Appendix A – Elaboration of the main text

A1 Search terms

Literature searches were performed in PubMed using Medical Subject Headings (MESH) terms and keywords "(?sequenc* OR metagenom* OR Illumina OR RNA-seq OR RNASeq OR (Roche 454) OR (Ion torrent) OR (Proton / PGM) OR MiSeq OR HiSeq OR NextSeq OR MinION OR Nanopore OR PacBio) AND (infectio* OR microorganism OR microorganisms OR pathogen OR pathogens OR bacteria* OR virus OR viral OR fungus OR fungi OR parasite OR parasites OR parasitic)", searching references of articles, and supplemented by expert opinion from within the group. Articles were limited to those in English language published between January 2000 and June 2019.

A2 Methods used for bioinformatics analysis (see section 4.3)

Taxonomic profiling and read classification methods fall into two categories, either using sequence composition (alignment-free) or sequence identity (similarity-based) compared against user-provided reference databases or aligning to marker genes (1). Similarity search-based methods, using algorithms such as megaBLAST, Bowtie2 (nucleotide) and DIAMOND, RAPSearch (translated nucleotide or protein), are considered the most sensitive methods for read classification, but require considerable computational power (2). Protein level taxonomic classification, based on translated nucleotide alignments is often considered the most appropriate for novel pathogen discovery and/or RNA virus detection (3), but may be prone to lower specificity than nucleotide-based classification for bacteria and eukaryotes due to higher sequence conservation in these organisms.

Methods using exact *k*-mer (sequences of *k* nucleotides) matching allow efficient analysis of samples and are typically accurate to at least the genus level but may identify false positive hits, for example with dinucleotide or homopolymer repeat regions, even if those can be removed prior to species-assignment, together with other low-complexity reads. To resolve ambiguous matches, several methods work either by partitioning reads with multiple possible assignments to the highest taxonomic level containing all matching species or their lowest

common ancestor (LCA) (4), or consider the unique *k*-mer content relative to the number of reads assigned (5). Other similarity-based methods use probabilistic read classification to formally resolve ambiguous matches (6, 7).

All taxonomic profiling methods are limited by false positive and false negative results (8, 9). These can be reduced through appropriate read filtering before classification, for example, by removing low complexity and/or host DNA and rRNA gene fragments, and depending on context, by penalising the addition of taxa in the results. Finally, a universal limitation of similarity-based methods is their reliance on public reference databases, even though masked databases can be used (10), which we describe in section 9.3. Several studies, including community-driven initiatives (11), have set out to benchmark the performance of the different classification methods. However, the necessary reliance on simulated data affects the evaluation of overall performance. The selection of appropriate classifiers is affected by the dataset, resource and research question. For instance, in microbiome research, accurate taxonomic classification to the species (or strain level) in a diverse community is critical, whereas lower resolution with more robust assignments may be required for clinical testing.

Rather than species composition, bioinformatic analyses can be targeted towards the detection of features of interest such as known AMR mutations/genes for one or multiple strains present in a metagenomic sample. An obvious application is the prediction of phenotype from sequence data, which is increasingly used in diagostics. However, implementation of phenotypic diagnostic tests based on whole-genome sequencing remains challenging even when applied to single species samples (12). The mixed nature of metagenomic data adds the further challenge of localising AMR genes and mutations to a particular species (13).

A3 Orthogonal methods to confirm pathogen identity, function and viability (see section 4.5)

Orthogonal methods should be considered for pathogen confirmation where immunohistochemistry or culture are not available or feasible (for example, species-specific qPCR, fluorescent in situ hybridisation ‘FISH’, and convalescent serology). Important tools include measuring a sample twice, preferably with a different extraction method in separate runs, or subsequent verification with a different sequencing setup. However, contamination

and incompleteness of reference databases can yield repeated false positive or false negative results (10) (section 9.3).

Sequence data, if applied only to DNA, should not automatically be assumed to indicate viability of an organism and/or functional expression of a gene, discussed in 4.3. In theory, follow-up proteomics could be used to confirm gene expression, although challenges and inconsistencies currently exist in correlating multi-omics data and this is rarely performed (14). Alternative methods may confirm functionality, such as the use of animal models to confirm findings of microbiome studies (15).

Detection of DNA (or proteins) alone does not establish the presence of a viable organism. The half-life of DNA and proteins is typically long, and DNA can be detected days, if not weeks, after clearance of an infection (16, 17). If the primary objective of a study is to detect viable organisms from tissue, sequencing of RNA is preferable due to its shorter half-life, i.e. in the order of minutes, however convincing evidence of viability would depend upon further experiments (18). An approach to detect metabolically active organisms (both dormant and replicating) is to treat the sample with Ethidium-monoazide or Propidium-monoazide (PMA) to selectively remove DNA from dead cells during downstream sequencing process (19), though these protocols remain challenging to optimise, especially for clinical use (20).

A4 Potential sources of bias (sampling, transport, storage, library preparation, and sequencing; see section 9.1)

It is recommended that the use of water and reagent blank controls (see section 4.4) should be reported, alongside checking for contamination from current and previous runs and cycle indexing (11).

Experimental bias that is caused at different stages of a metagenomics experiment is more challenging to control for than sources caused by selection bias or contamination. The fact that the microbiome is a mixture of different microorganisms means that a given protocol may favour certain groups being over-represented in the processed samples. For example, enrichment protocols may introduce bias for pathogen detection (21). Capture probe-targeted sequencing will limit detection to targeted sequences, and 16S has limitations with regard to the level of taxonomic classification. While this precise form of bias does not exist in

untargeted metagenomics, other experimental bias can occur at different protocol stages, including sampling, nucleic acid extraction (22) or during post-extraction steps (23). In 16S rRNA gene sequencing, different primers amplify different bacterial families to differing degrees due to mismatches, resulting in biased abundances which in turn bias diversity metrics (24) and cannot be completely corrected bioinformatically (25).

Validation procedures applying mock communities (26-29) or synthetic spike-ins (30, 31) prior to library preparation can assist in understanding experimental performance and identifying bias. Spike-in controls have also been shown to be useful for discriminating whether detected microbes likely represent reagent contaminants versus true infections given that the abundance of the former tends to be inversely correlated to the input mass of the nucleic acid (i.e., contaminants form a higher proportion of datasets from low input samples) (32-34). Spike-in controls are also useful for evaluating reduced analytic sensitivity (35).

A5 Limitations of reference database(s) (see section 9.3)

In cases where databases may be sufficiently comprehensive, the choice of protein or nucleotide reference database depends on the study setting and objectives. Protein (translated nucleotide) similarity searches recover more remote homology, allowing for the detection of divergent viruses that might be missed by nucleotide similarity searches (36), but incur increased risk of false-positives. In other cases, intraspecies or sub-strain differences may be clinically important, requiring nucleotide similarity searches against comprehensive reference databases, often employing computationally sophisticated methods (37, 38). Similarly, metagenomics gene screening approaches for specific genotypic features, such as for known AMR mutations, rely on nucleotide-scale resolution and updated databases facilitating rule-based prediction of phenotypes. Increasingly, nucleotide sequences are being used as comparative references, providing environment-specific taxonomic composition which may be useful to identify disease or host-specific niches (39).

Outputs of almost all metagenomics assigners are ranked as taxonomic hits relative to a reference database. These provide convenient classification units, however, it is acknowledged that within-taxon genetic diversity can generate misassignments. A well-known example is the apparent mis-classification of *Shigella* despite its high sequence similarity (>80%) to *Escherichia coli* compared to other *Escherichia* species **(40)**. It follows

that a single cut-off is not possible to delineate different taxonomic levels. Despite recent progress in the ability of metagenomic approaches for detection and targeted culturing of previously 'unculturable' bacteria **(41)**, classification of the latter remains challenging **(42)**. New microorganisms and viruses are being constantly described and naming conventions may change following progress in the taxonomy of previously described taxa (Figure 2d). It is also not uncommon for the same fungal organism to have been given multiple names due to the historic nomenclature of referring to different sexual phases with distinct names **(43)**. Though we note that a full discussion of the moving and complex field of taxonomic nomenclature is outside the scope of our review.

A6 Power calculations for metagenomics studies (see section 10.1)

The need for *a priori* power calculations is exemplified by a recent meta-analysis of gut microbiome studies; despite significant overall associations when combined, single studies with small sample sizes often had few or no significant associations, most likely due to being under-powered (44). Methods for different scenarios have been described. For example, Thompson's method has been suggested as appropriate for the calculation of power to detect α-diversity, as has a method to calculate the power of PERMANOVA to estimate β-diversity (45), whilst La Rosa *et al* (46) proposed a power method based on Dirichlet multinomial distribution for analysing relative abundance (47).

Power calculations depend on a number of variables including expected effect size, sequencing depth, level of taxonomical description, acceptable significance level (*p*-value) and power threshold, choice of the test statistics, how multiple comparisons were corrected for and the sample size (46). Parameters such as significance level and power threshold are usually chosen by convention as p-value ≤ 0.05 and power of 80% (46) and the level of taxonomical description will depend on the question asked. However, other parameters, such as effect size, require more careful consideration and a rationale for chosen options.

Published effect sizes are affected by other factors which differ between studies (section 9). Large effect sizes can overcome this issue - for example, differences between infants' and adults' microbiome composition have been observed even in studies using very different processing techniques because of the large effect size of age (48). This was demonstrated in a meta-analysis of the association between differences in the microbiome and obesity, which

found that most studies lacked the power to detect modest effect sizes of 0.9-6% changes in alpha diversity (49). Pilot studies are useful to determine the expected effect size, although small pilot studies may overestimate an effect, and lead to underestimation of the required sample (50). It may be challenging to anticipate a clinically or biologically relevant effect size, particularly for microbiota studies or studies with indeterminate endpoints, but some justification and elaboration should be attempted.

Multiple comparisons are inevitable in metagenomics studies (e.g. between hundreds of bacterial species and tens of clinical variables), and these increase the probability of false positive findings. For example, Vogtman *et al* conducted whole-genome shotgun sequencing on faecal samples from 52 cases with colorectal cancer and 52 matched controls and compared findings from a previously published 16S rRNA gene study. The study was underpowered to detect many statistically significant associations after correction for multiple testing (51). Power calculations need to incorporate correction for multiple testing, using approaches such as the Bonferroni method, or the less conservative Benjamini-Hochberg method (52).

Appendix B - Proposed criteria for causality

The use of metagenomics for infectious disease research and clinical medicine brings new opportunities for the attribution of disease causality. However, a critical approach is required to differentiate aetiological agents from commensals or contaminants and ensure the validity of interpretation. Prevailing concepts of disease causality and their application and extension to encompass metagenomics are discussed below.

Koch's postulates (1942) are the oldest criteria for evaluating disease causation (53). They proposed that an agent needs to be present in every case of the disease, be specific for that disease, that it can be grown in culture and replicate the disease when introduced into a naïve host.

The *Bradford Hill Criteria (1965)* added important parameters such as strength of association, temporality, consistency of results, coherence of studies, dose-response effect and plausibility (54).

Fedak *et al* updated these criteria in view of methodological advancements, including in molecular biology and statistics (55). These advancements enabled more nuanced exploration of the often multistep and complex pathways of disease causation, and it was suggested that the Bradford Hill Criteria needed to evolve. As such, while statistical significance was re-emphasised as a marker for the strength of an association, it was with the caveat that the underlying methods, evidence from the literature and other contextual factors should be explicitly taken into account in considering causality. More sophisticated integration of results from different study designs and with varying methods to demonstrate mechanistic links were proposed as extensions to the concept of consistency. Similarly, in recognition of the multifactorial and complex nature of disease progression, it was acknowledged that biological plausibility may require integration of data from different disciplines.

Since their conception, the limitations of Koch's postulates, such as for unculturable disease-causing microorganisms, have led to work-around modifications of the criteria, e.g. the use of immunological evidence to detect infection for unculturable pathogens (56, 57).

The unit of disease causation has been revised in some cases from organism to gene with the identification of virulence genes as both necessary and sufficient for disease causation (58) and, following the introduction of PCR, the presence of microbial nucleic acid sequences rather than intact organism has been used to attribute disease causality (59).

However, it has also been argued that virulence genes may not be readily identifiable if they are expressed at different times during an infection (60). This led to a revised set of 'Metagenomic Koch's Postulates' focussing on the identification of metagenomic traits, i.e. molecular markers such as sequence reads, assembled contigs, genes or full-genomes that can be used to distinguish disease-causing metagenomes from metagenomes obtained from healthy controls.

More recently, Neville *et al* developed the 'Commensal Koch's postulates' as a framework for establishing causality in microbiome studies and in recognition of the potential health-promoting effect of the microbiota (61). In a further extension, Vonaesch *et al* developed the 'Ecological Koch's postulates', which propose an ecological approach to disease causation, acknowledging the interactions between microorganisms, their hosts and environments, rather than focusing on a single pathogen or gene (62). The postulates propose that a 'dysbiotic' microbiota pattern, which represents a disease-promoting ecosystem, needs to be found in all diseased cases.

In accordance with Koch's original postulates, these modified criteria also require that inoculation of an unaffected host with an isolate or sample from a case and the observation of disease (or the prevention of disease) is necessary to establish causation.

Lipkin *et al,* developed a three-level scoring system based on laboratory, clinical and epidemiological data for establishing confidence in causation, ranging from possible to definitive (63). They recognised that ethical practice often precludes inoculation of healthy human hosts, and that animal models have their own limitations and may not be available. Moreover, the increasing use of molecular diagnostics rather than culture means that viable organisms may never be recovered. For a causal relationship to be confirmed, Lipkin therefore acknowledged the importance of reproduction of disease by inoculation, but also suggested that the attenuation or prevention of disease by microorganism-specific vaccines or drugs can provide alternative evidence of causality.

Similarly, according to Fredericks & Relman, evidence for a causal relationship is provided by the identification of microbial genomes in tissue samples with pathological changes, particularly where this is supported by comparisons of pathogen genome copy number in the tissue samples with and without pathology during different disease phases, including prior to onset (59).

Finally, the anatomical site of detection can add to the evidence for causality (64). For instance, in cases of encephalitis, detection of a single pathogen in the brain, compared to cerebrospinal fluid (CSF) or another sterile site, increases the likelihood of the organism being the causative agent (65, 66). Evidence of an organism-specific immune response adds further weight.

References

1. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods 9:811.
2. Patil KR, Haider P, Pope PB, Turnbaugh PJ, Morrison M, Scheffer T, McHardy AC. 2011. Taxonomic metagenome sequence assignment with structured output models. Nature Methods 8:191.
3. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Communications 7:11257.
4. Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. Genome Research 17:377-386.
5. Breitwieser FP, Baker DN, Salzberg SL. 2018. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biology 19:198.
6. Morfopoulou S, Plagnol V. 2015. Bayesian mixture analysis for metagenomic community profiling. Bioinformatics 31:2930-2938.
7. Francis OE, Bendall M, Manimaran S, Hong C, Clement NL, Castro-Nallar E, Snell Q, Schaalje GB, Clement MJ, Crandall KA, Johnson WE. 2013. Pathoscope: Species identification and strain attribution with unassembled sequencing data. Genome Research.
8. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. 2008. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 61:344-9.
9. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, Glynn JR, Gallo V, Ramsay M, Sonnenberg P, MacCannell D, Charlett A, Egger M, Green J, Vineis P, Abubakar I. 2014. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. The Lancet Infectious Diseases 14:341-352.
10. Lu J, Salzberg SL. 2018. Removing contaminants from databases of draft genomes. PLoS Comput Biol 14:e1006277.
11. Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, Bremges A, Fritz A, Garrido-Oter R, Jørgensen TS, Shapiro N, Blood PD, Gurevich A, Bai Y, Turaev D, DeMaere MZ, Chikhi R, Nagarajan N, Quince C, Meyer F, Balvočiūtė M, Hansen LH, Sørensen SJ, Chia BKH, Denis B, Froula JL, Wang Z, Egan R, Don Kang D, Cook JJ, Deltel C, Beckstette M, Lemaitre C, Peterlongo P, Rizk G, Lavenier D, Wu Y-W, Singer SW, Jain C, Strous M, Klingenberg H, Meinicke P, Barton MD, Lingner T, Lin H-H, Liao Y-C, et al. 2017. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. Nature Methods 14:1063.
12. Harris SR, Cole MJ, Spiteri G, Sánchez-Busó L, Golparian D, Jacobsson S, Goater R, Abudahab K, Yeats CA, Bercot B, Borrego MJ, Crowley B, Stefanelli P, Tripodo F, Abad R, Aanensen DM, Unemo M, Azevedo J, Balla E, Barbara C, Bergheim T, Borrego MJ, Bremer V, Buder S, Maikanti-Charalambous P, Chisholm S, Cowan S, Crowley B, Crucitti T, Diez M, Dudás M, Eastick K, Goubard A, Haller M, Hauksdóttir GS, Hoffmann S, Hughes G, Igoe D, Jeverica S, Klavs I, Kløvstad H, Kohl P, Konte V, Linde I, Mavcutko V, Melillo JM, Pakarna G, Pavlik P, Pieridou D, La Ruche G, et al. 2018. Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. The Lancet Infectious Diseases 18:758-768.
13. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, Martin S, Hoyles L, Clarke P, Hall LJ, Clark MD. 2018. Rapid profiling of the preterm infant gut microbiota using nanopore sequencing aids pathogen diagnostics. bioRxiv doi:10.1101/180406:180406.
14. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R. 2016. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Briefings in Bioinformatics 19:286-302.
15. Hugenholtz F, de Vos WM. 2018. Mouse models for human intestinal microbiota research: a critical evaluation. Cellular and Molecular Life Sciences 75:149-160.

16. Branger S, Casalta JP, Habib G, Collard F, Raoult D. 2003. Streptococcus pneumoniae Endocarditis: Persistence of DNA on Heart Valve Material 7 Years after Infectious Episode. Journal of Clinical Microbiology 41:4435-4437.
17. Selenska S, Klingmuller W. 1991. DNA recovery and direct detection of Tn5 sequences from soil. Lett Appl Microbiol 13:21-4.
18. Bernstein JA, Lin PH, Cohen SN, Lin-Chao S. 2004. Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. Proc Natl Acad Sci U S A 101:2758-63.
19. Nocker A, Camper AK. 2009. Novel approaches toward preferential detection of viable cells using nucleic acid amplification techniques. FEMS Microbiol Lett 291:137-42.
20. Li R, Tun HM, Jahan M, Zhang Z, Kumar A, Dilantha Fernando WG, Farenhorst A, Khafipour E. 2017. Comparison of DNA-, PMA-, and RNA-based 16S rRNA Illumina sequencing for detection of live bacteria in water. Scientific reports 7:5752-5752.
21. Pettengill JB, McAvoy E, White JR, Allard M, Brown E, Ottesen A. 2012. Using metagenomic analyses to estimate the consequences of enrichment bias for pathogen detection. BMC Research Notes 5:378.
22. Velásquez-Mejía EP, de la Cuesta-Zuluaga J, Escobar JS. 2018. Impact of DNA extraction, sample dilution, and reagent contamination on 16S rRNA gene sequencing of human feces. Applied Microbiology and Biotechnology 102:403-411.
23. Huggett JF, Laver T, Tamisak S, Nixon G, O'Sullivan DM, Elaswarapu R, Studholme DJ, Foy CA. 2013. Considerations for the development and application of control materials to improve metagenomic microbial community profiling. Accreditation and Quality Assurance 18:77-83.
24. Cai L, Ye L, Tong AHY, Lok S, Zhang T. 2013. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PloS one 8:e53649-e53649.
25. Edgar RC. 2017. UNBIAS: An attempt to correct abundance bias in 16S sequencing, with limited success. bioRxiv doi:10.1101/124149:124149.
26. Singer E, Andreopoulos B, Bowers RM, Lee J, Deshpande S, Chiniquy J, Ciobanu D, Klenk H-P, Zane M, Daum C, Clum A, Cheng J-F, Copeland A, Woyke T. 2016. Next generation sequencing data of a defined microbial mock community. Scientific Data 3:160081.
27. Rinke C, Low S, Woodcroft BJ, Raina J-B, Skarshewski A, Le XH, Butler MK, Stocker R, Seymour J, Tyson GW, Hugenholtz P. 2016. Validation of picogram- and femtogram-input DNA libraries for microscale metagenomics. PeerJ 4:e2486.
28. Bowers RM, Clum A, Tice H, Lim J, Singh K, Ciobanu D, Ngan CY, Cheng J-F, Tringe SG, Woyke T. 2015. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. BMC Genomics 16:856.
29. The Human Microbiome Project C, Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam DA, et al. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207.
30. Hardwick SA, Chen WY, Wong T, Kanakamedala BS, Deveson IW, Ongley SE, Santini NS, Marcellin E, Smith MA, Nielsen LK, Lovelock CE, Neilan BA, Mercer TR. 2018. Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nature Communications 9:3096.
31. Chiu CY, Miller SA. 2019. Clinical metagenomics. Nature Reviews Genetics 20:341-355.
32. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6:226.
33. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, Shah MP, Richie MB, Gorman MP, Hajj-Ali RA, Calabrese LH, Zorn KC, Chow ED, Greenlee JE, Blum JH, Green G, Khan LM, Banerji D, Langelier C, Bryson-Cahn C, Harrington W,

Lingappa JR, Shanbhag NM, Green AJ, Brew BJ, Soldatos A, Strnad L, Doernberg SB, Jay CA, Douglas V, Josephson SA, DeRisi JL. 2018. Chronic Meningitis Investigated via Metagenomic Next-Generation Sequencing. JAMA Neurol 75:947-955.
34. Zinter MS, Mayday MY, Ryckman KK, Jelliffe-Pawlowski LL, DeRisi JL. 2019. Towards precision quantification of contamination in metagenomic sequencing experiments. Microbiome 7:62.
35. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, Stryke D, Pham E, Fung B, Bolosky WJ, Ingebrigtsen D, Lorizio W, Paff SM, Leake JA, Pesano R, DeBiasi R, Dominguez S, Chiu CY. 2019. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. Genome Res 29:831-842.
36. Fancello L, Raoult D, Desnues C. 2012. Computational tools for viral metagenomics and their application in clinical research. Virology 434:162-174.
37. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. 2017. Microbial strain-level population structure and genetic diversity from metagenomes. Genome Research 27:626-638.
38. Quince C, Delmont TO, Raguideau S, Alneberg J, Darling AE, Collins G, Eren AM. 2017. DESMAN: a new tool for de novo extraction of strains from metagenomes. Genome Biology 18:181.
39. Pasolli E, Schiffer L, Manghi P, Renson A, Obenchain V, Truong DT, Beghini F, Malik F, Ramos M, Dowd JB, Huttenhower C, Morgan M, Segata N, Waldron L. 2017. Accessible, curated metagenomic data through ExperimentHub. Nature Methods 14:1023.
40. Lan R, Reeves PR. 2002. Escherichia coli in disguise: molecular origins of Shigella. Microbes and Infection 4:1125-1132.
41. Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533:543.
42. Konstantinidis KT, Rosselló-Móra R. 2015. Classifying the uncultivated microbial majority: A place for metagenomic data in the Candidatus proposal. Systematic and Applied Microbiology 38:223-230.
43. Taylor JW. 2011. One Fungus = One Name: DNA and fungal nomenclature twenty years after PCR. IMA fungus 2:113-120.
44. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. 2017. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun 8:1784.
45. Kelly BJ, Gross R, Bittinger K, Sherrill-Mix S, Lewis JD, Collman RG, Bushman FD, Li H. 2015. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. Bioinformatics 31:2461-8.
46. La Rosa P, Zhou Y, Sodergren E, Weinstock G, D. Shannon W. 2015. Hypothesis Testing of Metagenomic Data, p 81-96 doi:10.1016/B978-0-12-410472-3.00006-3.
47. Li H, Gao H, Meng H, Wang Q, Li S, Chen H, Li Y, Wang H. 2018. Detection of Pulmonary Infectious Pathogens From Lung Biopsy Tissues by Metagenomic Next-Generation Sequencing. Front Cell Infect Microbiol 8:205.
48. Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R. 2016. Tiny microbes, enormous impacts: what matters in gut microbiome studies? Genome biology 17:217-217.
49. Sze MA, Schloss PD. 2016. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. mBio 7:e01018-16.
50. Odintsova V, Tyakht A, Alexeev D. 2017. Guidelines to Statistical Analysis of Microbial Composition Data Inferred from Metagenomic Sequencing. Curr Issues Mol Biol 24:17-36.
51. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork P, Sinha R. 2016. Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. PLoS One 11:e0155362.
52. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, Conrad M, Collman RG, Baldassano R, Bushman FD, Bittinger K. 2017. Optimizing methods and dodging pitfalls in microbiome research. Microbiome 5:52.
53. Duhe RJ. 2011. Koch's Postulates, p 1957-1959. *In* Schwab M (ed), Encyclopedia of Cancer doi:10.1007/978-3-642-16483-5_3242. Springer Berlin Heidelberg, Berlin, Heidelberg.

54. Hill AB. 1965. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? Proceedings of the Royal Society of Medicine 58:295-300.
55. Fedak KM, Bernal A, Capshaw ZA, Gross S. 2015. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerging themes in epidemiology 12:14-14.
56. Evans AS. 1976. Causation and disease: the Henle-Koch postulates revisited. Yale J Biol Med 49:175-95.
57. Rivers TM. 1937. Viruses and Koch's Postulates. J Bacteriol 33:1-12.
58. Falkow S. 1988. Molecular Koch's postulates applied to microbial pathogenicity. Rev Infect Dis 10 Suppl 2:S274-6.
59. Fredricks DN, Relman DA. 1996. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. Clin Microbiol Rev 9:18-33.
60. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. Curr Opin Virol 2:63-77.
61. Neville BA, Forster SC, Lawley TD. 2018. Commensal Koch's postulates: establishing causation in human microbiota research. Curr Opin Microbiol 42:47-52.
62. Vonaesch P, Anderson M, Sansonetti PJ. 2018. Pathogens, microbiome and the host: emergence of the ecological Koch's postulates. FEMS Microbiol Rev 42:273-292.
63. Lipkin WI. 2013. The changing face of pathogen discovery and surveillance. Nature Reviews Microbiology 11:133.
64. Granerod J, Cunningham R, Zuckerman M, Mutton K, Davies NW, Walsh AL, Ward KN, Hilton DA, Ambrose HE, Clewley JP, Morgan D, Lunn MP, Solomon T, Brown DW, Crowcroft NS. 2010. Causality in acute encephalitis: defining aetiologies. Epidemiology & Infection 138:783-800.
65. Granerod J, Tam CC, Crowcroft NS, Davies NWS, Borchert M, Thomas SL. 2010. Challenge of the unknown: A systematic review of acute encephalitis in non-outbreak situations. Neurology 75:924-932.
66. Granerod J, Cunningham R, Zuckerman M, Mutton K, Davies NWS, Walsh aL, Ward KN, Hilton Da, Ambrose HE, Clewley JP, Morgan D, Lunn MP, Solomon T, Brown DWG, Crowcroft NS. 2010. Causality in acute encephalitis: defining aetiologies. Epidemiology and infection 138:783-800.