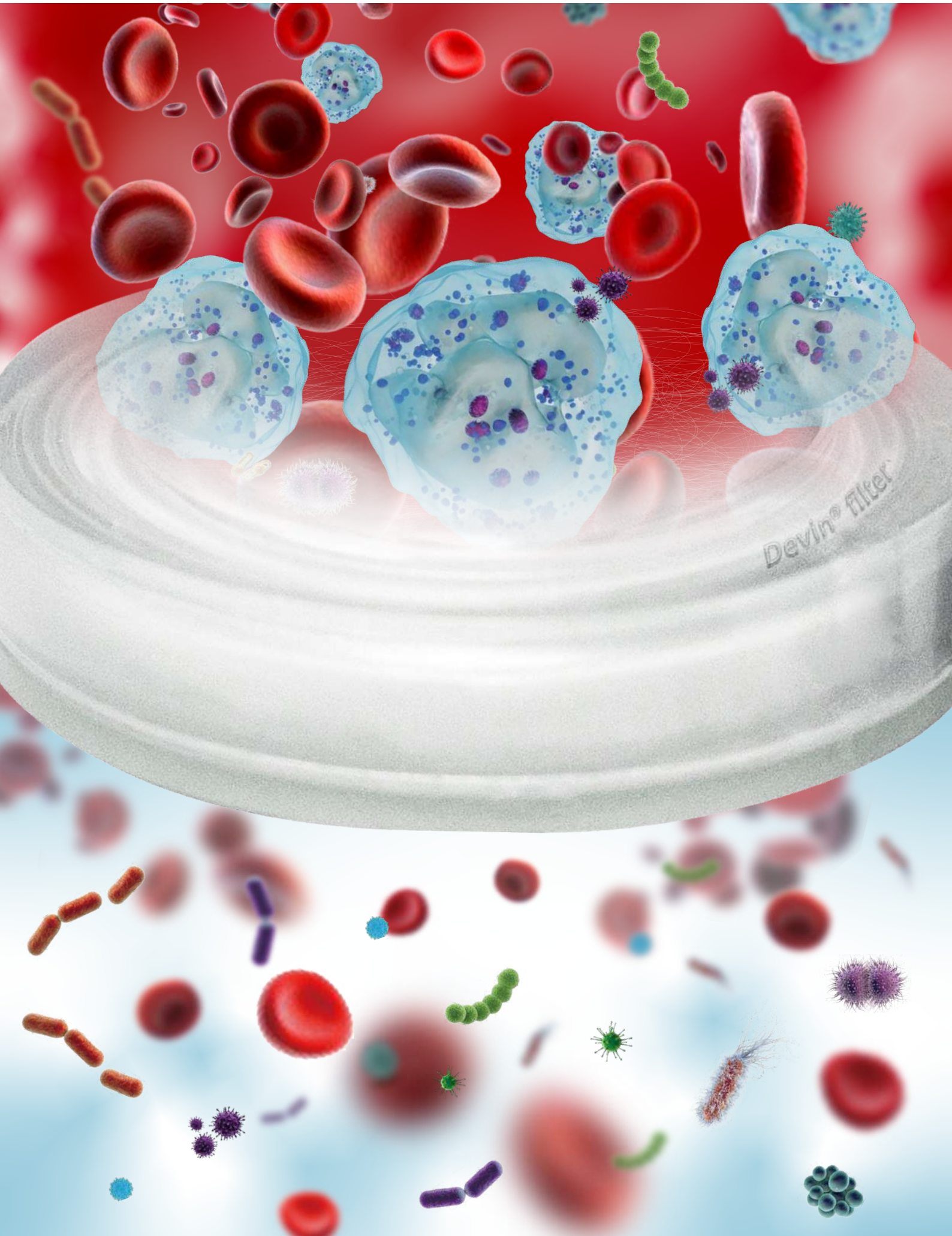
Needle in the Haystack: How to Remove Human Background When You Want to Detect Microorganisms



Development of rapid and accurate diagnostic tests for pathogens detection

The emergence of novel infectious diseases poses a huge challenge for conventional diagnostic tests, which are mainly based on *a priori knowledge* and sometimes time-consuming culturing steps. Making use of advances in sequencing technology and bioinformatics tools, metagenomics can elucidate microbial profiles from various samples with significantly faster turnaround time comparing to the conventional culture methods. Although sequencing of conserved gene families (16S rRNA, 18S rRNA, rpoB, etc.) can provide microbiome structure in a sample, metagenomic whole-genome sequencing is more advantageous in identifying undetectable pathogens with accompanied virulent factors and/or antibiotic resistance genes[1-3]. These information pieces are valuable in clinical settings such as choosing an appropriate treatment regimen or tracking an outbreak of a novel pathogen.

Metagenomic studies for clinical diagnosis have investigated a broad spectrum of specimens such as blood or positive blood cultures, urine, other body fluids, bone, joint, and tissues[4]. Body fluids are some of the most ubiquitous materials on account of their ready accessibility compared to diseased tissues and abundance of genetic information. Circulating cell-free DNA (cfDNA) in plasma has also been used in some tests (for example, Karius; https://kariusdx.com/) to detect infectious diseases and promoted for its sensitivity and non-invasive nature. However, this kind of tests have some main constraints which make it impossible to compete against other genomic-based tests. Because the material represents only fragments of microbial DNA, therefore it is impossible to identify antimicrobial resistance factors. Other drawbacks of using cfDNA are the relatively low specificity (63%) and the loss of RNA information. Moreover, speed (29 hours for sample receipt to result) and cost are also the limitations of the such tests and make it less likely to be the optimal choice of patients and healthcare providers[5-7]. Other metagenomic sequencing assays designed for intact nucleic acids analysis are still predominant with the ability to combine pathogen and antimicrobial resistance identification. These approaches have been developing to continually reduce the cost and sample-to-result time without a trade-off with comprehensiveness and depth of information.

Concurrently, a growing number of studies on validating these approaches will expedite the clinical utility of diagnostic metagenomics. *Despite the potential described above, some current key issues to overcome for implementation of metagenomics in mainstream laboratories are lack of user-friendly analytic pipelines and complete public databases, high reagent cost, and relatively low sensitivity with a high host DNA contamination background.*

The struggle with host DNA contamination

One of the most important wet lab technical limitations of metagenomics is the degree of host background.

A small fraction of human genetic materials contaminated in a clinical sample could lead to a high host-to-pathogen nucleic acid ratio due to the relative size of human versus pathogen genomes. The interfering host DNA could negatively affect the sensitivity of medical tests in detecting very low abundant microbes. Indeed, compared to traditional culture testing, recent studies using metagenomics to investigate diarrhea samples which were positive for Shiga-Toxigenic *E. coli* and cerebrospinal fluid (CSF) samples that were positive for a single primary pathogen showed the analytical sensitivity of only 67% and 70.2%, respectively[13, 14]. Other researchers also

demonstrated that when host reads exceed 90% of the total, the accuracy of the microbiome profiling and the sensitivity of low abundant species detection significantly decreased[15, 16]. Pereira-Marques *et al.* also indicated that increasing sequencing depth could improve the identification of minor species in the samples containing a high level of host DNA[15]. However, a higher sequencing depth will add a fraction of the costs for sequencing (currently ranging from $128 to $685)[4] and also lengthen the sample-to-answer time, consequently, decrease the clinically actionable likelihood of these approaches[17].

Although some *in silico* techniques to remove the influence of human sequences have been rigorously applied, such as Decontam, improved host DNA depletion and microbial enrichment methods during the biological sample processing steps are still in urgent need for diagnostic metagenomics[16].

Contemporary human DNA depletion techniques

A wide range of human DNA depletion methods can be categorized into 2 groups: prior DNA extraction and post DNA extraction.

1) Prior DNA extraction: these assays are widely used in many studies.

Physical approaches

They include centrifugation and filtration which remove DNA and cell components originated from patients based on their larger mass and size compared to microbes. Earlier studies show that these pre-steps were usually combined and applied to positive blood culture, blood, urine, sonication fluids from orthopaedic devices[8, 18-25] and could remove up to 89% human DNA (by centrifugation alone)[18]. However, they also reported that the human proportion still accounted for > 90% of total reads[24-26].

Chemical/enzymatic approaches

One common technique is to employ a nonionic surfactant like saponin to lyse human cells, then deplete human DNA using DNase treatment. The procedure could decrease up to 99.9% or approximately (approx.) 10^5 folds human DNA content in blood, respiratory samples, and cerebrospinal fluid[27, 28]. A previous study also indicated an impressive enrichment of pathogen DNA with a high pathogen-to-host ratio, approx. 30- to 100-fold[28]. The hands-on time for this technique took approximately 40 to 50 mins.

Osmotic lysis of mammalian cells followed by treatment with propidium monoazide (lyPMA) can also be utilized, this tactic remarkably depleted host reads to 8.53% of total reads while increasing microbial reads by ten times in human saliva samples. Moreover, lyPMA treatment requires fewer washing steps and less handling time and has a competitive reagent cost (~ 0.15$/sample) compared to other enzymatic alternatives[26]. However, PMA treatment seems to create a shift of viable community structure because of its impact on extracellular bacterial DNA[29].

Commercial nuclease-based kits are also available such as MolYsis® kits (Molzym, Bremen, Germany), QIAamp DNA Microbiome Kit (Qiagen, Hilden, Germany). MolYsis® kits differentially lyse host cells from human/animal body fluid samples and deplete host DNA using MolDNase. Previous investigations demonstrated that this method decreased the average proportions of human-originated reads in saliva, sputum, respiratory samples to ~62.88%[26], ~ 87%[29], < 10%[30], respectively, and significantly increased the microbial reads. However, under the chaotropic condition, not only human cells but some bacteria are likely to lyse as well, thus the method could affect the microbial composition[31]. The suggested protocol of MolYsis® takes at least 40 mins for depleting host DNA.

Similar mechanism is used by QIAamp DNA Microbiome Kit (Qiagen, Hilden, Germany) to lyse host cells from swab and body fluid samples and eliminate host DNA by Benzonase. The manufacturer declares that this kit could reduce human reads to <5% of total reads, while other studies for metagenomics in saliva and nasal samples reported an average human reads of 29.17%[26] and 57%[32] in treated samples, respectively. The performance time for a typical depletion procedure is approximately 100 mins.

Other commercial alternatives seem to be less popular than the above items such as HostZERO microbial DNA kit (Zymo Research, California, USA) to selectively degrade host cells and their DNA using the Host Depletion Solution and HetaSep® (STEMCELL Technologies Inc., Cambridge, MA) to isolate nucleated cells in the blood.

The main limitation of these pre-extraction depletion methods is the removal of intracellular viral materials[33] and the loss of genetic materials during DNA digestion and transferring between many tubes, which could lead to a serious effect on the microbial community. In addition, turnaround time is another challenge because most of these methods require additional time for incubation and washing steps (ranging from ~40 minutes to 1.6 hours).

2) Post DNA extraction:

Employing methylation-dependent restriction enzyme (MDRE)

Based on the characteristics of MDRE, they are used to cut human DNA at a specific site, and the digested DNA was separated by gel electrophoresis or magnetic bead-based size selection. Oyola *et al.* indicated that this treatment eliminated 80% of human DNA and enriched microbial reads from 10% to ~92%[34]. However, the incubation time for optimal enzyme activity is up to 16 hours at 37°C.

Targeting CpG domains

Some commercial kits utilize the different abundances of methyl-CpG domains between mammalian cells and microbes to deplete host DNA.

The Pureprover kit (previously available from SIRS-Lab GmbH, Jena, Germany) extracts whole genomic DNA from clinical sample, then uses a protein to bind non-methylated CpG sites in bacterial genomes. The removal of human DNA using this treatment was up to 90%[35]. The recovery rate of microbial DNA may vary due to the different contents of CpG motifs in distinct microbes, thus the relative abundance of species in the sample could be changed[31].

The NEBNext Microbiome DNA Enrichment Kit (New England Biolabs Inc., Ipswich, MA, USA) separates methylated DNA (mainly in human DNA) by taking advantage of a human methyl-CpG binding domain fused to the Fc-tail of human IgG (MBD-Fc) known to specially bind methyl-CpG dinucleiotides[31]. The kit removed 94-96% of human reads and increased from 8 to 43-fold reads from microbes compared to un-depleted samples[31, 36]. Nevertheless, these results were not in line with other studies which indicated a significantly lower efficiency[24, 26, 29], that is maybe due to the requirement of high molecular weight genomic DNA input (> 15kb)[31]. The pre-extraction steps of these studies could affect the DNA fragments (such as homogenization, sonication, freeze-thawing process).

Overview of all currently available host DNA depletion methods is illustrated below (Fig. 1)

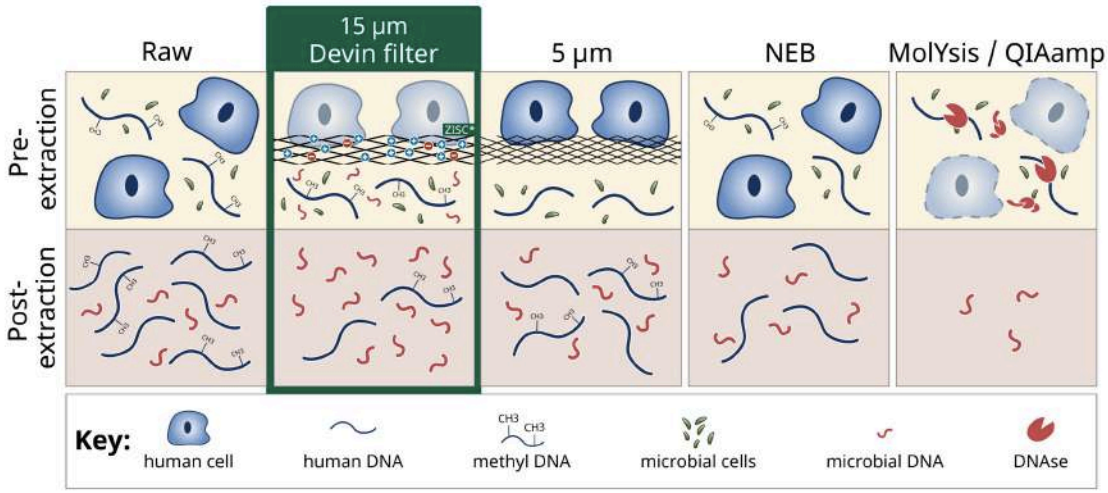


Fig. 1 Overview of contemporary depletion techniques

Host depletion kits and techniques are heterogeneous across studies. Most studies utilized commercial kits which are usually biased to some degree while non-commercial methods are quite hard to be validated. In brief, the above approaches have their pros and cons as described in Table 1. One of the major common drawbacks of these methods are labor-intensive lengthy workflow which is impossible to be practically applied to diagnostic metagenomic tests.

Type of approaches	Commercial kit(s)	Advantage(s)	Disadvantage(s)
Differential cell size-based methods	PluriStrainer®, CellTrics™, etc.	Cheap. Easy to perform	Significant quantities of human DNA residuals
Chemical/enzymatic treatments	MolYsis®, QIAamp, HostZERO, HetaSep®	Effectively deplete human cells and DNA	Time-consuming process. Washing and transfer steps may cause DNA loss.
Methylation-dependent restriction enzyme (MDRE)	N/a	Effective removal of methylated host DNA	Long enzyme incubation time
Non CpG-methylated microbial DNA	Pureprover	High quantity of human DNA depletion	Inconsistent microbial DNA recovery rate
CpG-methylated host DNA	NEBNext	Useful for unsheared high molecular weight DNA (> 15 kb fragments)	Expensive. Requirement of high molecular weight gDNA input

Table 1. Current approaches to deplete human DNA for metagenomics analysis

Micronbrane addresses the current bottleneck of human DNA depletion in clinical microbiology with ***Devin***® filter that can remove 95% leukocytes within just 5 minutes (fig. 2). ***Devin***®

membrane utilizes patented ZISC technology (Zwitterionic Interface Ultra-Self-assemble Coating Technology or ZISC) to specifically and efficiently bind and deplete white blood cells (WBCs) in the whole blood. Devin® filter fundamentally differ from the other filtration devices as it doesn't rely on the pore size of filter membrane to exclude white blood cells. In fact, the pore size of Devin® filter is 15-20 μm which is larger than most of the human cells. Thus, Devin® filter can allow red blood cells from whole blood pass through the filter without clogging it. The WBC depletion effect of Devin® filter for the various amount of whole blood is shown in Fig. 2.

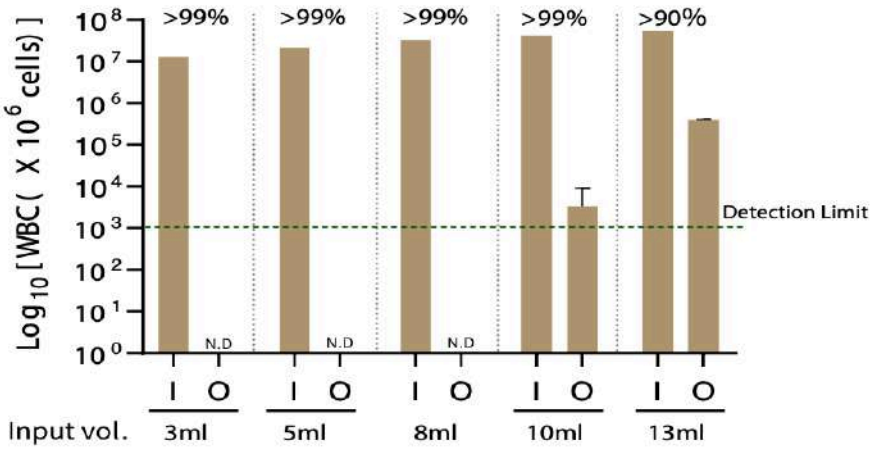


Fig. 2 Leukocytes reduction efficiency

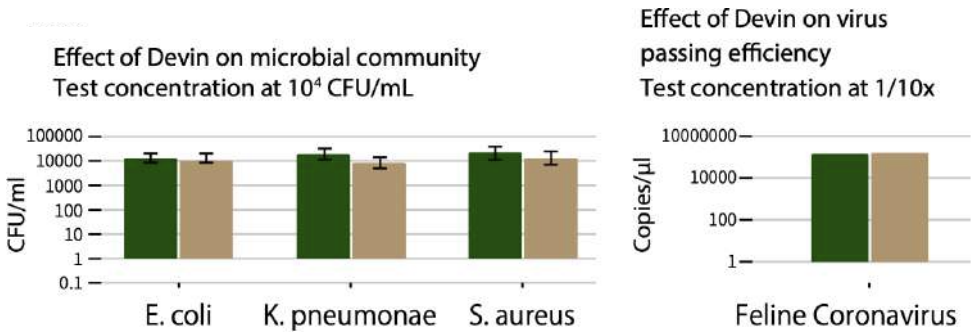


Fig. 3 Microbial passing efficiency

At the same time Devin® filter demonstrates high microbial passing efficiency including bacteria and viruses (Fig. 3)

The functional features of WBC depletion and microbial passing of Devin® filter makes it an ideal device for microbial enrichment for the biological fluids, especially whole blood to enable metagenomic test, comparing against other methods (Fig. 1). When Devin® filter was tested for simulated samples (the whole blood from healthy doners spiked with 10*4 Genome Copies/mL spike-in control from ZYMO Research), it not only takes less than 5 minutes, significantly faster than the other methods, but also shows much better enrichment results by various downstream detection methods including next generation sequencing (NGS), Nanopore sequencing as well as qPCR, as shown in Fig. 4.

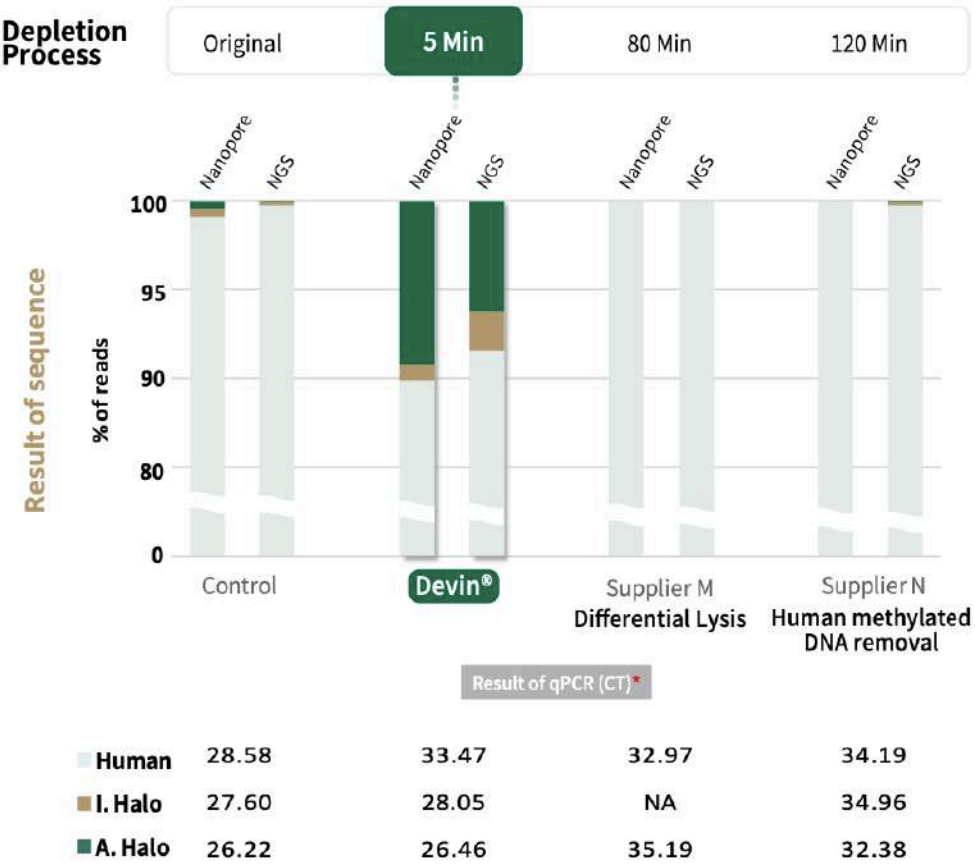


Fig 4. Results of human and microbial identification using qPCR, NGS and Nanopore*

*Human blood samples (5mL) were added 10*4 Genome Copies/mL spike-in control from ZYMO Research and then processed using different depletion methods. Test results show that Devin® filter increases ratio of microbial DNA and decreases the ratio of human (host) DNA.

PaRTI-Seq® test, a metagenomic sequencing workflow built upon Devin® filter as illustrated in Fig. 5, can demonstrated ultimate applicability to detect potential pathogens from blood samples within 24 hours with a sensitivity of 102 genome copies/mL.

Devin® together with PaRTI-Seq® delivers complete solution of rapid pathogen detection within 24 hours and introduce a promising clinical solution in future to be used in routine for rapid infectious disease diagnosis.

More details on Micronbrane's innovational technology on the website: www.Micronbrane.com

References

1. Zhou Y, Wylie KM, Feghaly REE, et al. Metagenomic Approach for Identification of the Pathogens Associated with Diarrhea in Stool Specimens. *Journal of Clinical Microbiology*. 2016;54(2):368-375. doi:doi:10.1128/JCM.01965-15
2. Wylie KM, Truty RM, Sharpton TJ, et al. Novel bacterial taxa in the human microbiome. *PLoS One*. 2012;7(6):e35294. doi:10.1371/journal.pone.0035294
3. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. Metagenomics for pathogen detection in public health. *Genome Medicine*. 2013/09/20 2013;5(9):81. doi:10.1186/gm485
4. Govender KN, Street TL, Sanderson ND, Eyre DW. Metagenomic sequencing as a clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis. *medRxiv*. 2020:2020.03.30.20043901. doi:10.1101/2020.03.30.20043901
5. Blauwkamp TA, Thair S, Rosen MJ, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nature Microbiology*. 2019/04/01 2019;4(4):663-674. doi:10.1038/s41564-018-0349-6
6. O'Grady J. A powerful, non-invasive test to rule out infection. *Nature Microbiology*. 2019/04/01 2019;4(4):554-555. doi:10.1038/s41564-019-0424-7
7. MacIntyre AT, Hirst A, Duttagupta R, Hollemon D, Hong DK, Blauwkamp TA. Budget Impact of Microbial Cell-Free DNA Testing Using the Karius® Test as an Alternative to Invasive Procedures in Immunocompromised Patients with Suspected Invasive Fungal Infections. *Applied Health Economics and Health Policy*. 2021/03/01 2021;19(2):231-241. doi:10.1007/s40258-020-00611-7
8. Grumaz S, Stevens P, Grumaz C, et al. Next-generation sequencing diagnostics of bacteremia in septic patients. *Genome Med*. Jul 1 2016;8(1):73. doi:10.1186/s13073-016-0326-8
9. Kowarsky M, Camunas-Soler J, Kertesz M, et al. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. *Proc Natl Acad Sci U S A*. 2017;114(36):9623-9628. doi:10.1073/pnas.1707009114
10. Hong DK, Blauwkamp TA, Kertesz M, Bercovici S, Truong C, Banaei N. Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. *Diagnostic Microbiology and Infectious Disease*. 2018/11/01/ 2018;92(3):210-213. doi:https://doi.org/10.1016/j.diagmicrobio.2018.06.009
11. Chiu K-P, Yu AL. Application of cell-free DNA sequencing in characterization of bloodborne microbes and the study of microbe-disease interactions. *PeerJ*. 2019;7:e7426-e7426. doi:10.7717/peerj.7426
12. Barrett SLR, Holmes EA, Long DR, et al. Cell free DNA from respiratory pathogens is detectable in the blood plasma of Cystic Fibrosis patients. *Scientific Reports*. 2020/04/23 2020;10(1):6903. doi:10.1038/s41598-020-63970-0
13. Loman NJ, Constantinidou C, Christner M, et al. A Culture-Independent Sequence-Based Metagenomics Approach to the Investigation of an Outbreak of Shiga-Toxigenic Escherichia coli O104:H4. *JAMA*. 2013;309(14):1502-1510. doi:10.1001/jama.2013.3231
14. Schlaberg R, Chiu CY, Miller S, et al. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. *Archives of Pathology & Laboratory Medicine*. 2017;141(6):776-786. doi:10.5858/arpa.2016-0539-RA
15. Pereira-Marques J, Hout A, Ferreira RM, et al. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Original Research. *Frontiers in Microbiology*. 2019-June-12 2019;10(1277)doi:10.3389/fmicb.2019.01277
16. McArdle AJ, Kaforou M. Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue. *Access Microbiology*. 2020;2(4)doi:https://doi.org/10.1099/acmi.0.000104
17. Wilson MR, Sample HA, Zorn KC, et al. Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis. *New England Journal of Medicine*. 2019;380(24):2327-2340. doi:10.1056/NEJMoa1803396
18. Anson LW, Chau K, Sanderson N, et al. DNA extraction from primary liquid blood cultures for bloodstream infection diagnosis using whole genome sequencing. *J Med Microbiol*. 2018;67(3):347-357. doi:10.1099/jmm.0.000664
19. McNaughton AL, Roberts HE, Bonsall D, et al. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). *Scientific Reports*. 2019/05/08 2019;9(1):7081. doi:10.1038/s41598-019-43524-9
20. Hasman H, Saputra D, Sicheritz-Ponten T, et al. Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. *Journal of Clinical Microbiology*. 2014;52(1):139-146. doi:doi:10.1128/JCM.02452-13
21. Delwart EL. Viral metagenomics. *Reviews in medical virology*. Mar-Apr 2007;17(2):115-31. doi:10.1002/rmv.532
22. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. Laboratory procedures to generate viral metagenomes. *Nature protocols*. 2009;4(4):470-83. doi:10.1038/nprot.2009.10
23. Horiba K, Kawada J-i, Okuno Y, et al. Comprehensive detection of pathogens in immunocompromised children with bloodstream infections by next-generation sequencing. *Scientific Reports*. 2018/02/28 2018;8(1):3784. doi:10.1038/s41598-018-22133-y
24. Street TL, Sanderson ND, Atkins BL, et al. Molecular Diagnosis of Orthopedic-Device-Related Infection Directly from Sonication Fluid by Metagenomic Sequencing. *Journal of Clinical Microbiology*. 2017;55(8):2334-2347. doi:doi:10.1128/JCM.00462-17
25. Sanderson ND, Street TL, Foster D, et al. Real-time analysis of nanopore-based metagenomic sequencing from infected orthopaedic devices. *BMC Genomics*. 2018/09/27 2018;19(1):714. doi:10.1186/s12864-018-5094-y
26. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*. 2018/02/27 2018;6(1):42. doi:10.1186/s40168-018-0426-3
27. Charalampous T, Richardson H, Kay GL, et al. Rapid Diagnosis of Lower Respiratory Infection using Nanopore-based Clinical Metagenomics. *bioRxiv*. 2018:387548. doi:10.1101/387548
28. Hasan MR, Rawat A, Tang P, et al. Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen Detection by Next-Generation Sequencing. *Journal of Clinical Microbiology*. 2016;54(4):919-927. doi:doi:10.1128/JCM.03050-15
29. Nelson MT, Pope CE, Marsh RL, et al. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Reports*. 2019;26(8):2227-2240.e5. doi:10.1016/j.celrep.2019.01.091
30. Votintseva AA, Bradley P, Pankhurst L, et al. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. *Journal of Clinical Microbiology*. 2017;55(5):1285-1298. doi:doi:10.1128/JCM.02483-16
31. Feehery GR, Yigit E, Oyola SO, et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One*. 2013;8(10):e76096-e76096. doi:10.1371/journal.pone.0076096
32. Bjerre RD, Hugerth LW, Boulund F, Seifert M, Johansen JD, Engstrand L. Effects of sampling strategy and DNA extraction on human skin microbiome investigations. *Scientific Reports*. 2019/11/21 2019;9(1):17287. doi:10.1038/s41598-019-53599-z
33. López-Labrador FX, Brown JR, Fischer N, et al. Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure. *Journal of Clinical Virology*. 2021/01/01/ 2021;134:104691. doi:https://doi.org/10.1016/j.jcv.2020.104691
34. Oyola SO, Gu Y, Manske M, et al. Efficient Depletion of Host DNA Contamination in Malaria Clinical Sequencing. *Journal of Clinical Microbiology*. 2013;51(3):745-751. doi:doi:10.1128/JCM.02507-12
35. Horz HP, Scheer S, Vianna ME, Conrads G. New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe*. Feb 2010;16(1):47-53. doi:10.1016/j.anaerobe.2009.04.009
36. Graham RMA, Doyle CJ, Jennison AV. Epidemiological typing of Neisseria gonorrhoeae and detection of markers associated with antimicrobial resistance directly from urine samples using next generation sequencing. *Sexually Transmitted Infections*. 2017;93(1):65-67. doi:10.1136/sextrans-2015-052422

Appendix 1. See how PaRTI-Seq® can reduce your sequencing cost

Imagine that each circle represents relative cost of mNGS pathogen detection test for ONE patient

Depending on Sequencing Platform, total cost of the test will be different. If you sequence with HiSeq it will be the lowest cost, but will require to run samples for 200 patients at the same time, not to mention extremely high capital cost of platform itself. With ~18 samples in a run, the cost will be the same as running the single sample on smaller platforms like MiSeq.

Smaller sequencing solutions make it possible to test just sample of 1 patient but sequencing cost in that case is still extremely high

Regardless of sequencing method each sample requires about 20 million reads as output

DNA extraction using Devin filter® allows to reduce standard output required for mNGS tests by 4 times (from 20M reads to 5M) which saves cost of sequencing in clinical setting more than a half

cost per sample $

Sample Prep & DNA Extraction

Library prep

Sequencing

1 sample = 1 patient

Analysis & Report

Capital cost

Other mNGS tests

×200

×18

×1

This is how many patients you need to test at once

$$$$$$$$$

$$$$$$$$$

$$$

PaRTI-Seq®

Devin Filter +

×4

3/4 of mNGS cost

$$$