Whole-genome sequencing analysis in families with recurrent pregnancy loss: A pilot study

Tsegaselassie Workalemahu[1], Cecile Avery[1], Sarah Lopez[1], Nathan Blue[1], Amelia Wallace[1], Aaron Quinlan[1], Hilary Coon[1], Derek Warner[1], Michael Varner[2], Ware Branch[1], Lynn Jorde[1], and Robert Silver (USA)[3]

[1]University of Utah Health
[2]University of Utah Health Sciences Center
[3]University of Utah

July 1, 2022

Abstract

Objective To conduct a feasibility whole-genome sequencing (WGS) study in families to identify genetic variants relevant to unexplained pregnancy loss. Methods We conducted a pilot WGS study of four families with recurrent pregnancy loss, including parents, healthy live births, and losses, which included an embryonic loss (<10 weeks' gestation), fetal deaths (10-20 weeks' gestation) and stillbirths ([?] 20 weeks' gestation). We used the Illumina platform for WGS and state-of-the-art protocols to identify single nucleotide variants (SNVs) following various modes of inheritance. Results We identified 87 SNVs involving 75 genes in embryonic loss (n=1), 370 SNVs involving 228 genes in fetal death (n=3), and 122 SNVs involving 122 genes in stillbirth (n=2). Of these, 22 de novo, 6 autosomal dominant and an X-linked recessive SNVs were pathogenic (probability of being loss-of-function intolerant >0.9), impacting known genes (e.g., DICER1, FBN2, FLT4, HERC1, and TAOK1) involved in embryonic/fetal development and congenital abnormalities. Further, we identified missense compound heterozygous SNVs impacting genes (e.g., VWA5B2) in two fetal death samples that were absent from live births and population controls, providing evidence for haplosufficient genes relevant to pregnancy loss. Conclusions In this pilot study, we provide evidence for de novo and inherited SNVs relevant to pregnancy loss. Our findings provide justification for conducting WGS using larger numbers of families and warrant validation by targeted sequencing to ascertain causal variants. Elucidating genes causing pregnancy loss may facilitate the development of risk stratification strategies and novel therapeutics.

Whole-genome sequencing analysis in families with recurrent pregnancy loss: A pilot study

Workalemahu T[1], Avery C[2], Lopez S[1], Blue NR[1,3], Wallace A[2], Quinlan AR[2,5], Coon H[4], Warner D[7], Varner MW[1,3], Branch DW[1,3], Jorde LB[2], Silver RM[1,3]

[1]University of Utah Health, Department of Obstetrics and Gynecology, Salt Lake City, Utah, United States

[2]University of Utah, Department of Human Genetics, Salt Lake City, Utah, United States

[3]Intermountain Healthcare, Maternal-Fetal Medicine, Salt Lake City, Utah, United States

[4]University of Utah, Department of Psychiatry, Salt Lake City, Utah, United States

[5]Department of Biomedical Informatics, University of Utah, Salt Lake City, United States

[7]DNA Sequencing Core, University of Utah, Salt Lake City, UT, USA

Corresponding author : Tsegaselassie Workalemahu, PhD MS; University of Utah, 30 North

1900 East, Suite 2B293, Salt Lake City, UT, 84132. Phone: 801-587-1607. Email: tsegaselassie.workalemahu@hsc.utah.edu

This work was supported by grant funding from the National Center for Advancing Translational Sciences of the National Institutes of Health (grant 1UL01TR002538). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Author TW is supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2 grant TR002549). Author LBJ is supported by National Institute of General Medical Sciences (grant R35GM118335).

Abstract

Objective

To conduct a feasibility whole-genome sequencing (WGS) study in families to identify genetic variants relevant to unexplained pregnancy loss.

Methods

We conducted a pilot WGS study of four families with recurrent pregnancy loss, including parents, healthy live births, and losses, which included an embryonic loss (<10 weeks' gestation), fetal deaths (10-20 weeks' gestation) and stillbirths ([?] 20 weeks' gestation). We used the Illumina platform for WGS and state-of-the-art protocols to identify single nucleotide variants (SNVs) following various modes of inheritance.

Results

We identified 87 SNVs involving 75 genes in embryonic loss (n=1), 370 SNVs involving 228 genes in fetal death (n=3), and 122 SNVs involving 122 genes in stillbirth (n=2). Of these, 22 *de novo* , 6 autosomal dominant and an X-linked recessive SNVs were pathogenic (probability of being loss-of-function intolerant >0.9), impacting known genes (e.g., *DICER1* , *FBN2* , *FLT4* , *HERC1* , and *TAOK1*) involved in embryonic/fetal development and congenital abnormalities. Further, we identified missense compound heterozygous SNVs impacting genes (e.g., *VWA5B2*) in two fetal death samples that were absent from live births and population controls, providing evidence for haplosufficient genes relevant to pregnancy loss.

Conclusions

In this pilot study, we provide evidence for *de novo* and inherited SNVs relevant to pregnancy loss. Our findings provide justification for conducting WGS using larger numbers of families and warrant validation by targeted sequencing to ascertain causal variants. Elucidating genes causing pregnancy loss may facilitate the development of risk stratification strategies and novel therapeutics.

Keywords

Pregnancy loss; whole-genome sequencing; families; genetics

Running title

Genetic variants relevant to pregnancy loss

Tweetable abstract

Whole-genome sequencing may allow identification of causes of pregnancy loss.

Introduction

Pregnancy loss is a common obstetric complication leading to significant economic and emotional burden for affected families and the health care system.[1] Women experiencing pregnancy loss are at increased risk of its recurrence, as well as other obstetric complications in subsequent pregnancies.[2-4] Recurrent pregnancy loss occurs in 1-2% of couples who are trying to conceive.[5,6] Recurrent pregnancy loss is commonly defined by the American Society of Reproductive Medicine as [?] 2 pregnancy losses,[7] and because the etiologies of

pregnancy loss vary across gestational age, more specific characterizations of losses by gestational age have been recommended.[8] Thus, pregnancy loss can be divided into three epochs: embryonic loss (<10 weeks' gestation), fetal death (10-20 weeks' gestation) and/or stillbirth ([?]20 weeks' gestation).

Though known and suspected causes of recurrent pregnancy loss include autoimmune, endocrine, uterine, and genetic abnormalities, over half are not currently explained by these mechanisms.[9-11] Among genetic abnormalities, the most clearly associated with recurrent pregnancy loss is parental balanced translocation.[12]However, this abnormality is found in fewer than 5% of couples with recurrent pregnancy loss.[13,14] Embryonic losses (<10 weeks) are often due to spontaneously-occurring aneuploidy which result from errors in maternal meiosis.[7] Such cases are identified by karyotype but often have a low recurrence risk.[15]

Many previous studies of pregnancy loss did not distinguish gestational ages of the losses and focused on sporadic losses <10 weeks.[7,16] However, systematic evaluation of unexplained embryonic loss, fetal death and stillbirth cases is critical to identify genetic abnormalities that are not detected by karyotype and may influence specific developmental epochs. Whole-genome sequencing (WGS) allows identification of previously unrecognized genetic abnormalities (e.g., copy number changes, single gene mutations, single nucleotide variants [SNVs] and/or structural variants [SVs]) that may cause unexplained pregnancy loss.[10] Few studies included DNA from parents, losses, and live births. The power of WGS technology can be further amplified by examining DNA from family pedigrees to clarify autosomal-dominant transmission of risk alleles and prove whether variants appeared in the germline of the probands as*de novo* , which will be critical for interpretation and determination of genetic causes of recurrent and sporadic pregnancy loss.

Therefore, we conducted a pilot WGS study of four families with several unexplained pregnancy losses, which included embryonic loss, fetal death and stillbirth. We applied best practice standards of WGS and analyses to identify variants using DNA from couples and their products of conception (pregnancy losses and live births). We hypothesized that pathogenic SNVs and/or SVs that may be inherited or occur *de novo*in the offspring will be relevant to the losses.

Methods

Description of study participants

Our pilot study included patients who received care at the University of Utah and had suffered at least two pregnancy losses with at least one uncomplicated live birth and in whom evaluation for accepted causes of sporadic and recurrent pregnancy loss had proven negative.[7,17] Not all cases had complete evaluations which were performed at the discretion of the providers. This study was approved by the Institutional Review Board (IRB) of the University of Utah (IRB #: 00055018; date: 3/13/2019) and all participants gave informed consent. Pregnancy losses in these patients included embryonic losses (<10 weeks), fetal deaths (10-20 weeks) and/or stillbirths ([?]20 weeks). Data regarding medical and reproductive examinations (e.g., uterine abnormalities, parental karyotypic and chromosomal microarray abnormalities, endocrinopathies including diabetes) were obtained by medical record abstraction and patient interview. In this pilot study, we included four families with available biospecimens from parents their products of conception (pregnancy losses and live births) for DNA sequencing.

Data and sample collection

Couples received saliva sample and buccal swab kits to collect cells for DNA sampling with instructions along with a brief questionnaire for demographic data collection. Research team and obstetricians examined patient clinical and demographic data and entered the data in REDcap. Couples provided spit saliva and buccal saliva from their live-born children. Placenta samples from pregnancies that resulted in fetal demise were processed by pathology within three days of delivery. One family with a known aneuploid stillbirth (Family 3) was included since they had five unexplained losses (**Table S1**). Placentas were processed using clinical protocols for placental pathology, and samples were obtained from formalin fixed and paraffin embedded (FFPE) blocks and stored at room temperature. In some cases, samples were collected for research

only. In these cases, placentas were washed and dissected from fetal villi and maternal decidual tissue to ensure sampling of fetal tissue. Tissue from these samples were divided into aliquots and stored at -80 °C.

DNA extraction and whole-genome sequencing

DNA from saliva and FFPE samples was purified and extracted using Qiagen Kit (Qiagen Systems) and Promega Kit, respectively. WGS libraries were prepared for Illumina 150bp paired-end reads sequencing using the NEBNext Ultra II DNA Library Prep Kit protocols. All libraries were sequenced on the Novaseq 6000 platform (Illumina, San Diego, CA, USA) using standard protocols. Whole-genome analysis was performed by the Utah Center for Genomic Discovery (UCGD) at the University of Utah. Germline SNVs and SVs for each sample (22 samples total) were detected following a Genome Analysis Tool Kit (GATK) best practices equivalent workflow for variant detection.[18]

Variant detection and quality control of WGS

Variant detection and quality control protocol details are provided in**Online Supplement** . Variant detection methods were tuned to detect low-frequency mutations (gnomAD allele frequency [AF]<0.001) to explore and compare germline variants in protein coding regions (potentially impactful variants) across samples.

Variant prioritization and selection of candidate genes relevant to pregnancy loss

We used Slivar[19] to prioritize and filter variants based on modes of inheritance (e.g., compound heterozygous, *de novo* , autosomal dominant and x-linked recessive). Slivar integrates population allele frequencies from the Trans-Omics for Precision Medicine (TopMED)[20] and spliceAI scores into a comprehensive variant filtering strategy to identify candidate genes.[19] Details on variant prioritization and exploratory analyses of variants relevant to recurrent pregnancy loss are provided in **Online Supplement** and **Table S2** . We evaluated SNVs across the families by modes of inheritance and highest impact on genes (in-frame deletion/insertion, missense [nonsynonymous], frameshift, stop gained, splice region).

Given the potential for identifying false positive germline SNVs due to DNA quality (e.g., prioritization of false positive autosomal dominant SNVs that differ by orders of magnitude from SNVs following other modes of inheritance[19]) and overwhelming majority of variants of unknown significance, we applied several approaches to interpret our main findings. First, we selected SNVs identified in any of the pregnancy losses but not live births within our data to interpret candidate genes relevant to recurrent pregnancy loss. Second, we interpreted rare (AF<.001) compound heterozygous SNVs, autosomal recessive variants, where both parents are heterozygous for the variant and the affected offspring receives two copies. We prioritized compound heterozygous SNVs that were identified in losses within our data but not observed in healthy controls (gnomeAD[21]) to highlight variants in haplosufficient genes relevant to embryonic/fetal lethality. Third, among SNVs that were identified in any of the pregnancy losses, we selected pathogenic SNVs (SNVs with pLI>0.90 and LOEUF<0.36) to highlight potentially damaging variants in candidate genes. Finally, we selected SNVs in genes that were involved in pregnancy loss-relevant phenotypes/diseases (e.g., embryonic/fetal death and developmental abnormalities[22-24]) to interpret candidate genes. Analyses were performed using Slivar and R, utilizing resources and support from the Center for High Performance Computing at the University of Utah.

Results

Summary of study participants

Study participants included four families with 3-6 losses and 2-4 live births (**Table 1**). Participants' maternal and paternal ages ranged between 25-34 and 34-36 years, respectively. All maternal and paternal participants self-identified as non-Hispanic White. Race/ethnicity inferred from the genotype of the participants suggested White/Hispanic, i.e., admixed Americans for the Family 2 mother and White/non-Hispanic, i.e., Western European ancestry for all other participants. Family 3 had an abnormal karyotype stillborn fetus in their second pregnancy. Samples were available from an embryonic loss at 5 weeks and 6 days (Family 3), fetal

deaths at 15 weeks and 6 days, 13-20 weeks, and 13 weeks and 6 days (Family 1), 17 weeks and 6 days, and 18 weeks and 6 days (Family 4), and stillbirths at 20 weeks (Family 1), 20-23 weeks (Family 2) and 20-40 weeks (Family 3). Samples from live births (n=10 from four families) were healthy babies born after 37 weeks.

SNVs relevant to recurrent pregnancy loss

After removing poor DNA quality samples and samples failing sex-check (five pregnancy losses samples and one family), 3,211,893 SNVs remained for further analysis. Finally, 28,485 impactful SNVs (i.e., missense, frameshift, insertion/deletion, stop gained/retained, and splice region) in all samples from the products of conception (n=16 in three families) were prioritized by Slivar. Using samples that passed quality control (n=16 in three families; **Online Supplement** and **Table S1**), we identified 87 SNVs involving 75 genes in an embryonic loss sample, 370 SNVs involving 228 genes in three fetal death samples, and 122 SNVs involving 122 genes in two stillbirth samples (**Figure 1** and **Table 2**). In Family 1, the SNVs included 11 compound heterozygous, 11 *de novo* and 92 autosomal dominant in the fetal death cases, and 1 compound heterozygous, 7 *de novo* and 35 autosomal dominant in stillbirth cases (**Figure 1**). In addition, the SNVs in Family 2 included 6 compound heterozygous, 41 *de novo*and 40 autosomal dominant in the embryonic loss case, 6 compound heterozygous, 15 *de novo* and 62 autosomal dominant in the fetal death case, and 6 compound heterozygous, 30 *de novo* and 43 autosomal dominant in the stillbirth case. Further, the SNVs in Family 4 included 12 compound heterozygous, 5 *de novo* and 155 autosomal dominant in the fetal death case. Several SNVs identified in our data impact genes that were known to be involved in the development of the embryo and fetus, and congenital abnormalities (e.g.,*DICER1,*[25] *FBN2,*[22]*FLT4,*[26] *HERC1,*[27,28] and*TAOK1*[29]).

Among the SNVs we identified, 29 SNVs are predicted as pathogenic (pLI>0.9; LOUEF<0.36), impacting 27 genes, several of which are involved in known diseases (**Table S3**). Specifically, we identified three autosomal dominant and three *de novo* pathogenic SNVs in fetal death and stillbirth from Family 1, one autosomal dominant and sixteen *de novo* pathogenic SNVs in embryonic loss, fetal death and stillbirth from Family 3, and one autosomal dominant, one X-linked recessive and three *de novo*pathogenic SNVs in fetal death from Family 4. Given the counts of*de novo* SNVs that are higher in losses than live births, we provided details, which included a table of loss-of-function *de novo* SNVs by pathogenicity and gene impact and exploratory *de novo* enrichment analysis (**Online Supplement** and **Table S4**). *De novo* SNVs were predominantly missense (nonsynonymous) followed by frameshift, splice region, *in-frame*deletion/insertion and stop gained. The observed mean *de novo*loss-of-function SNVs in pregnancy losses was higher than that of the expected (2 vs 0.2; p-value=0.01). Moreover, the SNVs were enriched in >1 protein altering genes (p-value<0.001).

Furthermore, among compound heterozygous SNVs we identified, four SNVs in three genes (*TM2D1, MUC16, VWA5B2*) were identified in fetal death from Family 1 but not in any of the live births (**Table 4**). The SNVs were not observed as homozygotes in healthy controls, highlighting their potential relevance to pregnancy loss in our samples. Finally, we conducted exploratory analyses to confirm and validate our findings, which included exploratory SNV rates comparison (**Table S2**), rare-variant association, and Sanger sequencing analyses. The methods and summary of results based on our exploratory analyses are provided in **Online Supplement** .

Discussion

Our pilot WGS study identified 87 SNVs involving 75 genes in embryonic loss (n=1), 370 SNVs involving 228 genes in fetal death (n=3), and 122 SNVs involving 122 genes in stillbirth (n=2) samples, as potentially related to pregnancy loss, across three families. The SNVs included twenty-two *de novo* , six autosomal dominant and one X-linked recessive mutation(s) that had high pathogenicity scores (pLI>0.9; LOUEF<0.36). In addition, our findings for higher counts of *de novo* SNVs in losses compared with live births, excess of genes with >1 loss-of-function *de novo* SNVs (p-value=0.01), and occurrence of multiple *de novo*events in a single gene in samples from losses, implicate *de novo*SNVs in the pathogenesis of pregnancy loss. Furthermore, several of the identified SNVs impact genes (e.g., *DICER1,*[25]*FBN2,*[22] *FLT4,*[26]*HERC1,*[27,28] and *TAOK1*[29]) that were known to be involved in the development of the embryo and fetus, and that are asso-

ciated with congenital abnormalities, highlighting the potential role of SNVs in phenotypes that may share a common pathway with recurrent pregnancy loss. Lastly, we identified missense compound heterozygous SNVs impacting genes (e.g., *VWA5B2*) in two fetal death samples that were absent from live births and population controls, providing evidence for haplosufficient genes relevant to recurrent pregnancy loss.

Previous genetic studies of pregnancy losses are limited for several reasons, including (1) lack of access to paternal DNA samples, which would make interpretations difficult without distinguishing inherited form from *de novo* variants,[30] (2) unavailability of pedigrees with products of conception from chromosomally normal losses and live-births, or (3) unavailability of high-quality data and protocols for DNA restoration and variant detection.[31-33] Loss-of-function risk variants and inherited variants in intolerant genes (i.e., genes that are critical for human development, conditions incompatible with life resulting in fetal demise)[16,23,34] were not identified, possibly due to limited sample size and focus on families with recurrent, rather than sporadic losses.

Recently, whole-exome sequencing of stillbirth in maternal-offspring duos was conducted to identify variants in intolerant genes that were impossible to ascertain with karyotype or microarray.[23] Though the study was limited in ascertaining *de novo* from inherited variants, due to unavailability of paternal DNA, genes were reported by the authors that are either lethal, known to cause disease, or increase stillbirth risk (e.g., *CCR5* , *FAT1* , *FLNB* , *INPP5K* ,*MYO1C* , *PLOD2*). Importantly, these genes overlap with findings in our data. For example, we identified a *de novo*missense chr3:58141895:C:T in *FLNB* , a gene known for its role in atelosteogenesis type 1, in a stillbirth of Family 3. In addition, we identified autosomal dominant missense chr17:1471262:C:T in*MYO1C* , a gene linked to deafness and cytoskeletal development, in a stillbirth of Family 1. Given that the SNVs were not identified in live births, the findings warrant validation to confirm potentially lethal variants in *FLNB* and *MYO1C* genes causing chromosomally normal stillbirths.

Recently, Kline *et . al .* similarly hypothesized that chromosomally normal losses are caused by rare variants in several different genes, some of which are incompatible with development to the fetal stage.[22] The authors reported damaging variants in several genes that are relevant to recurrent pregnancy loss, including *FBN2* (Fibrillin 2). Notably, we identified a *de novo* in-frame deletion involving *FBN2* in fetal death in Family 4 that was not identified in any of the live births across families. Although Kline *et . al .* identified compound heterozygous variants of *FBN2* in embryonic loss, our SV analysis in stillbirth in Family 4 (see **Online Supplement**) confirmed a *de novo*SVs (chr5:128335405) impacting *FBN2* , suggesting variants disrupting the *FBN2* gene may be incompatible with development to the fetal stage. Furthermore, *FBN2* is well described for its role in congenital contractual arachnodactyly and embryonic and fetal development, and may be a potential candidate worth investigating in larger studies.[35,36]

Given the small participant sample with WGS data in our pilot study, it is noteworthy that we identified variants in several genes (e.g.,*DICER1,*[25] *FBN2,*[22]*FLT4,*[26] *HERC1,*[27,28] and*TAOK1*[29]) that were previously identified by genetic studies of pregnancy loss. To confirm our findings, we conducted several validation and confirmatory analyses. First, we compared our data to a population of healthy controls (gnomAD) and identified rare (gnomAD AF<0.006) compound heterozygous SNVs in four genes (*TM2D1, MUC16, VWA5B2*) across two families that were not observed as homozygotes in healthy gnomeAD controls. This finding suggested that variants in haplosufficient genes may contribute to fetal demise in offspring of two healthy parent carriers. Given that our filtering approach is cantered on allele frequencies and predicted impact, and is agnostic to the phenotype of interest, the identification of gene candidates associated with congenital and developmental phenotypes is notable. Although we demonstrated some sharing of SNVs across families (e.g., compound heterozygous SNVs in four *TM2D1* , *MUC16* ,*VWA5B2* were shared across two families), losses may not have common etiologies.[22,37] As such, this finding suggests that different genes may play a role at different developmental epochs and across families.[16] Second, we explored validation of *VWA5B2* gene by Sanger sequencing and confirmed that WGS in our sample confidently called its compound heterozygous SNV (chr3:184236380:T:C). However, further interpretations from our Sanger sequencing results were hindered by the DNA extraction quality and require sequencing of additional samples with higher DNA quality.

Compound heterozygous variants have been previously implicated in pregnancy loss[38] and present a scenario in which each parent is purportedly healthy but carries variants in the same gene(s) that may be incompatible with life. As such, functional validation of compound heterozygous variants may provide a clearer picture of the genetic landscape of recurrent pregnancy loss, especially recurrent cases. *De novo* variants in highly conserved or constrained genes also may lead to pregnancy loss. However, a *de novo* mutation has a much lower recurrence rate than recessive or dominant inherited disorders.[39,40] Impactful X-linked recessive variants, for example, a missense X-linked SNV (chrX:108591181:C:A) impacting *COL4A5* (Alport syndrome 1 gene) and possibly relevant to fetal death (**Online Supplement Table S3**), may also serve as candidates for validation. Importantly, genetic diagnoses based on impactful variants following various modes of inheritance may be used to provide a prognosis based on data from other families with similar mutations.[41,42] Confirmation of genes relevant to pregnancy loss will also identify critical pathways and novel therapeutic targets for improving pregnancy outcomes.

Our study has several limitations. The higher counts of *de novo*SNVs we observed in pregnancy losses compared with live births could result from sequencing error, reflected from degradation of placenta samples due to FFPE. FFPE samples have small fragment sizes and very uneven coverage, contributing to false positive SNVs/SVs. For example, low quality libraries (high DNA degradation) from two samples may have contributed to the large number of *de novo* SNVs observed in losses in our data. To validate SNVs in our data, we conducted exploratory Sanger sequencing analysis. Results showed poor validation for *de novo* (data not shown) but confirmed several compound heterozygous calls (**Table 3**) that were not confidently called in our samples. Furthermore, we used Slivar, a method that is strictly a filtering strategy, and the utility of the output relies on high-quality input variants. Future studies utilizing freshly obtained placenta samples for WGS may address elevated sequencing error potentially contributed by FFPE.

Strengths of our study include some prospective collection of samples from losses and live births. This may improve strategies for determining the 'intolerome', conditions incompatible with life resulting in fetal demise, and potential to improve database of lethal genes and phenotypes, which are poorly represented. Although our study is underpowered to compare rates of SNVs/SVs between losses and live births, our study serves as a requisite feasibility step in exploring genes relevant to pregnancy loss. Thus, the findings from our pilot study will provide justification for conducting WGS using larger parent-offspring families with potential to identify SNVs causing pregnancy loss.

In summary, the findings reported herein provide evidence for genetic variants (including several in previously recognized genes) relevant to unexplained pregnancy loss in families. WGS of DNA from larger numbers of families (including parent-offspring DNA from affected and unaffected pregnancies) may help identify lethal genes contributing to sporadic and recurrent pregnancy loss. Elucidating pregnancy loss causing genes may lead to biomarkers useful for risk stratification, the identification of genes relevant to normal and abnormal pregnancy, and novel therapeutic targets.

Disclosure of interests

All authors report no conflict.

Contribution to authorship

RMS, TW, and DWB had a critical role in the conception and planning of the study. CA and TW had a critical role in the analysis, data management and interpretation of the analyses. SL, RMS and DWB had a critical role in patient recruitment, IRB and execution of the study. TW, CA, SL, NRB, AW, ARQ, HC, DW, CH, MWV, DWB, LBJ, and RMS, each, had a critical role in interpretation of the analyses and writing the manuscript.

Ethical approval

This study was approved by the Institutional Review Board of the University of Utah (IRB #: 00055018; date: 3/13/2019).

Acknowledgements

Sequence alignment and variant calling were performed at the Utah Center for Genetic Discovery Core facility, part of the Health Sciences Center Cores at the University of Utah. This work utilized resources and support from the Center for High Performance Computing at the University of Utah. The computational resources used were partially funded by the NIH Shared Instrumentation grant 1S10OD021644-01A1. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged.

Funding/support

This work was supported by grant funding from the National Center for Advancing Translational Sciences of the National Institutes of Health (grant 1UL01TR002538). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Author TW is supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2 grant TR002549) and the Utah Genome Project pilot grant.

Table 1. Characteristics of study participants.

	Family 1	**Family 2**	**Family 3**	**Family 4**
Maternal				
Age at first pregnancy, years	26	34	25	26
Genotype inferred race/ethnicity	Non-Hispanic White	Hispanic	Non-Hispanic White	Non-Hispanic White
Body-mass-index kg/m^2	26.6	27.1	19.1	32.3
Paternal				
Age at first pregnancy, years	34	36	35	35
Genotype inferred race/ethnicity	Non-Hispanic White	Non-Hispanic White	Non-Hispanic White	Non-Hispanic White
Pregnancy outcome				
1st Pregnancy	Male live birth (>37 wks) *	Male live birth (>37 wks) *	Female live birth (>37 wks) *	Male live birth (>37 wks) *
2nd Pregnancy	Male stillbirth (20 wks) *	Unknown sex fetal death (13 wks 6 days)	Male stillbirth (20-40 wks) *	Female live birth (>37 wks) *
3rd Pregnancy	Female embryonic loss (6 wks 6 days)	Male live birth (>37 wks) *	Unknown sex pre-embryonic loss (5 wks 6 days) *	Female live birth (>37 wks) *
4th Pregnancy	Male fetal death (15 wks 6 days) *	Unknown sex stillbirth (20-23 wks) *	Unknown sex pre-embryonic loss (6-9 wks)	Unknown sex embryonic loss (6-9 wks)
5th Pregnancy	Male live birth (>37 wks) *	Unknown sex embryonic loss (8 wks 6 days)	Female live birth (>37 wks)	Male fetal death (17 wks 6 days) *
6th Pregnancy	Male fetal death (13-20 wks) *	Unknown sex embryonic loss (9 wks 6 days)	Female live birth (>37 wks) *	Unknown sex stillbirth (20-23 wks)

	Family 1	**Family 2**	**Family 3**	**Family 4**
7th Pregnancy	Male fetal death (13 wks 6 days)	Unknown sex embryonic loss (7 wks 6 days)	Unknown sex fetal death (14 wks 6 days)	Male fetal death (18 wks 6 days) *
8th Pregnancy	Male fetal death (13 wks 6 days) *	-	Unknown sex live-birth (>37 wks)	-
Karyotype testing	Normal all pregnancies	-	Abnormal 2nd pregnancy	-
Microarray	Normal 7th pregnancy	-	Not ordered	-
Products of conception with WGS				
Embryonic loss, N	-	-	1	-
Fetal death, N	3	-	1	2
Stillbirth, N	1	1	1	-
Live-birth, N	2	2	3	3

* Obtained DNA from samples for WGS

Figure 1. Number of SNVs by modes of inheritance and products of conception.

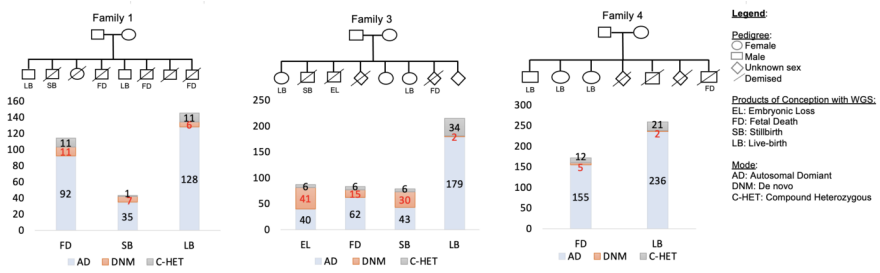

Table 2. Number of SNVs (and genes) in losses.

	Embryonic Loss	**Embryonic Loss**	**Fetal Death**	**Fetal Death**	**Stillbirth**	**Stillbirth**
	SNVs (N)	**Genes (N)**	**SNVs (N)**	**Genes (N)**	**SNVs (N)**	**Genes (N)**
Family 1	-	-	114	89	43	44
Family 3	87	75	83	75	79	68
Family 4	-	-	173	127	-	-
Total *	87	75	370	228	122	112

***** Total counts account for overlapping SNVs/genes across the samples.

Note : The crude counts also represent SNVs that may be present in live births.

Table 3 . Compound heterozygous SNVs identified in losses but not live births and gnomAD controls.

	SNV	**Highest impact**
Family 1 Male fetal death in the 6th pregnancy at 13-20 weeks' gestation	chr1:61723786:T:C	Splice region

	SNV	Highest impact
	chr1:61725096:G:A	Missense
	chr19:8939543:C:T	Missense
	chr19:8953864:G:T	Missense
Family 1 Male fetal death in the 8th pregnancy at 13 weeks and 6 days gestation	chr3:184236380:T:C	Missense

References

1. Feodor Nilsson S, Andersen PK, Strandberg-Larsen K, Nybo Andersen AM. Risk factors for miscarriage from a prevention perspective: a nationwide follow-up study. *BJOG: An International Journal of Obstetrics & Gynaecology.* 2014;121(11):1375-1385.

2. Lamont K, Scott NW, Jones GT, Bhattacharya S. Risk of recurrent stillbirth: systematic review and meta-analysis. *bmj.*2015;350:h3080.

3. Melve KK, Skjaerven R, Rasmussen S, Irgens LM. Recurrence of stillbirth in sibships: population-based cohort study. *American journal of epidemiology.* 2010;172(10):1123-1130.

4. Black M, Shetty A, Bhattacharya S. Obstetric outcomes subsequent to intrauterine death in the first pregnancy. *BJOG: An International Journal of Obstetrics & Gynaecology.* 2008;115(2):269-274.

5. Ford HB, Schust DJ. Recurrent pregnancy loss: etiology, diagnosis, and therapy. *Reviews in obstetrics and gynecology.* 2009;2(2):76.

6. Fritz R, Kohan-Ghadr H-R, Bolnick JM, et al. Noninvasive detection of trophoblast protein signatures linked to early pregnancy loss using trophoblast retrieval and isolation from the cervix (TRIC).*Fertility and sterility.* 2015;104(2):339-346. e334.

7. Medicine PCotASfR. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. *Fertility and sterility.*2012;98(5):1103-1111.

8. Silver RM, Branch DW, Goldenberg R, Iams JD, Klebanoff MA. Nomenclature for pregnancy outcomes: time for a change. *Obstetrics & Gynecology.* 2011;118(6):1402-1408.

9. Bardos J, Hercz D, Friedenthal J, Missmer SA, Williams Z. A national survey on public perceptions of miscarriage. *Obstetrics and gynecology.* 2015;125(6):1313.

10. Kasak L, Rull K, Sõber S, Laan M. Copy number variation profile in the placental and parental genomes of recurrent pregnancy loss families.*Scientific reports.* 2017;7:45327.

11. Li T, Spuijbroek MD, Tuckerman E, Anstie B, Loxley M, Laird S. Endocrinological and endometrial factors in recurrent miscarriage.*BJOG: An International Journal of Obstetrics & Gynaecology.*2000;107(12):1471-1479.

12. THARAPEL AT, THARAPEL SA, BANNERMAN RM. Recurrent pregnancy losses and parental chromosome abnormalities: a review. *BJOG: An International Journal of Obstetrics & Gynaecology.* 1985;92(9):899-914.

13. Kavalier F. Investigation of recurrent miscarriages. In. Vol 331: British Medical Journal Publishing Group; 2005:121-122.

14. Franssen MT, Korevaar JC, Leschot NJ, et al. Selective chromosome analysis in couples with two or more miscarriages: case-control study.*bmj.* 2005;331(7509):137-141.

15. Warburton D, Kline J, Stein Z, Hutzler M, Chin A, Hassold T. Does the karyotype of a spontaneous abortion predict the karyotype of a subsequent abortion? Evidence from 273 women with two karyotyped spontaneous abortions. *American journal of human genetics.*1987;41(3):465.

16. Carey AZ, Blue NR, Varner MW, et al. A systematic review to guide future efforts in the determination of genetic causes of pregnancy loss.*Frontiers in Reproductive Health.* 2021:97.

17. Christiansen OB, Elson J, Kolte AM, et al. ESHRE guideline: recurrent pregnancy loss. *Human Reproduction Open.*2018;2018(2):hoy004-hoy004.

18. Franke KR, Crowgey EL. Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. *Genomics & informatics.* 2020;18(1).

19. Pedersen BS, Brown JM, Dashnow H, et al. Effective variant filtering and expected candidate variant yield in studies of rare human disease.*NPJ Genomic Medicine.* 2021;6(1):1-8.

20. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature.*2021;590(7845):290-299.

21. Karczewski K, Francioli L, Tiao G, Cummings B, Alföldi J, Wang Q. Genome Aggregation Database, C.(2020). *The mutational constraint spectrum quantified from variation in.* 141:434-443.

22. Kline J, Vardarajan B, Abhyankar A, et al. Embryonic lethal genetic variants and chromosomally normal pregnancy loss. *Fertility and sterility.* 2021;116(5):1351-1358.

23. Stanley KE, Giordano J, Thorsten V, et al. Causal Genetic Variants in Stillbirth. *New England Journal of Medicine.* 2020.

24. Shehab O, Tester DJ, Ackerman NC, Cowchock FS, Ackerman MJ. Whole genome sequencing identifies etiology of recurrent male intrauterine fetal death. *Prenatal Diagnosis.* 2017;37(10):1040-1045.

25. Teijeiro V, Yang D, Majumdar S, et al. DICER1 is essential for self-renewal of human embryonic stem cells. *Stem cell reports.*2018;11(3):616-625.

26. Page DJ, Miossec MJ, Williams SG, et al. Whole exome sequencing reveals the major genetic contributors to nonsyndromic tetralogy of fallot. *Circulation research.* 2019;124(4):553-563.

27. Cubillos-Rojas M, Schneider T, Hadjebi O, et al. The HERC2 ubiquitin ligase is essential for embryonic development and regulates motor coordination. *Oncotarget.* 2016;7(35):56083.

28. Aggarwal S, Bhowmik AD, Ramprasad VL, Murugan S, Dalal A. A splice site mutation in HERC1 leads to syndromic intellectual disability with macrocephaly and facial dysmorphism: further delineation of the phenotypic spectrum. *American Journal of Medical Genetics Part A.*2016;170(7):1868-1873.

29. van Woerden GM, Bos M, de Konink C, et al. TAOK1 is associated with neurodevelopmental disorder and essential for neuronal maturation and cortical development. *Human mutation.* 2021;42(4):445-459.

30. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. *Nature genetics.* 2014;46(9):944-950.

31. Quintero-Ronderos P, Laissue P. Genetic variants contributing to early recurrent pregnancy loss etiology identified by sequencing approaches. *Reproductive Sciences.* 2019:1933719119831769.

32. Cochery-Nouvellon E, Chauleur C, Demattei C, et al. The A6936G polymorphism of the endothelial protein C receptor gene is associated with the risk of unexplained foetal loss in Mediterranean European couples. *Thrombosis and haemostasis.* 2009;102(10):656-667.

33. Alonso A, Soto I, Urgellés MF, Corte JR, Rodríguez MJ, Pinto CR. Acquired and inherited thrombophilia in women with unexplained fetal losses. *American journal of obstetrics and gynecology.*2002;187(5):1337-1342.

34. Gray KJ, Wilkins-Haug L. Special issue on "Feto-Maternal Genomic Medicine": a decade of incredible advances. In: Springer; 2020.

35. Putnam EA, Zhang H, Ramirez F, Milewicz DM. Fibrillin–2 (FBN2) mutations result in the Marfan–like disorder, congenital contractural arachnodactyly. *Nature genetics.* 1995;11(4):456-458.

36. Quondamatteo F, Reinhardt DP, Charbonneau NL, Pophal G, Sakai LY, Herken R. Fibrillin-1 and fibrillin-2 in human embryonic and early fetal development. *Matrix biology.* 2002;21(8):637-646.

37. Colley E, Hamilton S, Smith P, Morgan NV, Coomarasamy A, Allen S. Potential genetic causes of miscarriage in euploid pregnancies: a systematic review. *Human reproduction update.* 2019;25(4):452-472.

38. Rajcan-Separovic E. Next generation sequencing in recurrent pregnancy loss-approaches and outcomes. *European Journal of Medical Genetics.* 2020;63(2):103644.

39. Rahbari R, Wuster A, Lindsay SJ, et al. Timing, rates and spectra of human germline mutation. *Nature genetics.* 2016;48(2):126.

40. Campbell IM, Stewart JR, James RA, et al. Parent of origin, mosaicism, and recurrence risk: probabilistic modeling explains the broken symmetry of transmission genetics. *The American Journal of Human Genetics.* 2014;95(4):345-359.

41. Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of de novo mutations in health and disease.*Genome biology.* 2016;17(1):241.

42. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. *Cell.*2014;156(5):872-877.

Table S1 . Biospecimen obtained from parent-offspring samples for WGS

	Parent	**Parent**	**Products of Conception +**	**Products of Conception +**	**Pr**
	Maternal	Paternal	Embryonic loss (<10 weeks)	Fetal death (10-20 weeks)	Sti
Family 1	spit saliva	spit saliva	-	FFPE male placenta (15 weeks 6 days) FFPE male placenta (13 weeks 6 days) Frozen male placenta (13 weeks 6 days)*	FF
Family 2	spit saliva	spit saliva	-	-	FF
Family 3	spit saliva	spit saliva	FFPE male placenta (5 weeks 6 days)	FFPE female placenta (14 weeks 6 days	FF
Family 4	spit saliva	spit saliva	-	FFPE male placenta (17 weeks 6 days) FFPE male placenta (18 weeks 6 days)	-

* Sample excluded after sequencing read and genotype quality control (included sex-check, heterozygosity, Mendelian violation and call rate)

+ FFPE: Formalin-Fixed Paraffin-Embedded

Table S2 . Number (percent) of all Slivar prioritized SNVs in products of conception by pathogenicity and mode of inheritance

Mode of Inheritance	**Pregnancy loss**	**Pregnancy loss**	**Pregnancy loss**	**Live birth**	**Live birth**	**Live birth**	**Total**	p-value*
	Total	**pLI > 0.9; LOEUF < 0.35**	**pLI < 0.9; LOEUF > 0.35**	**Total**	**pLI > 0.9; LOEUF < 0.35**	**pLI < 0.9; LOEUF > 0.35**		

Mode of Inheritance	**Pregnancy loss**	**Pregnancy loss**	**Pregnancy loss**	**Live birth**	**Live birth**	**Live birth**	**Total**	p-value*
Autosomal Dominant, n (%)	296 (47.1)	56 (8.9)	240 (38.2)	332 (52.9)	68 (10.8)	264 (42.0)	628 (100.0)	0.15
De novo, n (%)	108 (92.3)	22 (18.8)	86 (73.5)	9 (7.7)	2 (1.7)	7 (6.0)	117 (100.0)	<0.001
Compound Heterozygous, n (%)	27 (42.2)	5 (7.8)	22 (34.4)	37 (57.8)	5 (7.8)	32 (50.0)	64 (100.0)	0.21
Total, n (%)	**432 (53.3)**	**84 (10.4)**	**348 (43.0)**	**378 (46.7)**	**75 (9.3)**	**303 (37.4)**	**810 (100.0)**	**0.06**

*P-values are from two-sided 1-degree of freedom Chi-squared test, comparing the proportions of SNVs between losses and live births

Note : The denominator is the total Slivar-picked autosomal dominant, *de novo* and compound heterozygous SNVs

Table S3. SNVs with high pathogenicity scores (pLI>0.9; LOUEF<0.36) identified in pregnancy losses but not live births.

Family	**Outcome**	**SNV**
Family 1	fetal death	chr3:63912887:C:A
	fetal death	chr3:63912887:C:A
	fetal death	chr19:17314331:C:A
	stillbirth	chr14:95129520:ATTTTCTCTAGTTTCTGAATC:A
	stillbirth	chr10:419184:G:A
	fetal death	
	fetal death	
	fetal death	chr9:83970202:G:A
	fetal death	chr10:17840773:C:T
	stillbirth	chr3:88139246:C:T
Family 4	fetal death	chrX:108591181:C:A
	fetal death	chr11:117458790:G:A
	fetal death	chr5:128335211:TAGCAGAGGCAGCGATACTCTCCAGGAATGTTGGTACACTGGCCGCC
	fetal death	chr8:41977214:C:T
	fetal death	chr9:33062105:CTGCCTTGCCTCGAAACAAATCTATGGTCATACCAGGAGGAAGCAATC
Family 3	Embryonic loss	chr7:140739949:GGA:G
	stillbirth	chr6:36200959:G:A
	Embryonic loss	chr17:47171999:C:T
	fetal death	chr5:180603262:G:A
	Embryonic loss	chr15:63774800:G:T
	Embryonic loss	chr15:63774801:A:C
	Embryonic loss	chr15:63774805:C:CT
	Embryonic loss	chr5:150134013:TAG:T
	Embryonic loss	chr15:43766828:TTGGAGAGCACTGC:T
	Embryonic loss	chr8:140668348:G:A

Family	Outcome	SNV
	Embryonic loss	chr9:127069303:G:A
	stillbirth	chr2:11249767:G:A
	Embryonic loss	chr12:27309142:G:A
	stillbirth	chr6:147316370:TC:T
	stillbirth	chr17:29507963:G:A
	fetal death	chr2:169994355:A:AG
	stillbirth	chr2:61227175:TCTTCTTCTTCCC:T

Table S4 . Pathogenicity and loss-of-function *de novo*SNVs in products of conception.

	Products of Conception	**Products of Conception**	**Products of Conception**	**Pro**
	Pregnancy loss * pLI > 0.9; LOEUF < 0.35	**Pregnancy loss *** pLI < 0.9; LOEUF > 0.35	**Pregnancy loss * Total**	**Liv** pLI
In-frame Deletion/Insertion	3	7	**10**	0
Missense	10	36	**46**	2
Frameshift	5	18	**23**	0
Stop gained	2	6	**8**	0
Splice region	2	19	**21**	0
Total	**22**	**86**	**108**	**2**

* The observed mean *de novo* loss-of-function SNVs in pregnancy losses was higher than that of the expected (2 vs 0.2; p-value=0.01); SNVs were enriched in >1 protein altering genes (p-value<0.001).