Supplementary Material

for

SIMULTANEOUS GENOMIC AND METAGENOMIC ANALYSIS FOR MISCARRIAGE AND STILLBIRTH FURTHER INCREASES THE DIAGNOSTIC UTILITY OF GENOME SEQUENCING

Alicia B. Byrne[1,2,*], Luis Arriola-Martinez[1,3,*], Leila Eshraghi[1,3,#], Thuong Ha[1,3,#], Jinghua Feng[2,3], Milena Babic[1], Andy Nguyen[4], Eleanor G. Ludington[5], Sarah L. King-Smith[1,6], Wendy T. Parker[3], Rosalie Kenyon[3], Brett Ritchie[7], Jill Lipsett[8], Nick Manton[8], T. Yee Khong[8,9], Lynette Moore[8,9], Peer Arts[1,9], Christopher P. Barnett[9,10], Andreas W. Schreiber[2,3,11], Hamish S. Scott[1-4,6,9]

SUPPLEMENTARY APPENDIX

SUPPLEMENTARY RESULTS AND DISCUSSION

Extended comparison of PathSeq metagenomic analysis results to conventional microbiology testing results

In 2 cases with chorioamnionitis (P7 and U1) where discordant results were obtained by standard testing methodologies, metagenomic analysis was supportive of the pathologist's conclusion. For P7, maternal serology testing identified *Toxoplasma* IgG and IgM which was determined to be a false positive as the finding was not supported by microscopy or PCR, and false positive IgM results are not uncommon. For U1, both fetal tissue microscopy and maternal serology testing at autopsy were positive for *Cytomegalovirus*, though this was inconsistent with the clinical picture of chorioamnionitis and PCR was negative. No evidence of *Toxoplasma* or *Cytomegalovirus*, at any taxonomic level, was identified by metagenomic analysis in these cases, respectively (Supplementary Figure 2, Supplementary Table 1).

Comparison of PathSeq metagenomic analysis of standard and modified extraction DNA

Genome sequencing of modified extracted DNA yielded an average of 676 million mapped reads (human and microbial) per case after quality filtering. Human-normalised microbial read counts per case ranged from 1,519 to 2,400,868 for Positive and Unknown Cases, and from 236 to 3,803 for Negative Cases (Supplementary Table 8). For modified extraction DNA, microbial reads represented, on average, 0.0353% of all mapped reads for Positive and Unknown Cases, and 0.0002% for Negative Cases. This was comparable to the results obtained from standard extraction DNA where microbial reads represented, on average, 0.0465% of all mapped reads for Positive and Unknown Cases, and 0.0004% for Negative Cases (Supplementary Tables 2 and 8).

Causal or suspected causal genera detected from analysis of standard extraction DNA were concordant with analysis of modified extraction DNA in 5/6 cases (P1-P3, U1, U2). The single discordant result was observed for P4, where the gram-negative genera, *Klebsiella*, identified as causative at autopsy and detected from standard extraction DNA, was not detected from modified extraction DNA. Gram-positive *Enterococcus*, expected to be present in this sample, was successfully detected. Fungal *Fusarium* was also, unexpectedly, identified in P4 at greater than 11,000 human-normalised reads, the significance of which is unclear. While it may be an environmental contaminant introduced through sample processing, its absence from other samples, processed similarly, makes this unlikely (Supplementary Table 8).

All microbes known or suspected as causal by conventional testing or metagenomic analysis, were present at greater than 3,000 human-normalised microbial reads, and were either absent in all other samples or

were present at human-normalised read counts less than 750. Similarly to the analysis of standard extraction DNA, no significant microbial load (<300 human-normalised reads) was identified in 8/8 Negative Cases. While the modified-extraction protocol was trialled to improve the recovery of gram-positive bacterial DNA, no improvement in the detection of these microbes was observed (Supplementary Table 8).

Extended description of human genetic findings

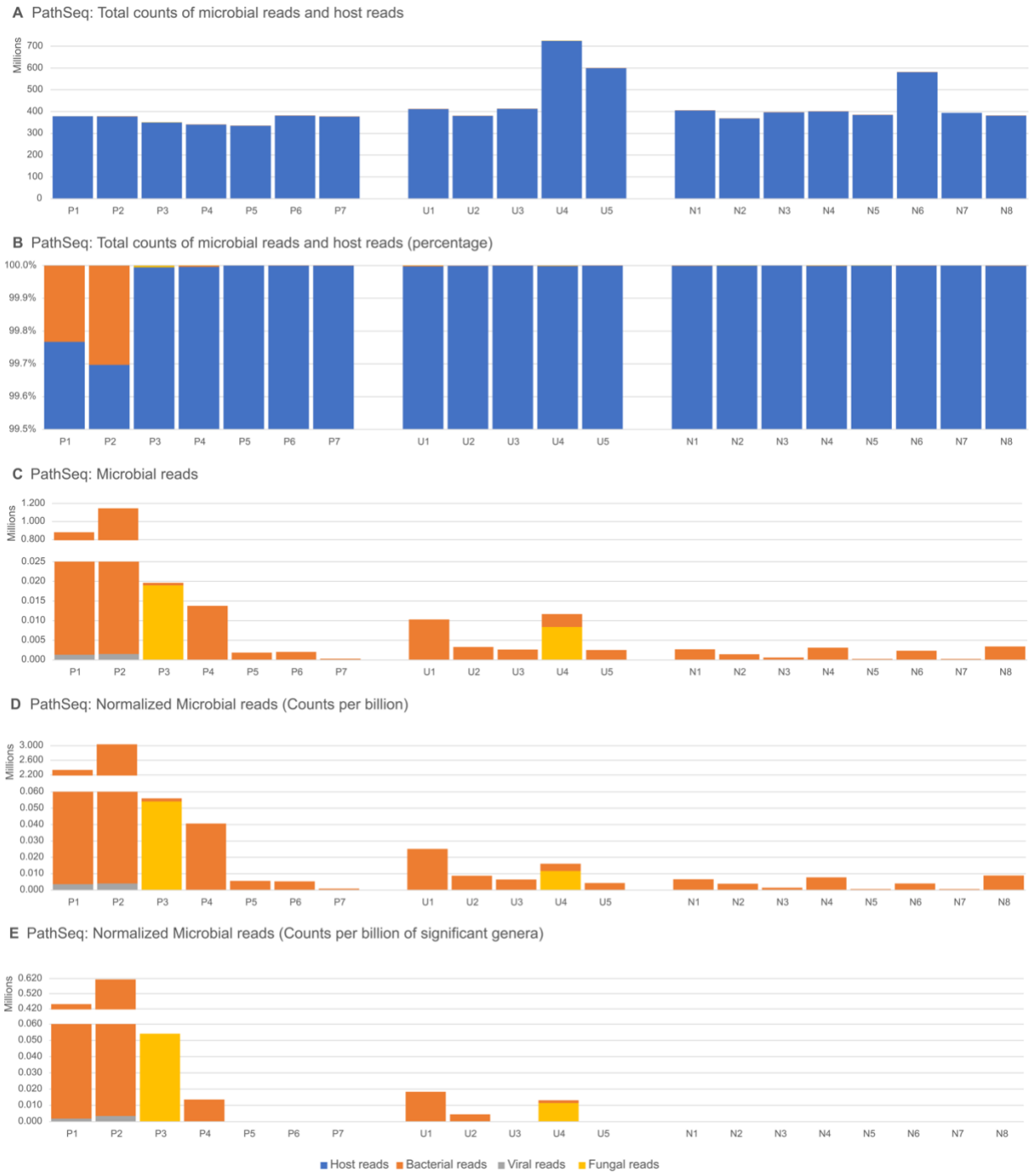
No relevant differences were observed in the sequencing quality metrics for human DNA isolated using the modified extraction protocol compared to the standard extraction protocol (Supplementary Table 5) and negligible genotype differences were observed between DNA from the two methods (Supplementary Table 10).

HPO term-based gene list analysis identified causative or candidate causative variants for 3/8 Negative Cases with congenital abnormalities. In 2 cases with Noonan-like features, heterozygous missense variants were identified in *LZTR1* (MIM: 600574; N3: p.Asp444Glu, N4: p.Arg284Cys). The p.Arg284Cys variant has been previously reported in cases of both Noonan Syndrome and Schwannomatosis, and is classified in ClinVar as pathogenic (variant ID: 209089). The novel p.Asp444Glu variant is absent from the population database gnomAD and computational prediction tools suggest the variant to be deleterious, with Asp444 being highly conserved (93/94 species). There are 2 alternate Asp444 missense variants present in gnomAD (p.Asp444Asn: 2 alleles, 0.0008%; p.Asp444Val: 1 allele, 0.003%), with the p.Asp444Asn variant also present in ClinVar, classified as a variant of uncertain significance (VUS; variant ID 1055416). Without parental data to confirm the p.Asp444Glu variant identified in P3 as *de novo*, it is classified as a VUS. In a third case (N7), a known pathogenic variant was identified in *RHD* (MIM: 111680; p.Leu110Pro, variant ID: 17712). This variant is associated with expression of an alternate RhD antigen ('partial D'), consistent with the cause of death being related to Rhesus incompatibility despite administration of anti-D (Supplementary Table 1, full variant details in Supplementary Table 6).

Analysis of genetic variants in immunodeficiency-related genes did not identify any variants of interest (Supplementary Table 7). The majority of variants identified were monoallelic variants in genes associated with autosomal recessive immune deficiencies. Of the heterozygous variants in genes associated with autosomal dominant disorders, the majority were not predicted by computational tools to be highly damaging. The remaining variants were in genes associated with autoimmunity without recorded predisposition to infection or were syndromes with distinctive phenotypes beyond immune

manifestations. However, in the future, expanded assessment of fetal DNA in cases of congenital infection may 1) reveal rare variants causative of monogenic immune deficiencies or, 2) through the analysis of larger cohorts, result in the identification of more common variants that together confer susceptibility of the fetus to infectious disease. The implementation of parent-proband trio genome sequencing would also allow identification of maternal sequence variants potentially responsible for a depressed immune response. Together, these analyses could eventually help elucidate why some pregnancies are affected by maternal vaginal carriage of certain microbes while others are not.

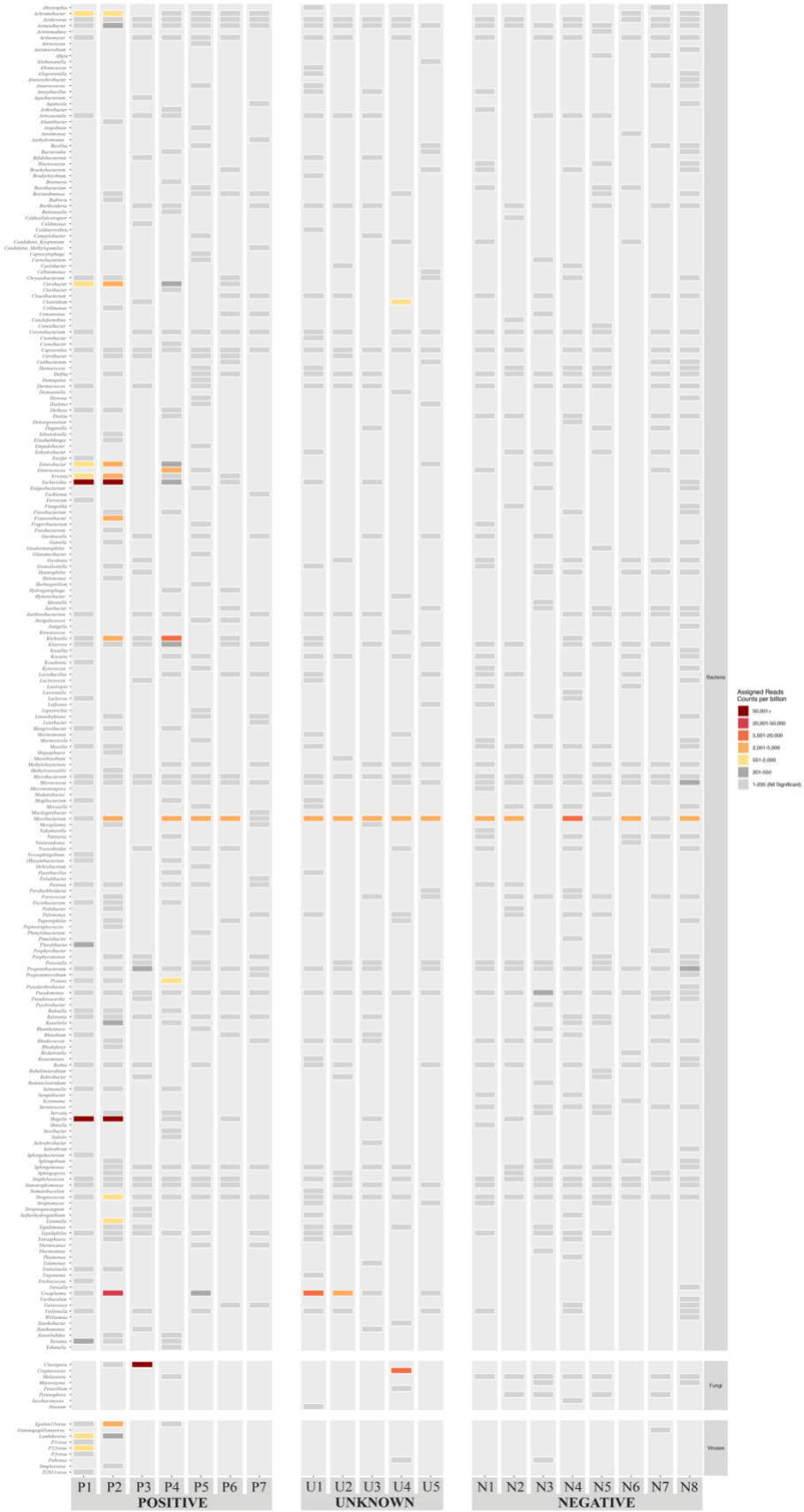
SUPPLEMENTARY FIGURES



Supplementary Figure 1: Summary of microbial reads identified from PathSeq.

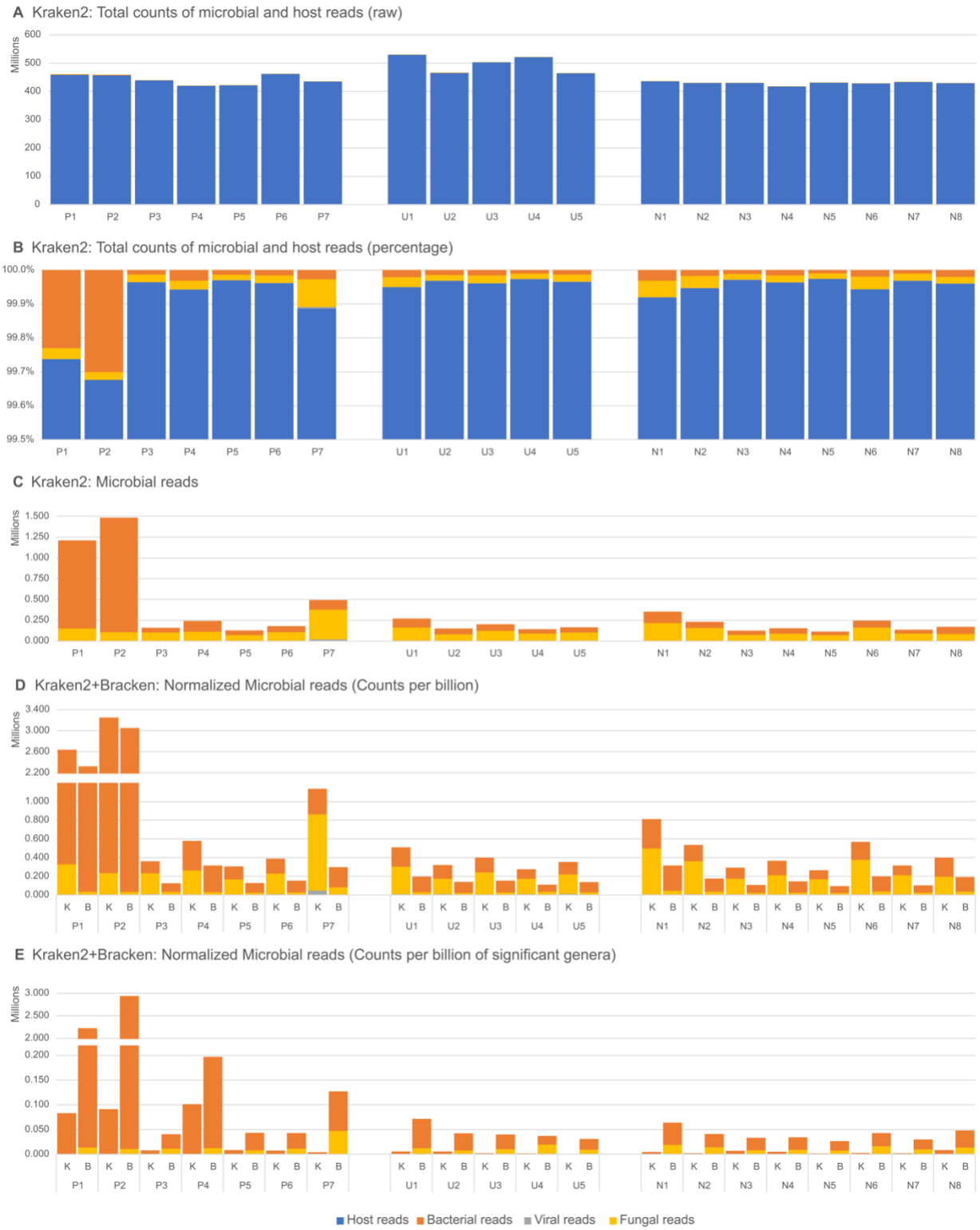
Bar graphs representing five quantitative measures of microbial reads (y-axis) from each fetal lung sample (x-axis), as identified from analysis using PathSeq. Blue bars represent human (host) reads, with orange, yellow, and grey bars representing bacterial, fungal and viral reads, respectively. (**A**) Total counts of mapped human and microbial sequencing reads, visualized together. (**B**) Proportion of mapped human

and microbial reads as a percentage of the total. (**C**) Total counts of microbial reads, visualized alone. (**D**) Microbial read counts normalised by human read count. (**E**) Human-normalised microbial read counts, with only individual genera with abundance greater than 550 human-normalised reads visualised. Genera known to be false positives, arising from misannotated reference genomes (Mycobacterium) are excluded.


Assigned Reads
Counts per billion
50,001+
20,001-50,000
5,001-20,000
2,001-5,000
551-2,000
201-550
1-200 (Not Significant)
Bacteria
Fungi
Viruses
P1
P2
P3
P4
P5
P6
P7
U1
U2
U3
U4
U5
N1
N2
N3
N4
N5
N6
N7
N8
POSITIVE
UNKNOWN
NEGATIVE

Supplementary Figure 2: Taxonomic profiling from PathSeq.

Heatmap showing the relative abundance of different microbes, identified using PathSeq, at the genus level (left side y-axis) and kingdom level (right side y-axis) in each fetal lung sample (x-axis). The abundance of each microbe is coloured according to discrete bins of number of human-normalised reads. Light grey colour represents genera that are present with less than 200 human-normalised reads, through to dark maroon colour representing genera with more than 50,000 human-normalised reads. Genera with less than 550 human-normalised reads were considered as background microbial load. Genera with greater than 2,000 human-normalised reads were considered as potentially causative. Genera known to be false positives, arising from misannotated reference genomes (Mycobacterium) were not considered further.



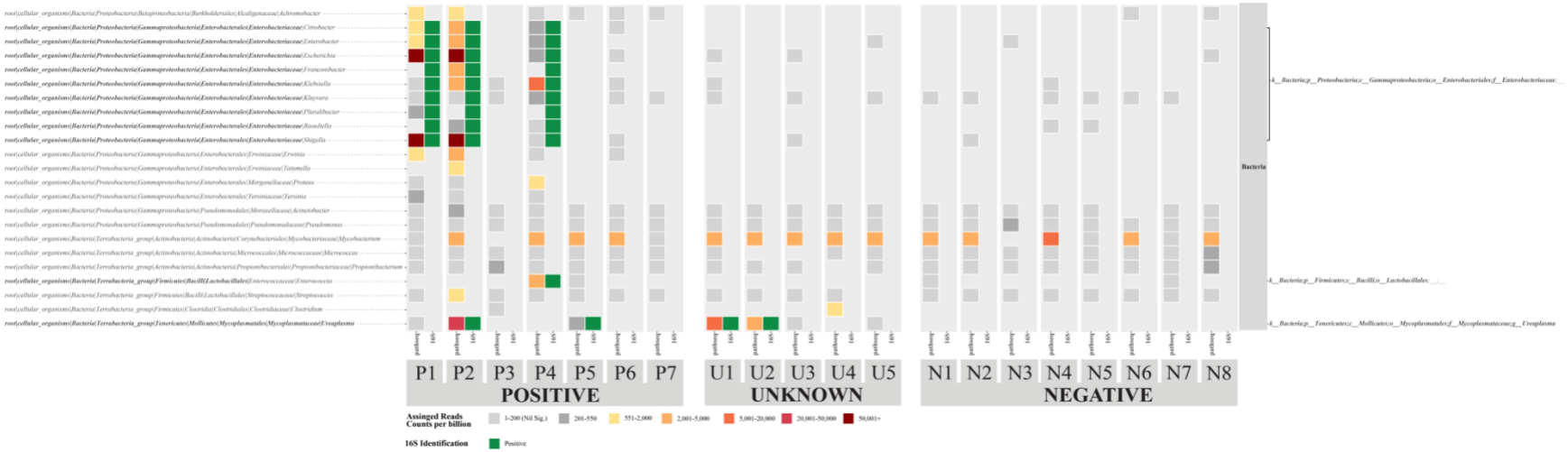
Supplementary Figure 3: Summary of microbial reads identified from Kraken2 and Bracken.

Bar graphs representing five quantitative measures of microbial reads (y-axis) from each fetal lung sample (x-axis), as identified from analysis using Kraken2 (**A-C**), and Kraken2 and Bracken post-processing (**D,E**).

The bars represent human (host; blue), bacterial (orange), fungal (yellow) or viral (grey) reads. (**A**) Total counts of aligned human and microbial sequencing reads, visualized together. (**B**) Proportion of aligned human and microbial reads as a percentage of the total. (**C**) Total counts of microbial reads, visualized alone. (**D**) Microbial read counts, normalised by human read count. (**E**) Human-normalised microbial read counts, with only individual genera with abundance greater than 550 human-normalised reads visualised.

Supplementary Figure 4: Taxonomic profiling from PathSeq, Kraken2 and Bracken.

Heatmap showing the relative abundance of microbes (genera: left side y-axis, kingdom: right side y-axis) identified using PathSeq compared to the relative abundance of microbes identified using Kraken2 and Bracken for each fetal lung sample (x-axis). The abundance of each microbe is coloured according to discrete bins of number of human-normalised reads. Light grey colour represents genera that are present with less than 200 human-normalised reads, through to dark maroon colour representing genera with more than 50,000 human-normalised reads. Only genera identified by PathSeq metagenomic analyses as having >200 human-normalised microbial reads in any one sample are displayed for comparison. Genera that are concordantly identified by PathSeq and Bracken +/- Kraken2 are considered a true positive result. Genera with <2,000 human-normalised microbial reads by PathSeq analysis, that are present in most or all samples at comparable abundance by Kraken2 and/or Bracken analysis, were not considered further.



Supplementary Figure 5: Taxonomic profiling from PathSeq and QIIME2 analysis of 16S rRNA amplicon sequencing.

Heatmap showing the relative abundance of microbes (genera: left side y-axis, family or order: right side y-axis) identified using novel PathSeq analysis compared to positive identification of microbes by established QIIME2 analysis of 16S rRNA amplicon sequencing for each fetal lung sample (x-axis). For PathSeq, the abundance of each microbe is coloured according to discrete bins of number of human-normalised reads (light grey colour: genera present with less than 200 human-normalised reads, through dark maroon colour: genera with more than 50,000 human-normalised reads). For QIIME2, positive identification (>50 normalised microbial reads) is represented by green colour. Only genera identified by PathSeq metagenomic analyses as having >200 human-normalised microbial reads in any one sample are displayed for comparison. 16S sequencing can only detect bacterial and archaeal organisms, so was not expected to identify the fungal organisms in P3 or U4.

SUPPLEMENTARY TABLES

Supplementary Table 1: Detailed cohort characteristics and comparison of microbiology and genetic testing results using conventional autopsy methods and metagenomic and genomic approaches.

Supplementary Table 2: Raw read numbers of unambiguously assigned reads (genus level) from PathSeq analysis of genome sequencing data - standard extraction.

Supplementary Table 3: Raw read numbers of unambiguously assigned reads (genus level) from Kraken2, and Bracken analysis of genome sequencing data, with >200 human-normalised microbial reads by PathSeq analysis - standard extraction.

Supplementary Table 4: Raw read numbers from QIIME2 analysis of 16S rRNA amplicon sequencing data - standard extraction.

Supplementary Table 5: Impact of modified-extraction method compared to standard- extraction method on performance of genome sequencing for human DNA.

Supplementary Table 6: Causative / candidate causative variants identified from human genetic analysis.

Supplementary Table 7: Prioritised variants in primary immunodeficiency-linked genes.

Supplementary Table 8: Raw read numbers of unambiguously assigned reads (genus level) from PathSeq analysis of genome sequencing data - modified extraction.

Supplementary Table 9: Raw read numbers from QIIME2 analysis of 16S rRNA amplicon sequencing data - modified extraction.

Supplementary Table 10: Inferred identity of standard- and modified- extraction samples by Peddy analysis of human genotypes.