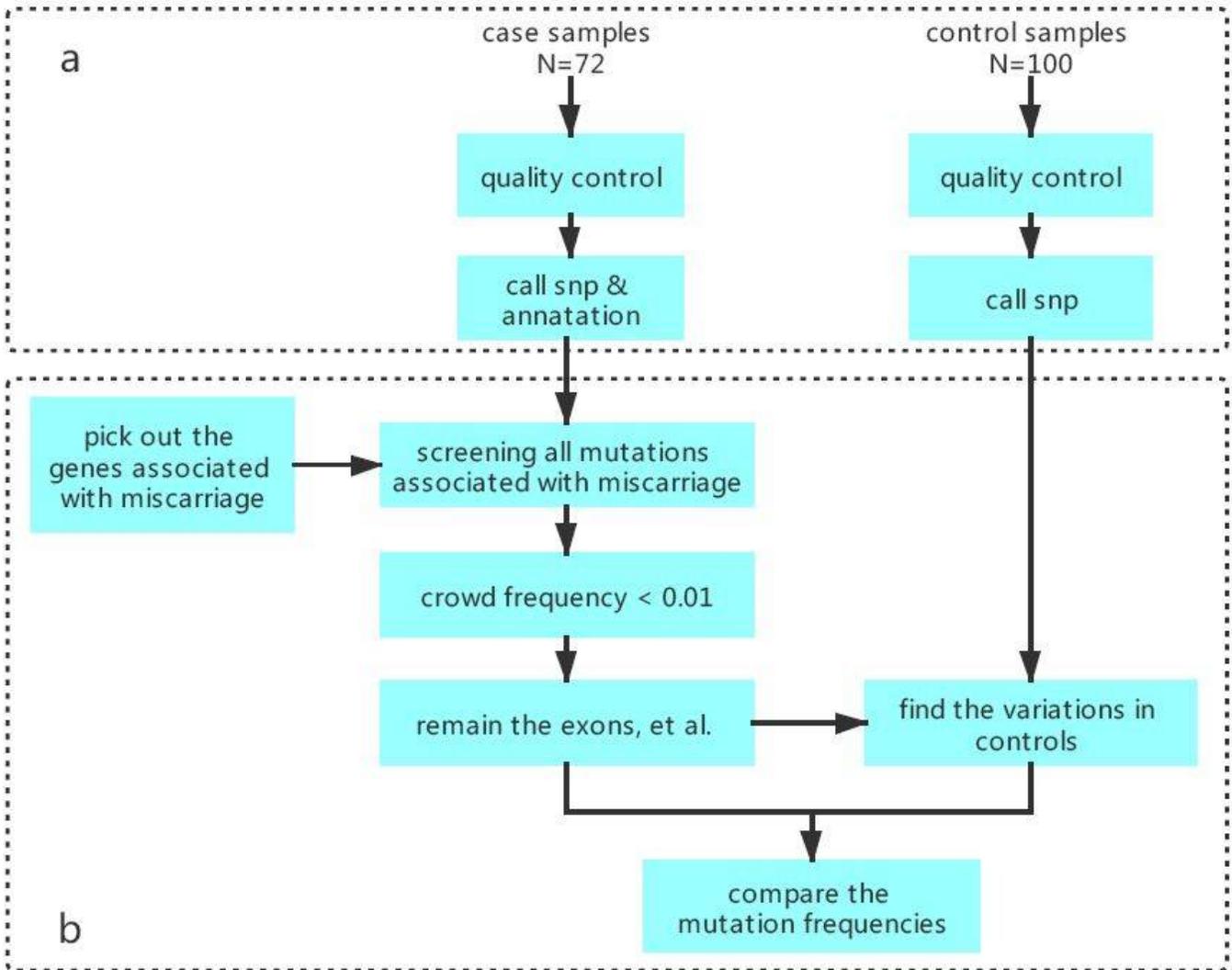| SampleID	Cell cycle(2)	Fetal development(5)	Genital correlation(3)	Kidneys(4)	Nerve(4)	Other(10)
S1	0.5(1/2)	0.8(4/5)	0(0/3)	0(0/4)	0(0/4)	0.2(2/10)
S2	0(0/2)	0.8(4/5)	0(0/3)	0(0/4)	0(0/4)	0(0/10)
S3	0.5(1/2)	0.2(1/5)	0(0/3)	0.25(1/4)	0.25(1/4)	0(0/10)
S4	0.5(1/2)	0.8(4/5)	0(0/3)	0.5(2/4)	0(0/4)	0.1(1/10)
S5	0(0/2)	0.6(3/5)	0(0/3)	0(0/4)	0(0/4)	0(0/10)
S6	0(0/2)	0.6(3/5)	0(0/3)	0(0/4)	0(0/4)	0(0/10)
S7	0(0/2)	0.4(2/5)	0.33(1/3)	0(0/4)	0.25(1/4)	0.1(1/10)

Figures

Figure 1

The main pipeline of analysis. (a) Sequence data analysis. (b) Filter and retain meaningful mutation sites related to abortion.

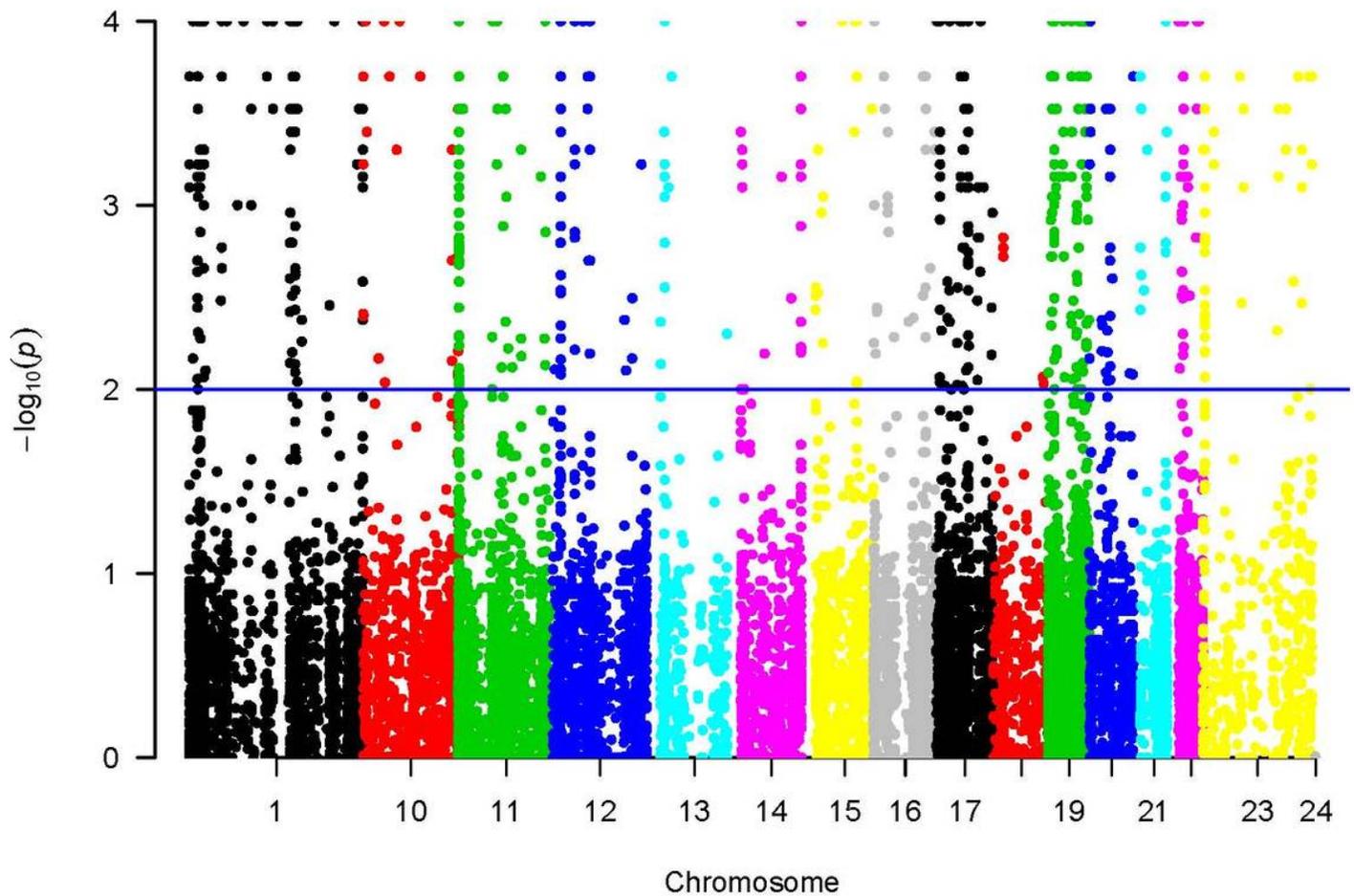## The distribution of the frequency of human presence in CHR

Figure 2

Distribution of the population frequency. Filtered out the variants with allele frequencies above 1% in the ExAC database or unknown population frequencies. The population frequency of mutation sites in chromosomes 2 to chromosomes 9 was not found in the database.
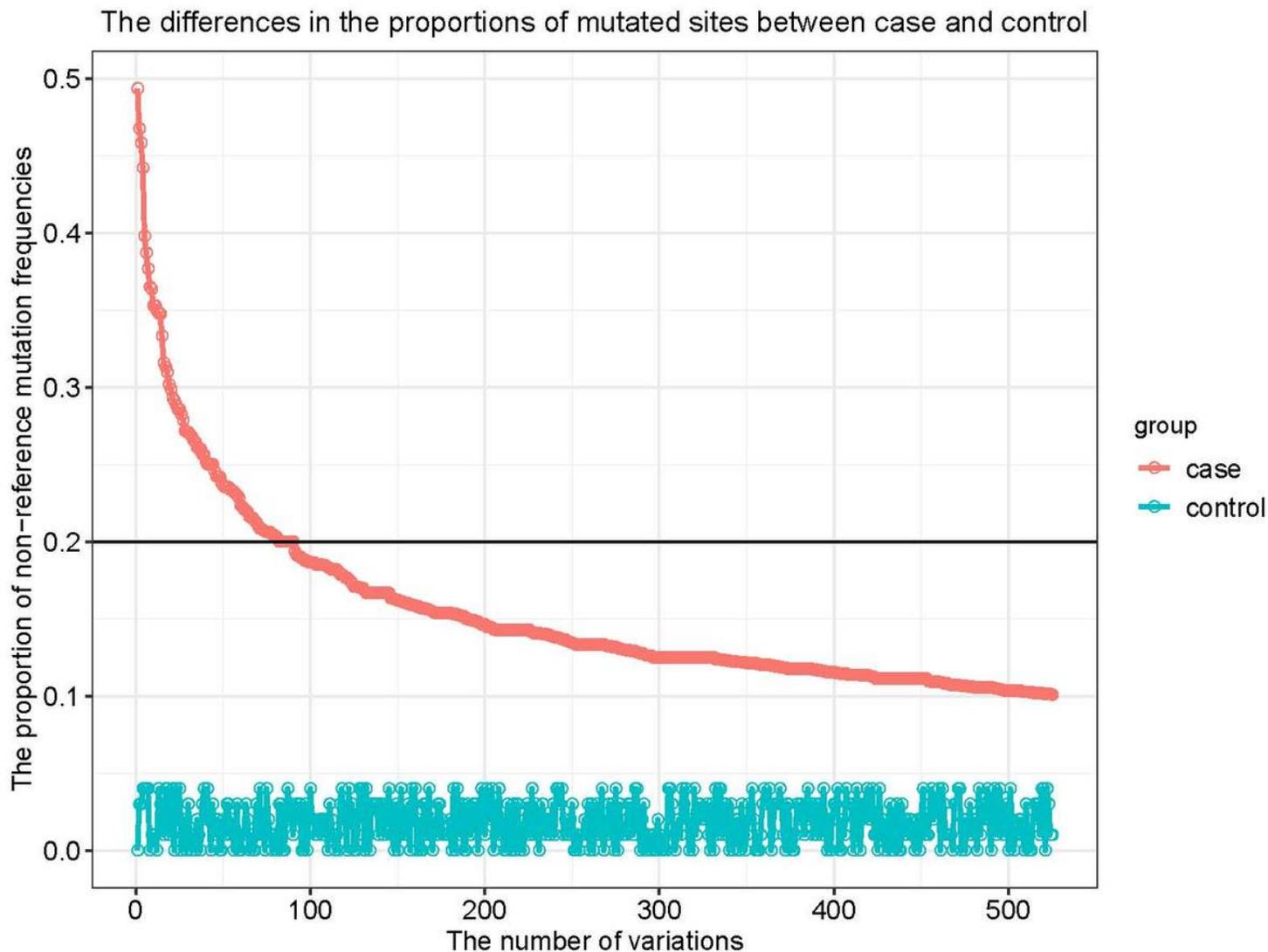
Figure 3

The comparison of non-reference mutations frequencies in case and control. There are significantly different between merged sample and controls of the non-ref mutation frequency of 525 mutation sites (Wilcoxon rank-sum test, P value = 2.2e-16). The red line: the non-reference mutation frequency is greater than 10% in the case; the blue line: the proportion of samples whose non-reference mutation frequency is greater than 10% in the control, retaining the point where the mutation frequency ratio is less than 0.05. The horizontal axis is the mutation site, with a total of 525; the vertical axis is the proportion of non-reference mutation frequencies. The black threshold line is to keep only non-reference mutation frequencies greater than 20%. After filtering, there are 81 mutation sites remaining, of which 28 sites are closely related to reproductive reproduction.
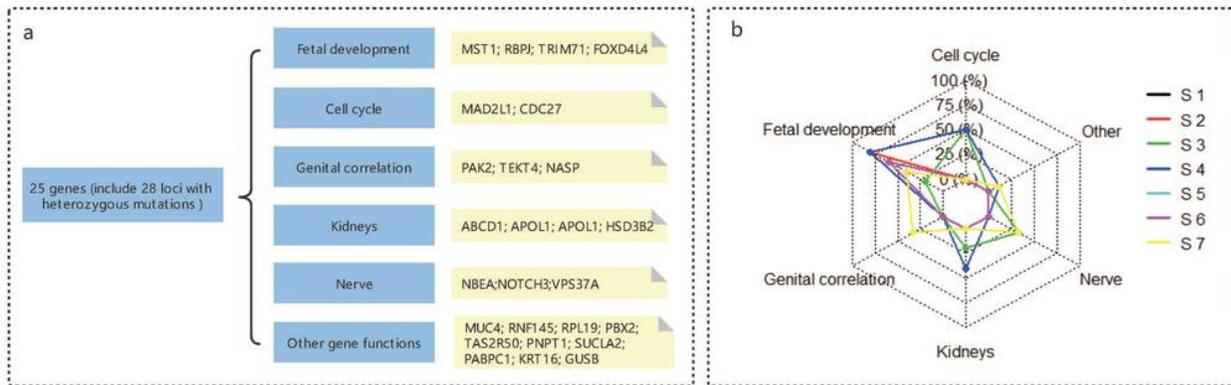
Figure 4

Functional classification of 25 genes and examples of mutations in these genes in abortion tissue samples. (a) These 28 mutation sites are located on 25 genes. According to the different annotation information of genes, these 25 genes are divided into 6 clusters, such as fetal development (5 SNP), cell cycle (2 SNP), genital correlation (3 SNP), kidneys (4 SNP), nerve (4 SNP) and other know functions (10 SNP). (b) These seven samples have mutation ratios in each classification at these 28 loci. The ratio of S1, S5 and S6 are exactly the same, and they overlap when displayed on the graph.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xls