# Metagenomic sequencing in clinical virology: advances in pathogen detection and future prospects

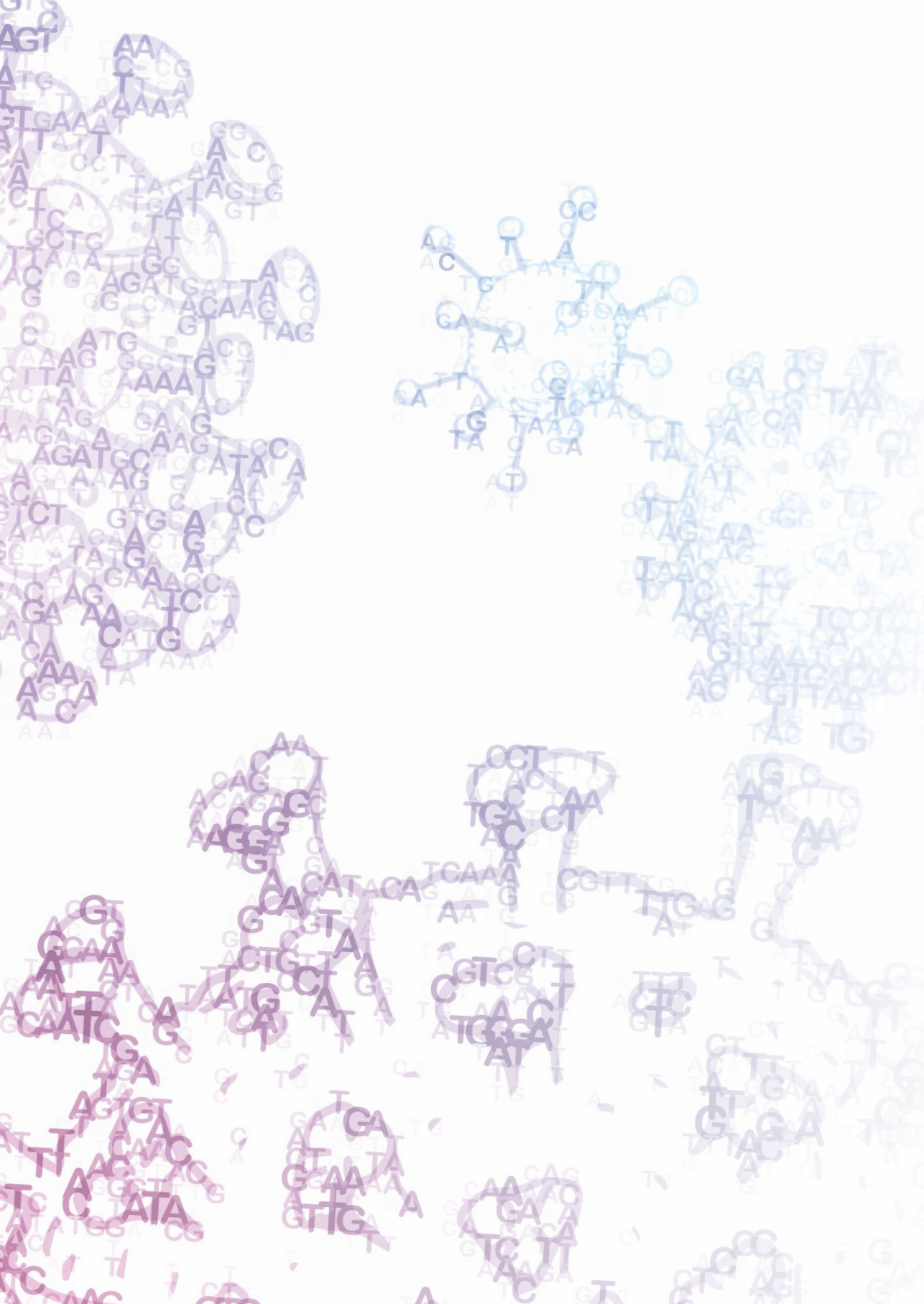
Carbo, E.C.

*The human mind treats a new idea*

*the same way the body treats*

*a strange protein; it rejects it.*

Peter Medawar, in “The Art of the Soluble”, 1960



# Chapter 9 **General Discussion**

**Viral metagenomic next-generation sequencing (mNGS), an approach to potentially identify all viral genomes in a sample at once, is a promising contribution to the current virus diagnostic repertoire in modern health care. With already more than 1,000 virus species known to be able to infect humans [1], a densely populated civilization and a constant threat of zoonotic infections [2], it is a worthwhile addition to the current methods in which either one virus is tested (traditional PCR test), or a limited number of viruses when combined PCR tests are used. This discussion will initially focus on the applications, diagnostic yield and potential of viral metagenomic sequencing. Further, mNGS diagnostic test accuracy advancement, both within the wet laboratory and using bioinformatics, will then be discussed. Additionally, in-depth advances in the genetics analysis of whole genome sequencing of a single-virus genome will be explained. An in-depth view on the limitations of metagenomic sequencing and an outlook on the future of molecular diagnostics will be presented in the last two sections.**

# Implementing viral metagenomic sequencing

## Viral metagenomics improves diagnostic yield

With various viruses that can infect humans and many undiagnosed cases [3-7], implementing metagenomic sequencing in a clinical setting will potentially lead to the identification of more viruses and an increased number of patients diagnosed with a viral infection. One of the aims of the research of this thesis was to assess the improved diagnostic yield using metagenomics: the proportion of additional potential pathogenic viruses that can be found after initial testing remained negative. In **chapter 2**, a systematic review and meta-analysis was conducted and an additional 10.88% (95% CI 4.6-17.15%) of viruses were detected that were not identified by traditional diagnostic testing in patients suffering from meningoencephalitis [8-17]. A selection of reports on patients from (sub)tropical climate regions revealed an additional diagnostic yield of 21.61% (95% CI 12.16-31.07%) partially since the initial test spectrum was more limited, the decreased vaccine administration in this region, and an increased risk of mosquito born viral diseases that are more frequent in (sub) tropical climates. In **chapter 3**, a cohort of hematologic patients suffering from encephalitis was tested and a corresponding additional diagnostic yield of 12.2% (95% CI 2.2-22.2%) was observed.

In **chapter 4**, patient sera were tested from a cohort of international travellers returning with febrile illness, resulting in 6.3% (95% CI -2.4-17.2%) of cases where additional pathogenic viruses were detected. This number seems comparable to the result of a similar study on travellers with febrile illness where in three out of 40 patients (7.5%) extra pathogenic viruses were detected based on mNGS results [18].

Longitudinal testing of transplantation patients by means of metagenomic sequencing is present in **chapter 6**. In this study, BKV, CMV, and HHV6B were additionally detected by mNGS in three out of six patients (50%), and all additional findings were confirmed either by qPCR or supported by auxiliary bioinformatic analysis.

A systematic review and meta-analysis presented in **chapter 1** showed a relatively high number of additional viral findings - 28.73% (CI [19.80, 37.63]) - when assessing studies of diverse patient types that were negative during initial testing and mNGS was used as a second step approach. In the research of this thesis, additional findings were found in 6.3% (95% CI [–2.4, 17.2]) of returning travellers with febrile

illness. Two prospective papers describing metagenomics in a clinical setting identified 13 out of 58 central nervous system infections by means of metagenomics that were not found by PCR (22%) [10], and an additional 24 (23%) pathogenic virus infections in 105 patients in a tertiary diagnostic unit [11]. The research of this thesis and available literature show that the use of metagenomics as a second step approach when initial testing is negative improves the diagnostic yield. This accounts for patients suffering from encephalitis where metagenomics can detect a neuroinvasive pathogen [10], for travellers returning with febrile illness, and for immunocompromised patients where unexpected viruses can be detected.

## Viral capture probes increase diagnostic test sensitivity

Most previously published studies focused on metagenomics and the test accuracy for the detection of bacteria, with a significant knowledge gap concerning viruses. Two systematic review studies have been published on the overall test performance, with one focusing on metagenomic sequencing for all pathogens including studies prior to August 2020 (note: including papers published before this date) [19], and one focusing on lower respiratory tract infections [20]. A combined overview of two papers focusing (partly) on viruses is shown in Table 1. Wilson et al. [10] showed a relatively low sensitivity of 0.55; however, when looking more into detail in the virus diagnoses missed by mNGS, most of these were found positive in IgM by serology testing while when followed up by qPCR testing these also remained negative. Only two out of 204 results were positive by means of qPCR due to low pathogen titers [10]. In the manuscript by Parize et al., a single viral pathogen was undetected by means of mNGS, attributable to the different sample type that was used: a sample positive human cytomegalovirus (CMV) was identified in whole blood, and for mNGS only plasma was used. When testing the plasma by means of qPCR, CMV was not detected, as CMV probably was residing in leukocytes and not accessible for amplification. Amending this finding would lead to a sensitivity of 100% in this particular study when incorporating only virus data [21]. In the study by Hong et al., a sensitivity of 0.74 was found; however, the portion of mNGS samples resulting negative were found positive only by serological testing [22].

Viral pathogens originally detected by means of PCR were confirmed by viral metagenomics as described in **chapter 3 and 4**, due to the usage of a more sensitive, capture probe-based enrichment, instead of solely performing shotgun metagenomics. In **chapter 6,** a 100% sensitivity was indicated; all initial positive qPCR results were positive by mNGS. Collectively, the majority of published studies and our findings illustrate the high sensitivity of mNGS to identify viruses in samples.

The results of an extensive comparison of shotgun metagenomics with metagenomics using viral capture probes are described in **chapter 3**. Data showed that with shotgun metagenomics several pathogens were marked as false negative, after having a positive diagnostic PCR result. In contrast, metagenomics with viral capture probes performed in a much more sensitive manner, with 1,283-38,749,926 sequence reads per pathogenic virus was found positive by means of PCR. The viral capture probe metagenomic method not only resulted in a sensitivity of 100%, but yielded 100-10,000-fold more sequence reads compared to shotgun metagenomics. An overview of technical aspects of protocols of the few European centres that offer viral metagenomics in a clinical setting is presented in **chapter 2.** It shows that these diagnostic laboratories offering viral metagenomics services are aiming at increased sensitivity by either using viral metagenomic probes, or by performing shotgun metagenomics in parallel for both DNA-based and RNA-based organisms. In **chapter 4,** travellers returning with febrile illness were tested by viral capture probe metagenomics, and all earlier positive PCR test results were confirmed resulting in a sensitivity of 100% of the mNGS method. Transplantation patients that were longitudinally sampled and sequenced using mNGS had positive mNGS results for the viruses that initially tested positive by means of qPCR (cytomegalovirus (CMV), Epstein-Barr virus (EBV), BK polyomavirus (BKV), adenovirus (ADV), parvovirus B19 (B19V), and torque teno-virus (TTV)), resulting in a sensitivity of 100%, as it is shown in **chapter 6**.

## Amino acid-based taxonomic classifying tools perform the most accurate

Taxonomic classifiers for virus identification are widely available and use different underlying algorithms [34,35]. A ring trial in Switzerland [36] reported that the chosen algorithms influenced the overall performance of mNGS, more than the chosen reference databases. Only a limited number of studies report benchmarking 'dry lab' protocols, despite bioinformatic protocol validation being equally important to wet laboratory validation for accurate performance. Many tools were specifically designed for bacterial detection – such as Kraken [37] and CLARK [38] – and it is especially important to validate these tools for virus identification prior to use for that aim. The limited amount of benchmark publications have focused more on bacterial analysis [39-45], mostly only performing in silico analysis of artificial sequence data [39,46,47], or NGS data for mock samples that are typically less diverse compared to real clinical samples [39,48].

Bioinformatic taxonomic classifiers were benchmarked, as described in **chapter 5**. Up to a billion sequence reads of 88 respiratory samples were used for benchmarking of five classifiers for performance based on results of 1,144 PCR tests used as the gold standard. Sensitivity and specificity of the classifiers tested ranged from 83% to 100% and 90% to 99%, respectively, and was dependent on the classification level and data pre-processing. The bioinformatic tool reaching the highest sensitivity was the Kaiju tool [40] with k-mer classification based on amino acids. Exclusion of human reads generally resulted in increased specificity. Normalization of read counts for genome length resulted in a minor effect on overall performance, however it negatively affected the detection of targets with read counts around detection level.

In a benchmark of the European network of next-generation sequencing [49], datasets from real clinical metagenomic samples (tested positive for viral pathogens) were distributed to thirteen collaborating centres. The optimal performing tool, both for sensitivity and specificity was the MetaMix classification tool [49,50]. This tool, like Kaiju [40] performing the most optimal in **chapter 5**, is based on amino acid identification which, due to lower mutation rates of amino acid compared to DNA/RNA, results in a higher sensitivity, mainly for highly divergent viruses [39,40]. To distinguish contamination from real clinical findings and to further enhance specificity, respectively, tools for removal of sequences detected in negative control samples can be used [51,52], and extra mapping/alignment steps can be added to assess the distribution of sequence reads over the viral genome.

**Table 1. Overview of sensitivity and specificity from reports on (viral) metagenomics.**

| Study | Type of sample | Sequencing technique | Gold standard | Sensitivity | Specificity |
|---|---|---|---|---|---|
| Hong et al. [22] | Cerebrospinal fluid | Illumina MiSeq | PCR | 0.74 | 0.66 |
| Miller et al. [23] | Cerebrospinal fluid | Illumina HiSeq | Conventional laboratory results and additional molecular testing | 0.89 | 0.99 |
| Wilson et al. [10] | Cerebrospinal fluid | Illumina HiSeq | Conventional laboratory results and additional molecular testing | 0.55 Higher for only viruses | 0.98 |
| Blauwkamp et al. [24] | Plasma (cfDNA) | Illumina NextSeq 500 | Conventional laboratory results and additional molecular testing | 0.93 | 0.63 |
| Parize et al. [21] | Plasma | Ion Proton | Culture, serological diagnosis and PCR | 0.63 1.0 (virus only) | 0.71 |
| Somasekar et al. [25] | Serum | Illumina HiSeq | PCR | 0.96 | 1 |
| Rossoff et al. [26] | Plasma | Illumina NextSeq 500 | Clinical review | 0.92 | 0.64 |
| Schlaberg et al. [27] | Respiratory | Illumina HiSeq 2500 | Culture, serological diagnosis and PCR | 0.90 | 0.64 |
| Doan et al. [28] | Intraocular fluid | Illumina HiSeq 4000 | PCR | 0.87 | 0.78 |
| Langelier et al. [29] | TA | Illumina HiSeq 4000 | Clinical microbiologic testing | 1.00 | 0.88 |
| Wang et al. [30] | Pulmonary biopsy and BALFs | NA | Conventional tests | 0.97 | 0.63 |
| Van Rijn et al. [31] | Nasopharyngeal samples | Illumina NextSeq 500 | PCR | 0.96 | 0.98 |
| Huang et al. [32] | Lung tissue, BALF, and PSB | BGISEQ-100 | Culture, microscopic examination | 0.88 | 0.81 |
| Van Boheemen et al. [33] | Nasopharyngeal washings, sputa, BALF, bronchial washing and throat swab | Illumina HiSeq 4000 and NextSeq 500 | PCR | 0.83 | 0.94 |

Adapted table of data of two papers focusing on test accuracy and only including papers focusing on viruses or when more than >1 virus found. [19,20]

Abbreviations; BALF, broncho-alveolar lavage fluid; NA, not applicable; PSB, protected specimen brushes; TA, tracheal aspirate. cfDNA, cell free DNA. Clinical review indicates that an organism was classified as clinically relevant by a treating physician, and if unclear was determined by a 2nd paediatric infectious disease (ID) physician, finally relying on the opinion of a third physician in case of discrepant opinions.

## Further advantages of metagenomics

A characteristic advantage of metagenomics is that it is a pathogen-agnostic test (Figure 1). No specific pathogen needs to be expected in contrast to a PCR test, or a multiplex of PCR tests. Additionally, mutations occurring in evolving viruses in the primer target regions lead to a false-negative PCR test result whereas viral metagenomic diagnostics would potentially pick up viruses with mutations. In addition, the host transcriptome can be interpreted straight from sequence data after certain shotgun metagenomic protocols.

Metagenomic sequencing results in information about the nucleotide sequences of virus species presented in a given sample, and these sequences can be used for typing and for phylogenetic analyses for these viruses. In **chapter 4,** subsequent typing of viruses detected in the serum samples of travellers resulted in characterization of serotypes and genotypes of the detected viruses, and enabled phylogenetic analysis of the Dengue viruses detected in the serum samples of travellers directly from the metagenomic test results. These results illustrate that viral metagenomic analysis is not only suitable in the detection of extra viruses, but additionally, viruses can be correctly typed, further aiding phylogenetic analysis. Once the nucleotide sequences are established, this information can be used for finding resistance mutations as well.

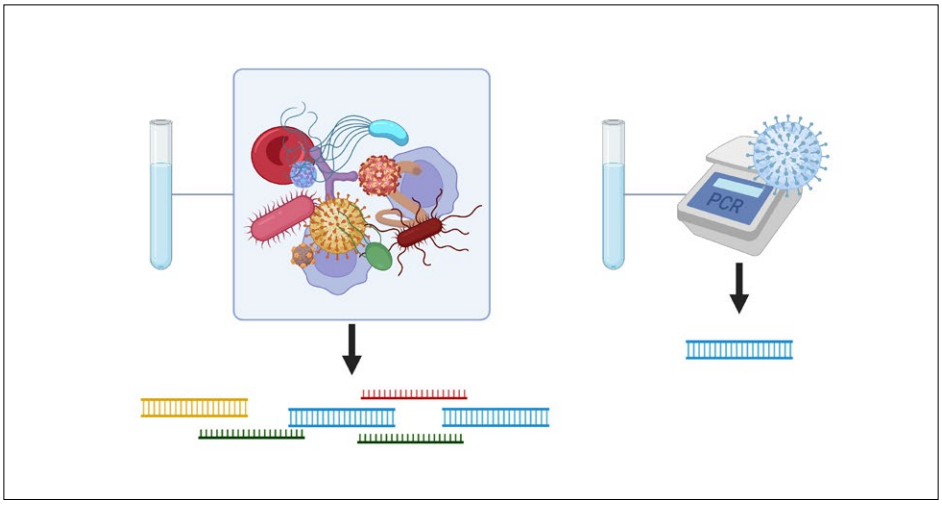



**Figure 1. Pathogen-agnostic and unbiased testing using metagenomics versus PCR, testing only known pathogens.**

Metagenomic sequencing giving information about all species present in a sample, versus PCR where one or a handful known viruses are tested on being present in a sample. Created using Biorender.com.

## Quantification by means of metagenomic sequencing

Quantification of viral load is possible based on metagenomics data. The precision of this greatly depends upon correct classification of the viral pathogen. After sequencing of clinical samples positive for various viruses, normalizing the sequence reads for total read count and genome length, a quantitative correlation between qPCR and metagenomic target reads was found of 62.7% (**chapter 5**). The coefficient of determination varied per bioinformatic tool, data pre-processing, and per virus, and $R^2$ ranged 15.1-63.4%, with 63.4% scored by amino acid-based classifier. Divergent viruses such as rhinoviruses were the most challenging in assessing correlation of sequence reads with Ct-values. Only a limited number of rhinoviruses are present in the underlying RefSeq database and it could be that precision was decreased as a result, as previously observed in the study of Menzel et al. [40]. In **chapter 6**, longitudinal plasma samples from six patients and qPCR positive for transplantation-related DNA viruses were tested using mNGS in combination with calibration samples. Viral loads as determined based on mNGS results correlated with the qPCR results, with inter-method differences in viral loads per virus ranging from 0.19 log10 IU/mL for EBV to 0.90 log10 copies/mL for ADV. The patterns of viral loads of patients tracked over time based on the metagenomic classifying results resembled that of the loads established by means of qPCR. This was in line with a mNGS report using calibration samples, where identical challenges with torque teno virus (TTV) quantification are discussed as in our study since there was no calibration material available [53]. The results that this paper of Shah et al. describes further imply that viral metagenomic sequencing can be used in a quantitative manner where viral loads are identified straight from metagenomic sequence data [53].

## Discovery of viruses directly from clinical samples

Viral metagenomics played a major role in the discovery of SARS-CoV-2 and the characterization of the viral genome when there were several patients in Wuhan presenting with fever and respiratory failure, and screening routine respiratory pathogens for these patients gave negative results [54-56]. **Chapter 7** illustrates that metagenomic sequencing in a clinical setting can be successfully used for virus discovery directly from patient samples. Mimicking virus discovery, using only viruses present in databases from before the discovery of SARS-CoV, MERS-CoV and SARS-CoV-2, revealed that these viruses could be labelled as indicative for a novel coronavirus. Bioinformatic tools Centrifuge and Genome Detective [57] showed classification of reads to the closest relative of the emerging coronavirus. Contig genome assemblies with lengths ranging from 2,503 to 30,097 nucleotides, created out of the patient sequence data, could be linked with low nucleotide identity to coronaviruses present before the emerging virus

by means of BLAST [58]. These results validate discovery of these novel viruses direct from clinical respiratory samples. Capture probes designed before the emergence of a virus can aid positive discovery findings, supported by the mismatches that are allowed during capture enrichment, or the presence of many homologic regions in a known closely related virus, resulting in effective virus discovery as long as a virus from the same genus or family is present in the probe kit.

## Whole genome sequencing (WGS) of SARS-CoV-2 for surveillance

The genomic surveillance of SARS-CoV-2 is of great importance for monitoring and detection of variants of concern, and for developing diagnostic, therapeutic and preventative strategies [59-61]. The most sequenced pathogen for surveillance currently is SARS-CoV-2 and worldwide consensus sequences can be uploaded to GISAID [62] guiding phylogenetics of a given sample, not only in local test sets but additionally in relation to sequences from globally. This kind of surveillance is mostly performed via WGS of patient samples targeting one specific virus.

In **chapter 7**, sequencing of two SARS-CoV-2 genomes using both a shotgun and a viral metagenomic capture probe method is described, and an increase in genome coverage when using capture probes is demonstrated. Few comparisons have been published, though WGS comparisons are usually limited to a single type of sequencing principle [63-65] whereas only two benchmark studies dealt with cross-platform protocols [66-67]. However, these studies for the most part indicate that amplicon-based methods yield the highest genome coverage. A more extensive comparison including viral probe metagenomic sequencing and several amplicon-based WGS protocols designed for SARS-CoV-2 is shown in **chapter 8**. Amplicon-based WGS protocols gave an overall median genome coverage of 81.6-99.8% (samples with CT-values of 30 and lower), with custom primers for Oxford Nanopore Technology (ONT) performing the lowest, and Illumina Ampliseq protocol resulting in the highest coverage. Amplicon distribution signatures differed across methods, illustrating the need to acquire coverage statistics when interested in certain genes or domains. Phylogenetic clustering of consensus sequences were independent of the workflow used, though in some cases it resulted in clustering per method when using settings where gaps were masked.

The usage of viral metagenomic probes showed an 86.7% median genome coverage, demonstrating that this method can indeed be of aid for limited surveillance when no specific genome amplicon kits are yet available, for instance when concerning novel or emerging viruses.

# Challenges in viral metagenomics

## General limitations

One of the challenges of current viral metagenomics protocols is the required turnaround time and costs of the NGS technique. Even with the current decline in sequencing costs, metagenomic sequencing is still more expensive compared to testing with PCR. Additionally, whereas PCR can provide results in only less than an hour, metagenomic sequencing takes approximately 2-6 days, depending on the protocol. However, diagnostic departments are becoming less reluctant to use more expensive and time-consuming NGS methods since the pandemic presented them with a great necessity for WGS for surveillance, spending more money on a metagenomic test could save money in other health care departments [68], due to an increased diagnostic yield.

For implementation in clinical settings, standardization of protocol validation is limited, although first attempts for establishing standardized guidelines have been reported [34,69]. Another limitation of metagenomic sequencing is the impairment of the data due to the high abundance of host cell material, and the potential threat of contamination, although contamination can partially be controlled for by sequencing an environmental control.

The research described in this thesis does not include bacterial, fungal or any other pathogenic classification of microorganisms, therefore it presents an overview of viruses and lacks a broader perspective that yields a higher diagnostic potential when looking at all organisms at once. Another general characteristic to take into account is that metagenomic sequencing can lead to incidental findings, such as the hepatitis C virus finding in the cohort of travellers described in **chapter 4** of this thesis, and the HIV findings in a Swiss study [11]. Even though these findings may not always be clinically expected based on the patient's syndrome, when using metagenomic sequencing, clinicians should be aware that there is always a possibility of finding unexpected viral pathogens as bystander infections.

## Platform-specific sequence errors

The metagenomic sequencing in this thesis was performed using Illumina sequencing, a platform in which index hopping – the swapping of sample indexes leading to incorrect assignment of reads to a neighbouring sample – can occur. Other sequence platforms not impaired by index swapping were not evaluated in this thesis.

Though this effect can be limited by using dual indexing, adding unique barcodes at both ends of the sequencing reads [70]. Illumina platforms are also known to have a median error rate of 0.109% for the NovaSeq 6000, 0.429% for the NextSeq 500 and 0.613% for the MiniSeq, of which Novaseq6000 and MiniSeq were included in the WGS comparison in **chapter 8**. [71] These error rates might impair a correct establishment of nucleotides, potentially leading to incorrect typing and mutation calling. One of the platforms currently most suited to determine minor variants would be PacBio, resulting in reliable sequenced long reads enabling detection and phasing of variants that are only present in low percentages in a sample [72]. Additional platforms can be used as well, ideally in combination with unique molecular identifiers applied during the library preparation, resulting in a unique label per every single molecule and allowing for amplification error filtering in subsequent bioinformatic analyses [73]. Alternatively, other sequence protocols for labelling unique molecules can be used, for instance single molecule molecular inversion probes [74,75].

## Bioinformatics: always a challenge

The performance of metagenomic sequencing is greatly dependent on accurate data analyses after the sequence reads are obtained from the sequencer. Various tools and pipelines exist, though standardized validation formats are lacking. As mentioned above, the majority of tools for classification and assembly are initially built for other organisms than viruses, rendering validation specifically for viruses of the utmost importance. Benchmarks are scarce or based on in silico data sets or mock samples with low abundance of the background sequences [39,46-48]. Misclassification of human genome sequence reads has been reported for several taxonomic classifiers [39], which is in line with our findings (**chapter 5**). This is most likely due to the presence of human genomic host reads in microbial assemblies uploaded to reference databases [76,77]. Other species can also lead to inaccurate uploads to GenBank, for instance the Illumina control phage PhiX174 that is present in many uploaded assemblies [78,79]. This viral phage is often used as a control for Illumina sequence runs and not always completely filtered out of sequence data [80]. Database curation should be improved when the database is used for metagenomic analyses, ideally by admitting only iterative assemblies based on long reads and by applying automated scripts to control for host material and contamination since research has shown that over 2,000,000 entries in Genbank contain cross-kingdom contamination [81]. Despite the high number of virus genome sequences available publicly, the list is incomplete and many virus genomes, especially those of bacteriophages, need to be sequenced and assembled to be added to public databases. Lower numbers of reference genomes available for specific targets lead to decreased

sensitivity and specificity [40]. To expand databases, the viral dark matter needs to be identified. Viral dark matter is sequence data resembling viruses though currently not immediately identified by regular classifiers [82,83]. Further bioinformatic issues may arise from the fact that many microbial laboratories lack bioinformaticians or lack the access to a high-performance computing cluster. With several cloud- or web-based user-friendly software tools for viral metagenomic analysis [57,84-86], local removal of human host reads is required before uploading the data, as even with viral metagenomic target probes as with amplicon WGS protocols there are usually human reads present in a sample after sequencing (**chapter 8**).

# The future of viral metagenomic sequencing

## Viral metagenomic as an add-on test for difficult to diagnose cases

Applying viral metagenomic sequencing, as described in this thesis, resulted in additional viral pathogenic findings, from 6.3 and 10.88% in two of our own cohorts of patients to 28.73% (95% CI [19.80-37.63]) in a systematic review as described in the introduction. To identify causes of infections in, for instance, the 20-62% [87-89] of patients suspected for acute respiratory infection where no microbial agent is detected, or in the up to 63% of encephalitis patients that remain without a causal pathogen [3], viral metagenomics can aid as an add-on test to the current diagnostic repertoire of clinical testing. In the formal diagnostic algorithm of the Dutch Society of Medical Microbiology (NVMM) for the paediatric patients suffering from acute hepatitis in 2022, viral metagenomics is officially advised on biopsies (and plasma or feces) in cases where other results are inconclusive [90]. The use of viral metagenomics may be additionally justified in severely affected infectious patients where no causal pathogenic viral pathogen is found by traditional testing methods. Metagenomic sequencing currently is more expensive compared to traditional tests when including lab costs only; however, recurrent or sequential negative test results in the microbiology department lead to extra costs elsewhere in the health care system. A cost analysis performed on the detection of infectious diseases by mNGS in cases with pyrexia of unknown origin justified implementation of metagenomic sequencing minimally as a second line investigation [68].

Sensitivity rates of pathogen detection have been published of >83% and often >90% (Table 1), and the 100% sensitivity in our research (**chapter 3, 4 and 6**) using viral capture probes indicates that this technique is becoming a trustworthy method to begin to implement in diagnostics. With limits of detection between 10-1,000 copies/ml, viral pathogens do not need to be highly abundant in a patient sample to be detected. With the additional information that can be retrieved from the metagenomic sequencing data for typing, resistance, phylogenetic information and virus discovery, it provides extra information for antiviral treatment, outbreak monitoring and surveillance.

## Overcoming technical challenges

Sequencing costs constantly decline, and as of this year the sequencing of a full human genome is possible for $100 [91]. Workflows for WGS library preparation have been made faster by adding sequence adapters in a two-step amplification protocols. However, for shotgun libraries such advances still have to be developed, and these protocols currently take six hours hands-on time (**chapter 8**). Sequencing instruments are becoming much faster in sequencing: whereas the first NGS machines were running for days, Illumina NextSeq 500 now has a minimum runtime of 12 hours, MiSeq minimally four hours, and a recent paper shows that pathogens can be detected from the ONT platform in combination with real-time analysis 30-38 minutes after the start of the Minion sequencer for highly abundant pathogens [92,93]. Besides being fast, ONT sequencers are handheld devices that are relatively cheap for laboratories compared to the investment needed for other sequence platforms. The size is also compact making them more even suitable for remote locations and – in the future – for patient bedside sequencing.

A challenge in metagenomic sequencing is the background level of host sequence reads, and with proper enrichment for viruses or depletion of host material it is easier to detect a potential viral pathogen. Centrifugation, filtration, and DNase treatment have not proven to be effective in every case [32,34,35,95]. Ribosomal RNA depletion and poly-A tail enrichment is sometimes used, though the latter may lead to false negative results in detection of viruses in a non-replicative state or those that translate without poly-A tail [34]. Another comparison of human genome depletion methods has been performed in a microbiome study where selective lysis of cells and endonuclease digestion worked well, and where benzonase increased metagenomic sequencing coverage [95]. However, sizeable benchmarks of host depletion methods are lacking specifically for viral metagenomics.

## ‘Virome in a bottle’ as a validation sample

Other aspects currently lacking with regard to the implementation of viral mNGS are uniform metagenomic validation samples. Only benchmark samples with limitations, such as cultured mock samples with limited sample sizes, or only in silico samples, are available. However, widely available and uniform benchmark samples resembling backgrounds reflecting real patient samples and containing several viral pathogens with different established viral loads are needed. Such benchmark material, like the “Genome in a Bottle” samples [96] is available for clinical genetics and used for validation in clinical genetic laboratories around the world, would be of great benefit to the metagenomics community [96-98]. The Genome in a Bottle materials are reference samples sold as vials containing human DNA. These samples contain an entire human genome, and even a combination of three human genomes can be bought, of which every known SNP and indel is additionally available in several different file formats. This allows a true reference so that every single mutation found in the lab’s diagnostic process can be accurately checked. An initiative like creating a uniform ‘microbiome/virome in a bottle’ (Figure 2) is greatly needed in the microbiology field, also making benchmarks more comparable within the field.

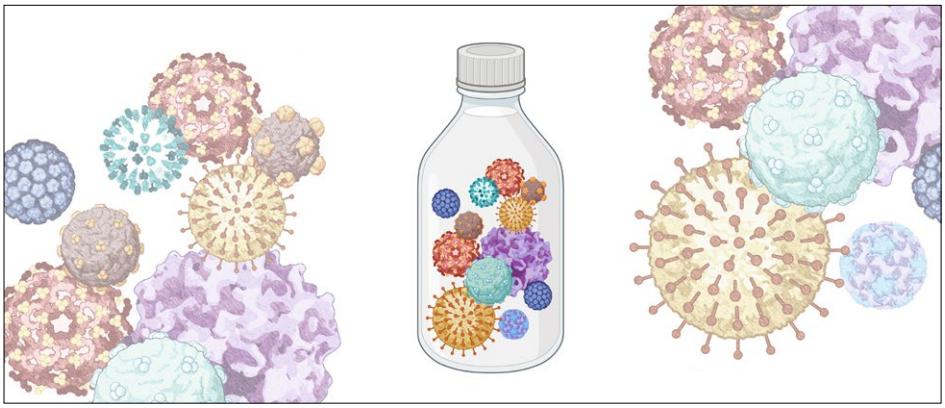


**Figure 2. ‘Virome in a bottle’.**

Example of a uniform mNGS benchmark sample containing different kinds of viruses. Created using Biorender.com.

## Solving computational challenges and solving challenges computationally

Concerning bioinformatics, the microbiology community would benefit from training a higher number of more specialized bioinformaticians or data analysts, not just to only work on microbiology, but also to educate them on FAIR principles [99], and

on privacy issues. Laboratory technicians could be trained in performing simple data analysis using a graphical user interface. Departments should not limit their focus on implementing the wet lab part of NGS, but they should additionally think about the hardware, and the costs of running and storing of analysis data. Long read sequencing will aid in distinguishing viral quasi-species, the same species of virus being present in a sample but with variating genomes due to high mutation rates. Bioinformatic tools like haplotype aware variant callers can aid this detection, though these are designed for human genomes and need to be benchmarked for detection of viral quasi-species. High quality quasi-species detection can additionally aid in tracking and surveillance of recombination in viruses [100].

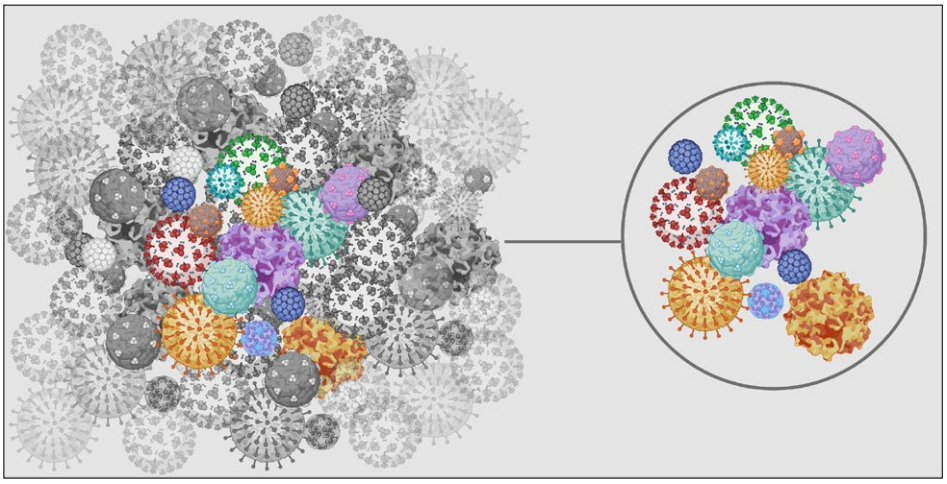


**Figure 3. 'Viral dark matter' and relatively few viruses identified.**

The coloured viruses represent the identified viruses, the dark grey coloured viruses represent the viruses that show resemblance to the viruses that we already are aware of and are present in our reference databases. The light grey faded viruses represent the viral dark matter, unidentified sequences that make op 40-90% of certain sample types [83]. Expanding databases and methods for detection are needed to further identify these. Created using Biorender.com.

## Viral dark matter

Viral dark matter (Figure 3), the uncharacterized sequences, should be explored more to expand public databases, since there are still many sequences resembling viral genetic material though sequences are not directly identified as a virus [82,83]. There is a lot about the microbial world that remains unknown. Creation of databases and tools to identify the unknown are needed, however, some of this information could already be supplied from existing data. We now have unravelled almost all species

and substances on land, and humans even explored deep oceans and even space, though a large part of the microorganisms right in front of us and in our bodies are still unidentified. In some samples, 40-90% of the sequences remain unidentified [83] and are thought to make up the viral dark matter, as these sequences resemble viral material though they do not match the reference sets. Detection of these viral dark matter sequences can be performing using a tool called VirSorter [101], detects viral signals based on viral protein resemblance in assembled contigs without sequence data directly matching a known virus family though not with a large resemblance percentage. Many sequences that resemble viruses that are already present in our reference databases are most likely undiscovered bacteriophages. The tools needed for deciphering this data are based on virus discovery tools, though an extensive benchmark for these discovery tools is currently lacking.

## Expanding virome databases

Investments would be beneficial to create reference databases of the virome of healthy and affected individuals in various sample types, and this can aid the differentiation between pathogenic and non-pathogenic viruses. This can be partly done based on sequence material that is publicly available and of which the raw data are shared within the science community. Using such a methodology, novel coronaviruses were recently found [102], and for blood a DNA virome [103] and cell free DNA virome [104] is already assembled, still leaving a variety of sample types to be explored. Another opportunity is to classify viral sequences of available data at the cancer genome atlas (TCGA) database [105] or in other public databases with sequence data of cancer patients to find associations between certain types of cancers and viruses [106,107].

## Artificial intelligence aiding viral health care

State-of-the-art artificial intelligence (AI) implementation in virology has greatly increased since the SARS-CoV-2 pandemic started. AI models are used for outbreak epidemiology: to provide information on the infection rate, number of cases, transmission dynamics and predicting the development and outcome of an outbreak [108,109]. Low-income countries lacking PCR data could even use mobile health technology [110] by applying AI on survey and sensor data from smart devices to predict the number of positive virus as was done for COVID-19 cases [111,112]. Extensive research has been performed on how AI can aid COVID-19 diagnosis, with a review reporting an accuracy as high as 70.00-99.92% in 46 studies on AI-assisted diagnosis. This included an accuracy of 74.4-95.20% on prognosis of critical COVID-19 patients [108]. AI can additionally be applied for developing therapeutics

and vaccines strategies. In the recently published review, an overview of eight studies using AI on COVID-19 is provided, mainly focusing on drug discovery or drug redirecting [108]. One study utilized reverse vaccinology and machine learning to find a vaccine for COVID-19 [108,113], while another study used the data of the GISAID [62] database to find vaccine targets [114]. These AI-assisted methods can be translated to potentially other viruses and outbreaks in the future. Likely, AI techniques may additionally be used for mutation prediction, further exploring and identifying viruses and viral dark matter, and perhaps to predict clinical outcome of pathogens present in metagenomic samples.

## Collaborations within the field of diagnostic microbiology

Within the field of microbiology, collaboration is needed with partners worldwide to bridge the gaps that currently exist due to insufficient virome databases, the lack of a suggested validation 'virome in a bottle' sample and to establish a standardized validation approach for NGS protocols. At the beginning of the SARS-CoV-2 pandemic, surveillance in the Netherlands was slightly behind when compared to the UK and Denmark, where a larger number of samples were sequenced compared to the proportion of cases. Currently, a relatively similar number of samples are sequenced in the Netherlands compared to the UK, Iceland, Denmark or Australia [61,115]. The organization of sequencing was scattered early in the pandemic: sequencing largely depended on local initiatives mainly organised by University Medical Centres. On one hand it was positive that these centres thrived in such a fast way, as it was needed to sequence extensively for surveillance. On the other hand, the efficiency of the implementation was questionable since centres were individually testing, optimising and validating a WGS SARS-CoV-2 lab protocol and creating a bioinformatic pipeline for analysis, while better collaboration could have saved time and effort.

## The metagenome aiding personalized medicine

Within hospitals, interdisciplinary laboratory departments can combine standardized approaches for isolating nucleotides and sequencing. This type of collaboration can also be applied in bioinformatics, since a variety of tools used in microbiology originate from the human genetics field. There is a variety of clinical information present in patients with samples when looking at the metagenome (Figure 4). One of the potential collaborations based on this approach would be with the pharmacy department for utilizing the pharmacogenetic data to predict an individual's drug response on both a pharmacokinetic level, for instance predicting the metabolising enzyme capacities of an individual, and pharmacodynamic level [116]. Currently, there are already collaborations between microbiology and pharmaceutics departments/

companies for the development of vaccine and anti-viral treatments, though pharmacogenetics is not implemented in daily clinical decision making. With sequencing more samples of patients with infectious diseases, the pharmacogenetics [117] should not be forgotten to facilitate future prescribed drugs in a diagnostic setting: joint use of NGS data can aid in not only selecting drugs targeting specific pathogens, but can additionally be based on genetic variations in the drug-metabolising enzyme genes of the human host for personalized medicine [117].

## A pathogen-agnostic test for both pathogens and pathogenic host gene variants

It would be useful to ascertain disease severity and investigate why certain people get more severely ill compared to others with joint insight from human genetics and metagenome sequencing. For instance, genetic variation in the ACE2 gene of the human genome is associated with disease severity in individuals with SARS-COV-2 infection [118-120], and most likely many other associations are yet to be found. Genome-wide sequencing revealed that immunity genes are under pathogen-imposed selection pressure [121], and some differences resulted from specific outbreaks like tuberculosis [122].

Since the course of infection can be influenced by (inherited) autoimmune or autoinflammatory disorders, it would be of great value to have the combined knowledge on both the infection of a patient and any hereditary disorder present interfering with the patient's immune system. Regarding metagenomics, it would be, for instance, beneficial when using a shotgun metagenomics approach for undiagnosed but suspected cases of encephalitis in young patients, to additionally diagnostically screen for genetic pathogenic variants for immunological disorders. Coexistence between autoimmune encephalitis and other systemic auto immune diseases have been previously described [123]. This type of dual application of metagenomics can be used in various kinds of infectious diseases and sample types. Research has already shown that using NGS in severely diseased infants for detection of hereditary mutations will lead to lower morbidity and mortality [124-126]. A more expanded approach, taking into account both the genome of the individual and the metagenomic sequence data, could in the future also lead to similar reduction of morbidity and mortality.

The field of clinical genetics is leaning to a genome-first approach, where genetic variants of interests are agnostically linked to the associated phenotype. Metagenomic sequencing as a combined agnostic test for both pathogens and

pathogenic host variants could be the next step in molecular diagnostics (Figure 4). In the future, the wide availability of shotgun metagenomic sequence data of many different sample types and locations that are tested can be of help to the clinical genetics field as well. This is particularly relevant for mosaic mutations. Mosaic mutations are present in very minor fractions in whole blood, but fully penetrant in certain body parts. These mutations can perhaps be earlier detected when testing different sample types from different locations, instead of only testing DNA isolated from whole blood samples, the common utilized sample type in clinical genetics research. [74,75].

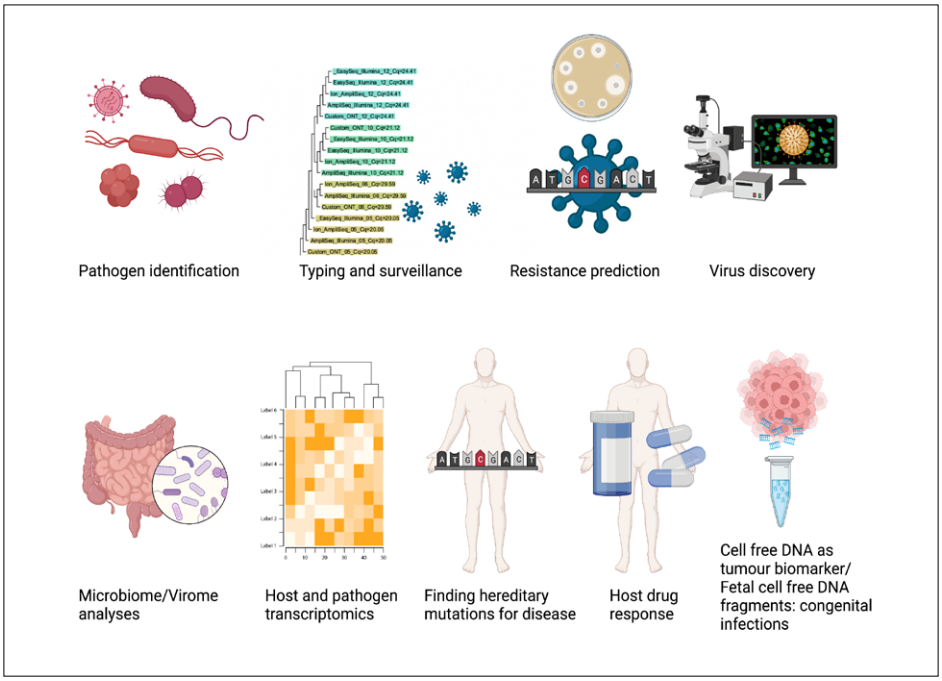



**Figure 4. One sequence combination test for all departments: metagenomic sequencing as a potential combined clinical application.**

Sequencing the complete metagenome, all the genetic material present in a sample, enables identification of pathogens and in addition provides detailed information used for the typing, surveillance and identification of resistance mutations. This metagenomic test can be used for virus discovery and microbiome/virome analyses. The host-pathogen interaction can be interpreted from transcriptome data, providing information about what genes are activated or repressed. In collaboration with other health care departments, hereditary mutations can be identified in a combined agnostic test for both pathogens and pathogenic variants in the host genome as the next step in molecular diagnostics. Drug metabolizing enzyme information can be retrieved, and the cell-free DNA (cfDNA) can be used as a biomarker for tumour detection, and in addition for the identification of congenital infections. Created using Biorender.com.

## Metagenomic sequencing aiding tumour detection

Cell-free DNA (cfDNA) sequenced in the metagenomic sequencing process is useful to detect congenital infection in the fetal cfDNA, especially CMV [127] and parvovirus [104,128]. Sequencing of cfDNA, also considered a liquid biopsy, is additionally a potential biomarker for detecting cancerous tumours within patients for the pathology field. Patients with fast dividing tumour cells have cfDNA in large proportions in serum or plasma as a result of cellular necrosis and apoptosis [129-132]. The detection of cancerous small DNA particles is based on finding specific mutations in the circulating tumour DNA that are not present in the DNA of white blood cells [133-137]. The viral target enrichment panel used in the research in this thesis (**chapter 3 and 4**) can be further enriched using probes of the Cancer Personalized Profiling (CAPP-Seq) sequencing kit, a method proven to be successful in finding and identifying the circulating tumour DNA particles in the cfDNA liquid biopsies [129,138].

## Time to update Koch's postulates

With a growing number of viruses detected by means of viral metagenomics and other sequencing techniques, modern thoughts about causality have to be explored as viruses can be found that are not always known to be causal for disease. Around 1880, Robert Koch postulated criteria to establish a causal relationship between a microbe and a disease [139,140]. Over the years these criteria have been updated to four criteria (Figure 5) [141-144], as inoculating an organism with the potentially pathogenic microorganism was not included in the original postulates [139,140]. However, some of Koch's postulates are hard to bring into research practice. Some pathogens cannot be grown in pure culture and Koch's criteria to have "no abundance of the disease-causing organism in healthy individuals" is difficult to prove as per identified virus it is difficult to test many healthy patients efficiently, shortly after the moment a novel or unexpected virus is detected using viral metagenomics. Furthermore, it will be difficult to receive medical ethical approval to follow up in humans on criteria 3 by introducing the cultured disease-causing microorganism into a healthy individual. Additionally, multi-factorial causes, like host health circumstances and dose of infection, play an additional role, and make it harder to rule out any confounding factors.

To investigate microorganism prevalence in the sequence era, it is required to create virome databases that are made publicly available and that can be filtered on abundance of microorganism for clinical syndrome and sample types. In the data that is currently publicly available in sequencing databases, the virome information can also be retrieved [102-104]. In human genetics, mutations are checked

for abundance in databases such as GNOMAD [145,146], 1000 genomes [147] and GoNL [148], as genetic disease can never be present in large percentages of healthy individuals. The microbiology community needs similar databases with both prevalence and disease information in order to enable a better differentiation between a healthy and unhealthy microorganism population, and it is recommended that the microbiology community get these in place to be ready for the future.

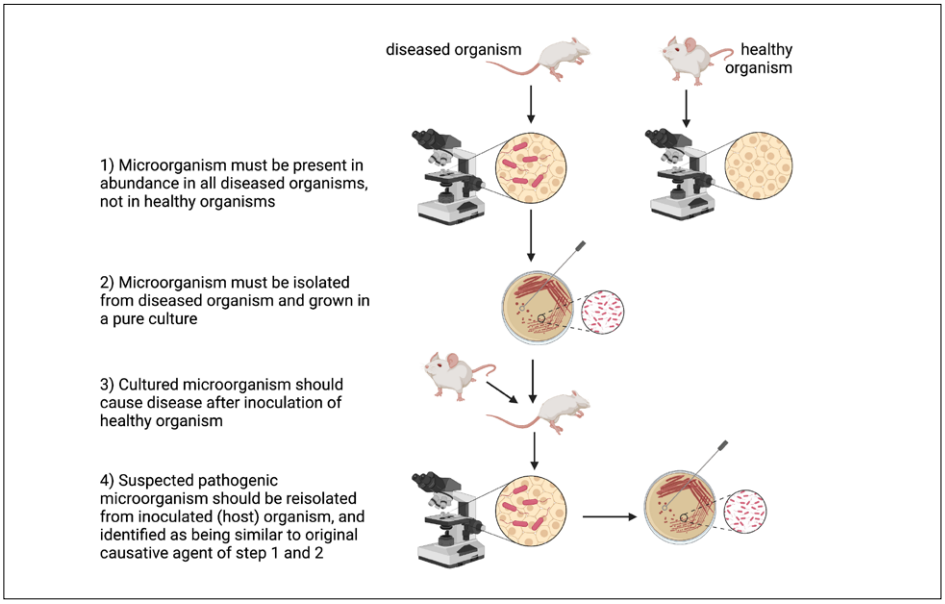



**Figure 5. Koch's postulates from the 19th century.**

The expanded criteria of the outdated Koch's postulates. The four criteria were designed to establish a causal relationship between a microorganism and a disease. In diseased organisms the suspected microorganism must be found in abundance, and not in healthy organisms (1), and the suspected microorganism must be isolated and grown in pure culture (2). The cultured microorganism, when inoculated into a healthy organism, should cause disease (3) and after infection and disease the suspected pathogenic microorganism had to be reisolated and identified as the original causative agent (4). Created using Biorender.com.

When analysing NGS data from metagenomic sequencing, host-pathogen interactions and virus activity within a host can be investigated to find proof of causality and virulence directly from the available metagenomic sequence data. Some RNA library protocols can differentiate between plus and minus strands, thus providing information about viral activity [149]. Virus transcription is known to differ between stages of infection [150-152] and with this information virus activity can be taken into account in the diagnosis of a patient. Additionally, with transcriptomic sequence read analyses the

differential and co-expression of the host immune transcriptome can be mapped [153]. In this way, transcriptomics will give information on viral gene activity and a host response [154-156]. Identifying pathogens using viral metagenomics, retrieving information on prevalence in sequence databases and investigating virulence will lead to novel sequence proof of Koch's postulates (Figure 6). The evaluation of the activity of a virus and corresponding host response will result in an evolution of viral molecular diagnostics. Most likely simply stating if a virus is present or absent and in what quantity will in the future be outdated by novel findings, when precise viral and host immunity activity can be predicted in an infection site based on metagenomic sequencing.

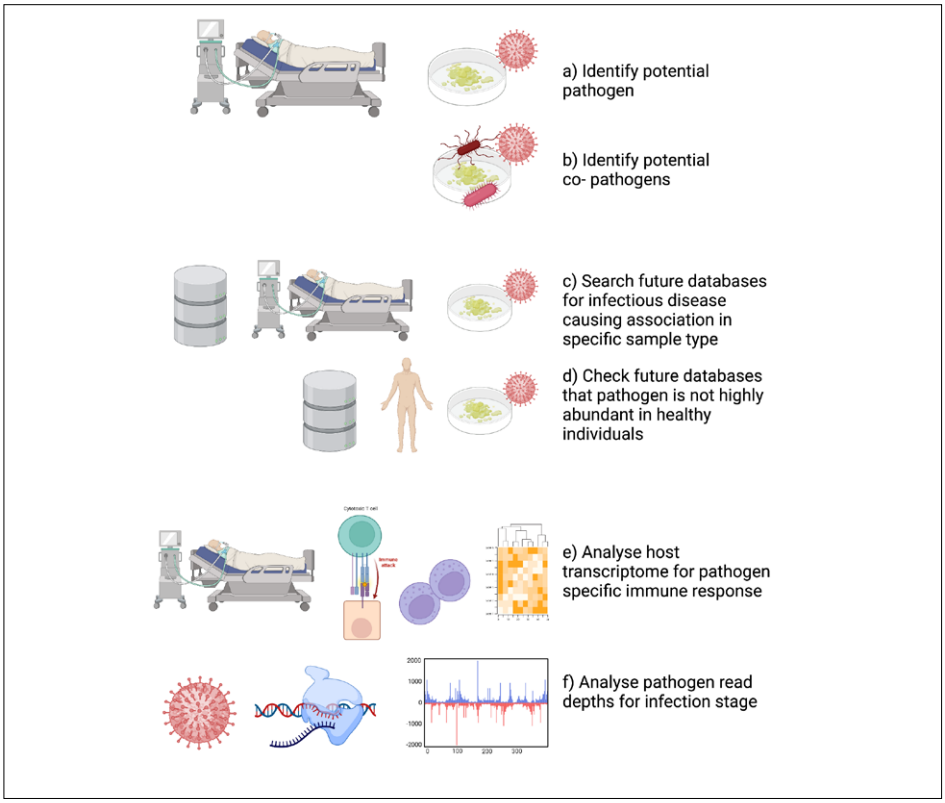



**Figure 6. The updated Koch's postulates for the 'next-generation sequencing era'.**

When applying the updated Koch's postulates directly on metagenomic NGS data to establish a causal relationship between a microbe and a disease, the first step is to identify a pathogen (a) and to rule out any other potential pathogen within a sample (b). After these steps, future databases should be checked to see if this pathogen is linked to disease-causing symptoms (c), and whether the pathogen is present in the same sample type in healthy individuals (d). In the future, the host-pathogen interaction can be followed by looking at the immune response in a host by analysing the host transcriptome (e) and the pathogen activity and infection stage can be tracked (f). Created using Biorender.com.

# Concluding remarks

In the near future of diagnosing infectious diseases, metagenomic sequencing will be implemented at larger scale as a secondary test to evaluate all organisms present at once, by sequencing all available genetic material in a sample. This is the pathogen-agnostic way of identifying a virus that might be pathogenic. Though currently an expensive and time-consuming test, metagenomic sequencing will ultimately improve the diagnostic yield and potentially lead to lower costs when other diagnostic and treatment areas are included in the consideration of costs. Sensitivity of mNGS can be increased by the use of capture probes and more optimal taxonomic classifications tools. With the information available on resistance mutations and typing, metagenomic sequencing provides useful data for virus surveillance. Viruses can be discovered directly from patient samples as exemplified in the beginning of the SARS-CoV-2 pandemic. In the future, sequencing protocols are expected to be faster and more applicable, and with improved filtering of genetic host sequences and addressing causation problems, disease-causing viruses can be differentiated from the regular virome.

A combined metagenomic sequencing test can be used to detect infecting organisms, but can be additionally useful for looking at pathogen activity, pathogen resistance, host transcriptome activity, host pharmacogenetics, genetic inherited pathogenic defaults of a host, and tumour surveillance. By combining all these data from metagenomic sequencing, this test has the promise to function as a multidimensional future diagnostic test aiding multiple clinical disciplines.

# References


[1] G. Lasso et al., 'A Structure-Informed Atlas of Human-Virus Interactions', Cell, vol. 178, no. 6, pp. 1526-1541.e16, Sep. 2019, doi: 10.1016/j.cell.2019.08.005.

[2] L. H. Taylor, S. M. Latham, and M. E. J. woolhouse, 'Risk factors for human disease emergence', Phil. Trans. R. Soc. Lond. B, vol. 356, no. 1411, pp. 983–989, Jul. 2001, doi: 10.1098/rstb.2001.0888.

[3] J. Granerod and N. S. Crowcroft, 'The epidemiology of acute encephalitis', Neuropsychological Rehabilitation, vol. 17, no. 4–5, pp. 406–428, Aug. 2007, doi: 10.1080/09602010600989620.

[4] P. Kennedy, P.-L. Quan, and W. Lipkin, 'Viral Encephalitis of Unknown Cause: Current Perspective and Recent Advances', Viruses, vol. 9, no. 6, p. 138, Jun. 2017, doi: 10.3390/v9060138.

[5] S. Jain et al., 'Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults', N Engl J Med, vol. 373, no. 5, pp. 415–427, Jul. 2015, doi: 10.1056/NEJMoa1500245.

[6] T. Heikkinen and A. Järvinen, 'The common cold', The Lancet, vol. 361, no. 9351, pp. 51–59, Jan. 2003, doi: 10.1016/S0140-6736(03)12162-9.

[7] M. Ieven et al., 'Aetiology of lower respiratory tract infection in adults in primary care: a prospective study in 11 European countries', Clinical Microbiology and Infection, vol. 24, no. 11, pp. 1158–1163, Nov. 2018, doi: 10.1016/j.cmi.2018.02.004.

[8] E. C. Carbo et al., 'Improved diagnosis of viral encephalitis in adult and pediatric hematological patients using viral metagenomics', Journal of Clinical Virology, vol. 130, p. 104566, Sep. 2020, doi: 10.1016/j.jcv.2020.104566.

[9] J. C. Haston et al., 'Prospective Cohort Study of Next-Generation Sequencing as a Diagnostic Modality for Unexplained Encephalitis in Children', Journal of the Pediatric Infectious Diseases Society, vol. 9, no. 3, pp. 326–333, Jul. 2020, doi: 10.1093/jpids/piz032.

[10] M. R. Wilson et al., 'Clinical Metagenomic Sequencing for Diagnosis of Meningitis and Encephalitis', N Engl J Med, vol. 380, no. 24, pp. 2327–2340, Jun. 2019, doi: 10.1056/NEJMoa1803396.

[11] Kufner et al., 'Two Years of Viral Metagenomics in a Tertiary Diagnostics Unit: Evaluation of the First 105 Cases', Genes, vol. 10, no. 9, p. 661, Aug. 2019, doi: 10.3390/genes10090661.

[12] S. L. Salzberg et al., 'Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system', Neurol Neuroimmunol Neuroinflamm, vol. 3, no. 4, p. e251, Aug. 2016, doi: 10.1212/NXI.0000000000000251.

[13] H. E. Ambrose et al., 'Diagnostic Strategy Used To Establish Etiologies of Encephalitis in a Prospective Cohort of Patients in England', Journal of Clinical Microbiology, vol. 49, no. 10, pp. 3576–3583, Oct. 2011, doi: 10.1128/JCM.00862-11.

[14] S. Saha et al., 'Unbiased Metagenomic Sequencing for Pediatric Meningitis in Bangladesh Reveals Neuroinvasive Chikungunya Virus Outbreak and Other Unrealized Pathogens', mBio, vol. 10, no. 6, pp. e02877-19, /mbio/10/6/mBio.02877-19.atom, Dec. 2019, doi: 10.1128/mBio.02877-19.

[15] P. Turner et al., 'The aetiologies of central nervous system infections in hospitalised Cambodian children', BMC Infect Dis, vol. 17, no. 1, p. 806, Dec. 2017, doi: 10.1186/s12879-017-2915-6.

[16] J. Kawada et al., 'Next-Generation Sequencing for the Identification of Viruses in Pediatric Acute Encephalitis and Encephalopathy', Open Forum Infectious Diseases, vol. 3, no. suppl_1, p. 1172, Dec. 2016, doi: 10.1093/ofid/ofw172.875.

[17] S. L. Smits et al., 'Novel Cyclovirus in Human Cerebrospinal Fluid, Malawi, 2010–2011', Emerg. Infect. Dis., vol. 19, no. 9, Sep. 2013, doi: 10.3201/eid1909.130404.

[18] H. Jerome et al., 'Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers', Journal of Infection, vol. 79, no. 4, pp. 383–388, Oct. 2019, doi: 10.1016/j.jinf.2019.08.003.

[19] K. N. Govender, T. L. Street, N. D. Sanderson, and D. W. Eyre, 'Metagenomic Sequencing as a Pathogen-Agnostic Clinical Diagnostic Tool for Infectious Diseases: a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies', J Clin Microbiol, vol. 59, no. 9, pp. e02916-20, Aug. 2021, doi: 10.1128/JCM.02916-20.

[20] Z. Diao, D. Han, R. Zhang, and J. Li, 'Metagenomics next-generation sequencing tests take the stage in the diagnosis of lower respiratory tract infections', Journal of Advanced Research, vol. 38, pp. 201–212, May 2022, doi: 10.1016/j.jare.2021.09.012.

[21] P. Parize et al., 'Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study', Clinical Microbiology and Infection, vol. 23, no. 8, p. 574.e1-574.e6, Aug. 2017, doi: 10.1016/j.cmi.2017.02.006.

[22] N. T. T. Hong et al., 'Performance of Metagenomic Next-Generation Sequencing for the Diagnosis of Viral Meningoencephalitis in a Resource-Limited Setting', Open Forum Infectious Diseases, vol. 7, no. 3, p. ofaa046, Mar. 2020, doi: 10.1093/ofid/ofaa046.

[23] S. Miller et al., 'Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid', Genome Res., vol. 29, no. 5, pp. 831–842, 2019, doi: 10.1101/gr.238170.118.

[24] T. A. Blauwkamp et al., 'Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease', Nat Microbiol, vol. 4, no. 4, pp. 663–674, Apr. 2019, doi: 10.1038/s41564-018-0349-6.

[25] S. Somasekar et al., 'Viral Surveillance in Serum Samples From Patients With Acute Liver Failure By Metagenomic Next-Generation Sequencing', Clinical Infectious Diseases, vol. 65, no. 9, pp. 1477–1485, Oct. 2017, doi: 10.1093/cid/cix596.

[26] J. Rossoff et al., 'Noninvasive Diagnosis of Infection Using Plasma Next-Generation Sequencing: A Single-Center Experience', Open Forum Infectious Diseases, vol. 6, no. 8, p. ofz327, Aug. 2019, doi: 10.1093/ofid/ofz327.

[27] R. Schlaberg et al., 'Viral Pathogen Detection by Metagenomics and Pan-Viral Group Polymerase Chain Reaction in Children With Pneumonia Lacking Identifiable Etiology', The Journal of Infectious Diseases, vol. 215, no. 9, pp. 1407–1415, May 2017, doi: 10.1093/infdis/jix148.

[28] T. Doan et al., 'Metagenomic DNA Sequencing for the Diagnosis of Intraocular Infections', Ophthalmology, vol. 124, no. 8, pp. 1247–1248, Aug. 2017, doi: 10.1016/j.ophtha.2017.03.045.

[29] C. Langelier et al., 'Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults', Proc. Natl. Acad. Sci. U.S.A., vol. 115, no. 52, Dec. 2018, doi: 10.1073/pnas.1809700115.

[30] J. Wang, Y. Han, and J. Feng, 'Metagenomic next-generation sequencing for mixed pulmonary infection diagnosis', BMC Pulm Med, vol. 19, no. 1, p. 252, Dec. 2019, doi: 10.1186/s12890-019-1022-4.

[31] A. L. van Rijn et al., 'The respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease', PLoS ONE, vol. 14, no. 10, p. e0223952, Oct. 2019, doi: 10.1371/journal.pone.0223952.

[32] J. Huang et al., 'Metagenomic Next-Generation Sequencing versus Traditional Pathogen Detection in the Diagnosis of Peripheral Pulmonary Infectious Lesions', IDR, vol. Volume 13, pp. 567–576, Feb. 2020, doi: 10.2147/IDR.S235182.

[33] S. van Boheemen et al., 'Retrospective Validation of a Metagenomic Sequencing Protocol for Combined Detection of RNA and DNA Viruses Using Respiratory Samples from Pediatric Patients', The Journal of Molecular Diagnostics, vol. 22, no. 2, pp. 196–207, Feb. 2020, doi: 10.1016/j.jmoldx.2019.10.007.

[34] J. J. C. de Vries et al., 'Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting', Journal of Clinical Virology, vol. 138, p. 104812, May 2021, doi: 10.1016/j.jcv.2021.104812.

[35] S. Nooij, D. Schmitz, H. Vennema, A. Kroneman, and M. P. G. Koopmans, 'Overview of Virus Metagenomic Classification Methods and Their Biological Applications', Front. Microbiol., vol. 9, p. 749, Apr. 2018, doi: 10.3389/fmicb.2018.00749.

[36] Junier et al., 'Viral Metagenomics in the Clinical Realm: Lessons Learned from a Swiss-Wide Ring Trial', Genes, vol. 10, no. 9, p. 655, Aug. 2019, doi: 10.3390/genes10090655.

[37] D. E. Wood and S. L. Salzberg, 'Kraken: ultrafast metagenomic sequence classification using exact alignments', Genome Biol, vol. 15, no. 3, p. R46, 2014, doi: 10.1186/gb-2014-15-3-r46.

[38] R. Ounit, S. Wanamaker, T. J. Close, and S. Lonardi, 'CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers', BMC Genomics, vol. 16, no. 1, p. 236, Dec. 2015, doi: 10.1186/s12864-015-1419-2.

[39] S. H. Ye, K. J. Siddle, D. J. Park, and P. C. Sabeti, 'Benchmarking Metagenomics Tools for Taxonomic Classification', Cell, vol. 178, no. 4, pp. 779–794, Aug. 2019, doi: 10.1016/j.cell.2019.07.010.

[40] P. Menzel, K. L. Ng, and A. Krogh, 'Fast and sensitive taxonomic classification for metagenomics with Kaiju', Nat Commun, vol. 7, no. 1, p. 11257, Sep. 2016, doi: 10.1038/ncomms11257.

[41] K. Mavromatis et al., 'Use of simulated data sets to evaluate the fidelity of metagenomic processing methods', Nat Methods, vol. 4, no. 6, pp. 495–500, Jun. 2007, doi: 10.1038/nmeth1043.

[42] F. Meyer, A. Bremges, P. Belmann, S. Janssen, A. C. McHardy, and D. Koslicki, 'Assessing taxonomic metagenome profilers with OPAL', Genome Biol, vol. 20, no. 1, p. 51, Dec. 2019, doi: 10.1186/s13059-019-1646-y.

[43] A. Sczyrba et al., 'Critical Assessment of Metagenome Interpretation — a benchmark of metagenomics software', Nat Methods, vol. 14, no. 11, pp. 1063–1071, Nov. 2017, doi: 10.1038/nmeth.4458.

[44] A. B. R. McIntyre et al., 'Comprehensive benchmarking and ensemble approaches for metagenomic classifiers', Genome Biol, vol. 18, no. 1, p. 182, Dec. 2017, doi: 10.1186/s13059-017-1299-7.

[45] Z. Sun et al., 'Challenges in benchmarking metagenomic profilers', Nat Methods, vol. 18, no. 6, pp. 618–626, Jun. 2021, doi: 10.1038/s41592-021-01141-3.

[46] A. Escobar-Zepeda et al., 'Analysis of sequencing strategies and tools for taxonomic annotation: Defining standards for progressive metagenomics', Sci Rep, vol. 8, no. 1, p. 12034, Dec. 2018, doi: 10.1038/s41598-018-30515-5.

[47] A. Brinkmann et al., 'Proficiency Testing of Virus Diagnostics Based on Bioinformatics Analysis of Simulated In Silico High-Throughput Sequencing Data Sets', J Clin Microbiol, vol. 57, no. 8, pp. e00466-19, /jcm/57/8/JCM.00466-19.atom, Jun. 2019, doi: 10.1128/JCM.00466-19.

[48] N. Couto et al., 'Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens', Sci Rep, vol. 8, no. 1, p. 13767, Dec. 2018, doi: 10.1038/s41598-018-31873-w.

[49] J. J. C. de Vries et al., 'Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples', Journal of Clinical Virology, p. 104908, Jul. 2021, doi: 10.1016/j.jcv.2021.104908.

[50] S. Morfopoulou and V. Plagnol, 'Bayesian mixture analysis for metagenomic community profiling', Bioinformatics, vol. 31, no. 18, pp. 2930–2938, Sep. 2015, doi: 10.1093/bioinformatics/btv317.

[51] J. M. Martí, 'Recentrifuge: Robust comparative analysis and contamination removal for metagenomics', PLoS Comput Biol, vol. 15, no. 4, p. e1006967, Apr. 2019, doi: 10.1371/journal.pcbi.1006967.

[52] V. C. Piro and B. Y. Renard, 'Contamination detection and microbiome exploration with GRIMER', Bioinformatics, preprint, Jun. 2021. doi: 10.1101/2021.06.22.449360.

[53] D. Shah, J. R. Brown, J. C. D. Lee, M. L. Carpenter, G. Wall, and J. Breuer, 'Use of a sample-to-result shotgun metagenomics platform for the detection and quantification of viral pathogens in paediatric immunocompromised patients', Journal of Clinical Virology Plus, vol. 2, no. 2, p. 100073, Jun. 2022, doi: 10.1016/j.jcvp.2022.100073.

[54] C. Huang et al., 'Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China', The Lancet, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[55] P. Zhou et al., 'Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin', Microbiology, preprint, Jan. 2020. doi: 10.1101/2020.01.22.914952.

[56] N. Zhu et al., 'A Novel Coronavirus from Patients with Pneumonia in China, 2019', N Engl J Med, vol. 382, no. 8, pp. 727–733, Feb. 2020, doi: 10.1056/NEJMoa2001017.

[57] M. Vilsker et al., 'Genome Detective: an automated system for virus identification from high-throughput sequencing data', Bioinformatics, vol. 35, no. 5, pp. 871–873, Mar. 2019, doi: 10.1093/bioinformatics/bty695.

[58] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 'Basic local alignment search tool', Journal of Molecular Biology, vol. 215, no. 3, pp. 403–410, Oct. 1990, doi: 10.1016/S0022-2836(05)80360-2.

[59] W. T. Harvey et al., 'SARS-CoV-2 variants, spike mutations and immune escape', Nat Rev Microbiol, vol. 19, no. 7, pp. 409–424, Jul. 2021, doi: 10.1038/s41579-021-00573-0.

[60] K. Tao et al., 'The biological and clinical significance of emerging SARS-CoV-2 variants', Nat Rev Genet, vol. 22, no. 12, pp. 757–773, Dec. 2021, doi: 10.1038/s41576-021-00408-x.

[61] Z. Chen et al., 'Global landscape of SARS-CoV-2 genomic surveillance and data sharing', Nat Genet, vol. 54, no. 4, pp. 499–507, Apr. 2022, doi: 10.1038/s41588-022-01033-y.

[62] 'https://www.gisaid.org/', Apr. 2022.

[63] D. Liu et al., 'Development and Multicenter Assessment of a Reference Panel for Clinical Shotgun Metagenomics for Pathogen Detection', In Review, preprint, Feb. 2021. doi: 10.21203/rs.3.rs-208796/v1.

[64] J. A. Nasir et al., 'A Comparison of Whole Genome Sequencing of SARS-CoV-2 Using Amplicon-Based Sequencing, Random Hexamers, and Bait Capture', Viruses, vol. 12, no. 8, p. 895, Aug. 2020, doi: 10.3390/v12080895.

[65] M. Xiao et al., 'Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples', Genome Med, vol. 12, no. 1, p. 57, Dec. 2020, doi: 10.1186/s13073-020-00751-4.

[66] F. Wegner et al., 'External Quality Assessment of SARS-CoV-2 Sequencing: an ESGMD-SSM Pilot Trial across 15 European Laboratories', J Clin Microbiol, vol. 60, no. 1, pp. e01698-21, Jan. 2022, doi: 10.1128/JCM.01698-21.

[67] J. Plitnick, S. Griesemer, E. Lasek-Nesselquist, N. Singh, D. M. Lamson, and K. St. George, 'Whole-Genome Sequencing of SARS-CoV-2: Assessment of the Ion Torrent AmpliSeq Panel and Comparison with the Illumina MiSeq ARTIC Protocol', J Clin Microbiol, vol. 59, no. 12, pp. e00649-21, Nov. 2021, doi: 10.1128/JCM.00649-21.

[68] J. H. Chai et al., 'Cost-benefit analysis of introducing next-generation sequencing (metagenomic) pathogen testing in the setting of pyrexia of unknown origin', PLoS ONE, vol. 13, no. 4, p. e0194648, Apr. 2018, doi: 10.1371/journal.pone.0194648.

[69] F. X. López-Labrador et al., 'Recommendations for the introduction of metagenomic high-throughput sequencing in clinical virology, part I: Wet lab procedure', Journal of Clinical Virology, vol. 134, p. 104691, Jan. 2021, doi: 10.1016/j.jcv.2020.104691.

[70] M. Costello et al., 'Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms', BMC Genomics, vol. 19, no. 1, p. 332, Dec. 2018, doi: 10.1186/s12864-018-4703-0.

[71] N. Stoler and A. Nekrutenko, 'Sequencing error profiles of Illumina sequencing instruments', NAR Genomics and Bioinformatics, vol. 3, no. 1, p. lqab019, Jan. 2021, doi: 10.1093/nargab/lqab019.

[72] Z. Feng, J. C. Clemente, B. Wong, and E. E. Schadt, 'Detecting and phasing minor single-nucleotide variants from long-read sequencing data', Nat Commun, vol. 12, no. 1, p. 3032, Dec. 2021, doi: 10.1038/s41467-021-23289-4.

[73] R. Kou et al., 'Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations', PLoS ONE, vol. 11, no. 1, p. e0146638, Jan. 2016, doi: 10.1371/journal.pone.0146638.

[74] J. B. Hiatt, C. C. Pritchard, S. J. Salipante, B. J. O'Roak, and J. Shendure, 'Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation', Genome Res., vol. 23, no. 5, pp. 843–854, May 2013, doi: 10.1101/gr.147686.112.

[75] B. Kant et al., 'Gene Mosaicism Screening Using Single-Molecule Molecular Inversion Probes in Routine Diagnostics for Systemic Autoinflammatory Diseases', The Journal of Molecular Diagnostics, vol. 21, no. 6, pp. 943–950, Nov. 2019, doi: 10.1016/j.jmoldx.2019.06.009.

[76] K. Kryukov and T. Imanishi, 'Human Contamination in Public Genome Assemblies', PLoS ONE, vol. 11, no. 9, p. e0162424, Sep. 2016, doi: 10.1371/journal.pone.0162424.

[77] M. S. Longo, M. J. O'Neill, and R. J. O'Neill, 'Abundant Human DNA Contamination Identified in Non-Primate Genome Databases', PLoS ONE, vol. 6, no. 2, p. e16410, Feb. 2011, doi: 10.1371/journal.pone.0016410.

[78] S. Mukherjee, M. Huntemann, N. Ivanova, N. C. Kyrpides, and A. Pati, 'Large-scale contamination of microbial isolate genomes by Illumina PhiX control', Stand in Genomic Sci, vol. 10, no. 1, p. 18, Dec. 2015, doi: 10.1186/1944-3277-10-18.

[79] M. Laurence, C. Hatzis, and D. E. Brash, 'Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes', PLoS ONE, vol. 9, no. 5, p. e97876, May 2014, doi: 10.1371/journal.pone.0097876.

[80] A. Rhie et al., 'Towards complete and error-free genome assemblies of all vertebrate species', Nature, vol. 592, no. 7856, pp. 737–746, Apr. 2021, doi: 10.1038/s41586-021-03451-0.

[81] M. Steinegger and S. L. Salzberg, 'Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank', Genome Biol, vol. 21, no. 1, p. 115, Dec. 2020, doi: 10.1186/s13059-020-02023-1.

[82] S. Roux, S. J. Hallam, T. Woyke, and M. B. Sullivan, 'Viral dark matter and virus–host interactions resolved from publicly available microbial genomes', eLife, vol. 4, p. e08490, Jul. 2015, doi: 10.7554/eLife.08490.

[83] S. R. Krishnamurthy and D. Wang, 'Origins and challenges of viral dark matter', Virus Research, vol. 239, pp. 136–142, Jul. 2017, doi: 10.1016/j.virusres.2017.02.002.

[84] T. G. Burland, 'DNASTAR's Lasergene Sequence Analysis Software', in Bioinformatics Methods and Protocols, vol. 132, New Jersey: Humana Press, 1999, pp. 71–91. doi: 10.1385/1-59259-192-2:71.

[85] S. S. Minot, N. Krumm, and N. B. Greenfield, 'One Codex: A Sensitive and Accurate Data Platform for Genomic Microbial Identification', Bioinformatics, preprint, Sep. 2015. doi: 10.1101/027607.

[86] S. Flygare et al., 'Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling', Genome Biol, vol. 17, no. 1, p. 111, Dec. 2016, doi: 10.1186/s13059-016-0969-1.

[87] R. Lozano et al., 'Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010', The Lancet, vol. 380, no. 9859, pp. 2095–2128, Dec. 2012, doi: 10.1016/S0140-6736(12)61728-0.

[88] H. Nair et al., 'Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis', The Lancet, vol. 381, no. 9875, pp. 1380–1390, Apr. 2013, doi: 10.1016/S0140-6736(12)61901-1.

[89] M. Bates, V. Mudenda, P. Mwaba, and A. Zumla, 'Deaths due to respiratory tract infections in Africa: a review of autopsy studies', Current Opinion in Pulmonary Medicine, vol. 19, no. 3, pp. 229–237, May 2013, doi: 10.1097/MCP.0b013e32835f4fe4.

[90] 'NVMM', Accessed: Aug. 03, 2022. [Online]. Available: www.nvmm.nl/media/4618/220422_diagnostisch-algoritme_nvmm-nwkv-nvk.pdf

[91] G. Almogy et al., 'Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics platform', Genomics, preprint, May 2022. doi: 10.1101/2022.05.29.493900.

[92] R. M. Leggett et al., 'Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens', Nat Microbiol, vol. 5, no. 3, pp. 430–442, Mar. 2020, doi: 10.1038/s41564-019-0626-z.

[93] M. D. Cao, D. Ganesamoorthy, A. G. Elliott, H. Zhang, M. A. Cooper, and L. J. M. Coin, 'Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinIONTM sequencing', GigaSci, vol. 5, no. 1, p. 32, Dec. 2016, doi: 10.1186/s13742-016-0137-2.

[94] C. P. Oechslin et al., 'Limited Correlation of Shotgun Metagenomics Following Host Depletion and Routine Diagnostics for Viruses and Bacteria in Low Concentrated Surrogate and Clinical Samples', Front. Cell. Infect. Microbiol., vol. 8, p. 375, Oct. 2018, doi: 10.3389/fcimb.2018.00375.

[95] M. T. Nelson et al., 'Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles', Cell Reports, vol. 26, no. 8, pp. 2227-2240.e5, Feb. 2019, doi: 10.1016/j.celrep.2019.01.091.

[96] 'GIAB', Accessed: Aug. 03, 2022. [Online]. Available: https://www.nist.gov/programs-projects/genome-bottle

[97] J. M. Zook et al., 'Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls', Nat Biotechnol, vol. 32, no. 3, pp. 246–251, Mar. 2014, doi: 10.1038/nbt.2835.

[98] J. M. Zook et al., 'An open resource for accurately benchmarking small variant and reference calls', Nat Biotechnol, vol. 37, no. 5, pp. 561–566, May 2019, doi: 10.1038/s41587-019-0074-6.

[99] 'GoFair', Accessed: Aug. 03, 2022. [Online]. Available: https://www.go-fair.org/fair-principles/

[100] E. Simon-Loriere and E. C. Holmes, 'Why do RNA viruses recombine?', Nat Rev Microbiol, vol. 9, no. 8, pp. 617–626, Aug. 2011, doi: 10.1038/nrmicro2614.

[101] S. Roux, F. Enault, B. L. Hurwitz, and M. B. Sullivan, 'VirSorter: mining viral signal from microbial genomic data', PeerJ, vol. 3, p. e985, May 2015, doi: 10.7717/peerj.985.

[102] R. C. Edgar et al., 'Petabase-scale sequence alignment catalyses viral discovery', Nature, vol. 602, no. 7895, pp. 142–147, Feb. 2022, doi: 10.1038/s41586-021-04332-2.

[103] A. Moustafa et al., 'The blood DNA virome in 8,000 humans', PLoS Pathog, vol. 13, no. 3, p. e1006292, Mar. 2017, doi: 10.1371/journal.ppat.1006292.

[104] J. Linthorst, M. M. M. Baksi, M. R. A. Welkers, and E. A. Sistermans, 'The cell‐free DNA virome of 108,349 Dutch pregnant women', Prenatal Diagnosis, p. pd.6143, Apr. 2022, doi: 10.1002/pd.6143.

[105] 'TCGA Research Network', Accessed: Sep. 04, 2022. [Online]. Available: https://www.cancer.gov/tcga

[106] D. M. Parkin, 'The global health burden of infection-associated cancers in the year 2002', Int. J. Cancer, vol. 118, no. 12, pp. 3030–3044, Jun. 2006, doi: 10.1002/ijc.21731.

[107] M. Plummer, C. de Martel, J. Vignat, J. Ferlay, F. Bray, and S. Franceschi, 'Global burden of cancers attributable to infections in 2012: a synthetic analysis', The Lancet Global Health, vol. 4, no. 9, pp. e609–e616, Sep. 2016, doi: 10.1016/S2214-109X(16)30143-7.

[108] L. Wang et al., 'Artificial Intelligence for COVID-19: A Systematic Review', Front. Med., vol. 8, p. 704256, Sep. 2021, doi: 10.3389/fmed.2021.704256.

[109] Q.-V. Pham, D. C. Nguyen, T. Huynh-The, W.-J. Hwang, and P. N. Pathirana, 'Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts', IEEE Access, vol. 8, pp. 130820–130839, 2020, doi: 10.1109/ACCESS.2020.3009328.

[110] B. M. C. Silva, J. J. P. C. Rodrigues, I. de la Torre Díez, M. López-Coronado, and K. Saleem, 'Mobile-health: A review of current state in 2015', Journal of Biomedical Informatics, vol. 56, pp. 265–272, Aug. 2015, doi: 10.1016/j.jbi.2015.06.003.

[111] A. S. R. Srinivasa Rao and J. A. Vazquez, 'Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine', Infect Control Hosp Epidemiol, vol. 41, no. 7, pp. 826–830, Jul. 2020, doi: 10.1017/ice.2020.61.

[112] H. S. Maghdid, K. Z. Ghafoor, A. S. Sadiq, K. Curran, D. B. Rawat, and K. Rabie, 'A Novel AI-enabled Framework to Diagnose Coronavirus COVID 19 using Smartphone Embedded Sensors: Design Study'. arXiv, May 30, 2020. Accessed: Aug. 04, 2022. [Online]. Available: http://arxiv.org/abs/2003.07434

[113] E. Ong, M. U. Wong, A. Huffman, and Y. He, 'COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning', Front. Immunol., vol. 11, p. 1581, Jul. 2020, doi: 10.3389/fimmu.2020.01581.

[114] S. F. Ahmed, A. A. Quadeer, and M. R. McKay, 'Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies', Viruses, vol. 12, no. 3, p. E254, Feb. 2020, doi: 10.3390/v12030254.

[115] Y. Furuse, 'Genomic sequencing effort for SARS-CoV-2 by country during the pandemic', International Journal of Infectious Diseases, vol. 103, pp. 305–307, Feb. 2021, doi: 10.1016/j.ijid.2020.12.034.

[116] K. Krebs and L. Milani, 'Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good', Hum Genomics, vol. 13, no. 1, p. 39, Dec. 2019, doi: 10.1186/s40246-019-0229-z.

[117] M. Pirmohamed, 'Pharmacogenetics and pharmacogenomics: Editorial', British Journal of Clinical Pharmacology, vol. 52, no. 4, pp. 345–347, Oct. 2001, doi: 10.1046/j.0306-5251.2001.01498.x.

[118] C. I. van der Made et al., 'Presence of Genetic Variants Among Young Men With Severe COVID-19', JAMA, vol. 324, no. 7, p. 663, Aug. 2020, doi: 10.1001/jama.2020.13719.

[119] D. Burgner, S. E. Jamieson, and J. M. Blackwell, 'Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better?', The Lancet Infectious Diseases, vol. 6, no. 10, pp. 653–663, Oct. 2006, doi: 10.1016/S1473-3099(06)70601-6.

[120] L. Kachuri et al., 'The landscape of host genetic factors involved in immune response to common viral infections', Genome Med, vol. 12, no. 1, p. 93, Dec. 2020, doi: 10.1186/s13073-020-00790-x.

[121] L. Quintana-Murci, 'Human Immunology through the Lens of Evolutionary Genetics', Cell, vol. 177, no. 1, pp. 184–199, Mar. 2019, doi: 10.1016/j.cell.2019.02.033.

[122] G. Kerner et al., 'Homozygosity for TYK2 P1104A underlies tuberculosis in about 1% of patients in a cohort of European ancestry', Proc. Natl. Acad. Sci. U.S.A., vol. 116, no. 21, pp. 10430–10434, May 2019, doi: 10.1073/pnas.1903561116.

[123] J. Zhao et al., 'Coexistence of Autoimmune Encephalitis and Other Systemic Autoimmune Diseases', Front. Neurol., vol. 10, p. 1142, Oct. 2019, doi: 10.3389/fneur.2019.01142.

[124] L. Farnaes et al., 'Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization', npj Genomic Med, vol. 3, no. 1, p. 10, Dec. 2018, doi: 10.1038/s41525-018-0049-4.

[125] H. Daoud et al., 'Next-generation sequencing for diagnosis of rare diseases in the neonatal intensive care unit', CMAJ, vol. 188, no. 11, pp. E254–E260, Aug. 2016, doi: 10.1503/cmaj.150823.

[126] C. J. Saunders et al., 'Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units', Sci. Transl. Med., vol. 4, no. 154, Oct. 2012, doi: 10.1126/scitranslmed.3004041.

[127] V. Chesnais et al., 'Using massively parallel shotgun sequencing of maternal plasmatic cell-free DNA for cytomegalovirus DNA detection during pregnancy: a proof of concept study', Sci Rep, vol. 8, no. 1, p. 4321, Dec. 2018, doi: 10.1038/s41598-018-22414-6.

[128] J. Linthorst, M. R. A. Welkers, and E. A. Sistermans, 'Distinct fragmentation patterns of circulating viral cell-free DNA in 83,552 non-invasive prenatal testing samples', EVCNA, 2021, doi: 10.20517/evcna.2021.13.

[129] A. M. Newman et al., 'An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage', Nat Med, vol. 20, no. 5, pp. 548–554, May 2014, doi: 10.1038/nm.3519.

[130] G. D. Sorenson, D. M. Pribish, F. H. Valone, V. A. Memoli, D. J. Bzik, and S. L. Yao, 'Soluble normal and mutated DNA sequences from single-copy genes in human blood', Cancer Epidemiol Biomarkers Prev, vol. 3, no. 1, pp. 67–71, Feb. 1994.

[131] V. Vasioukhin, P. Anker, P. Maurice, J. Lyautey, C. Lederrey, and M. Stroun, 'Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia', Br J Haematol, vol. 86, no. 4, pp. 774–779, Apr. 1994, doi: 10.1111/j.1365-2141.1994.tb04828.x.

[132] A. J. Bronkhorst, V. Ungerer, and S. Holdenrieder, 'The emerging role of cell-free DNA as a molecular marker for cancer management', Biomolecular Detection and Quantification, vol. 17, p. 100087, Mar. 2019, doi: 10.1016/j.bdq.2019.100087.

[133] J. D. Cohen et al., 'Detection and localization of surgically resectable cancers with a multi-analyte blood test', Science, vol. 359, no. 6378, pp. 926–930, Feb. 2018, doi: 10.1126/science.aar3247.

[134] J. Vendrell, F. Mau-Them, B. Béganton, S. Godreuil, P. Coopman, and J. Solassol, 'Circulating Cell Free Tumor DNA Detection as a Routine Tool forLung Cancer Patient Management', IJMS, vol. 18, no. 2, p. 264, Jan. 2017, doi: 10.3390/ijms18020264.

[135] E. Kidess and S. S. Jeffrey, 'Circulating tumor cells versus tumor-derived cell-free DNA: rivals or partners in cancer care in the era of single-cell analysis?', Genome Med, vol. 5, no. 8, p. 70, Aug. 2013, doi: 10.1186/gm474.

[136] J. Phallen et al., 'Direct detection of early-stage cancers using circulating tumor DNA', Sci Transl Med, vol. 9, no. 403, p. eaan2415, Aug. 2017, doi: 10.1126/scitranslmed.aan2415.

[137] S. Cristiano et al., 'Genome-wide cell-free DNA fragmentation in patients with cancer', Nature, vol. 570, no. 7761, pp. 385–389, Jun. 2019, doi: 10.1038/s41586-019-1272-6.

[138] F. Scherer et al., 'Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA', Sci. Transl. Med., vol. 8, no. 364, Nov. 2016, doi: 10.1126/scitranslmed.aai8545.

[139] Koch R. The etiology of anthrax, based on the life history of Bacillus anthracis. Beiträge zur Biologie der Pflanzen. 1876;2(2):277-310.

[140] Koch R. Die Atiologie der Tuberculose. Berl. Klin. Wochenschr. 1882;19:221-30.

[141] Loeffler F (1884) Mitt. Aus dem Kaiserl. Gesundheitsamte, 2, 421e499.

[142] T. M. Rivers, 'Viruses and Koch's Postulates', J Bacteriol, vol. 33, no. 1, pp. 1–12, Jan. 1937, doi: 10.1128/jb.33.1.1-12.1937.

[143] Evans AS. Causation and disease: the Henle-Koch postulates revisited. Yale J Biol Med. 1976 May;49(2):175-95.

[144] S. Falkow, 'Molecular Koch's Postulates Applied to Microbial Pathogenicity', Clinical Infectious Diseases, vol. 10, no. Supplement 2, pp. S274–S276, Aug. 1988, doi: 10.1093/cid/10.Supplement_2.S274.

[145] R. L. Collins et al., 'A structural variation reference for medical and population genetics', Nature, vol. 581, no. 7809, pp. 444–451, May 2020, doi: 10.1038/s41586-020-2287-8.

[146] K. J. Karczewski et al., 'The mutational constraint spectrum quantified from variation in 141,456 humans', Nature, vol. 581, no. 7809, pp. 434–443, May 2020, doi: 10.1038/s41586-020-2308-7.

[147] The 1000 Genomes Project Consortium et al., 'A global reference for human genetic variation', Nature, vol. 526, no. 7571, pp. 68–74, Oct. 2015, doi: 10.1038/nature15393.

[148] D. I. Boomsma et al., 'The Genome of the Netherlands: design, and project goals', Eur J Hum Genet, vol. 22, no. 2, pp. 221–227, Feb. 2014, doi: 10.1038/ejhg.2013.118.

[149] D. P. Depledge, I. Mohr, and A. C. Wilson, 'Going the Distance: Optimizing RNA-Seq Strategies for Transcriptomic Analysis of Complex Viral Genomes', J Virol, vol. 93, no. 1, pp. e01342-18, Jan. 2019, doi: 10.1128/JVI.01342-18.

[150] S. Boersma et al., 'Translation and Replication Dynamics of Single RNA Viruses', Cell, vol. 183, no. 7, pp. 1930-1945.e23, Dec. 2020, doi: 10.1016/j.cell.2020.10.019.

[151] J. Sun et al., 'Comparative Transcriptome Analysis Reveals the Intensive Early Stage Responses of Host Cells to SARS-CoV-2 Infection', Front. Microbiol., vol. 11, p. 593857, Nov. 2020, doi: 10.3389/fmicb.2020.593857.

[152] Z. Yang, D. P. Bruno, C. A. Martens, S. F. Porcella, and B. Moss, 'Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing', Proc. Natl. Acad. Sci. U.S.A., vol. 107, no. 25, pp. 11513–11518, Jun. 2010, doi: 10.1073/pnas.1006594107.

[153] P. Khatri, M. Sirota, and A. J. Butte, 'Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges', PLoS Comput Biol, vol. 8, no. 2, p. e1002375, Feb. 2012, doi: 10.1371/journal.pcbi.1002375.

[154] A. J. Westermann, S. A. Gorski, and J. Vogel, 'Dual RNA-seq of pathogen and host', Nat Rev Microbiol, vol. 10, no. 9, pp. 618–630, Sep. 2012, doi: 10.1038/nrmicro2852.

[155] A. J. Westermann et al., 'Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions', Nature, vol. 529, no. 7587, pp. 496–501, Jan. 2016, doi: 10.1038/nature16547.

[156] A. J. Westermann, L. Barquist, and J. Vogel, 'Resolving host–pathogen interactions by dual RNA-seq', PLoS Pathog, vol. 13, no. 2, p. e1006033, Feb. 2017, doi: 10.1371/journal.ppat.1006033.