Simultaneous Genomic and Metagenomic Analysis for Miscarriage and Stillbirth Further Increases the Diagnostic Utility of Genome Sequencing

Hamish Scott (✉ hamish.scott@sa.gov.au)

SA Pathology <https://orcid.org/0000-0002-5813-631X>

Alicia Byrne

Centre for Cancer Biology, an alliance between UniSA and SA Pathology <https://orcid.org/0000-0002-8141-1818>

Luis Arriola

Centre for Cancer Biology, an alliance between UniSA and SA Pathology

Leila Eshraghi

Centre for Cancer Biology, an alliance between UniSA and SA Pathology

Thuong Ha

Centre for Cancer Biology, an alliance between UniSA and SA Pathology

Jinghhua Feng

ACRF South Australian Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology

Milena Babic

Centre for Cancer Biology, an alliance between UniSA and SA Pathology

Hung Nguyen

SA Pathology

Eleanor Ludington

Women's and Children's Hospital

Sarah King-Smith

Centre for Cancer Biology, an alliance between UniSA and SA Pathology

Wendy Parker

Centre for Cancer Biology, An alliance between SA Pathology and the University of South Australia

Rosalie Kenyon

Centre for Cancer Biology, An alliance between SA Pathology and the University of South Australia

Brett Ritchie

SA Pathology

Jillian Lipsett

Anatomical Pathology, SA Pathology, Women's and Children's Hospital

Nick Manton

Anatomical Pathology, SA Pathology, Women's and Children's Hospital

T Khong

Anatomical Pathology, SA Pathology, Women's and Children's Hospital <https://orcid.org/0000-0002-2404-007X>

Lynette Moore

Anatomical Pathology, SA Pathology, Women's and Children's Hospital

Peer Arts

Centre for Cancer Biology <https://orcid.org/0000-0002-6742-6239>

Christopher Barnett

South Australian Clinical Genetics Service

Andreas Schreiber

SA Pathology <https://orcid.org/0000-0002-9081-3405>

Brief Communication**Keywords:**

Posted Date: February 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1366032/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Congenital abnormalities and infections are the predominant cause of perinatal death in developed countries. While standard-of-care autopsy investigations can ascertain the event that caused the death, the aetiology cannot always be established. Genetic approaches have demonstrated utility in identifying the underlying causes of congenital anomalies and infectious diseases, but the feasibility of simultaneous metagenomic and genomic analyses has not been examined. To determine the potential of this consolidated approach, genome sequencing was performed on DNA from lung in 20 cases of fetal death. Causative microbes were correctly identified in 4/7 known infection cases and plausible causative microbes identified for 2/5 cases unresolved from standard investigations. No significant microbial load was identified in 8/8 cases due to congenital abnormality. While methods for concurrent human and microbial genome analyses still require optimisation, this dual approach should improve the diagnostic yield compared to current standard-of-care investigations, enabling more accurate counselling and appropriate care in subsequent pregnancies.

Full Text

Perinatal death, defined as the stillbirth of a fetus from 20 weeks' gestation or the neonatal death of a liveborn baby within 28 days of birth¹, is a global health problem that has a devastating impact on families and the wider community. Accurate determination of the aetiology of the death is essential for appropriate counselling and the development of strategies to reduce occurrence.

Congenital abnormalities are the most common cause of perinatal death (spontaneous death or elective termination) in developed nations, accounting for ~ 33%^{1,2}. While the majority of cases are expected to have a genetic underpinning, caused by chromosomal abnormalities or monogenic disorders^{3,4}, there is currently limited genetic investigation to identify this molecular basis performed as standard of care⁵. An additional challenge to the diagnosis of the cause of congenital anomalies also lies in the ambiguity of the aetiology. For disorders such as non-immune hydrops fetalis (NIHF), or those involving abnormalities of the haematopoietic or central nervous system, it can be unclear clinically whether the underlying cause has genetic or infectious origin^{6,7}.

Perinatal infections are the second leading cause of perinatal death, either directly causing or indirectly contributing to ~ 15% of deaths^{1,2}, with a more significant impact observed in low- and middle-income countries. The majority of perinatal infections are acquired antenatally, leading to stillbirth through activation of the fetal inflammatory response and resultant preterm labour, or as a consequence of the infection itself⁸. Ascending bacterial infections are the most common cause of prenatal infection in developed countries, with microbes entering the amniotic sac across the membranes and infecting fetal lung through aspiration of contaminated amniotic fluid⁸.

Identification of the cause of infection at perinatal autopsy is currently performed using a combination of culture, molecular testing (PCR), and microscopy performed on fetal/placental samples, and serology

performed on maternal blood, coupled with clinical evidence⁵. However, the utility of these methods is hindered by the necessitation for *a priori* knowledge, the inherently limiting requirement of organism viability for culture, and fastidious organisms⁹. Together, these limitations prevent identification of a causative pathogen in ~ 60% of infection-related cases^{2,10}.

DNA sequencing and other genetic testing approaches have clear utility in identifying the causes underlying congenital abnormalities and infections, but their clinical use in perinatal autopsy is currently limited, and generally restricted to hypothesis-based tests. Genome sequencing (GS) however, offers an unbiased and capture-free approach to sequence all DNA, human and microbial, present in a sample¹¹. Genomic analysis has revolutionised the identification of genetic variation causative of monogenic disorders¹², and while the efficacy of metagenomic analysis for the identification of pathogens underlying infectious disease is comparatively less well established, it has been demonstrated in a number of recent reports⁹. GS therefore represents a single test, with superior clinical validity, capable of consolidating the multiple individual tests currently used to investigate the causes underlying congenital abnormality- and infection-related perinatal deaths.

Aiming to determine the feasibility of a GS-based diagnostic approach, with concurrent interrogation of metagenomic and genomic data, we developed a testing pipeline (Fig. 1) and performed a proof-of-principle study on 20 cases of fetal (13–40 weeks' gestation) death.

The study cohort comprised 12 cases with chorioamnionitis; 7 with confirmed pathogen (Positive Cases) and 5 with no cause identified (Unknown Cases) via standard-of-care microbiology/virology practises, and 8 cases with congenital abnormalities and no histological evidence of infection (Negative Cases) (Fig. 2, detailed summary in Supplementary Table 1). Following extraction of genomic DNA from fetal lung tissue, GS was performed per the standard protocol for a 30x human genome and data processed for human genomic and metagenomic analysis as summarised in Fig. 1. On average, after quality filtering, there were 417 million mapped reads (human and microbial) per case (Supplementary Fig. 1A).

Initial metagenomic analysis was performed using GATK's PathSeq¹³ (Supplementary Fig. 2, Supplementary Table 2). Human-normalised microbial read counts per case ranged from 831 to 3,057,244 for Positive and Unknown Cases, and 553 to 8,880 for Negative Cases (Supplementary Fig. 1D). Microbial abundances were analysed at the genus-level, with PathSeq able to identify causative pathogens (gram-negative and fungal organisms) in 4/7 Positive Cases concordant with standard-of-care testing (Fig. 2, Supplementary Table 1). The predominant genera present in P1 and P2 were gram-negative *Escherichia* and *Shigella*. The *Shigella* genus is, however, genetically similar to the *Escherichia coli* species, so this result is likely a false classification due to misalignment¹⁴. *Ureaplasma* was also identified in P2, concordant with results from conventional testing. Fungal *Clavispora* was correctly identified in P3, and gram-negative *Klebsiella* was the predominant genus identified in P4, with gram-positive *Enterococcus* also present (Fig. 2, Supplementary Fig. 2). While the genera identified in this case were concordant with conventional testing, the relative abundances were reversed (Supplementary Table 1), suggestive of a failure to accurately detect gram-positive genera. No significant microbial load

was identified by metagenomic analysis for the remaining 3 Positive Cases (P5-P7) determined at autopsy to be due to gram-positive Streptococcal species (Fig. 2, Supplementary Fig. 2).

In contrast to standard-of-care testing, a plausible causative pathogen was identified from metagenomic analysis for 2/5 Unknown Cases (Fig. 2, Supplementary Table 1). Significant loads of *Ureaplasma* were identified in both U1 and U2 (Fig. 2, Supplementary Fig. 2), with the fastidious nature of this bacterium consistent with it being missed by standard testing¹⁵. Fungal *Cryptococcus* was also identified in U4, however the significance of this is unclear as the finding is not consistent with the clinical picture of chorioamnionitis¹⁶ (Fig. 2, Supplementary Table 1). All microbes known or suspected as causal by conventional testing or metagenomic analysis, were present at greater than 2,000 human-normalised microbial reads. Background microbial loads, including organisms residing as normal skin or vaginal flora, were only identified at low abundance (< 550 human-normalised reads), and were present in multiple cases (Supplementary Fig. 2). No significant microbial load was identified in 8/8 Negative Cases, concordant with results from standard autopsy (Fig. 2, Supplementary Fig. 2, Supplementary Table 1).

To provide orthogonal validation of PathSeq results, metagenomic analysis was also performed using Kraken2 with re-estimation of abundance performed using Bracken^{17,18}, and 16S rRNA amplicon sequencing data was analysed using QIIME2¹⁹. Comparable numbers of human and microbial reads were aligned per case by Kraken2 as for PathSeq (Supplementary Fig. 3), with the predominant genera identified by PathSeq also being identified by Bracken (+/- Kraken2) (Supplementary Fig. 4, Supplementary Table 3). 16S sequencing data was also concordant for all cases with bacterial taxa present (Fig. 2, Supplementary Fig. 5, Supplementary Tables 1 and 4).

Genomic analysis of human-mapping reads was undertaken for cases with anatomical abnormalities, unrelated to infection (P2, P7, U5, N1-N8), in order to identify putative causative genetic variants. An average read depth of 40X was achieved, with 95% of bases covered at 20X (Supplementary Table 5). In two Negative Cases (N3 and N4), presenting with NIHF and Noonan-like features, heterozygous variants were identified *LZTR1*; a previously described pathogenic variant (p.Arg284Cys), and a novel variant classified as a variant of uncertain significance (VUS; p.Asp444Glu). A known pathogenic variant in *RHD* (p.Leu110Pro) was also identified in N7, consistent with death related to Rhesus incompatibility (Fig. 2, Supplementary Results, Supplementary Tables 1 and 6). No likely causal variants were identified in the remaining Negative Cases, or the Positive and Unknown Cases with unexplained congenital abnormalities (Fig. 2, Supplementary Table 1). Analysis of variants in primary immunodeficiency-linked genes was also performed, for all cases, to identify potential cryptic susceptibility to infection, however, no likely causal or contributory variants were identified (Supplementary Results, Supplementary Table 7).

While both genomic and metagenomic analysis have, separately, been demonstrated to have utility in the diagnosis of genetic disorders and infectious disease, this is the first study to demonstrate the feasibility of simultaneous analysis to identify the causes underlying fetal death (Fig. 1). This dual analysis approach has recently begun to be investigated in cancer research²⁰, but is unlikely to yield results for other disorders occurring past the perinatal period, even those with clinical overlap between genetic and

infectious disease, owing to the different sample types required to investigate each aetiology. The approach is, however, facilitated for prenatal disease by the unique uterine environment where fetal lung is exposed to amniotic fluid, and for ascending infections is the first and most severely affected organ⁸. As congenital abnormalities and congenital infections are the two leading factors causing or contributing to perinatal death^{1,2}, the potential to combine their assessment into a single test is expected to be of significant clinical utility.

In this pilot study, using the GATK PathSeq pipeline for initial metagenomic analysis, known causative microbes were correctly identified in 4/7 Positive Cases with known infectious cause (P1-P4), and plausible causative microbes were identified in 2/5 Unknown Cases, where cause of infection was unable to be resolved from standard-of-care testing (U1-U2). No significant microbial load was identified in 8/8 Negative Cases, consistent with conventional testing (N1-N8), and concurrent human genomic analysis, using the GATK HaplotypeCaller pipeline, enabled the detection of two causative (pathogenic) and one candidate causative (VUS) variants underlying hydrops in 3 cases (N3, N4 and N7) which were not identified by testing performed as current standard of care (Fig. 2). Together, the results from this study highlight the enormous potential of this dual metagenomic and genomic analysis approach in identifying the causes underlying perinatal death, and in identifying diagnoses beyond those made by current testing methodologies.

Although metagenomic analysis provided concordant results with routine microbiology/virology investigations for 12/15 (80%) resolved cases, it failed to detect known gram-positive Streptococcal species in 3 Positive Cases, likely due to limitations of the DNA extraction method to adequately disrupt their thick peptidoglycan cell wall²¹. The concordance of results between metagenomic and 16S rRNA amplicon analysis supports that no known diagnoses were missed due to unforeseen failings arising from the dual analysis approach (Fig. 2, Supplementary Fig. 5). A modified extraction method, using bead-beating, was trialled to address this shortcoming but no increase in the detection of gram-positive genera was observed (Supplementary Results, Supplementary Tables 8 and 9). Future iterations of this methodology will therefore require optimisation to facilitate extraction of DNA from gram-positive organisms, being cautious that increasing disruption efficiency does not also cause shearing of human DNA, and should also remove heated steps to enable recovery of RNA. Although viral RNA species (e.g. *Rubella*) are a less significant cause of congenital infection in developed countries⁸, the ability to detect them will be important for completeness of testing. Another methodological improvement required will be the inclusion of appropriate negative (no template) controls to detect and account for potential contamination⁹. This may currently be prohibitively expensive, but will be important for defining background microbial loads and to determine clinical significance thresholds.

Comparison of data between the two sequencing runs (standard and modified extraction DNA) allowed for preliminary assessment of reproducibility and identification of potential test parameters. Although there were differences in the total number of mapped (human and microbial) reads, there was consistency in the human-normalised microbial read counts and complete concordance for 5/6 cases in

the predominant genera identified (Supplementary Results). Over both runs, Positive and Unknown Cases (with successful extraction of the implicated pathogen) had total microbial read percentages greater than 0.001%, which could be used as a threshold to flag samples with high microbial load for further analysis. While some Negative Cases had total microbial read percentages approaching this value, human-normalised read counts mapping to individual genera did not exceed 550 and were generally less than 200, suggesting that they represented low-level environmental contamination rather than clinically significant colonisation. In contrast, the number of human-normalised reads mapping to individual causal or putative causal genera in Positive and Unknown Cases was greater than 2,000 reads (Supplementary Tables 2 and 8).

As part of the GATK suite of tools, PathSeq was chosen for metagenomic analysis in this setting due to its seamless integration into existing human genetic analysis pipelines (Fig. 1). While its utility in identifying samples with a high percentage of microbial mapping reads and in genus-level assignment has been demonstrated in this study, its ability to assign reads at the species-level is less reliable²². An ideal analysis pathway in a diagnostic setting will therefore utilise PathSeq as a preliminary screening tool, followed by secondary analysis of samples with high microbial load using a clinically validated tool, such as Kraken2 with Bracken²³, to provide accurate classification at the species level (Fig. 1). Additionally, alignment of reads to the NCBI microbial reference genomes of candidate causative organisms will help refine the diagnosis in cases where there are reads mapping to multiple, genetically similar genera/species within a family. Another complementary approach to further improve the metagenomics workflow, would be the combined use of a *de novo* assembler, such as metaSPAdes²⁴, to determine the identity of any unclassified reads, potentially representing novel microbial strains.

This proof-of-principle study highlights the potential utility of a GS-based investigation to consolidate multiple individual tests, with disparate workflows, into a single test capable of identifying the causes underlying congenital abnormalities and infections leading to perinatal death. While considerable optimisation, of both experimental protocols and bioinformatics pipelines, is still required to implement this dual analysis methodology to its full potential, its feasibility has been conceptually demonstrated. The simultaneous genomic and metagenomic analysis facilitated by this approach has the potential to identify the underlying aetiology in the two leading causes of perinatal death in developed countries, enabling more accurate recurrence risk counselling for affected families and aiding in the prevention of recurrence.

Declarations

WEB RESOURCES

AnnotSV: <https://github.com/lgmgeo/AnnotSV>

Bracken: <https://github.com/jenniferlu717/Bracken>

ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar>

GATK: <https://github.com/broadinstitute/gatk>

gnomAD: <https://gnomad.broadinstitute.org>

Greengenes reference database: <https://docs.qiime2.org/2019.10/data-resources/#marker-gene-reference-databases>

The Human Phenotype Ontology: <https://hpo.jax.org/app>

Kraken2: <https://github.com/DerrickWood/kraken2>

Manta: <https://github.com/Illumina/manta>

NCBI RefSeq: <https://www.ncbi.nlm.nih.gov/refseq>

OMIM: <https://omim.org>

PathSeq: <https://github.com/gatk-workflows/gatk4-pathseq>

Peddy: <https://github.com/brentp/peddy>

QIIME2: <https://github.com/qiime2/qiime2>

Radboud UMC gene lists (primary immunodeficiency gene list):
<https://order.radboudumc.nl/en/products/conditions>

SnEff: <http://snpeff.sourceforge.net>

VariantGrid: <https://github.com/SACGF/variantgrid>

ACKNOWLEDGEMENTS

We thank all staff of the SA Pathology Anatomical Pathology Department, Microbiology and Infectious Diseases Department, and the Genetics and Molecular Pathology Department, in particular Sui Yu, for their input with routine autopsy investigations. Thanks also to Mark Walker and Hera Vlamakis, Broad Institute of MIT and Harvard, for their input with initial analyses using PathSeq and microbial extraction methods, respectively. Additional thanks to the staff of the Australian Genomics Research Facility, Melbourne and the staff of the Centre for Cancer Biology ACRF Genomics Facility, in particular David Lawrence and Paul Wang. Finally, thank you to the families for their involvement in research, and to all members of the Genetics and Molecular Pathology Research Laboratory for their support.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING

This research was supported by NHMRC grant (APP1123341), Genomics Health Futures Mission - Medical Research Futures Fund (GHFM76777), and the Australian Genomic Health Alliance NHMRC Targeted Call for Research into Preparing Australia for the Genomics Revolution in Healthcare (GNT1113531) to Hamish Scott and Chris Barnett; and the Australian Cancer Research Foundation to Hamish Scott. Additional support provided by Cancer Council SA's Beat Cancer Project on behalf of its donors and the State Government of South Australia through the Department of Health, and NHMRC Fellowship (APP1023059) to Hamish Scott; the Australian Government Research Training Program Scholarship, the Australian Genomics Health Alliance PhD Award & NHMRC (GNT1113531), and the Maurice de Rohan International Scholarship to Alicia Byrne; and The Hospital Research Foundation Fellowship and the Royal Adelaide Hospital Research Foundation Fellowship to Peer Arts.

AUTHOR CONTRIBUTIONS

ABB, LA-M, TH and PA drafted the manuscript. ABB designed and coordinated the study. ABB, MB, AN, EGL, SLK-S and CPB coordinated cases and biological samples. LA-M and LE processed GS data for metagenomics analysis with ABB, LA-M and LE performing data analysis. TH and JF processed GS data for human genomic analysis with ABB, TH and PA performing data analysis. WTP and RK generated 16S amplicon data with LA-M processing the data and ABB and LA-M performing data analysis. AWS supervised bioinformatic analyses. BR, JL, NM, TYK and LM performed routine autopsy investigations. ABB, LA-M, LE, TH, LM, PA, CPB, AWS and HSS contributed to interpretation and discussion of results. HSS conceived the study and CPB, AWS and HSS supervised. All authors read and approved the manuscript.

DATA AVAILABILITY STATEMENT

Human and microbial sequence data has been deposited at the European Genome-phenome Archive (currently preparing upload) which is hosted by the European Bioinformatics Institute. All unique materials and datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Online Methods

Ethics declaration and consent to participate

The study was approved by the Human Ethics Committee of the Women's and Children's Health Network, South Australia, Australia (HREC/15/WCHN/35), as part of the NHMRC-funded Genomic Autopsy Study. Consent for medical research was given by next of kin at autopsy, and all research was conducted in accordance with the Declaration of Helsinki.

Routine autopsy investigations

Cases were delivered at a range of metropolitan and rural maternity hospitals, with all autopsies performed at a single metropolitan referral hospital. Delay between death and autopsy ranged from 2 to

10 days (Supplementary Table 1). Standard-of-care autopsy encompassed review of family and obstetric history, macroscopic and histological placental examination, and examination of the fetus, including: external examination, radiological investigations, macroscopic and histological organ examination, and a range of haematology, biochemistry, microbiology/virology, and genetic testing as indicated. Microbiology/virology testing was performed using combinations of culture and PCR performed on fetal/placental tissue and serology performed on maternal blood. To assign a microbe as causative, positive results from these methods needed to be supported by consistent clinical evidence and histological findings. Genetic investigations included microarray and karyotype, with biochemical tests and single gene or panel testing performed where a specific phenotype was suspected.

DNA isolation

Genomic DNA was isolated from fresh frozen lung tissue (Figure 1A) using both a standard human tissue DNA extraction method (standard extraction) and a modified method, endeavouring to improve the yield of gram-positive bacterial DNA (modified extraction). For the standard extraction, DNA was isolated using the DNeasy blood and tissue kit (Qiagen, Cat #69504) following the manufacturer's instructions. For the modified extraction, DNA was isolated using the DNeasy blood and tissue kit with the addition of a mechanical pre-treatment for the lysis of gram-positive bacteria. In brief, lung tissue was lysed in Buffer ATL and proteinase K, then vortexed with PowerBead Pro tubes (Qiagen DNeasy PowerSoil Pro Kit, Cat #41014) for 2 minutes (6x 20 second bursts). After centrifugation, the supernatant was transferred to a new tube and the extraction was continued as per the DNeasy blood and tissue kit instructions.

Sequencing

Genome sequencing was performed at the Melbourne node of the Australian Genome Research Facility, Australia. DNA was prepared using Illumina TruSeq chemistry and libraries sequenced on an Illumina NovaSeq 6000 as 150bp paired-end reads to a mean human coverage of 30x (Figure 1B, Supplementary Table 5). Sequencing data was processed at the Centre for Cancer Biology ACRF Genomics Facility, Adelaide, Australia (adaptor trimming, filtration of low quality and/or complexity reads, and read deduplication) using a pipeline based on Picard.

16S rRNA amplicon sequencing was performed at the Centre for Cancer Biology ACRF Genomics Facility following the Illumina 16S Metagenomic Sequencing Library Preparation protocol. The V3 and V4 regions of the 16S rRNA gene were amplified by PCR using region of interest specific primers (Forward Primer: 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG, Reverse Primer = 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC). Amplicon libraries were sequenced on an Illumina MiSeq as paired-end reads, and sequencing data was processed (quality filtering, paired read merging, denoising and chimeric removal) using the DADA2 pipeline.

Human genetic analysis

Genome sequencing data was aligned to the human genome build 37 (hg19) using BWA (v0.7.12). Single Nucleotide Variants (SNVs) and small insertions/deletions (indels) were jointly called across all samples using GATK HaplotypeCaller (v3.8) and structural variants were called using Manta (v1.6.0). Joint genotyping was then performed using GATK GenotypeGVCF and variant quality scores were re-calibrated using GATK's VQSR. SNVs and indels were annotated using SnpEff (v4.1) and CNVs were annotated using AnnotSV (v2.1) (Figure 1C). Peddy was used to confirm biological replicates and ensure correctly paired standard and modified extraction samples for metagenomic analysis by inference from the individual human genotypes (Supplementary Table 10).

Mendeliome variant analysis (from standard extraction DNA) was performed in VariantGrid for all cases with unexplained anatomical abnormalities. HPO term-based gene lists were generated for each case in order to narrow the search space as no parental data was available. Analysis selected for rare, protein altering variants (SNVs and indels: gnomAD and in-house frequencies $\leq 1\%$, max 5 homozygotes for recessive, and $\leq 0.01\%$, max 5 heterozygotes for dominant; CNVs: no reciprocal overlap with known benign CNVs $\geq 70\%$), and prioritised variants for interpretation based on *in silico* pathogenicity predictions and sequence conservation scores. Variants of interest were classified according to ACMG guidelines (Figure 1C, Supplementary Table 6). Familial DNA samples were not available for segregation analysis. A gene list-based analysis of variants in primary immunodeficiency-linked genes (393 genes, Radboud UMC) was performed for all cases to identify potential cryptic susceptibility to infection (Supplementary Table 7).

PathSeq microbial analysis

Human-aligned BAMs from genome sequencing data were used as input files, and non-human reads aligned to microbial reference genomes using GATK's PathSeq (GATK v4.1.1.0). The default microbe database, pathseq_microbe (archaea, bacteria, fungi, protozoa, viruses; 2017/27/04 version) was used with default parameters. The relative abundance of each genera was determined as counts per billion by normalising the number of unambiguous reads mapping to each genus by the number of human mapping reads (Figure 1D). Normalisation was performed to allow comparison between samples, and was based on the hypothesis that each tissue sample contained equal amounts of human DNA. Read counts are listed in Supplementary Table 2 (standard extraction) and Supplementary Table 8 (modified extraction).

Kraken2 and Bracken microbial analysis

In parallel, genome sequencing data was classified against human and microbial reference genomes using Kraken2 (v2.0.7-beta) with a custom database built from the following libraries: archaea, bacteria, fungi, human, protozoa, viral and UniVec_Core (complete or representative) (April 2020 version). Bracken (v2.2) was used as a post-processing step to improve abundance estimates and remove false positive results. The relative abundance of each genera was determined as counts per billion, as performed for PathSeq (Figure 1E). Read counts are listed in Supplementary Table 3 (standard extraction).

QIIME2 microbial analysis

16S rRNA amplicon sequencing data was taxonomically classified using QIIME2 (v2019.1) with the Greengenes reference database (bacteria and archaea; version 13_8) and default parameters. The relative abundance of each taxa was determined by normalising the number of unambiguous reads assigned to each taxon by the total number of reads passing all filters. Read counts are listed in Supplementary Table 4 (standard extraction) and Supplementary Table 9 (modified extraction).

References

1. Australian Institute of Health and Welfare; *Stillbirths and neonatal deaths in Australia 2017–2018*. <https://www.aihw.gov.au/reports/mothers-babies/stillbirths-and-neonatal-deaths-in-australia-2017> (2021) doi:10.25816/91w8-1933.
2. Prevention and Population Health Branch, Wellbeing SA. *Maternal and Perinatal Mortality in South Australia 2007–2017*. (2019).
3. Reddy, U. M. *et al.* Karyotype versus microarray testing for genetic abnormalities after stillbirth. *N. Engl. J. Med.* **367**, 2185–2193 (2012).
4. Quinlan-Jones, E. *et al.* Molecular autopsy by trio exome sequencing (ES) and postmortem examination in fetuses and neonates with prenatally identified structural anomalies. *Genet. Med.* **0**, 1–9 (2018).
5. Nijkamp, J. W. *et al.* Perinatal death investigations: What is current practice? *Semin. Fetal Neonatal Med.* **22**, 167–175 (2017).
6. Lamont, R. F. *et al.* Parvovirus B19 infection in human pregnancy. *BJOG: An International Journal of Obstetrics & Gynaecology* vol. 118 175–186 (2011).
7. van Zuylen, W. J. *et al.* Congenital cytomegalovirus infection: Clinical presentation, epidemiology, diagnosis and prevention. *Obstet Med* **7**, 140–146 (2014).
8. Goldenberg, R. L. & Thompson, C. The infectious origins of stillbirth. *Am. J. Obstet. Gynecol.* **189**, 861–873 (2003).
9. Simner, P. J., Miller, S. & Carroll, K. C. Understanding the Promises and Hurdles of Metagenomic Next-Generation Sequencing as a Diagnostic Tool for Infectious Diseases. *Clin. Infect. Dis.* **66**, 778–788 (2018).
10. Page, J. M. *et al.* Stillbirth Associated With Infection in a Diverse U.S. Cohort. *Obstetrics & Gynecology* **134**, 1187 (2019).
11. Gihawi, A. *et al.* SEPATH: benchmarking the search for pathogens in human tissue whole genome sequence data leads to template pipelines. *Genome Biology* vol. 20 (2019).
12. Retterer, K. *et al.* Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.* **18**, 696–704 (2016).

13. Walker, M. A. *et al.* GATK PathSeq: a customizable computational tool for the discovery and identification of microbial sequences in libraries from eukaryotic hosts. *Bioinformatics* **34**, 4287–4289 (2018).
14. Lan, R. & Reeves, P. R. *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect.* **4**, 1125–1132 (2002).
15. Viscardi, R. M. *Ureaplasma* species: role in diseases of prematurity. *Clin. Perinatol.* **37**, 393–409 (2010).
16. Rahimi, K., Chetty, R. & Clarke, B. Cryptococcemia resulting in an incomplete abortion in an HIV-positive patient. *Can. J. Infect. Dis. Med. Microbiol.* **20**, e97–9 (2009).
17. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
18. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* vol. 3 e104 (2017).
19. Bolyen, E. *et al.* Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 1091 (2019).
20. Feng, Y. *et al.* Metagenomic and metatranscriptomic analysis of human prostate microbiota from patients with prostate cancer. *BMC Genomics* **20**, 146 (2019).
21. Gill, C., van de Wijkert, J. H. H. M., Blow, F. & Darby, A. C. Evaluation of Lysis Methods for the Extraction of Bacterial DNA for Analysis of the Vaginal Microbiota. *PLoS One* **11**, e0163148 (2016).
22. Ye, S. H., Siddle, K. J., Park, D. J. & Sabeti, P. C. Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell* **178**, 779–794 (2019).
23. Lepuschitz, S. *et al.* Analytical Performance Validation of Next-Generation Sequencing Based Clinical Microbiology Assays Using a K-mer Analysis Workflow. *Front. Microbiol.* **11**, 1883 (2020).
24. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

Figures

Figure 1

Overview of concurrent human genomic and metagenomics workflow.

Schematic diagram outlining the experimental and computational processes involved in simultaneous genomic and metagenomic analysis. (A) DNA extraction from fetal lung isolates both human and microbial DNA which is sequenced per the standard protocol for a 30x human genome. (B) Raw fastq files are pre-processed, for both human and microbial reads, with adapter trimming and fastQC performed to generate unaligned sequencing reads. (C) Mapping to hg19 and variant calling for human genomic analysis is performed according to GATK best practice guidelines. Variant prioritization and interpretation is then performed to identify causative genetic variants. (D) After subtraction of human reference-aligned reads, mapping to microbial reference genomes of the remaining reads is performed by PathSeq (part of GATK; red box) for primary metagenomic analysis. Following normalisation, relative abundance is interrogated to identify putative causative pathogens. (E) Raw fastq files are processed and aligned to microbial reference genomes using Kraken2, with post-processing by Bracken. Following normalisation,

relative abundance is interrogated then compared to PathSeq results for orthogonal validation of causative pathogens. Results generated from each analysis phase are available in supplementary tables and figures (referenced in figure). Additional methodological details are available in the online methods.

Figure 2

Microbial and genetic findings from investigation of 20 cases of fetal death

Summary of microbial findings from metagenomic analysis (A), 16S rRNA amplicon analysis (B), and conventional investigation (C), and human genetic findings from genome sequencing analysis (D). Detailed summary available in Supplementary Table 1. Cases (x-axis) are separated into 3 groups with respect to infection-status: Positive, Unknown, and Negative, based on standard-of-care autopsy findings. (A) Heatmap showing relative abundance of genera (left side y-axis) identified by PathSeq metagenomic analyses with >2,000 human-normalised microbial reads in any one sample. The abundance of each microbe is coloured according to discrete bins of number of human-normalised reads (right legend). (B) Genera (left side y-axis) with >50 normalised microbial reads identified using QIIME2 analysis of 16S rRNA amplicon sequencing data, providing orthogonal validation as an established DNA testing approach for bacterial and archaeal organisms. Green colour indicates positive identification (right legend). (C)

Causative species identified using conventional microbiology/virology testing (culture, serology, PCR) at autopsy. Red, blue, and green coloured text indicates gram-negative bacteria, gram-positive bacteria and fungal species, respectively. 'Negative' indicates no microbe identified and 'non-contributory' indicates microbes identified were post-mortem artifacts. **(D)** Causative/candidate causative genetic variants identified from human genome analysis, with ACMG classification provided. Clinical picture is indicated in parentheses. Genomic analysis was not performed for cases without congenital abnormalities.

CA: congenital abnormality. IUGR: intrauterine growth restriction. N/A: not assessed. NAD: no abnormality detected. NCA: no congenital abnormality. NIHF: non-immune hydrops fetalis.

Supplementary Files

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

- [MetagenomicsPaperSupplementaryTablesFINAL.xlsx](#)
- [MetagenomicsPaperSupplementaryMaterialFINALX.pdf](#)