CENTRO DE CIÊNCIAS MATEMÁTICAS E DA NATUREZA
INSTITUTO DE QUÍMICA

Processo Seletivo para o Curso de Mestrado do Programa de Pós-Graduação em Bioquímica 2021-1

Artigos e/ou revisões científicas a serem abordados durante a prova de conhecimentos em Bioquímica:

(1) Proteínas: estrutura, função e fracionamento

Geyer PE, Mann SP, Treit PV, Mann M. (2021) Plasma proteomes can be re-identifiable and potentially contain personally sensitive and incidental findings. *Mol Cell Proteomics.* **20:** 100035. doi: 10.1074/mcp.RA120.002359.

(2) Enzimas: regulação e cinética de Michaelis-Menten

Nemestóthy N, Megyeri G, Bakonyi P, Lakatos P, Koók L, Polakovic M, Gubicza L, Bélafi-Bakó K. (2017) Enzyme kinetics approach to assess biocatalyst inhibition and deactivation caused by [bmim][Cl] ionic liquid during cellulose hydrolysis. *Bioresour Technol.* **229:** 190-195. doi: 10.1016/j.biortech.2017.01.004.

(3) Fluxo da informação gênica e técnicas de biologia molecular

Yoo HM, Kim IH, Kim S. (2021) Nucleic Acid Testing of SARS-CoV-2. *Int J Mol Sci.* **22(11):** 6150. doi: 10.3390/ijms22116150.

(4) Metabolismo: bioenergética, glicólise, ciclo do ácido cítrico e fosforilação oxidativa.

Vaupel P, Schmidberger H, Mayer A. (2019) The Warburg effect: essential part of metabolic reprogramming and central contributor to cancer progression. *Int J Radiat Biol.* **95(7):** 912-919. doi: 10.1080/09553002.2019.1589653.

Plasma Proteomes Can Be Reidentifiable and Potentially Contain Personally Sensitive and Incidental Findings

Authors

Philipp E. Geyer, Sebastian Porsdam Mann, Peter V. Treit, and Matthias Mann

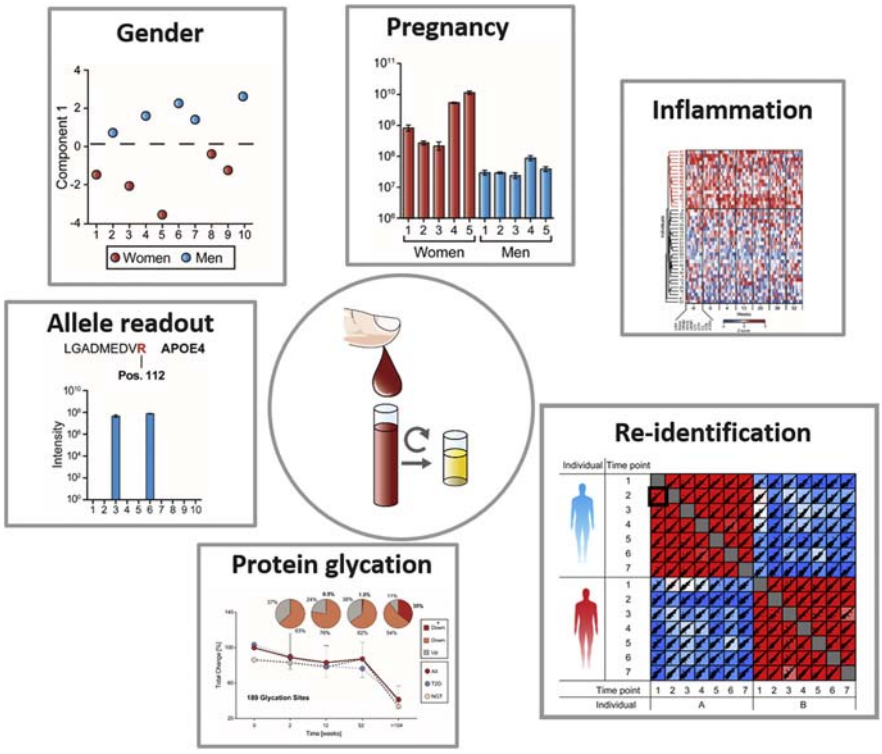
Correspondence

geyer@biochem.mpg.de;
mmann@biochem.mpg.de

In Brief

Due to its unbiased nature, proteomics may raise ethical and regulatory concerns. Plasma proteome samples can be reidentified given genomic information. Furthermore, plasma proteomes contain information of potential sensitive nature. Incidental findings can help diagnose unrelated but actionable disease states.

Graphical Abstract



Highlights

- Plasma proteomes can be reidentified based on protein expression levels.
- Plasma proteomes can be reidentified based on variant peptides.
- Plasma proteomes can contain actionable and nonactionable incidental findings.
- Plasma proteomes can contain sensitive information such as pregnancy status.

2021, Mol Cell Proteomics *20*, 100035

https://doi.org/10.1074/mcp.RA120.002359

Plasma Proteomes Can Be Reidentifiable and Potentially Contain Personally Sensitive and Incidental Findings

Philipp E. Geyer[1,2,3,*], Sebastian Porsdam Mann[4,5], Peter V. Treit[1], and Matthias Mann[1,2,*]

The goal of clinical proteomics is to identify, quantify, and characterize proteins in body fluids or tissue to assist diagnosis, prognosis, and treatment of patients. In this way, it is similar to more mature omics technologies, such as genomics, that are increasingly applied in biomedicine. We argue that, similar to those fields, proteomics also faces ethical issues related to the kinds of information that is inherently obtained through sample measurement, although their acquisition was not the primary purpose. Specifically, we demonstrate the potential to identify individuals both by their characteristic, individual-specific protein levels and by variant peptides reporting on coding single nucleotide polymorphisms. Furthermore, it is in the nature of blood plasma proteomics profiling that it broadly reports on the health status of an individual—beyond the disease under investigation. Finally, we show that private and potentially sensitive information, such as ethnicity and pregnancy status, can increasingly be derived from proteomics data. Although this is potentially valuable not only to the individual, but also for biomedical research, it raises ethical questions similar to the incidental findings obtained through other omics technologies. We here introduce the necessity of—and argue for the desirability for—ethical and human-rights-related issues to be discussed within the proteomics community. Those thoughts are more fully developed in our accompanying manuscript. Appreciation and discussion of ethical aspects of proteomic research will allow for deeper, better-informed, more diverse, and, most importantly, wiser guidelines for clinical proteomics.

Omics technologies can characterize biological materials, leading to a wealth of information useful for addressing a broad range of scientific and medical questions. Genomics has benefited from rapid technological progress over many decades, and large-scale DNA analysis now increases our knowledge of genetic diversity and the relation of genes to various phenotypical and disease-relevant traits (1). Genomic data are usually acquired in a broad and untargeted manner. Typically, many genes are assayed simultaneously. Some of these could contain individually identifiable and sensitive information, raising several ethical questions. The relatively long-lasting and widespread application of genomics in research and medicine has led to numerous and sometimes acrimonious debates concerning what to do with such information, which in turn has resulted in guidelines and frameworks (2–4). However, analogous discussions have not yet been fielded in proteomics or metabolomics. Partly this is because the ability to analyze a large number of human samples at great proteomic depth is a comparatively new and resource-intensive development, which as of this writing still requires highly specialized technology.

In contemporary medical practice, the majority of diagnostic decisions are based on tests quantifying biological parameters, generally referred to as “biomarkers” (5). The foremost bodily fluids used for this purpose are the blood (whole, plasma, or serum), urine, and cerebrospinal fluid (CSF). Of the various classes of clinically measured parameters such as cells, electrolytes, DNA, RNA, and small molecules, most tests target proteins (6). Thousands of these proteins circulate throughout the body and their levels report on a wide variety of systemic diseases, organ damage, and general health status (7–9).

From an analytical biochemistry perspective, the biomarker discovery task has traditionally consisted of accurately measuring the level of one or a few proteins in disease and control cohorts and then developing a robust clinical test—typically involving an antibody against a specific protein. Considering the fundamental role of laboratory tests, it is remarkable that most biomarkers were discovered more than 20 years ago and often lack specificity or sensitivity and that

From the [1]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany; [2]Faculty of Health Sciences, NNF Center for Protein Research, [4]Department of Media, Cognition and Communication, University of Copenhagen, Copenhagen, Denmark; [3]OmicEra Diagnostics GmbH, Planegg, Germany; [5]Uehiro Center for Practical Ethics, Oxford University, Oxford, UK

This article contains supporting information.

*For correspondence: Philipp E. Geyer, geyer@biochem.mpg.de; Matthias Mann, mmann@biochem.mpg.de.

Present address for Sebastian Porsdam Mann: Faculty of Law, University of Oxford, Oxford, UK.

https://doi.org/10.1074/mcp.RA120.002359

only a very small number of the more than 14,000 human diseases have corresponding tests (10).

Proteins control and execute the vast majority of biological processes, and mass spectrometry (MS)-based proteomics is the technology of choice to investigate the entirety of all proteins in a biological system—its proteome (11, 12). MS-based proteomics has continuously developed over the past 20 years, enabling the holistic, detailed, and quantitative investigation of diverse biological systems. Discovery of new and potentially better biomarkers and biomarker panels is facilitated by these rapid developments in MS-based proteomics, promising widespread medical applications (6, 13).

Traditionally, proteomics biomarker studies have analyzed small sample numbers in depth to discover one or a few potential biomarkers that were then to be validated in larger cohorts. In contrast, we have developed a "rectangular approach for biomarker discovery," in which all samples from large-scale cohorts are analyzed in as much depth as possible in discovery and validation cohorts together (6). The goal is to derive panels of regulated proteins that contain much more information than is reflected in the level of any single protein. Over the last few years, we have focused on technological developments enabling the analysis of large-scale plasma proteomic cohorts with a robust and automated pipeline (14) and have already analyzed clinical studies with over a thousand samples (15–17).

It will soon be possible to collect large-scale and increasingly comprehensive proteomics data sets, and this ability will not long be restricted to only a few specialized laboratories. Clinical proteomics is already benefitting from rapid advances in information technologies, including machine learning and big data analytics. The information that can be extracted from such powerful data sets has myriad applications. Clearly, the expanding data volume, scope, and quality of clinical studies analyzed by MS-based proteomics raise ethical issues, as they have in other fields. For instance, a plasma proteomic measurement may be used to uniquely identify a person if matching genomic information is available. In addition, such a measurement may contain incidental findings—findings unrelated to the primary aim of the study or procedure, but which may still contain information relevant to health or well-being. This potential is well recognized in other diagnostic fields. In a meta-analysis of tens of thousands of asymptomatic persons receiving body or brain MRI revealed a potentially serious incidental findings rate of 3.9% and another found a median clinically significant rate of 17% (18, 19). It has been pointed out that incidental findings, "which can occur in large numbers from genomic sequencing, are a potential barrier to the utility of this new technology due to their relatively high prevalence and the lack of evidence or guidelines available to guide their clinical interpretation" (20).

We argue that clinical proteomics, too, will soon face these challenges. Researchers and clinicians will have to deal with personally sensitive and incidental findings. How they should go about this task is a discussion that must be had. It is also a discussion that involves applied ethics and bioethical principles—topics that the proteomics community is currently not prepared for. As a first step in this direction and to enable our community to begin an open discussion about proteome ethics, we describe some ethical implications of proteomics data. Below, we reanalyze a previous plasma proteomics study (15) to investigate whether and to what extent ethically relevant information can be extracted. We show that samples can uniquely be assigned to individuals by both the individual-specific levels of plasma proteins and their individual-specific allele variations (coding SNPs). Furthermore, plasma proteomes inherently report on a broad range of health and disease parameters such as cardiovascular and metabolic risks. We discuss what would have to be stripped from results obtained in order only to retain narrowly disease-relevant information. While this might be necessary in certain diagnostic settings, it also negates one of the principal attractions of plasma proteomics. In an accompanying paper (21), we introduce bioethical and human rights principles that instead argue for deriving the maximum information and therefore benefit from clinical proteomics data for research, disease diagnosis, and general health and well-being.

EXPERIMENTAL PROCEDURES

Identification of Individual-Specific Alleles

All peptides from the variant FASTA file analysis were filtered to generate a set of reliable peptides suitable to separate individuals. For this purpose, we filtered the data set for peptides that were present at least once in six out of the seven time points in at least one individual. We excluded all peptides that were always or never identified as they do not contain information that could be used to distinguish between individuals. Next, we filtered for peptides that had at least one overlapping peptide from another allele. This resulted in 83 peptides. Peptides containing a missed cleavage site of Arginine or Lysine will contain the same information as their fully cleaved form. Hence, the information of the presence of the allele was only counted once, resulting in a set of 67 peptides. This set also contains alleles that were very randomly distributed with very high variation. The peptides with the highest variation were excluded. If the number of the identified peptides in the study was ten times larger than the sum of the peptides that were identified six or seven times within an individual, both alleles of the same gene were excluded, resulting in 53 peptides for the analysis.

Data Analysis

MS raw files were analyzed by MaxQuant software, version 1.6.1.9 (22), and peptide lists were searched against the human Uniprot FASTA databases. A regular FASTA file was downloaded from the UniProt database in May 2019 (https://www.uniprot.org/). Variant sequence entries were downloaded in text format from the UniProt database in May 2019 (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/taxonomic_divisions/). The Swissknife PERL module (http://swissknife.sourceforge.net/docs/) with the varsplic PERL script from ftp://ftp.ebi.ac.uk/pub/software/uniprot/varsplic/varsplic.pl was applied to generate the variant text formats for single sequences. The output produced includes the sequence for the variants.

A contaminant database generated by the Andromeda search engine (23) was configured with cysteine carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate (FDR) to 0.01 for protein and peptide levels with a minimum length of seven amino acids for peptides, and the FDR was determined by searching a reverse database. Enzyme specificity was set as C terminal to arginine and lysine as expected using trypsin and LysC as proteases. A maximum of two missed cleavages were allowed. Peptide identification was performed with an initial precursor mass deviation up to 7 ppm and a fragment mass deviation of 20 ppm. All proteins and peptides matching the reversed database were filtered out.

All bioinformatics analyses were performed with the Perseus software of the MaxQuant computational platform (22, 24).

RESULTS

Individuals Can Be Identified by Protein Levels in Blood Plasma

In clinical studies, samples are usually blinded and pseudonymized in order to avoid bias by the experimenter and to protect study participants from potentially sensitive findings. In this way, results from a sample cannot directly be tied to a person, *i.e.*, unusually high cholesterol levels could not be used to deny insurance. In our previous work, we noticed that the levels of hundreds of plasma proteins varied much more between participants than within the same participant over time (15). We therefore speculated that the "individual-specific" levels of many plasma proteins together could enable association of a given plasma proteome sample to a previous measurement on the same person. Although this would prevent the pervasive problem of sample mix-up (25), it could conceivably raise ethical issues with regard to reidentification of participants.

To test our hypothesis, we reinvestigated a plasma proteomics weight loss study in which samples of 42 individuals were obtained over 1 year (15). We defined a protein as "individual-specific" if its level in a participant was more than 1.5-fold different from the population median for at least a quarter of all study participants, and it had a coefficient of variation below 20% over time. In our study, 71% of all proteins fulfilled these criteria. When we compared the levels of individual-specific proteins over time, we found that these intra-individual correlations were much higher than the correlations of the same or different time points between two different persons (91 different correlation values; median Pearson R = 0.971 within and 0.926 between the two individuals in the example) (Fig. 1, *A* and *B*).

Global correlation of the 294 weight loss samples resulted in a matrix of 43,071 values (Fig. 1*C*). Pearson correlation coefficients of these individual-specific proteins had much higher intraindividual correlations even over a whole year, compared with interindividual correlations (Fig. 1*B*). Across the entire study, the median intraindividual correlation was 0.974 and the interindividual correlation was 0.928.

We next tried to identify individuals solely by the Pearson correlation coefficients of their individual-specific proteins. For this purpose, we defined the preweight loss time point as a reference and asked whether the others could be uniquely related to it *via* their plasma proteome correlations. Out of these 252 comparisons, all but one were assigned to the correct individual (Error rate of 0.4%; Fig. 1*D*). Even in the one misassignment (possibly caused by experimental issues; Experimental Procedures), the correct individual was ranked second. Moreover, highly individual-specific proteins such as the apolipoprotein(a) can be more than 100-fold different between individuals, but very constant over time. In the case of apolipoprotein(a), this can be explained by the genetically determined number of so-called Kringle domains affecting the concentration of this protein. Such proteins have a higher value for identifying or excluding an individual. In general, quantitative trait loci (pQTLs) link protein levels and genetic variants and explain part of the individual-specific protein levels in addition to lifestyle and life history (26). However, many proteins have individual-specific levels that are modifiable by lifestyle changes, disease, medication, and even preanalytical processing, adding uncertainty regarding identifiability (27–29).

Individuals Can Be Identified by Allelic Information

MS-based proteomics identifies peptides by matching experimental to theoretical spectra calculated from protein sequence databases. Therefore, the proteomics community relies heavily on protein sequence data and associated metadata supplied by consortia such as UniProt (30). As a default to reduce redundancy, UniProt provides all the proteins encoded by one gene as a "canonical sequence" that is usually the most prevalent or likely form. These can be downloaded as FASTA files (human, May 2019). In the case of well-studied species, UniProt provides additional information about variant sequences, including single amino acid polymorphisms from nonsynonymous single nucleotide polymorphisms (SNPs) or polymorphisms from multinucleotide exchange. We generated a human variant FASTA file from UniProt enabling us to identify polymorphisms using the Swissknife PERL module (Experimental Procedures).

In principle, the generated data could uniquely identify individuals using the combination of variant peptides in different proteins. They could also be used to link such variants to disease risks by connecting proteins to the UniProt–SwissProt protein knowledgebase. It currently contains a "human polymorphisms and disease mutations index," describing 30,706 disease variants, 40,091 polymorphisms without disease implications, and 8085 variants of uncertain medical significance.

To assess a database search with the variant FASTA (a sequence file derived from the database), we revisited the weight loss study data set again. The complete study resulted in 5888 and 6094 peptides when searching against the canonical and the variant FASTA file, respectively, containing 134 and 340 unique peptides. The unique sequences for the

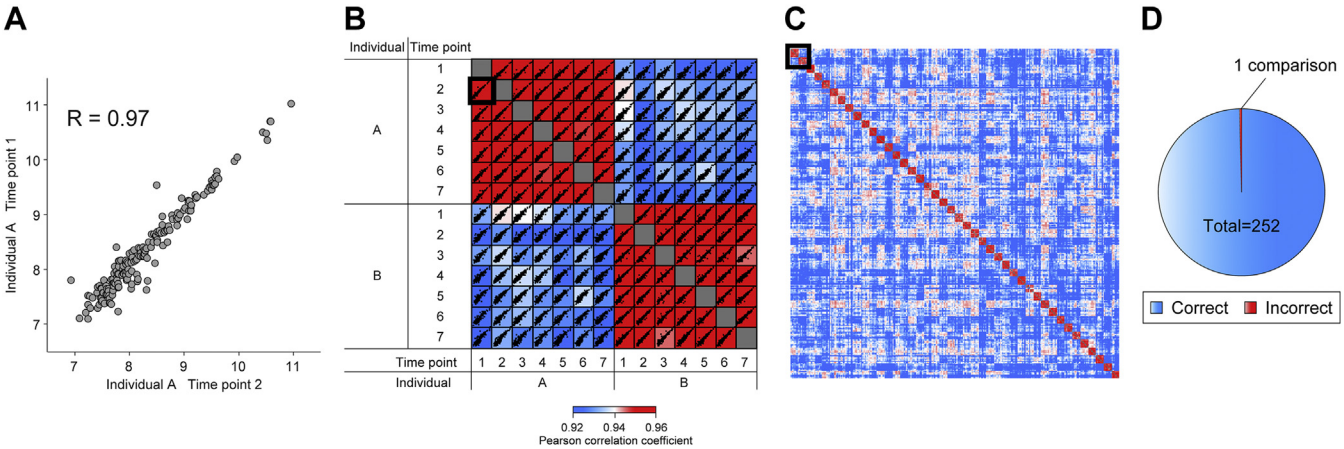


FIG. 1. **Identifying participants in a longitudinal study by correlation of individual-specific proteins.** *A*, correlation of individual-specific proteins of time points 1 and 2 of individual *A*. *B*, seven longitudinal samples of two individuals, *A* and *B*, are correlated with each other (Pearson correlation coefficient is color-coded, with color bar below). The comparison displayed in (*A*) is highlighted by a *black frame*. *C*, cross-correlated individual-specific proteins of all samples of the weight loss study. The correlation matrix shown in (*B*) is highlighted by a *black frame* (Pearson value coded according to the same color bar). *D*, identification of individuals by correlating individual-specific proteins. Proteomes of the reference time point were compared with all other time points in turn. The percentage of correctly and incorrectly assigned participants is color-coded.

canonical FASTA might be explained by the larger search space of the variant FASTA and a general variation due to the MaxQuant search algorithm.

Similar to the unblinding experiment described above, we aimed to extract a panel of peptides from the variant FASTA file experiment, which would allow us to gain additional confidence in the identification of individuals. Peptides were only considered if at least two alleles represented by one peptide for each allele were identified for a protein, resulting in 83 peptides. Additionally, we combined information concerning allele-specific peptides with missed cleavages with the fully cleaved peptide. We further applied a filter step to exclude peptides with very strong variation in their identification (Experimental Procedures).

This resulted in a set of 53 peptides of which the allele pattern can be seen across the 42 individuals and the seven time points of the weight loss study (Fig. 2*A*; supplemental Table S1). Next, we identified the study participants by a simple calculation of matching present and absent variant peptides, utilizing the allele information contained in their proteome. Figure 2*B* shows the comparison of each sample to all time point (TP) 1 samples. The median number of matches between two samples within the same individual was 46 and 37 between different individuals. One individual strongly separated from the rest of the population with a median of only 33 matches to the other individuals (blue vertical line in the heat map, Fig. 2*B*). The correct samples were identified in 89% of all comparisons (Fig. 2*C*). In 4998 of 5166 pairwise comparisons matching was correct (1.5% error rate).

The semistochastic sampling of peptides—especially in data-dependent acquisition (dda)—results in missing values of peptides, which decrease the probability of correctly reidentifying an individual. It follows that both data-independent acquisition (dia) and greater data completeness (for instance, in tissue measurements) would result in higher certainty. Note that our calculations are only a proof-of-principle at this point and more advanced experiments including larger and more representative populations have to be done. These issues are and the influence of laboratory errors, which are a frequent issue in DNA technology in forensic science, will have to be taken into account to calculate the true likelihood of reidentifying an individual by variant peptides (31, 32).

Untargeted Plasma Proteomics Delivers Incidental Diagnostic Findings

As clinical proteomics generates a broad overview of protein levels in a sample, it typically reports on many more conditions than the one under investigation, making "incidental findings" an inherent feature of the technology. If they are related to diseases, this may be sensitive information that raises ethical issues. When dealing with incidental findings, a clear line needs to be drawn between the medical benefits that can be obtained through their return and the principle of respecting an individual's capacity to choose for themselves whether they wish to have information returned or not. The latter is obtained by informed consent prior to participation in a study or a medical test.

Even short 20 min MS-based proteomics measurements quantify about 50 proteins that were approved by the U.S. Food and Drug Administration (FDA) as biomarkers, resulting in a multilayer reflection of the human health state (14). These MS-quantified markers include C-reactive protein (CRP), which reports primarily on inflammation and is one of the most frequently requested protein measurements in clinical practice. The coregulated protein serum amyloid alpha 1 (SAA1) shows highly similar fold changes upon infection and is

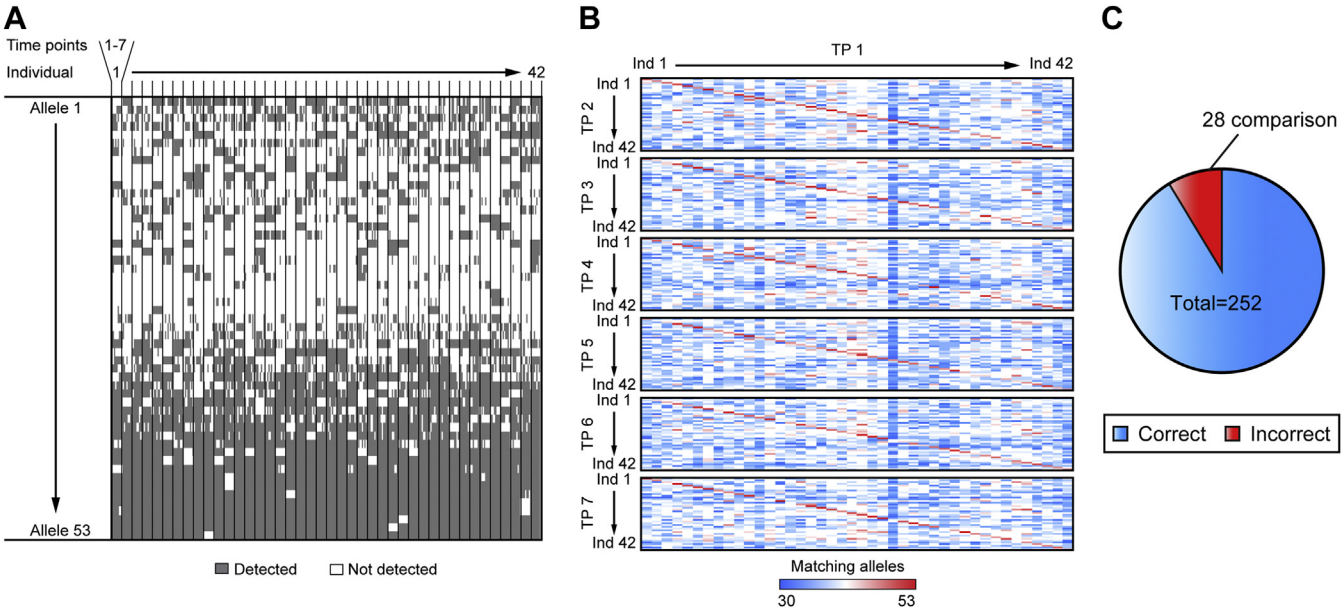


FIG. 2. **Allele information in the plasma proteome.** *A*, the *grayscale* indicates which peptides were detected across the seven time points (TP) for the 42 individuals in the weight loss study of the 53 variants considered. *B*, heat map for the number of matching present or absent variant peptides between individuals at the first time point (TP 1) and all individuals on the other time points. The *red diagonal lines* reflect the high number of matches of the same individual in adjacent time points. *C*, proportion of correctly identified individuals in the 252 comparisons.

likewise covered (Fig. 3*A*), resulting in a more inclusive reflection of an individual's inflammation status than the routine assessment of CRP alone. While an infection is usually short-term and benign, other MS-identifiable proteins may indicate life-threatening diseases, such as the cancer marker MUC16.

The health consequences of diabetes, one of the leading causes of death worldwide, can be minimized if the disease is diagnosed early and managed properly (33). In clinical practice, diagnosis of diabetes involves measuring HbA1c, the glycated form of hemoglobin. The concentration-dependent Maillard reaction of sugars with the amine groups of proteins is responsible for this glycation. In addition to hemoglobin, all other blood proteins and in particular the long-term circulating high-abundant ones are also glycated. In our experience, hundreds of glycated peptides are easily detectable in all plasma proteome profiling experiments (14, 16, 34). Their levels are an indicator of prediabetic or diabetic status, which is particularly relevant as up to a third of the population that has these conditions are not aware of it.

Apart from acute disease, proteomics covers proteins connected to the risk of a future disease. The lipid homeostasis system is covered by more than 20 factors (16), several of which—including apolipoprotein A1 (APOA1), apolipoprotein B (APOB), and apolipoprotein(a) (LPA)—are important predictors of cardiovascular diseases (35). MS-derived APOB intensities correlate to LDL-levels as shown in a data set of 142 samples in our weight loss study (Fig. 3*B*). This indicates that these apolipoproteins can be used to determine risk of cardiovascular diseases similarly to cholesterol, for which they are carriers. It is well known that decreasing high cholesterol levels by lifestyle changes or medication is beneficial for health outcomes (15, 35). Thus, an apolipoprotein panel derived from plasma proteomics likely provides at least the level of actionable health information than a routine cholesterol test. Furthermore, as individuals may respond quite differently to various treatments as shown in the regulation of APOB levels upon weight reduction (Fig. 3*C*), proteomics may provide more individualized and detailed information. This is an example where the risk is known and treatment options to significantly reduce future medical conditions may be significantly improved by proteomics.

MS-based proteomics can also report on risks for which no treatment option is currently available, making them "non-actionable." For instance, the three APOE alleles—APOE2, APOE3, and APOE4, can be differentiated by sequence specific peptides. Knowledge concerning the status of the APOE4 allele (7.5–15.6% of the population), which strongly increases risk of Alzheimer's disease, is medically unactionable information as there is currently no available treatment (36). Return of such severe unactionable information could leave some individuals with psychological trauma, frustrate others, and might potentially negatively influence future personal decisions in a negative manner. The APOE2 allele (6.7–10.0% of the population) increases cholesterol levels and cardiovascular pathologies (36). In contrast to APOE4, this knowledge is actionable and could lead to the decision to take cholesterol lowering medication or dietary interventions to decrease

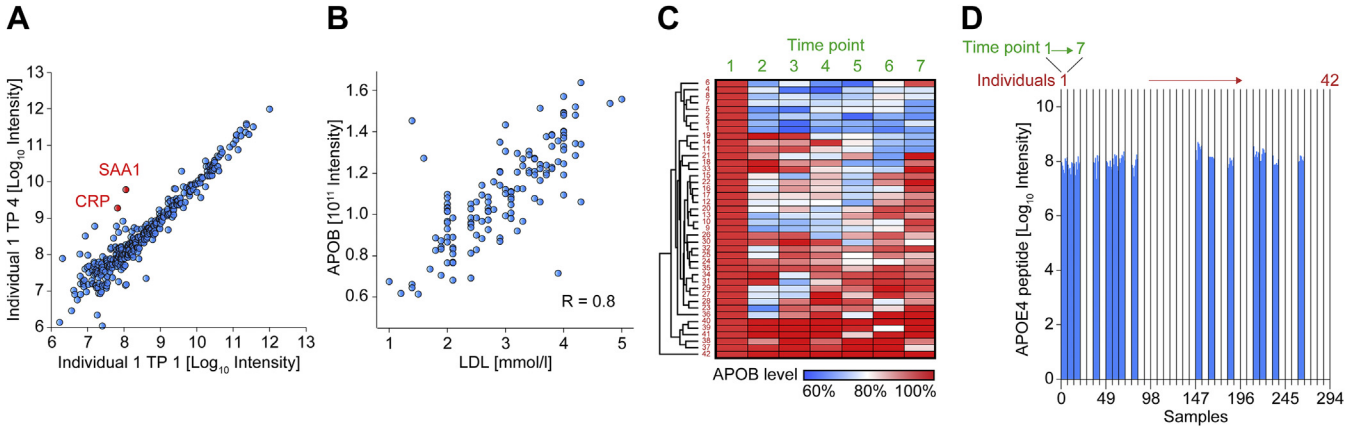


FIG. 3. **Disease diagnostics and disease risk assessment relevant proteins.** *A*, comparison of the plasma proteomes of one individual, indicating an infection at time point 4 (TP 4). *B*, correlation of LDL levels and proteomic measurements of APOB. *C*, hierarchical clustering illustrating individual-specific responses of APOB levels to weight loss and weight maintenance over seven time points (1–7). *D*), intensities of the quantified APOE4 allele determining peptide across individuals.

cardiovascular disease risk. There are therefore much more persuasive reasons for the return of information concerning APOE2 status.

Untargeted Plasma Proteomics Delivers Personally Sensitive Findings

The recognition of the equal worth and dignity of all members of the human family as enshrined in universally adopted and near-universally accepted normative and legal codes globally represents the outcome of centuries of struggle for legal and moral equality before the law and society. These efforts aim at stopping and preventing discrimination, which refers to unequal treatment on the basis of various morally irrelevant attributes such as gender, race, color, or national or ethnic origin, all of which still occur in both explicit and implicit forms in modern society. Therefore, it is morally noteworthy that one can distinguish the proteomes of men and women and of reproductive status (14). Comparing the longitudinal proteomes of women and men of our weight loss study revealed that the levels of estrogen-regulated proteins such as the sex hormone binding globulin (SHBG) and the pregnancy zone protein (PZP) were significantly elevated in women (Fig. 4*A*). While the levels of these proteins depend on several additional factors such as age and lifestyle, they are clear indicators of gender. Applying a one-dimensional principal component analysis clustered almost all women and men separately, although one of the eight men and one of the 34 women were assigned to the opposite clusters due to untypical PZP levels (Fig. 4, *B* and *C*). It is known that SHBG and especially PZP increase more than tenfold during pregnancy. Additionally, there are highly specific measures for determining pregnancy by quantifying placental proteins such as the family of pregnancy-specific glycoproteins (PSGs), which are the most abundant trophoblastic proteins in maternal blood during pregnancy (37, 38), and they are in our experience readily detectable by MS-based proteomics. Their concentration can increase over 1000-fold, even exceeding placental peptide hormone human chorionic gonadotropin (hCG), which is usually determined in pregnancy tests. This is significant given that pregnancy status is information that individuals often seek to keep private, as much of the discrimination faced by women is related to the unique status of (potential) motherhood.

Given that body weight is connected to discrimination, we investigated the weight loss study for pertinent markers. Interestingly, SHBG and proteins of the innate immune system that we readily quantified are among the proteins most affected by body mass (15, 16). Although individual-specific differences, acute inflammation, and gender might influence these parameters, they could conceivably be used to predict a person's weight based on their plasma proteome or more pertinently likely negative health effects for that person.

Similarly, it is also possible to discern information regarding the ethnic background of an individual using alleles in a similar way as described above (39). Furthermore, ethnic differences can complicate analytical results. A prominent example of this is the abundant plasma protein vitamin D-binding protein (GC), which has three common alleles Gc1f, Gc1s, and Gc2 with very different allele distribution depending on ethnic background. Gc1f is most frequent in West Africans and African Americans and least common in Caucasians (40, 41). All three alleles should result in potential MS-detectable peptides, and we identified the Gc1f peptide in nine and the Gc2 peptide in 21 of 42 study participants at each of the seven time points (Fig. 4*D*). However, the third peptide was not detectable, presumably due to poor ionization.

How to Render the Plasma Proteome Ethically Unproblematic and General Data Protection Regulation (GDPR) Compliant

The legal act of the European Union for the protection of personal information—the General Data Protection Regulation

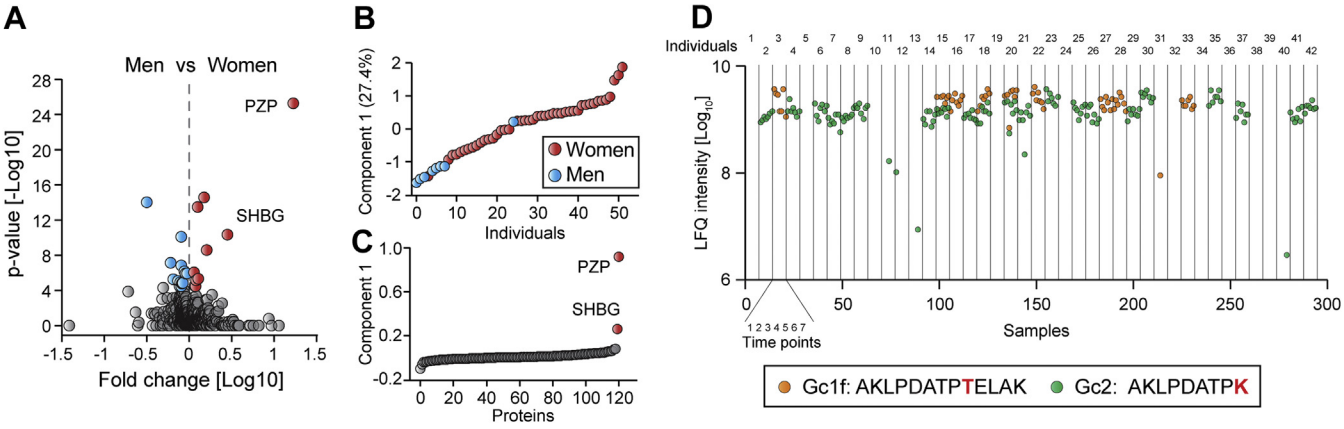


FIG. 4. **Information with the potential to discriminate individuals.** *A*, comparing the plasma proteomes of all individuals and samples in the weight loss study (15). Proteins with elevated levels in women and men are highlighted in *red* and *blue*, respectively. *B*, one-dimensional principal component analysis for plasma samples at one time point. *C*, proteins and their distribution to the separation in (*B*). *D*, vitamin D-binding protein determining alleles quantified by MS-based proteomics.

(GDPR)—regulates data protection and privacy with one of the central aims of giving control of personal data to the corresponding individuals (Directive 95/46/EC, (42)). Implemented in 2018 across the EU, GDPR supplies rules relating to the protection of an individual with regard to the processing of personal data and to the movement of personal data. It also regulates flow of information into "third countries," outside the sphere of GDPR. In the United States, the HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule is a similar directive on data protection. However, HIPAA focuses on the protection of individually identifiable health information, whereas GDPR is a regulation on all data that can be related to a person. Furthermore, there is an international trend toward GDPR style rules across many jurisdictions.

Study participants and patients must be informed about the potential information content of their proteome in advance of study participation. Ideally, they would be given the opportunity to state whether and to what extent their data can be used for the benefit of research and third parties. Furthermore, individuals should be informed about the potential and range of incidental findings and their preferences as to return of such information should be ascertained. However, in studies this is only possible under specific circumstances, for example, where data is pseudonymized or nonanonymized instead of anonymized. However, nonanonymization of samples and data collides with the protection of personal data as the proteome contains sensitive information about individuals and therefore adequate rules have to be promulgated and adhered to when working with this kind of data in research and clinical contexts. In genomics, samples and data from individuals have to be pseudonymized before analysis to comply with the GDPR. Pseudonymization still allows for the linking of an individual to their data under specific conditions and would act as a safeguard against third-personal access to personal and sensitive information. In contrast, this would not be possible in anonymized data and would hinder to link new studies and their results to previous findings.

Clearly, individuals have to decide to what extent their samples and personal data should be processed, and they have to be aware of the range of information that can be extracted. Informed consent must be given, which can happen either in an opt-out or opt-in modality. The choice between these will have a significant impact on the extent of information available for research and processing (see our accompanying paper, (21)). These issues exist along spectra of severity, such that their associated ethical questions do not admit of binary answers and degrees of likelihood, risks, benefits, and harms must be considered. Next to organizational measures and access to the samples and data, the extent further analysis, storage, and integration of proteomics data for research and clinical purposes are a topic of great importance. In proteomics, mass spectra of peptides are generated and saved in raw data files. In principle, allele-specific peptides could be deleted from the raw data, but this would itself go against good laboratory practice, which strongly discourages data manipulation. However, the sharing of raw data is beneficial due to several reasons including general research reproducibility. Therefore, in the research context instead of having controlled access to raw data, data-sanitization procedures might be applied to have the optimal trade-off between data utilization and privacy protection as it has recently been proposed in genomics (43). To prevent the existence of such data, one could instead direct data acquisition accordingly, which would be difficult or impossible for data-independent acquisition (dia), challenging for data-dependent acquisition (dda) and easily accomplished for targeted proteomics.

Once data is acquired, the interpretable information still depends on the processing workflow. For example, allele information is easily extracted when a sequence database

(FASTA) file with allele information is used. FASTA files without allelic information would render the proteomic readout non-identifiable for many purposes, as protein levels alone cannot identify an individual to the degree of certitude necessary in forensic and similar contexts. Therefore, tightly restricting access to the raw data might be warranted in many circumstances. We suggest that in clinical environments, the handling and analysis of data should be done in an automatic fashion. Access could be secured by technologies such as blockchain, ensuring that only approved information is read out and reported to the patient, depending on their former written consent.

MS-based proteomics is uniquely information-rich because it accurately and quantitatively measures thousands of peptides in clinical samples. However, affinity-based methods are also increasingly applied to the plasma proteome (26). As only a single or few epitopes are probed, there is much less chance of revealing allele-specific information. Furthermore, pre-analytical factor such as freeze–thaw cycles can affect epitope recognition, whereas this does not introduce variations in MS-based methods (29). Thus, identifiability is currently not a particular issue in affinity-based methods, and incidental findings could be stripped out by removing them from the reported protein level measurements.

DISCUSSION

Exploring the plasma proteome in various clinically relevant areas, we have encountered several categories of findings with potential ethical implications. In this paper, we systematically investigated the main ones using a plasma proteome profiling study (15), which served as a prototypical case. Quantitative protein levels allowed us to detect acute diseases and markers for future events that raise ethical issues. Individual-specific levels of plasma proteins enabled us to identify individuals in this longitudinal study. We further showed that MS-based proteomics delivered broad knowledge about allele distribution with the potential to identify individuals, predict disease risks, and detect alleles characterizing ethnic groups. Even though insights based on allele variation are not very extensive in plasma yet, they will become more so through further technological progress. This is already the case for the proteomic analysis of tissue samples because tens of thousands of peptides are routinely measured and large-scale efforts are already on the horizon (44–47). Much remains to be learned about plasma proteome alterations in response to variables such as lifestyle, disease, and medication. If these variables change the plasma proteome, they might complicate the identification of an individual. However, the continuous exploration of the plasma proteome will increasingly allow us to acquire information about plasma proteome modulators and take them into account. Moreover, as technology progresses, it will enable the investigation of increasingly large studies while expanding the information that can be extracted from a single plasma proteome.

As the power of proteomics increases (11), so do insights that may be derived from an individual's sample. Indeed this is necessary for the medical application of proteomics. Yet the hypothesis-free nature of proteome profiling also yields information not directly relevant to the study or intervention at hand, which may nevertheless be relevant for other medical reasons or for other purposes entirely. Since not all uses of this additional information are benign, it is important to discuss how the benefits of additional knowledge can be reaped, maximized, and shared, while avoiding the potential for harm and exploitation which that knowledge may also bring.

These almost exclusively relate to the knowledge that can be obtained from proteomic data and thus have much in common with issues familiar from broader health and research contexts such as data storage, data sharing, and adequate consent. However, the proteomic context is unique, and the progress made in other fields does not necessarily translate easily. Therefore, it is important that the proteomics community is aware of, raises, and discusses these issues to protect data subjects and facilitate research in ethical and legal ways.

We hereby wish to start a discussion within our proteomics community about potential uses of the data in our hands. We believe this is important not only to maximize benefit and minimize harm, but crucially also because our community is best positioned to explore questions not only of an ethical and legal nature but are also intertwined with technical and scientific issues related to the reliability, accuracy, storage, sharing, propriety, and kinds of inferences that may be surmised from proteomics.

We hope this discussion can serve as a first step toward serious scientific self-regulation in the proteomic community, which is acceptable to regulators and society, as has happened with recombinant DNA techniques at the Asilomar conference already in 1975 (48), and as is currently being attempted in the context of editing the human germline (49).

To achieve this, the proteomics community will need to become aware of ethical issues involved and begin to discuss them seriously. Luckily, the effort will not have to be made from scratch as structurally similar issues have been extensively debated in other fields. Nevertheless, these debates are far from settled and the unique research context of proteomics means that potential answers cannot be imported wholesale but rather need to have their fit and applicability assessed and thoroughly discussed.

In conclusion, the increasing power of MS-based proteomics for generating both medical and nonmedical insights brings with it a concomitant increase in the importance of associated ethical issues. We have presented some of these issues by showcasing a single plasma proteomics study. The purpose of this exploratory contribution has been to point out that these issues already exist and to initiate discussion; but more

importantly, we hope to stimulate others to point the attentions of ourselves and the community to issues that are not discussed here. In our accompanying paper, we further elucidate and develop ethical implications from a bioethical and philosophical perspective in the hope of starting a discussion concerning the first potential guidelines for the community: (21).

DATA AND MATERIAL AVAILABILITY

The MaxQuant output files of the searches have been deposited at the ProteomeXchange Consortium *via* the PRIDE partner repository and are available *via* the identifier PXD021677.

Acknowledgments—We thank all members of the Proteomics and Signal Transduction Group for help and discussions and in particular Johannes B. Müller-Reif.

Funding and additional information—The work carried out in this project was partially supported by the Max Planck Society for the Advancement of Science, the European Union´s Horizon 2020 research and innovation program with the MicrobPredict project (no. 825694), by the Novo Nordisk Foundation (NNF15CC0001; NNF15OC0016692), by the DFG project "Chemical proteomics inside us" (no. 412136960), and the Carlsberg Foundation Grant CF16-0925.

Author contributions—P. E. G. designed the experiments, performed and interpreted the MS-based proteomics analysis, did bioinformatics analysis, discussed ethical implications of proteomics data, and generated the text and figures for the manuscript. S. P. D. and P. V. T. designed the experiments and interpreted the proteomics data with respect to ethical aspects, discussed these findings, and generated the text. M. M. supervised and guided the project, designed the experiments, interpreted MS-based proteomics data, discussed ethical implications of proteomics data, and wrote the article.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article

Abbreviations—The abbreviations used are: CRP, C-reactive protein; CSF, cerebrospinal fluid; FDA, Food and Drug Administration; FDR, false discovery rate; GDPR, General Data Protection Regulation; hCG, human chorionic gonadotropin; HIPAA, Health Insurance Portability and Accountability Act; QTL, quantitative trait loci; MS, mass spectrometry; PSG, pregnancy-specific glycoprotein; PZP, pregnancy zone protein; SAA1, serum amyloid alpha 1; SNP, single nucleotide polymorphism.

Received September 25, 2020, and in revised from, December 2, 2020 Published, MCPRO Papers in Press, January 11, 2021, https://doi.org/10.1074/mcp.RA120.002359

REFERENCES

1. Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., and Waterston, R. H. (2017) DNA sequencing at 40: Past, present and future. *Nature* **550**, 345–353
2. Lunshof, J. E., Chadwick, R., Vorhaus, D. B., and Church, G. M. (2008) From genetic privacy to open consent. *Nat. Rev. Genet.* **9**, 406–411
3. McGuire, A. L., Caulfield, T., and Cho, M. K. (2008) Research ethics and the challenge of whole-genome sequencing. *Nat. Rev. Genet.* **9**, 152–156
4. O'Brien, S. J. (2009) Stewardship of human biospecimens, DNA, genotype, and clinical data in the GWAS era. *Annu. Rev. Genomics Hum. Genet.* **10**, 193–209
5. Biomarkers Definitions Working Group. (2001) Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95
6. Geyer, P. E., Holdt, L. M., Teupser, D., and Mann, M. (2017) Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13**, 942
7. Schwenk, J. M., Omenn, G. S., Sun, Z., Campbell, D. S., Baker, M. S., Overall, C. M., Aebersold, R., Moritz, R. L., and Deutsch, E. W. (2017) The human plasma proteome draft of 2017: Building on the human plasma PeptideAtlas from mass spectrometry and complementary assays. *J. Proteome Res.* **16**, 4299–4310
8. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: History, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867
9. Surinova, S., Schiess, R., Hüttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011) On the development of plasma protein biomarkers. *J. Proteome Res.* **10**, 5–16
10. Anderson, N. L. (2010) The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* **56**, 177–185
11. Aebersold, R., and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355
12. Larance, M., and Lamond, A. I. (2015) Multidimensional proteomics for cell biology. *Nat. Rev. Mol. Cell Biol.* **16**, 269–280
13. Omenn, G. S. (2005) Exploring the human plasma proteome. *Proteomics* **5**, 3223
14. Geyer, P. E., Kulak, N. A., Pichler, G., Holdt, L. M., Teupser, D., and Mann, M. (2016) Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195
15. Geyer, P. E., Wewer Albrechtsen, N. J., Tyanova, S., Grassl, N., Iepsen, E. W., Lundgren, J., Madsbad, S., Holst, J. J., Torekov, S. S., and Mann, M. (2016) Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* **12**, 901
16. Albrechtsen, N. J. W., Geyer, P. E., Doll, S., Treit, P. V., Bojsen-Møller, K. N., Martinussen, C., Jørgensen, N. B., Torekov, S. S., Meier, F., Niu, L., Santos, A., Keilhauer, E. C., Holst, J. J., Madsbad, S., and Mann, M. (2018) Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after Roux-en-Y gastric bypass surgery. *Cell Syst.* **7**, 601–612.e3
17. Bruderer, R., Muntel, J., Müller, S., Bernhardt, O. M., Gandhi, T., Cominetti, O., Macron, C., Carayol, J., Rinner, O., Astrup, A., Saris, W. H. M., Hager, J., Valsesia, A., Dayon, L., and Reiter, L. (2019) Analysis of 1508 plasma samples by capillary-flow data-independent acquisition profiles proteomics of weight loss and maintenance. *Mol. Cell. Proteomics* **18**, 1242–1254
18. Gibson, L. M., Paul, L., Chappell, F. M., Macleod, M., Whiteley, W. N., Al-Shahi Salman, R., Wardlaw, J. M., and Sudlow, C. L. M. (2018) Potentially serious incidental findings on brain and body magnetic resonance imaging of apparently asymptomatic adults: Systematic review and meta-analysis. *BMJ* **363**, k4577
19. Kay, F. U., Kandathil, A., Batra, K., Saboo, S. S., Abbara, S., and Rajiah, P. (2017) Revisions to the tumor, node, metastasis staging of lung cancer (8(th) edition): Rationale, radiologic findings and clinical implications. *World J. Radiol.* **9**, 269–279
20. Krier, J. B., and Green, R. C. (2015) Management of incidental findings in clinical genomic sequencing. *Curr. Protoc. Hum. Genet.* **87**, 9.23.1–9.23.16
21. Porsdam Mann, S., Geyer, P. E., Treit, P. V., Omenn, G. S., and Mann, M. (2020) Ethical Principles, Constraints and Opportunities in Clinical Proteomics. *Mol. Cell. Proteomics*. https://doi.org/10.1016/j.mcpro.2021.100046

22. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
23. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805
24. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M., Geiger, T., Mann, M., and Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740
25. Boja, E., Težak, fnm, Zhang, B., Wang, P., Johanson, E., Hinton, D., and Rodriguez, H. (2018) Right data for right patient-a precisionFDA NCI-CPTAC multi-omics mislabeling challenge. *Nat. Med.* **24**, 1301–1302
26. Suhre, K., McCarthy, M. I., and Schwenk, J. M. (2020) Genetics meets proteomics: Perspectives for large population-based studies. *Nat. Rev. Genet.* **22**, 19–37
27. Dodig-Crnkovic, T., Hong, M.-G., Thomas, C. E., Häussler, R. S., Bendes, A., Dale, M., Edfors, F., Forsström, B., Magnusson, P. K. E., Schuppe-Koistinen, I., Odeberg, J., Fagerberg, L., Gummesson, A., Bergström, G., Uhlén, M., *et al*. (2020) Facets of individual-specific health signatures determined from longitudinal plasma proteome profiling. *EBioMedicine* **57**, 102854
28. Ramsey, J. M., Cooper, J. D., Penninx, B. W., and Bahn, S. (2016) Variation in serum biomarkers with sex and female hormonal status: Implications for clinical tests. *Sci. Rep.* **6**, 26947
29. Geyer, P. E., Voytik, E., Treit, P. V., Doll, S., Kleinhempel, A., Niu, L., Müller, J. B., Buchholtz, M.-L., Bader, J. M., Teupser, D., Holdt, L. M., and Mann, M. (2019) Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **11**, e10427
30. UniProt, C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515
31. *DNA Technology in Forensic Science* (1992). National academy press, Washington, DC
32. Bruijns, B., Tiggelaar, R., and Gardeniers, H. (2018) Massively parallel sequencing techniques for forensics: A review. *Electrophoresis* **39**, 2642–2654
33. Khan, M. A. B., Hashim, M. J., King, J. K., Govender, R. D., Mustafa, H., and Al Kaabi, J. (2020) Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. *J. Epidemiol. Glob. Health* **10**, 107–111
34. Keilhauer, E. C., Geyer, P. E., and Mann, M. (2016) HCD fragmentation of glycated peptides. *J. Proteome Res.* **15**, 2881–2890
35. Jensen, M. K., Bertoia, M. L., Cahill, L. E., Agarwal, I., Rimm, E. B., and Mukamal, K. J. (2014) Novel metabolic biomarkers of cardiovascular disease. *Nat. Rev. Endocrinol.* **10**, 659–672
36. McKay, G. J., Silvestri, G., Chakravarthy, U., Dasari, S., Fritsche, L. G., Weber, B. H., Keilhauer, C. N., Klein, M. L., Francis, P. J., Klaver, C. C., Vingerling, J. R., Ho, L., De Jong, P. T., Dean, M., Sawitzke, J., *et al*. (2011) Variations in apolipoprotein E frequency with age in a pooled analysis of a large group of older people. *Am. J. Epidemiol.* **173**, 1357–1364
37. Moore, T., and Dveksler, G. S. (2014) Pregnancy-specific glycoproteins: Complex gene families regulating maternal-fetal interactions. *Int. J. Dev. Biol.* **58**, 273–280
38. Gordon, Y. P., Grudzinskas, J. G., Jeffrey, D., and Chard, T. (1977) Concentrations of pregnancy-specific beta 1-glycoprotein in maternal blood in normal pregnancy and in intrauterine growth retardation. *Lancet* **1**, 331–333
39. Schandorff, S., Olsen, J. V., Bunkenborg, J., Blagoev, B., Zhang, Y., Andersen, J. S., and Mann, M. (2007) A mass spectrometry-friendly database for cSNP identification. *Nat. Methods* **4**, 465–466
40. Kamboh, M. I., and Ferrell, R. E. (1986) Ethnic variation in vitamin D-binding protein (GC): A review of isoelectric focusing studies in human populations. *Hum. Genet.* **72**, 281–293
41. Constans, J., Hazout, S., Garruto, R. M., Gajdusek, D. C., and Spees, E. K. (1985) Population distribution of the human vitamin D binding protein: Anthropological considerations. *Am. J. Phys. Anthropol.* **68**, 107–122
42. Shabani, M., and Borry, P. (2018) Rules for processing genetic data for research purposes in view of the new EU general data protection regulation. *Eur. J. Hum. Genet.* **26**, 149–156
43. Gursoy, G., Emani, P., Brannon, C. M., Jolanki, O. A., Harmanci, A., Strattan, J. S., Cherry, J. M., Miranker, A. D., and Gerstein, M. (2020) Data sanitization to reduce private information leakage from functional genomics. *Cell* **183**, 905–917.e916
44. Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., Wang, X., Qiao, J. W., Cao, S., Petralia, F., Kawaler, E., Mundt, F., Krug, K., Tu, Z., Lei, J. T., Gatza, M. L., Wilkerson, M., Perou, C. M., Yellapantula, V., Huang, K. L., Lin, C., McLellan, M. D., Yan, P., Davies, S. R., Townsend, R. R., Skates, S. J., Wang, J., Zhang, B., Kinsinger, C. R., Mesri, M., Rodriguez, H., Ding, L., Paulovich, A. G., Fenyö, D., Ellis, M. J., and Carr, S. A. (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62
45. Zhang, H., Liu, T., Zhang, Z., Payne, S. H., Zhang, B., McDermott, J. E., Zhou, J.-Y., Petyuk, V. A., Chen, L., Ray, D., Sun, S., Yang, F., Chen, L., Wang, J., Shah, P., *et al*. (2016) Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* **166**, 755–765
46. Eckert, M. A., Coscia, F., Chryplewicz, A., Chang, J. W., Hernandez, K. M., Pan, S., Tienda, S. M., Nahotko, D. A., Li, G., Blaženović, I., Lastra, R. R., Curtis, M., Yamada, S. D., Perets, R., McGregor, S. M., Andrade, J., Fiehn, O., Moellering, R. E., Mann, M., and Lengyel, E. (2019) Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. *Nature* **569**, 723–728
47. Coscia, F., Doll, S., Bech, J. M., Schweizer, L., Mund, A., Lengyel, E., Lindebjerg, J., Madsen, G. I., Moreira, J. M., and Mann, M. (2020) A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J. Pathol.* **251**, 100–112
48. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O., and Singer, M. F. (1975) Summary statement of the Asilomar conference on recombinant DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 1981–1984
49. Gregorowius, D., Biller-Andorno, N., and Deplazes-Zemp, A. (2017) The role of scientific self-regulation for the control of genome editing in the human germline: The lessons from the Asilomar and the Napa meetings show how self-regulation and public deliberation can lead to regulation of new biotechnologies. *EMBO Rep.* **18**, 355–358

Bioresource Technology 229 (2017) 190–195

Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech

Enzyme kinetics approach to assess biocatalyst inhibition and deactivation caused by [bmim][Cl] ionic liquid during cellulose hydrolysis

Nándor Nemestóthy [a], Gábor Megyeri [a], Péter Bakonyi [a], Patrik Lakatos [a], László Koók [a], Milan Polakovic [b], László Gubicza [a], Katalin Bélafi-Bakó [a,*]

[a] Research Institute on Bioengineering, Membrane Technology and Energetics, University of Pannonia, Egyetem ut 10, 8200 Veszprém, Hungary
[b] Department of Chemical and Biochemical Engineering, Institute of Chemical and Environmental Engineering, Faculty of Chemical and Food Technology, Slovak University of Technology, Radlinského 9, Bratislava 81237, Slovakia

HIGHLIGHTS

- Cellulose hydrolysis process was studied by enzyme kinetics approach.
- [bmim][Cl] ionic liquid was found to behave as a competitive inhibitor.
- High concentration [bmim][Cl] caused cellulase enzyme deactivation.

ARTICLE INFO

Article history:
Received 30 November 2016
Received in revised form 4 January 2017
Accepted 5 January 2017
Available online 7 January 2017

Keywords:
Lignocellulose
Pretreatment
Ionic liquid
Enzymatic hydrolysis
Inhibition
Deactivation

ABSTRACT

The aim of this work was to study the inhibition and deactivation of commercial enzyme cocktail (Cellic® Htec2) in the presence of [bmim][Cl] ionic liquid employing model cellulosic substrate, carboxymethyl cellulose (CMC). It turned out from the experiments – relying on enzyme kinetics approach – that [bmim][Cl] could act as a competitive inhibitor. Furthermore, depending on the process conditions i.e. contact of enzyme solution with high concentration [bmim][Cl], severe biocatalyst inactivation should be also taken into account as a potential risk during the enzymatic cellulose hydrolysis even in as short process times as few minutes.

1. Introduction

Lignocellulose, as an inexpensive and abundant, renewable material is a trending feedstock for new generation energy production technologies. Among its constituents, cellulose is the most important one to be considered for biotechnological applications (Liu et al., 2012). Cellulose is a well-known polymer molecule, built-up by individual glucose monomers, which represent the primary source of fermentable sugars. Unfortunately, the direct conversion of lignocelluloses to valuable products i.e. biofuels is hindered by the limited access of biocatalysts to the cellulose regions (Kumar et al., 2015; Xu et al., 2016a). Therefore, to open up the complex, recalcitrant structure and enhance the solubilization of such organic matter, a pretreatment is recommended (Vancov et al., 2012). To accomplish this step, the use of ionic liquids (IL) has been widely proposed because of their recognized potential for the efficient structural transformation of the crystalline cellulose, which increases the efficiency of its consecutive enzymatic hydrolysis and thus, the glucose yield (Brandt et al., 2013; Maki-Arvela et al., 2010; Raj et al., 2016). Basically, two main routes to realize enzyme-catalyzed hydrolyzis after dissolving cellulose in IL have been suggested. In the first one, the cellulose is separated from the IL solution and this, so-called regenerated cellulose (having reduced degree of crystallinity and increased porosity) is subjected for subsequent enzymatic hydrolysis (Tan et al., 2011; Zhao et al., 2009). In the other one, referred as *one-pot or single-step* approach, the enzymatic hydrolysis of dissolved cellulose is conducted by adding cellulase enzymes directly to the IL phase in a water-based buffer solution (Gunny et al., 2014; He

* Corresponding author.
E-mail address: bako@almos.uni-pannon.hu (K. Bélafi-Bakó).

http://dx.doi.org/10.1016/j.biortech.2017.01.004
0960-8524/

et al., 2016; Shi et al., 2013; Xu et al., 2014, 2016b), after which the glucose released can be converted to biofuels or other alternative products such as glucose esters (Findrik et al., 2016).

To make sure that the best IL is chosen for the pretreatment of a particular lignocellulose, a screening is advised (Zavrel et al., 2009) since its appropriateness is dependent on factors such as the physicochemical properties, i.e. the anion and cation constituents of the particular IL (Raj et al., 2016). It was found that among the various ILs, those consisting of an imidazolium-ring, such as [bmim][Cl] are promising candidates to dissolve considerable amount of cellulose (Engel et al., 2012; Tan et al., 2011; Zhao et al., 2009). However, besides that high capacity, ILs should meet additional criteria e.g. biocompatibility with cellulase enzymes performing the cellulose saccharification (Li et al., 2010). In this regard, issues with ILs (including [bmim][Cl]) were observed attributed to their reportedly negative influence on hydrolytic enzyme activity and stability (Engel et al., 2012; Li et al., 2013; Lozano et al., 2011; Ouellet et al., 2011; Park et al., 2012; Salvador et al., 2010; Turner et al., 2003; Xiao et al., 2012; Zhao et al., 2009).

Although the enzyme inhibition and deactivation by ILs from these examples are known to occur, to our knowledge, no thorough study has been dedicated so far to evaluate these phenomena i.e. in terms of the mechanisms involved for the inhibition. Therefore, in this work, comprehensive enzyme kinetics (well-known approach of the biocatalysis area) was applied to get new insights to the enzymatic cellulose hydrolysis process using a model cellulose substrate (carboxymethyl cellulose, CMC) and a commercial enzyme solution (Cellic® Htec2) in the absence and presence of [bmim][Cl], as one of the most widely employed ILs for lignocellulose pretreatment. The novelty of this investigation is the findings presented for the first time using the enzyme kinetic approach, which can thus have the potential to contribute to the international knowledge and further expansion of the research area regarding the application of ionic liquids in the (ligno)cellulose-based biorefinery concept.

2. Materials and methods

2.1. Enzyme kinetics

2.1.1. Cellulose hydrolysis in the absence of [bmim][Cl] ionic liquid – Michaelis-Menten kinetics

The enzymatic cellulose hydrolysis was considered as a multi-step process: (i) the free (cellulase) enzyme (E) binds to the substrate (S) as reversible reaction (E + S), subsequently (ii) an enzyme-substrate complex (ES) is formed and afterwards, as a results of irreversible (ES) breakdown, (E) and product (P) are released (Zhang et al., 2010), in accordance with the classical Michaelis-Menten kinetics (Johnson and Goody, 2011). This model was thus applied to assess the enzymatic hydrolysis reaction (Yeh et al., 2010) in the absence of [bmim][Cl] ionic liquid (Eq. (1)).

$$V = \frac{Vmax[S]}{Ks + [S]} \tag{1}$$

where V and V_{max} are the actual (initial) and maximal (initial) product formation rates (g product/g enzyme-min), respectively, and [S] is the actual (initial) substrate concentration (g/L). K_s is denoted as the half-saturation constant (g/L), equaling to an [S] where $V = V_{max}/2$.

2.1.2. Cellulose hydrolysis in the presence of [bmim][Cl] ionic liquid – inhibition kinetics

Reversible enzyme inhibition may occur via four different reaction mechanisms such as competitive (Eq. (2)), uncompetitive (Eq. (3)), linear-mixed (Eq. (4)) and non-competitive (special case of Eq. (4), where $K_i = K'_i$) (Marangoni, 2003).

$$V = \frac{V_{max}[S]}{Ks\left(1 + \frac{[I]}{K_i}\right) + [S]} \tag{2}$$

$$V = \frac{V_{max}[S]}{Ks + \left(1 + \frac{[I]}{K'_i}\right)[S]} \tag{3}$$

$$V = \frac{V_{max}[S]}{Ks\left(1 + \frac{[I]}{K_i}\right) + \left(1 + \frac{[I]}{K'_i}\right)[S]} \tag{4}$$

where K_i and K'_i are the enzyme-inhibitor and (enzyme-substrate)-inhibitor dissociation constants (g/L), respectively, while [I] is the actual (initial) inhibitor concentration (g/L).

To determine which of the above is actually involved when an inhibition phenomenon is observed, a comprehensive kinetic study was carried out, requiring sets of experiments at various initial substrate and inhibitor (in this investigation, bmim][Cl] ionic liquid) concentrations (Marangoni, 2003).

2.1.3. Enzyme deactivation by undiluted [bmim][Cl] ionic liquid

To study whether [bmim][Cl] ionic liquid – besides its potential inhibitory effect on the cellulose hydrolysis process – causes the deactivation of the enzyme used thoroughly in this work, a first-order kinetics was taken (Eq. (5)). It is a widely used model in the literature to investigate enzyme deactivation (Lencki et al., 1992; Sadana, 1988) and in particular, deactivation of cellulase in the course of cellulose hydrolysis (Zhang et al., 2010).

$$A = \frac{a_t}{a_0} = e^{-k_{de1}t} \tag{5}$$

where the relative activity of the enzyme is A. a_t refers to the activity measured in the presence of the ionic liquid (with a particular concentration, herewith 50 mg/L) after treating the enzyme with undiluted ionic liquid for various times (t). a_0 *is* the activity measured in the presence of the ionic liquid (with a particular concentration, herewith 50 mg/L) without previous incubation of enzyme with undiluted ionic liquid ($t = 0$, to be taken into account as baseline activity). k_{de1} is denoted as the first-order rate constant of cellulase deactivation (min^{-1}) and e is the exponential term (2.718). By taking the natural logarithm of both sides in Eq. (5) and plotting ln A against t, k_{de1} is derived from the slope of the straight line (Lencki et al., 1992).

2.1.4. Cellulose hydrolysis assays

All cellulose hydrolysis measurements were performed in closed-top 250 mL Erlenmeyer flasks (to prevent evaporation) with 100 mL total working volume, ensuring vigorous stirring (400 rpm by magnetic bar) and constant 60 °C temperature. Citrate buffer (100 mM, pH of 4.5) served as bulk phase thoroughly. The model substrate used was low-viscosity carboxymethyl cellulose (CMC) (Sigma-Aldrich, USA) with phisyco-chemical properties reported in the specification sheet available for download at the manufacturer's official website, while Cellic® Htec2 (Novozymes®, Denmark) was utilized as enzyme source, which has an already reported potential for the hydrolysis of cellulosic materials (Benjamin et al., 2014; Joe et al., 2015; Song et al., 2014). Cellic® Htec2 is characterized with optimal working pH and temperature of 4.5–5.5 and 60–75 °C, respectively (according to the application sheet issued by the producer) and was provided in 600(±48) mg/L concentrations for each tests. Data were always evaluated by considering the exact, actual mass of enzyme supplemented to the particular reaction vessel.

During the experiments without ionic liquid (Michaelis-Menten kinetics), substrate (CMC) concentrations investigated were 0.5, 2, 5 and 25 g/L.

In case of enzyme inhibition tests, definite amounts of [bmim][Cl] (99% purity, IoLiTec, Germany) were added (without further purification) to the citrate buffer to get the desired ionic liquid concentration (50, 100, 150, 200 and 250 mg/L) along with substrate concentrations of 1, 2.5 and 5 g/L.

In the course of enzyme deactivation study, 60(±4.8) mg enzyme was first mixed with 5 mg (undiluted) [bmim][Cl] ionic liquid and held together for treatment (contact) times such as 1, 2, 3 and 10 min at 60 °C. Subsequently, the whole mixture was loaded to the above described citrate buffer-based reaction medium (resulting in an initial [bmim][Cl] and enzyme concentrations of 50 mg/L and 600(±48) mg/L, respectively) containing 2.5 g substrate/L.

Quantitative analysis of glucose (as final end-product of multistep cellulose hydrolysis) released from cellulose hydrolysis was monitored based on the refractive index (RI) change of the reaction mixture over time – relative to the RI of initial reaction mixture (background) – pumped continuously (10 mL min^{-1}) through Merck-Hitachi (RI-71) differential refractometer (at 37 °C), attached to the reaction vessel in a closed-loop design. Preliminary calibration – to establish the relationship of RI change and glucose concentrations – was accomplished by measuring the RI of reaction mixture with well-defined glucose (reagent grade, Sigma-Aldrich, USA) contents.

Experiments were always commenced by the injection of the enzyme solution to the reaction mixture. Kinetic parameters (i.e. K_s, V_{max}, K_i, kd_{e1}) were estimated by linear regression using the least squares method in Matlab software.

3. Results and discussion

3.1. Cellulose hydrolysis process kinetics without [bmim][Cl] ionic liquid

The progress curves obtained during the measurements can be seen in Fig. 1. It seems to infer that product (glucose) formation showed a directly proportional trend with time, as reflected by the appreciably high R^2 values (Table 1). Though to some extent, various terms of experimental error e.g. data recording, observations, etc. (Bakonyi et al., 2015) may make the determination of apparent, initial reaction rates vague, replicates undertaken (Fig. 1) indicated that the results were fairly reproducible and consequently, basically reliable for further analysis. The confidential evaluation of V is supposed to be performed using a good mass of data acquired within a limited time, otherwise it can be misleading due to fact that initial cellulose hydrolysis rates decline remarkably as time passes (Carrillo et al., 2005). Hence, information recorded in the first 10–15 min of the enzymatic hydrolysis tests was only taken into account in this study. Similar period of time was considered by other researchers working in this field, as well (Yeh et al., 2010).

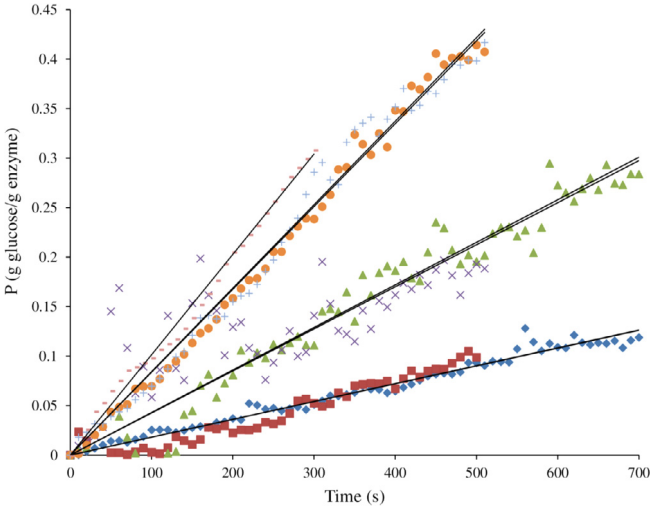


Fig. 1. Typical progress curves of cellulose hydrolysis (actual experimental conditions: absence of [bmim][Cl] ionic liquid, varied initial substrate concentration) blue diamond and red square: [S] = 0.5 g/L; green triangle and purple cross: [S] = 2 g/L; blue asterisk and orange dot: [S] = 5 g/L; pink dash: [S] = 25 g/L. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

It can be concluded from the time profiles of cellulose hydrolysis experiments (Fig. 1) that enhanced V values (slope of the trendlines fitted, Table 1) could be achieved along with increasing substrate concentrations. Applying the well-known Lineweaver-Burk double reciprocal plot (V^{-1} vs. $[S]^{-1}$), the kinetic parameters of the Michaelis-Menten model could be computed and as a result, K_s and V_{max} were found as 2.57 g/L and 0.068 g glucose/g enzyme-min, respectively. The value of K_s is in the same order of magnitude reported by Yeh et al. (2010) for a range of microcrystalline cotton cellulose substrates and a cellulase enzyme produced by *T. reesei* ATCC 26921.

Generally speaking, the use of Michaelis-Menten kinetics for discussing the experimental data requires a "mass of enzyme to mass of substrate ratio" <0.15, which ensures that the mechanistic model is applicable (Bezerra and Dias, 2004, 2005) and the substrate is not limiting. In our current investigation, apparent "mass of enzyme mass to mass of substrate ratios" such as 0.024, 0.12, 0.3 and 1.2 were obtained, being solely dependent on the initial substrate concentration because of the constant enzyme loadings (600 mg/L). Although seemingly not all of them meet the criteria referred, a decent correlation of the experimental outcomes with the fitted Michaelis-Menten kinetics was established. This contradiction can be explained by considering the properties of the enzyme source. Cellic® Htec2 is a commercial enzyme cocktail preparation and not a highly purified cellulase. Such products can be seen as a mixture of accessory enzymes such as cellulases and hemicellulases (Samayam and Schall, 2010), however, their exact composition and relative proportion of ingredients is unknown (Barr et al., 2012). Thus, once a given mass of this enzyme solution is used (e.g. 600 mg for every liter of reaction mixture in this study), only its limited portion was actually the cellulase enzyme participating in the hydrolysis reaction, while the rest consisted mainly of water, other enzymes i.e. endoxylanases (Benjamin et al., 2014; Joe et al., 2015; Song et al., 2014), etc.

Consequently, the real "cellulase enzyme mass to substrate mass ratio" was (probably orders of magnitudes) below the above defined threshold value (<0.15) even in the lower initial substrate concentration range (0.5–2 g/L), making the Michaelis-Menten theory valid within the respective experimental boundaries and leading to an adequate prediction of the hydrolysis process.

3.2. On the inhibition of enzymatic cellulose hydrolysis by [bmim][Cl] ionic liquid

It turned out in previous works examining cellulose hydrolysis in the mixture of aqueous buffer and IL that the cellulose hydrolysis can be negatively affected by ionic liquids (Kamiya et al., 2008; Turner et al., 2003). For instance, Kamiya et al. (2008) summarized that biocatalyzed cellulose saccharification (applying commercialized cellulase from *Trichoderma reesei*) was fully stopped in a citrate-buffer based reaction mixture composing of >40 vol.% [emim][dep], meaning that the proportion of IL is a determining

Table 1
Example for progress curve analysis to deliver the initial reaction rates (V) using the dataset of Fig. 1.

[S] (g/L)	Fitted trendline properties				V (g glucose/g enzyme-min)
	1. Repetition		2. Repetition		
	Slope (g glucose/g enzyme-s)	R^2	Slope (g glucose/g enzyme-s)	R^2	
0.5	0.000180	0.98	0.000181	0.91	0.011
2	0.000430	0.19	0.000425	0.95	0.026
5	0.000837	0.99	0.000843	0.99	0.051
25	0.001013	0.99			0.060

factor of the process. Moreover, another research carried out by Turner et al. (2003) also confirmed the existence of IL-induced enzyme inhibition working with [bmim][Cl], which was associated with the high, dissociated Cl^- ion concentration in the media, causing the interference of cellulose with ionic liquid.

Overall, it is clear by the literature examples that enzyme-mediated cellulose hydrolysis can be sensitive to the presence of ILs, including [bmim][Cl] employed thoroughly herewith. However, it is still of question which of the possible inhibition mechanisms (i.e. competitive, uncompetitive, mixed, non-competitive) stands behind once decreased reaction rates (as a side-effect of IL) are encountered. Therefore, a throughout kinetic study was devoted to get new insights on this subject.

The analysis of the progress curves and consecutive use of Lineweaver-Burk double reciprocal technique (Lineweaver and Burk, 1934) (V^{-1} vs. $[S]^{-1}$, along with various inhibitor concentrations [I], 50–250 mg [bmim][Cl]/L) yielded Fig. 2, which is a clear identification of the competitive-type inhibition (Marangoni, 2003). Thus, in the light of this diagnosis, the deviation of initial reaction velocities from those projected by the Michaelis-Menten model (no inhibition occuring) took place by the mechanism established in Eq. (2). When the data in Fig. 2 are subjected for an estimation of the catalytic parameters, it can be observed that V_{max} remains unaffected in accordance with the feature of competitive inhibition, meanwhile K_S is influenced by the inhibitor concentration ([I]) and the enzyme-inhibitor dissociation constant (K_i) (Eq. (2)). Since the experimental results suggest competitive inhibition caused by the [bmim][Cl] ionic liquid, it means that the inhibitor competes for the substrate-binding site of the enzyme and when

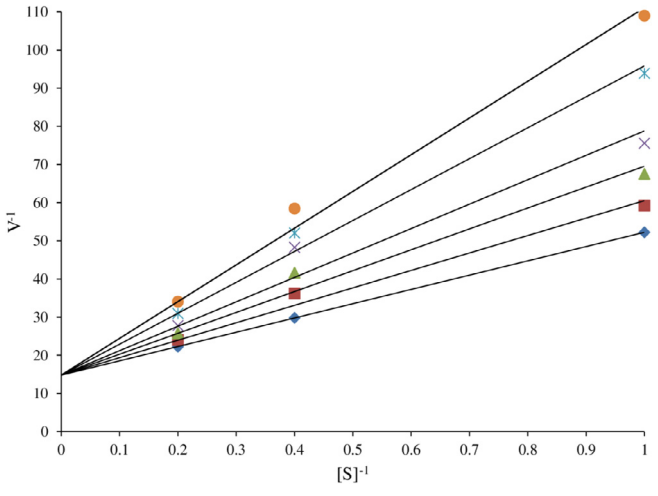


Fig. 2. Kinetic evaluation on enzyme inhibition test results obtained during cellulose hydrolysis with different initial [bmim][Cl] ionic liquid (inhibitor) and substrate concentrations. Blue diamond: no [bmim][Cl] added; red square: 50 mg/L [bmim][Cl]; green triangle: 100 mg/L [bmim][Cl]; purple cross: 150 mg/L [bmim][Cl]; blue asterisk: 200 mg/L [bmim][Cl]; orange dot: 250 mg/L [bmim][Cl]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

an enzyme-inhibitor complex is formed rather than an enzyme-substrate one, the enzyme cannot express its catalytic activity. Overall, it seems to be the case that the cellulase enzymes present in Cellic® Htec2 suffer from an inhibition in the presence of [bmim][Cl], even at low [I]. General strategies to suppress this type of inhibition include the use of higher enzyme as well as substrate loadings to increase the probability of enzyme and substrate interactions rather than that of the enzyme and the inhibitor. In summary of the kinetic analysis on enzyme inhibition tests (Fig. 2, Table 2), numerical estimate for K_i was obtained as 0.163 g/L. In general, the increase of [I]/K_i will more notably reduce the affinity of the enzyme to the substrate, causing the concomitant raise of K_S. For instance, once the [bmim][Cl] concentration in the reaction mixture reaches to K_i, it will double the original (uninhibited) K_S value, in accordance with the mechanistic model for competitive inhibition in Eq. (2).

3.3. Cellulase enzyme deactivation by [bmim][Cl] ionic liquid

The important work done by Turner et al. (2003) implied that not only an inhibition, but even deactivation of cellulase may occur when they are contacted with ionic liquids in the solution. As it was deduced, enzyme denaturation can be a possible threat and in such a case, overcoming strategies may be unable to refold, reactivate the cellulase. Deactivation (or in other words, irreversible inhibition) is one of the major phenomena known to slow the cellulose hydrolysis rate down (Bansal et al., 2009), leading essentially to a process of limited efficiency.

To judge whether the inactivation of cellulase enzymes in Cellic® Htec2 is to be taken into account, tests were undertaken and the enzyme solution was mixed and incubated with [bmim][Cl] for various durations, followed by the determination of actual, initial reaction rates to reveal residual enzyme activity. It can be drawn from Fig. 3 that the treatment time of enzyme by IL considerably affected the achievable V values, providing a good indication that contact of the cellulase with undiluted [bmim][Cl] – even when it happened for short times only e.g. 1 min – was accompanied by the relative loss of enzyme activity. Taking the advantage of the formula expressed in Eq. (5), kinetic analysis was conducted to compute the rate constant of deactivation. Since in the bottom part of Fig. 3 a well-fitting linear relationship can be found, the first-order kinetics to predict biocatalyst deactivation seems to be valid (Lencki et al., 1992) and k_{de1} as 0.132 min^{-1} could be determined. A correlation of similar shape between relative cellulase activity and various residence times in 10% [bmim][Cl] was communicated by Salvador et al. (2010), as well. However, in that investigation (Salvador et al., 2010), residual enzyme activities were much closer to the original, since even after 40 min of contact with IL, reduction was only 13–14%, suggesting that the inhibition was mostly of reversible nature and ascribed primarily to the change of thermodynamic water activity under the different conditions. However, it was also shown that increased IL concentrations are severely disadvantageous from an enzyme activity point of view (Salvador et al., 2010; Xiao et al., 2012). Significant loss of Celluclast® cellulase enzyme activity was observed by Engel et al.

Table 2
Characteristics of fitted trendlines (Fig. 2) as the function of [bmim][Cl] concentration to estimate the maximal initial reaction rate (*Vmax*) and enzyme-inhibitor dissociation constant (K_i) for competitive inhibition model. The corresponding $[S]^{-1}$ range studied is presented in Fig. 2.

	[bmim][Cl] concentration (mg/L)				
	50	100	150	200	250
Slope	45.71	54.71	63.99	81.04	96.27
Intercept	14.82	14.82	14.82	14.82	14.82
R^2	0.98	0.97	0.94	0.99	0.99

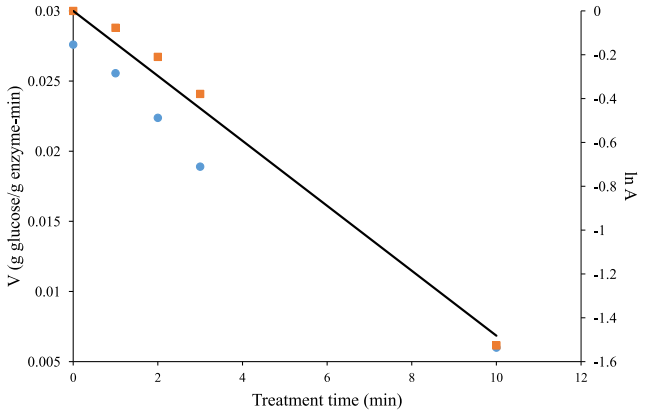


Fig. 3. Kinetic evaluation on enzyme deactivation experiments blue dots: initial reaction rate; red square: natural logarithm of relative enzyme activity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(2010) too in reaction mixtures comprising of commercialized ionic liquids (10%), including [bmim][Cl].

Keeping this in mind together with our results, it would appear that when cellulase enzymes (even for short times) are exposed to concentrated ILs, deactivation effect may take over. This would explain why in our experiments the reaction rates have notably dropped, causing that the enzyme lost approximately 25% of its activity relatively to its control (a_0) even after only 1 min of pre-incubation with undiluted [bmim][Cl]. This phenomenon can be interesting when cellulase hydrolysis is performed *in situ*, referred as *one-pot* process design. In this arrangement, the cellulosic raw material is first pretreated with concentrated ionic liquid to help the dissolution of the polysaccharide fractions and in that way, provide better accessibility for the enzymes participating in their hydrolysis. Thereafter, the hydrolytic enzymes – carried by an aqueous buffer solution – are loaded to the same vessel containing the cellulose being dissolved in the IL (Shi et al., 2013). Many ionic liquids, depending on their structural features, are hydrophilic (including [bmim][Cl]) and therefore soluble in water to certain degrees (Huddleston et al., 2001). In these cases, ILs can be more or less homogenously distributed in the whole reaction medium, which makes it possible to decrease their concentrations to a sufficiently low threshold level. However, appropriate dilution factor should be selected since even its low concentrations could exhibit a hindering impact on the hydrolysis for the particular example of [bmim][Cl] used in this work, as demonstrated by our results in Section 3.2. Besides dilution method (Li et al., 2013), the engineering and application of cellulases with satisfactory robustness to work under harsh conditions can be proposed (Nordwald et al., 2014; Raddadi et al., 2013) i.e. via the development of halophilic, IL-tolerant cellulases (Gunny et al., 2014; Xu et al., 2014; 2016b), their immobilization (Xu et al., 2016c) which may better withstand the negative impact of ILs. In other cases, when the IL is basically immiscible with the water-based solution, a biphasic (separated) system is formed (Kuroda et al., 2016), and saccharification can take place on the phase boundary, where the probability of an enzyme-concentrated IL encounter is higher, representing a threat on the time-stability of the process. For cellulose saccharification and subsequent utilization technologies relying on regenerated cellulose, ionic liquid residues should be removed as much as possible to avoid inhibition by a washing process (Li et al., 2013; Ouellet et al., 2011)

4. Conclusions

In this work, the effect of [bmim][Cl] ionic liquid on cellulose hydrolysis catalyzed by Cellic® Htec2 enzyme solution was studied. It was found that the enzymatic reaction could be fairly described by the Michaelis-Menten kinetics when [bmim][Cl] ionic liquid was absent. However, the presence of this IL in low concentrations significantly hindered the process via competitive inhibition, supported by the kinetic evaluation. Furthermore, it was demonstrated that contacting the enzyme with highly-concentrated IL even for short times could induce irreversible inhibition (deactivation), which should be considered as an important aspect of technology design.

Acknowledgements

The support of this work by the Slovakian-Hungarian cooperation 2013-0008 is appreciated. Nándor Nemestóthy was supported by the ÚNKP-2016-4-04 "New National Excellence Program of the Ministry of Human Capacities". Péter Bakonyi acknowledges the support received from National Research, Development and Innovation Office (Hungary) under Grant number PD 115640.

References

Bakonyi, P., Nemestóthy, N., Lankó, J., Rivera, I., Buitrón, G., Bélafi-Bakó, K., 2015. Simultaneous biohydrogen production and purification in a double-membrane bioreactor system. Int. J. Hydrogen Energy 40, 1690–1697.
Bansal, P., Hall, M., Realff, M.J., Lee, J.H., Bommarius, A.S., 2009. Modeling cellulase kinetics on lignocellulosic substrates. Biotechnol. Adv. 27, 833–848.
Barr, C.J., Mertens, J.A., Schall, C.A., 2012. Critical cellulase and hemicellulase activities for hydrolysis of ionic liquid pretreated biomass. Bioresour. Technol. 104, 480–485.
Benjamin, Y., García-Aparicio, M.P., Görgens, J.F., 2014. Impact of cultivar selection and process optimization on ethanol yield from different varieties of sugarcane. Biotechnol. Biofuels 7, 60. http://dx.doi.org/10.1186/1754-6834-7-60.
Bezerra, R.M.F., Dias, A.A., 2004. Discrimination among eight modified Michaelis–Menten kinetics models of cellulose hydrolysis with a large range of substrate/enzyme ratios. Appl. Biochem. Biotechnol. 112, 173–184.
Bezerra, R.M.F., Dias, A.A., 2005. Enzymatic kinetic of cellulose hydrolysis. Appl. Biochem. Biotechnol. 126, 49–59.
Brandt, A., Grasvik, J., Hallett, J.P., Welton, T., 2013. Deconstruction of lignocellulosic biomass with ionic liquids. Green Chem. 15, 550–583.
Carrillo, F., Lis, M.J., Colom, X., López-Mesas, M., Valldeperas, J., 2005. Effect of alkali pretreatment on cellulase hydrolysis of wheat straw: kinetic study. Process Biochem. 40, 3360–3364.
Engel, P., Krull, S., Seiferheld, B., Spiess, A.C., 2012. Rational approach to optimize cellulase mixtures for hydrolysis of regenerated cellulose containing residual ionic liquid. Bioresour. Technol. 115, 27–34.
Engel, P., Mladenov, R., Wulfhorst, H., Jager, G., Spiess, A.C., 2010. Point by point analysis: how ionic liquid affects the enzymatic hydrolysis of native and modified cellulose. Green Chem. 12, 1959–1966.

Findrik, Z., Megyeri, G., Gubicza, L., Bélafi-Bakó, K., Nemestóthy, N., Sudar, M., 2016. Lipase catalyzed synthesis of glucose palmitate in ionic liquid. J. Cleaner Prod. 112, 1106–1111.
Gunny, A.A.N., Arbain, D., Gumba, R.E., Jong, B.C., Jamal, P., 2014. Potential halophilic cellulases for in situ enzymatic saccharification of ionic liquids pretreated lignocelluloses. Bioresour. Technol. 155, 177–181.
He, C., Liu, F., Gong, L., Di, J.H., Ding, Y., Ma, C.L., et al., 2016. Enzymatic in situ saccharification of chestnut shell with high ionic liquid-tolerant cellulases from Galactomyces sp. CCZU11-1 in a biocompatible ionic liquid-cellulase media. Bioresour. Technol. 201, 133–139.
Huddleston, J.G., Visser, A.E., Reichert, W.M., Willauer, H.D., Broker, G.A., Rogers, R. D., 2001. Characterization and comparison of hydrophilic and hydrophobic room temperature ionic liquids incorporating the imidazolium cation. Green Chem. 3, 156–164.
Joe, M.H., Kim, J.Y., Lim, S., Kim, D.H., Bai, S., Park, H., et al., 2015. Microalgal lipid production using the hydrolysates of rice straw pretreated with gamma irradiation and alkali solution. Biotechnol. Biofuels 8, 125. http://dx.doi.org/10.1186/s13068-015-0308-x.
Johnson, K.A., Goody, R.S., 2011. The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. Biochemistry 50, 8264–8269.
Kamiya, N., Matsushita, Y., Hanaki, M., Nakashima, K., Narita, M., Goto, M., et al., 2008. Enzymatic in situ saccharification of cellulose in aqueous-ionic liquid media. Biotechnol. Lett. 30, 1037–1040.
Kumar, G., Bakonyi, P., Periyasamy, S., Kim, S.H., Nemestóthy, N., Bélafi-Bakó, K., 2015. Lignocellulose biohydrogen: practical challenges and recent progress. Renewable Sustainable Energy Rev. 44, 728–737.
Kuroda, K., Miyamura, K., Satria, H., Takada, K., Ninomiya, K., Takahashi, K., 2016. Hydrolysis of cellulose using an acidic and hydrophobic ionic liquid and subsequent separation of glucose aqueous solution from the ionic liquid and 5-(Hydroxymethyl)furfural. ACS Sustainable Chem. Eng. 4, 3352–3356.
Lencki, R.W., Arul, J., Neufeld, R.J., 1992. Effect of subunit dissociation, denaturation, aggregation, coagulation, and decomposition on enzyme inactivation kinetics: 1. First-order behavior. Biotechnol. Bioeng. 40, 1421–1426.
Li, C., Tanjore, D., He, W., Wong, J., Gardner, J.L., Sale, K.L., et al., 2013. Scale-up and evaluation of high solid ionic liquid pretreatment and enzymatic hydrolysis of switchgrass. Biotechnol. Biofuels 26, 154. http://dx.doi.org/10.1186/1754-6834-6-154.
Li, Q., Jiang, X., He, Y., Li, L., Xian, M., Yang, J., 2010. Evaluation of the biocompatibile ionic liquid 1-methyl-3-methylimidazolium dimethylphosphite pretreatment of corn cob for improved saccharification. Appl. Microbiol. Biotechnol. 87, 117–126.
Lineweaver, H., Burk, D., 1934. The determination of enzyme dissociation constants. J. Am. Chem. Soc. 56, 658–666.
Liu, C.Z., Wang, F., Stitles, A.R., Guo, C., 2012. Ionic liquids for biofuel production: opportunities and challenges. Appl. Energy 92, 406–414.
Lozano, P., Bernal, B., Bernal, J.M., Pucheault, M., Vaultier, M., 2011. Stabilizing immobilized cellulase by ionic liquids for saccharification of cellulose solutions in 1-butyl-3-methylimidazolium chloride. Green Chem. 13, 1406–1410.
Maki-Arvela, P., Anugwom, I., Virtanen, P., Sjöholm, R., Mikkola, J.P., 2010. Dissolution of lignocellulosic materials and its constituents using ionic liquids – a review. Ind. Crops Prod. 32, 175–201.
Marangoni, A.G., 2003. Enzyme Kinetics: A Modern Approach. John Wiley & Sons, New Jersey, USA.
Nordwald, E.M., Brunecky, R., Himmel, M.E., Beckham, G.T., Kaar, J.L., 2014. Charge engineering of cellulases improves ionic liquid tolerance and reduces lignin inhibition. Biotechnol. Bioeng. 111, 1541–1549.
Ouellet, M., Datta, S., Dibble, D.C., Tamrakar, P.R., Benke, P.I., Li, C., et al., 2011. Impact of ionic liquid pretreated plant biomass on *Saccharomyces cerevisiae* growth and biofuel production. Green Chem. 13, 2743–2749.
Park, J.I., Steen, E.J., Burd, H., Evans, S.S., Redding-Johnson, A.M., Batth, T., et al., 2012. A thermophilic ionic liquid-tolerant cellulase cocktail for the production of cellulosic biofuels. PLoS One 7, e37010. http://dx.doi.org/10.1371/journal.pone.0037010.
Raddadi, N., Cherif, A., Daffonchio, D., Fava, F., 2013. Halo-alkalitolerant and thermostable cellulases with improved tolerance to ionic liquids and organic solvents from Paenibacillus tarimensis isolated from the Chott El Fejej, Sahara desert, Tunisia. Bioresour. Technol. 150, 121–128.
Raj, T., Kapoor, M., Semwal, S., Sadula, S., Pandey, V., Gupta, R.P., et al., 2016. The cellulose structural transformation for higher enzymatic hydrolysis by ionic liquids and predicting their solvating capabilities. J. Cleaner Prod. 113, 1005–1014.
Sadana, A., 1988. Enzyme deactivation. Biotechnol. Adv. 6, 349–446.
Salvador, A.C., Santos, M.D.A., Saraiva, J.A., 2010. Effect of ionic liquid [bmim][Cl] and high pressure on the activity of cellulase. Green Chem. 12, 632–635.
Samayam, I.P., Schall, C.A., 2010. Saccharification of ionic liquid pretreated biomass with commercial enzyme mixtures. Bioresour. Technol. 101, 3561–3566.
Shi, J., Gladden, J.M., Sathitsuksanoh, N., Kambam, P., Sandoval, L., Mitra, D., et al., 2013. One-pot ionic liquid pretreatment and saccharification of switchgrass. Green Chem. 15, 2579–2589.
Song, Q., Winter, W.T., Bujanovic, B.M., Amidon, T.E., 2014. Nanofibrillated cellulose (NFC): a high-value co-product that improves the economics of cellulosic ethanol production. Energies 7, 607–618.
Tan, H.T., Lee, K.T., Mohamed, A.R., 2011. Pretreatment of lignocellulosic palm biomass using a solvent-ionic liquid [BMIM]Cl for glucose recovery: an optimisation study using response surface methodology. Carbohydr. Polym. 83, 1862–1868.
Turner, M.B., Spear, S.K., Huddleston, J.G., Holbrey, J.D., Rogers, R.D., 2003. Ionic liquid salt-induced inactivation and unfolding of cellulase from Trichoderma reesei. Green Chem. 5, 443–447.
Vancov, T., Alston, A.S., Brown, T., McIntosh, S., 2012. Use of ionic liquids in converting lignocellulosic material to biofuels. Renewable Energy 45, 1–6.
Xiao, W., Yin, W., Xia, S., Ma, P., 2012. The study of factors affecting the enzymatic hydrolysis of cellulose after ionic liquid pretreatment. Carbohydr. Polym. 87, 2019–2023.
Xu, J., He, B., Wu, B., Wang, B., Wang, C., Hu, L., 2014. An ionic liquid tolerant cellulase derived from chemically polluted microhabitats and its application in in situ saccharification of rice straw. Bioresour. Technol. 157, 166–173.
Xu, J., Xiong, P., He, B., 2016a. Advances in improving the performance of cellulase in ionic liquids for lignocellulose biorefinery. Bioresour. Technol. 200, 961–970.
Xu, J., Wang, X., Liu, X., Xia, J., Zhang, T., Xiong, P., 2016b. Enzymatic in situ saccharification of lignocellulosic biomass in ionic liquids using an ionic liquid-tolerant cellulases. Biomass Bioenergy 93, 180–186.
Xu, J., Sheng, Z., Wang, X., Liu, X., Xia, J., Xiong, P., et al., 2016c. Enhancement in ionic liquid tolerance of cellulase immobilized on PEGylated graphene oxide nanosheets: application in saccharification of lignocellulose. Bioresour. Technol. 200, 1060–1064.
Yeh, A.I., Huang, Y.C., Chen, S.H., 2010. Effect of particle size on the rate of enzymatic hydrolysis of cellulose. Carbohydr. Polym. 79, 192–199.
Zavrel, M., Bross, D., Funke, M., Büchs, J., Spiess, A.C., 2009. High-throughput screening for ionic liquids dissolving (ligno-)cellulose. Bioresour. Technol. 100, 2580–2587.
Zhang, Y., Xu, J.L., Xu, H.J., Yuan, Z.H., Guo, Y., 2010. Cellulase deactivation based kinetic modeling of enzymatic hydrolysis of steam-exploded wheat straw. Bioresour. Technol. 101, 8261–8266.
Zhao, H., Jones, C.L., Baker, G.A., Xia, S., Olubajo, O., Person, V.N., 2009. Regenerating cellulose from ionic liquids for an accelerated enzymatic hydrolysis. J. Biotechnol. 139, 47–54.

International Journal of
Molecular Sciences

MDPI

Review

Nucleic Acid Testing of SARS-CoV-2

Hee Min Yoo [1,2,†], Il-Hwan Kim [1,†] and Seil Kim [1,2,3,*]

[1] Microbiological Analysis Team, Biometrology Group, Korea Research Institute of Standards and Science (KRISS), Daejeon 34113, Korea; hmy@kriss.re.kr (H.-M.Y.); ihkim@kriss.re.kr (I.-H.K.)
[2] Department of Bio-Analytical Science, University of Science & Technology (UST), Daejeon 34113, Korea
[3] Convergent Research Center for Emerging Virus Infection, Korea Research Institute of Chemical Technology, Daejeon 34114, Korea
* Correspondence: stapler@kriss.re.kr
† Both authors are equally contributed.

Abstract: The coronavirus disease 2019 (COVID-19) has caused a large global outbreak. It is accordingly important to develop accurate and rapid diagnostic methods. The polymerase chain reaction (PCR)-based method including reverse transcription-polymerase chain reaction (RT-PCR) is the most widely used assay for the detection of SARS-CoV-2 RNA. Along with the RT-PCR method, digital PCR has emerged as a powerful tool to quantify nucleic acid of the virus with high accuracy and sensitivity. Non-PCR based techniques such as reverse transcription loop-mediated isothermal amplification (RT-LAMP) and reverse transcription recombinase polymerase amplification (RT-RPA) are considered to be rapid and simple nucleic acid detection methods and were reviewed in this paper. Non-conventional molecular diagnostic methods including next-generation sequencing (NGS), CRISPR-based assays and nanotechnology are improving the accuracy and sensitivity of COVID-19 diagnosis. In this review, we also focus on standardization of SARS-CoV-2 nucleic acid testing and the activity of the National Metrology Institutes (NMIs) and highlight resources such as reference materials (RM) that provide the values of specified properties. Finally, we summarize the useful resources for convenient COVID-19 molecular diagnostics.

Keywords: SARS-CoV-2; PCR; isothermal amplification; genome sequencing; nucleic acid testing; reference materials

Citation: Yoo, H.M.; Kim, I.-H.; Kim, S. Nucleic Acid Testing of SARS-CoV-2. *Int. J. Mol. Sci.* **2021**, *22*, 6150. https://doi.org/10.3390/ijms22116150

Academic Editor: Vasso Apostolopoulos

Received: 16 May 2021
Accepted: 4 June 2021
Published: 7 June 2021

1. Background

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is a novel coronavirus that first appeared in Wuhan, Hubei Province, China in December 2019, connected to a seafood market [1,2]. Seven coronaviruses have been reported to infect humans; four of them, human CoV-NL63 [3], HCoV-OC43 [4–6], HCoV-229E [7,8], and HCoV-HKU [9,10], cause mild and seasonal respiratory tract disease, whereas SARS-CoV [11–15], MERS-CoV [16–19], and SARS-CoV-2 can cause severe symptoms. In particular, SARS- CoV-2 is suited to human-to-human transmission and spreads rapidly to other locations, causing lung injury, multiorgan failure, and death [20,21]. As of this date, the number of confirmed cases is still increasing, as is the number of deaths [22,23]. Therefore, an understanding of the SARS-CoV-2 host and pathogen biology is important to offer valuable insights into the diagnosis and treatment of the disease including the development of new therapies [24,25]. Here, we review the basic biology of SARS-CoV-2 including the origin, pathophysiology, and diagnosis methods.

1.1. Nomenclature of SARS-CoV-2

Currently, almost a million sequences of the SAR-CoV-2 genome are publicly available via the Global Initiative on Sharing All Influenza Data (GISAID) and GenBank [26,27]. Based on these genome sequences, the phylogenetic classification of SARS-CoV-2 was performed and the nomenclature of GISAID, Phylogenetic Assignment of Named Global

Outbreak LINeages (PANGO lineage), and Nextstrain are widely used in scientific and clinical communities [26,28,29] The major lineages of each nomenclature are summarized in Table 1.

Table 1. Nomenclature of SARS-CoV-2.

GISAID Clades	PANGO Lineage	Nextstrain Clades	Notable Variants
S	A	19B	A.23.1
L	B	19A	Wuhan-Hu-1
V			
G	B.1	20A	B.1.525, B.1.627
GH	B.1	20C	B.1.427, B.1.429, B.1.526
	B.1.2	20G	
	B.1.596		
	B.1.351	20H/501.Y.V2	B.1.351
GR	B.1.1.1	20B	
	P.3		P.3
	C	20D	
	D.2	20F	
	P.1	20J/501.Y.V3	P.1
GV	B.1.177	20E (EU1)	B.1.177
GRY	B.1.1.7	20I/501.Y.V1	B.1.1.7

GISAID introduced the nomenclature system of SARS-CoV-2 based on marker mutations and named the clade with actual letters of marker mutations [30]. For example, the clade G has a characteristic mutation in the spike protein gene, D614G. In the nomenclature of GISAID, the initial strains of SARS-CoV-2 were grouped as S and L clades and the current strains of SARS-CoV-2 were classified as eight major clades (S, L, V, G, GH, GR, GV, and GRY) [30,31]. The L clade contains the reference strain WIV-04 and was the dominant lineage in early 2020. The L clade later diverged into clades V and G, and clade G diverged into clades GH, GR, GV, and GRY.

The PANGO nomenclature systems focused on the active virus lineage [29]. This nomenclature is dynamic and the lineages of the PANGO nomenclature are marked as three statuses: active, unobserved, or inactive. The lineages documented within a month are marked as active. The lineages documented within three months are marked as unobserved, and the lineages that were not documented for more than three months are regarded as inactive. The lineages of PANGO nomenclature are named with a letter and numerical values. The initial lineages are denoted as lineages A and B. Although clade B includes the first genome sequenced strain, the phylogenetic analysis suggested that the most recent common ancestor of SARS-CoV-2 was close to early lineage A [29]. The descendent lineages from initial lineages were assigned with numerical labels. The descendent lineages can be designated with the phylogenetic evidence that the descendent emerged from parental lineages and the descendants showed significant transmission to geographically distinct populations. The designated descendent lineages can also be parental of new emerging lineages and these new lineages have been labeled as parental lineages with additional numerical values. For example, a new emerging lineage from lineage B1 can be labeled as B1.1. The lineages can have a maximum of three sublevels and newly designated lineages emerging from a lineage with three sublevels will be labeled with new alphabetical letters. For example, the parental lineage of lineage C.1 is the lineage B.1.1.1.

The clades of Nextstrain nomenclature were initially named according to year–letter combinations [32]. Major clades were designated as the clade reached more than 20% of

global frequency for more than two months. Based on this criteria, the initial clades were designated as 19A, 19B, 20A, 20B, and 20C. However, due to the global travel restriction, no more clades were designated according to the criteria. For this reason, Nextstrain updated their major clade designation criteria with regional frequency (>30%) and recognized variants of concern. Currently, 12 major clades (19A, 19B, 20A, 20B, 20C, 20D, 20E, 20F, 20G, 20H/501Y.V2, 20I/501Y.V1, 20J/501Y.V3) are designated in the nomenclature of Nextstrain.

1.2. Notable Variants of SARS-CoV-2

The first strain of SARS-CoV-2 was discovered in Wuhan, China and designated as Wuhan-Hu-1 or WIV-04 [1,33]. The comparison of whole genome sequences showed that the strain was closest to the SARS-like coronavirus RATG13 found in bats (*Rhinolophus affinis*) in China [2]. The overall genomic sequence similarity of RATG13 to SARS-CoV-2 was 96.1%. However, the spike protein gene of RATG13 lacked the furin cleavage site that is essential for the cell entry of SARS-CoV-2, indicating that RATG13 was not the immediate ancestor of SARS-CoV-2 [34,35]. After discovery of the first SARS-CoV-2 strains, SARS-CoV-2 like viruses were found in pangolins and bats [36–38]. The genome sequences of pangolin-derived CoVs also showed high similarity to those of SARS-CoV-2, but the furin cleavage site was missing in the spike gene sequences of pangolin-derived CoVs [35,37,38]. A bat-derived CoV, RmYN02, was identified and the genome of the virus showed high similarity to that of SARS-CoV-2 [36]. Although the sequence similarity of RmYN02 was slightly lower than those of RATG13 and pangolin-derived CoVs, the furin cleavage site was inserted, indicating that the addition of the cleavage site can occur naturally [35,36].

The D614G variants had a change in spike gene and replaced the initial strains of SARS-CoV-2 [39]. The studies on the variant D614G showed that the infectivity of the variant was increased without increased severity [39–41]. The engineered variants containing the D614G substitution showed more efficient infection in human cells and animal models without altering antibody neutralization and pathogenicity [41]. A population genetic analysis of COVID-19 also showed that the transmissibility of the variant was increased but there was no sign of increased mortality or clinical severity of the variants [40].

As new variants with increased pathogenicity, reduced neutralization, and/or increased transmissibility emerged, the U.S. Centers for Disease Control and Prevention (CDC) and Public Health England (PHE) classified some notable variants according to the attributes of the variants [42,43]. The CDC classified the variants according to the evidence and significance of the variants into Variant of Interest, Variant of Concern, and Variant of High Consequences. PHE classified variants as Variant Under Investigation (VUI) and Variants Of Concern (VOC). When the variants are considered to have concerning characteristics, they are designated as VUI. After a risk assessment of VUI is conducted, they can be re-designated as VOC. These notable variants are summarized in Table 2.

VOC-20DEC-01, also known as 20I/501Y.V1 or B.1.1.7, was first discovered in the United Kingdom in December 2020 [42] and is defined by 13 mutations [42]. Recent studies have estimated that the transmissibility of VOC-20DEC-01 is increased by 43–90% and a similar transmission increase was observed globally [44]. VOC-20DEC-01 was also detected in domestic cats and dogs, raising concern over human-to-animal transmission or vice versa [45]. Previously reported animal infections were asymptomatic to mild symptomatic, but VOC-20DEC-01 infection in animals showed relatively severe symptoms such as myocarditis [45]. In February 2021, the variants with a spike gene E484K mutation were reported and designated as VOC-202102/02 [42]. Another variant with N501Y mutation is VOC-20DEC-02 (20H/501Y.V2, or B.1.351), which was first discovered in South Africa. VOC-20DEC-02 is defined by 17 mutations including the E484K mutation, K417N mutation, and two deletions. The variant also showed increased transmissibility (approximately 50%) compared to previous variants [44]. The third variant with N501Y is VOC-21JAN-02 (P.1 or 20J/501Y.V3), discovered in Brazil [42,46]. The genome of VOC-21JAN-02 is defined with 17 non-synonymous mutations, four synonymous mutations, three deletions, and four insertions [42]. VOC-202101/02 almost fully replaced its parental variant within

two months, indicating increased transmissibility of VOC-21JAN-02 [47,48]. Molecular clock analysis showed that the variants emerged in mid-November 2020 at which time hospitalizations rapidly increased [49].

Table 2. Notable variants of SARS-CoV-2.

PANGO Lineage	CDC Designation	PHE Designation	First Detected	Spike Protein Substitutions
B.1.1.7	VOC	VOC-20DEC-01, VOC-21FEB-02 *	United Kingdom	69del, 70del, 144del, (E484K), (S494P), N501Y, A570D, D614G, P681H, T716I, S982A, D1118H (K1191N)
B.1.351	VOC	VOC-20DEC-02	South Africa	D80A, D215G, 241del, 242del, 243del, K417N, E484K, N501Y, D614G, A701V
P.2	VOI	VUI-21JAN-01	Brazil	E484K, (F565L), D614G, V1176F
P.1	VOC	VOC-21JAN-02	Brazil	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I
A.23.1	-	VUI-21FEB-01 *	Uganda	F157L, V367F, (E484K), Q613H, P681R
B.1.525	VOI	VUI-21FEB-03	United Kingdom	A67V, 69del, 70del, 144del, E484K, D614G, Q677H, F888L
B.1.1.318	-	VUI-21FEB-04	United Kingdom	D614G, D796H, E484K, P681H, T95I, 144del
P.3	-	VUI-21MAR-02	Philippines	E484K, N501Y, P681H
B.1.617	VOI	VUI-21APR-01	India	L452R, E484Q, D614G
B.1.617.2	VOI	VOC-21APR-02	India	T19R, (G142D), 156del, 157del, R158G, L452R, T478K, D614G, P681R, D950N
B.1.617.3	VOI	VUI-21APR-03	India	T19R, G142D, L452R, E484Q, D614G, P681R, D950N
AV.1	-	VUI-21MAY-01	United Kingdom	D80G, T95I, G142D, 144del, N439K, E484K, D614G, P681H, I1130V, D1139H
B.1.617.1	VOI	-	India	(T95I), G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H
B.1.526	VOI	-	United States	(L5F), T95I, D253G, (S477N), (E484K), D614G, (A701V)
B.1.526.1	VOI	-	United States	D80G, 144del, F157S, L452R, D614G, (T791I), (T859N), D950H
B.1.427	VOC	-	United States	L452R, D614G
B.1.429	VOC	-	United States	S13I, W152C, L452R, D614G

* with E484K; ( ) detected in some sequences but not all; VOC; Variant of Concern, VOI; Variant of Interest, VUI; Variant under Investigation.

The characteristic mutations (N501Y, E484K, and K417N) of the variants with N501Y are mutations in binding sites to viral receptor ACE2 and were already a concern prior to the discovery of these variants [50–53]. The studies on these variants showed that they had impacts on neutralization by immunity [54–60]. However, recent research showed that the residual immunity still provided protection, although variants reduced the efficacy of the vaccine [61].

There were also emerging variants without N501Y, E484K, and/or K417N. The characteristic mutations of A.23.1 are F157L, V367F, Q613H, and P681R [62]. A.23.1 with E484K was designated as VUI-21FEB-01 in the United Kingdom. These strains were first identified in Uganda and are spreading. One of the characteristic mutations, Q613H, is regarded as functionally equivalent to the D614G mutation of 'G' clade strains. B.1.427 and B.1.429 were first discovered and designated as Variants of Concern in the United States [43]. The characteristic mutations of both lineages are L452R and D614G; these variants showed

increased transmissibility and reduced neutralization by convalescent and post-vaccination sera [63].

B.1.617 was the emerging lineage in India and also designated as VUI-21APR-01 in the United Kingdom [42]. B.1.167 has two characteristic mutations of different lineages: L452R and E484Q [64]. The variants were neutralized with convalescent sera of COVID-19 patients and vaccine of BBV152, although the efficacy was low [64].

2. PCR-Based SARS-CoV-2 Detection

2.1. Reverse Transcription Quantitative PCR (RT-qPCR) Method

Detection of the SARS-CoV-2 viral genome, consisting of single-stranded RNA, is effectively done by reverse transcription quantitative polymerase chain reaction (RT-qPCR), which is the gold standard technique widely used in molecular diagnostics [65,66]. There are several practical considerations when performing diagnostic assays using RT-qPCR.

(1) Sample quality: RT-qPCR tests are presently being used for the identification of SARS-CoV-2 in clinical specimens such as upper respiratory tract specimens (saliva, oropharyngeal swab-OPS, nasopharyngeal swab-NPS, nasal swabs), lower respiratory specimens (sputum, bronchoalveolar lavage-BAL, endotracheal aspirate-ET, fibrobronchoscope brush biopsy-FBB), blood (serum, plasma), urine, feces, rectal/anal swabs, stool, and corneal secretion [67,68]. To check the sample quality of clinical specimens from different origins, an RNA isolation procedure is required to obtain purified high-quality RNA from the samples, which then needs to be analyzed using chip-based capillary electrophoresis (such as the Agilent Bioanalyzer system), electrophoretic separation on a high-resolution agarose gel, and spectrophotometry [69].

(2) Reference curve: Data processing can critically affect the analysis of RT-qPCR results [70]. PCR data processing is based on standard curves or on PCR efficiency assessments [70]. Standard curves are used to assess RT-qPCR efficiency without theoretical and practical problems [70]. The estimation of RT-qPCR efficiency using standard curves usually involves the serial dilution of a concentrated stock solution, after which standard samples are analyzed through RT-qPCR by measuring the quantification cycle (Cq) using standard procedures [70]. The most widely used Cq value is the threshold cycle (Ct), the cycle at which the expression of a target gene first exceeds a calculated fluorescence threshold level [71]. For example, to detect low amounts of SARS-CoV-2 RNA, a series of diluted RNA templates are used to determine the Ct value, which can provide a standard curve for evaluating the reaction efficiency [72]. However, the Ct value itself cannot be directly explained as viral load without a standard curve using reference materials [73]. When interpreting the results of SARS-CoV-2 RT-qPCR, the validity of the standard curve should be proved using reference materials with accurate viral copy numbers to interpret Ct values as viral loads [73].

(3) Viral load: The success of virus isolation depends on the viral load [74]. Viral loads in sputum samples and throat swabs are high when obtained within seven days after initial symptoms are observed, ranging from 10^4 to 10^7 copies per mL. This pattern is broken as low quantity of virus are obtained from samples taken after day 8 [75]. In general, sputum samples show higher viral loads than throat swab samples, whereas low viral RNA is detected in urine or stool samples [75]. The two main factors that influence the quantitative measurement of viral roads are Cq values that are repeatable with acceptable uncertainty and a reliable means of converting from the Ct value to viral load [76–78]. For molecular diagnostic assays, a limit of detection (LoD) and a limit of quantification (LoQ) are also considered the lowest concentrations of target RNA that can be detected by RT-qPCR [79].

(4) Sampling methods: RT-qPCR tests for SARS-CoV-2 have shown a high variation of false-negative rates (FNR) and false-positive rates (FPR) [80,81]. Numerous methods have been developed with the goal of improving the sensitivity, safety, and rapidity of COVID-19 tests by RT-qPCR. For example, one group tested the efficiency and sensitivity of SARS-CoV-2 detection of clinical specimens collected directly in nucleic acid stabilization

and lysis buffer (NSLB), a mixture of lysis buffer and RNA preservative, instead of a viral transport medium (VTM), thus inactivating the virus immediately after sampling [82].

(5) Sample source: To improve the expandability of SARS-CoV-2 testing, several sampling approaches have been developed including nasal, pooled nasal, and throat (oropharyngeal) swabs as well as saliva. Different clinical sampling methods affect the diagnostic performance of SARS-CoV-2 infection tests by RT-qPCR including sensitivity and specificity, and thus should be carefully considered [83–86]. The combined swab is largely recommended as a more appropriate specimen for diagnosis by RT-qPCR [87–89].

(6) Sensitivity: The conserved regions, ORF 1ab (RNA-dependent RNA polymerase, RdRp), envelope (E), and nucleocapsid (N) genes of SARS-CoV-2, are usually selected as the standard target genes for primer and probe design [90,91]. However, initial reports of SARS-CoV-2 and other coronavirus sequences gave rise to an incorrect degenerate base that did not align with the SARS-CoV-2 RNA sequence found, and there were reports regarding the decreased sensitivity of using RdRp as a target gene for RT-qPCR assays [90,92]. As the pandemic continues, many laboratories around the world rely on routine diagnostic primers and probes. Thus, proper assays can increase the sensitivity of SARS-CoV-2 detection and help prevent the further spread of the virus [92–95].

(7) Pooling technologies: The pooling of multiple swab samples before RNA isolation and RT-qPCR analysis has been proposed as a promising solution to reduce costs and time as well as elevate the throughput of SARS-CoV-2 tests for large-scale testing as in the case of schools [96–99]. For example, batch testing of over 100,000 hospital-collected nasopharyngeal swab samples from patients alleviated three quarters of testing reactions with a minor reduction in sensitivity, indicating the effectiveness of the pooling approach in the field [100,101]. Current studies suggest that the pooling of individual samples before testing should be considered to increase the reliability of SARS-CoV-2 testing throughput.

Once all practical considerations have been evaluated, there are two ways that RT-qPCR can be performed. The two-step RT-qPCR method is required to convert RNA into complementary DNA (cDNA) [102]. On the other hand, the one-step RT-qPCR method combines reverse transcription and PCR in a single tube and uses reverse transcriptase as well as a DNA polymerase [103]. The schematic procedure of RT-qPCR is shown in Figure 1A.

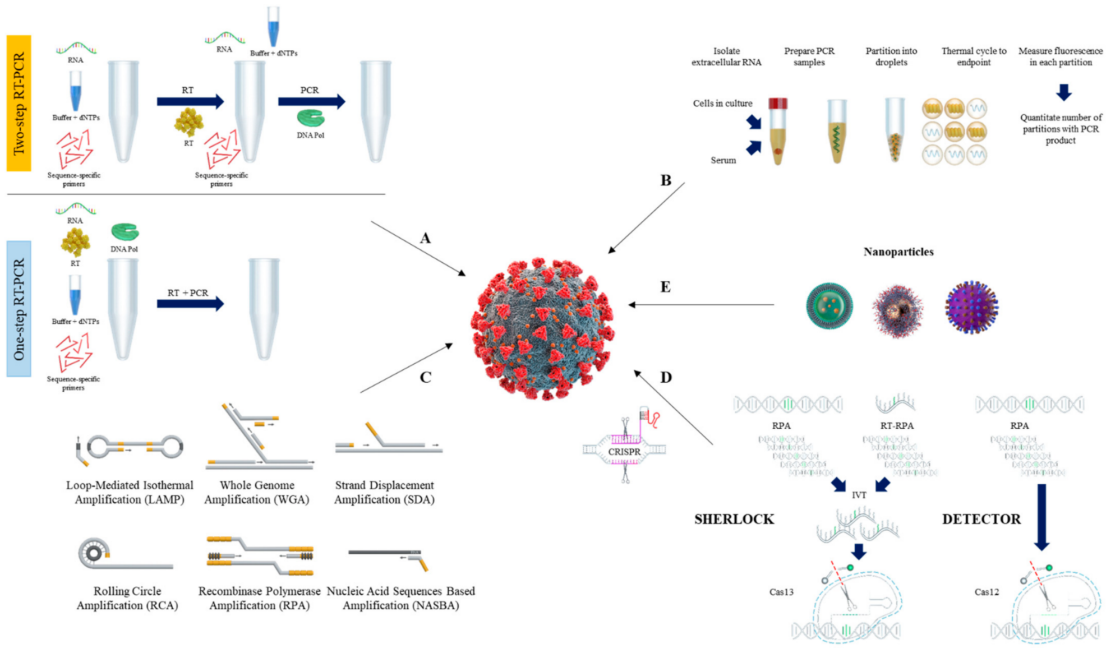


Figure 1. Overview of nucleic acid testing for SARS-CoV-2. The schematic procedure of RT-qPCR (**A**), and dPCR (**B**). Current isothermal amplification methods (**C**), CRISPR detection systems (**D**), and nanoparticles (**E**) are also shown.

A critical need for rapid and accurate diagnostic methods has emerged in the clinic and public health organizations. Several PCR-based assays have been developed and are currently being used in clinical, research, and public health laboratories [104–106]. However, it is not clear which PCR condition they should adopt or whether the data are comparable. In response to the growing need and the lack of publicly available information, several research groups have optimized real-time PCR-based primer sets, protocols, and PCR conditions [107,108].

Independent evaluations of the designed primer–probe sets used in SARS-CoV-2 RT–qPCR detection assays are necessary to compare and select appropriate assays [90,109]. Additionally, several studies have utilized serum and stool specimens for the RT-qPCR-based detection method [110–113].

2.2. Reverse Transcription Digital PCR (RT-dPCR) Method

In recent years, we have seen the advance of digital PCR (dPCR) as a complementary approach for measuring nucleic acids, a technique that is highly accurate and reproducible when targeting the viral genes of SARS-CoV-2 [114–117]. The advantages of digital PCR compared to quantitative PCR include quantification without the need for calibration curves, higher accuracy, and sensitivity that may arise from sub-optimal amplification efficacy because dPCR can detect low amounts of nucleic acid [118,119]. The schematic procedure of dPCR is shown in Figure 1B.

Reverse transcriptase quantitative PCR (RT-qPCR) and digital PCR (dPCR) have been widely used for quantitative analyses of clinical samples. Recently, many groups have developed a reverse transcription droplet digital PCR (RT-ddPCR) assay for sensitive detection of the SARS-CoV-2 virus [120–123]. Optimization of the primer. and probe assays is necessary to remove false negatives or positives for clinical diagnosis of viral infection [72,124]. Multiple molecular diagnostic kits have been developed and validated for use nationwide [125]. However, the analytical sensitivity and the relative sensitivity of different kits to detect low copy number of SARS-CoV-2 viral RNA are variable [126,127].

3. Isothermal Nucleic Acid Amplification Methods

Although the RT-qPCR method is considered the 'gold standard' for SARS-CoV-2 detection [128], its limitations have stimulated the development of simple, rapid yet sensitive nucleic acid detection methods [129]. As a result, isothermal nucleic acid amplification has emerged as an alternative detection method for SARS-CoV-2 viral RNA from clinical samples [130]. In general, isothermal amplification techniques increase the analytical signal by increasing the target nucleic acid concentration through enzymatic activities at a fixed temperature, and simultaneously detecting the signal with colorimetric or fluorescence indicators [131]. Changes in color, fluorescence level, or turbidity indicate the presence of SARS-CoV-2 RNA or DNA [131]. Therefore, unlike RT-qPCR, isothermal amplification methods do not require thermal cycling instruments or specialized technicians for disease diagnosis [132]. Current isothermal nucleic acid amplification methods used for SARS-CoV-2 detection include, but are not limited to, loop-mediated isothermal amplification (LAMP), recombinase polymerase amplification (RPA), nucleic acid sequence-based amplification (NASBA), strand-displacement amplification (SDA), and rolling circle amplification (RCA) [133–137] (Figure 1C). Herein, we describe the general procedures and components of isothermal amplification methods commonly used for diagnosis of SARS-CoV-2.

3.1. Loop-Mediated Isothermal Amplification (LAMP)

The loop-mediated isothermal amplification method, coupled with reverse transcription (RT-LAMP), is the most widely used isothermal amplification technique for SARS-CoV-2 nucleic acid detection. First described by Notomi et al. [138], this method uses strand displacement activity of DNA polymerase and a set of inner and outer primers (four or six specific primer sequences) to amplify the target nucleic acids. LAMP is carried out at a single temperature between 60 and 65 °C, and generates up to 10^9 copies of DNA in

less than an hour [133,137,139]. The LAMP procedure is initiated by hybridization of the forward inner primer (FIP) toward the target DNA template, which synthesizes the complementary strand. Then, the outer primer hybridizes to the target DNA, which initiates DNA synthesis by strand displacement. Subsequently, a FIP-hybridized complementary strand is released and forms a loop structure at one end of the sequence. The corresponding sequence becomes the template for the backward inner primer (BIP), which initiates another DNA synthesis by strand displacement, and then produces a 'dumb-bell' like DNA structure. Self-primed DNA synthesis of the corresponding sequence then converts the 'dumb-bell' like structure into a 'stem-loop' like DNA structure. Corresponding stem-loop DNA then becomes the template for LAMP cycling, and the target DNA sequence exponentially amplifies until the reaction is completed [138]. Amplified products are detected by changes of color as the accumulation of DNA changes, pH levels, or by changes in turbidity as magnesium pyrophosphate level increases [140–142]. Amplified products are also detected by Calcein fluorescent dye or fluorescent intercalating dye [129,139]. The schematic procedure of RT-LAMP is shown as Figure 2.

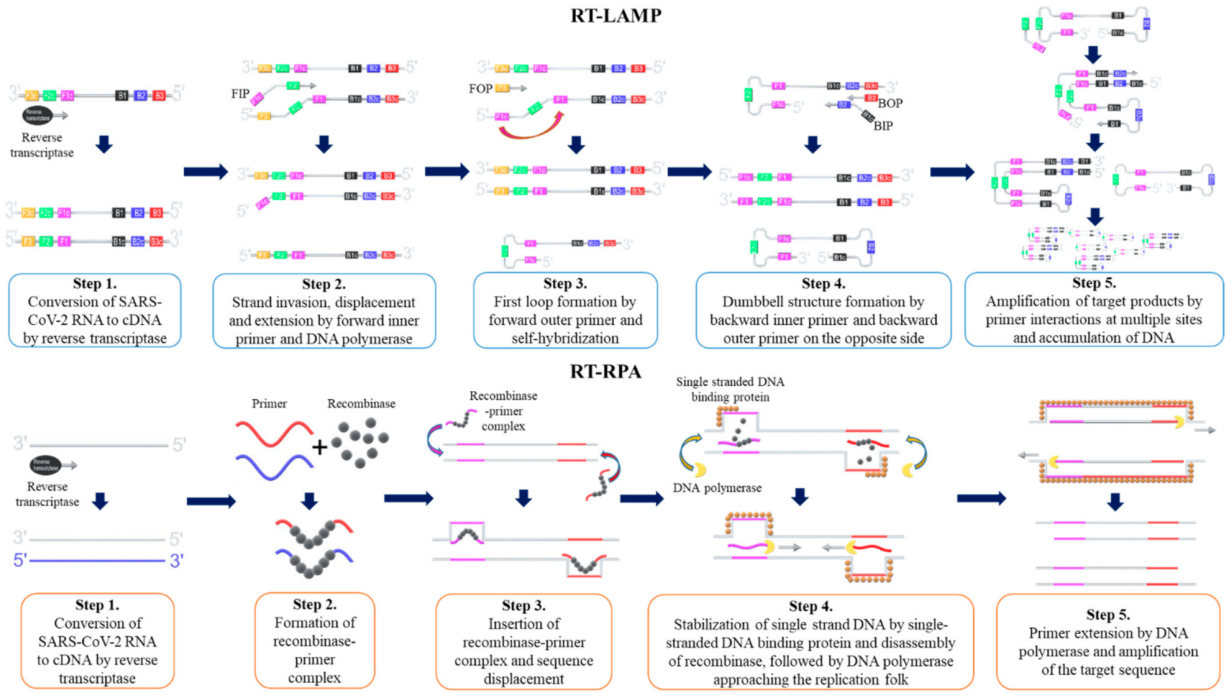


Figure 2. Schematic procedure of reverse transcription loop-mediated isothermal amplification (RT-LAMP) and reverse transcription recombinase polymerase amplification (RT-RPA). FIP = Forward Inner Primer, FOP = Forward Outer Primer, BIP = Backward Inner Primer, BOP = Backward Outer Primer.

Researchers have made efforts to optimize RT-LAMP for the development of rapid and sensitive detection of SARS-CoV-2. Several studies have evaluated the experimental parameters for RT-LAMP such as incubation temperature, incubation time, LoD, target genes, and primer sequences [143–145]. Aside from optimizing the experimental parameters, researchers have developed modified RT-LAMP procedures including methods without prior RNA extraction steps and high-throughput colorimetric assay methods using a 96-well plate format [146,147]. Modified RT-LAMP procedures also include methods coupled with Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) technology, a nanoparticle-based biosensor, and artificial intelligence [148–151].

3.2. Recombinase Polymerase Amplification (RPA)

Recombinase polymerase amplification is another isothermal amplification method that is widely used for SARS-CoV-2 detection. First described by Piepenburg et al. [152],

RPA uses a complex of recombinase and two target specific primers (forward and reverse primers) to amplify the target nucleic acids [152]. Once the target nucleic acids are identified, recombinase-primer complex unwinds the target DNA and allows forward and reverse primers to hybridize [153]. The displaced DNA strand is amplified in the presence of DNA polymerase as primers elongate, and the template DNA is exponentially amplified until the reaction is completed [153]. RPA reaction is carried out at a single temperature between 37 and 42 °C, and the reaction is completed when ATPs are depleted, typically in less than an hour [154]. Amplified products are detected by gel electrophoresis, antigenic tags on primers and tag-specific antibodies, or fluorescent signals produced by a conjugated fluorophore and quencher on primers [152–154]. The schematic procedure of reverse transcription RPA (RT-RPA) is shown as Figure 2.

For SARS-CoV-2 detection, researchers have optimized RPA procedures by testing various experimental parameters that can now detect less than five viral copies from patient samples within 45 min from the sample collection [134]. RPA methods have also been optimized for SARS-CoV-2 detection by coupling RPA-based amplification with various CRISPR-based detection methods [155–157].

3.3. Other Isothermal Nuleic Acid Amplification Methods

Other than LAMP and RPA, isothermal amplification methods such as NASBA, SDA, and RCA have been used for the detection of SARS-CoV-2 [136,137,158]. Although we will not describe each technique in detail here, Table 3 presents the general features and components of each isothermal amplification method.

Table 3. General features of the isothermal amplification techniques for SARS-CoV-2 detection.

Method	Components	Temperature	Time	Detection Method	Advantages *	Disadvantages *
Loop-mediated isothermal amplification (LAMP)	DNA polymerase, forward inner primer, backward inner primer, forward outer primer, backward outer primer	60–65 °C	>1 h	Colorimetric, turbimetric, fluorescence probe, intercalating dye	High specificity. Less sensitive to inhibitors in biological samples	False positive in negative control
Recombinase polymerase amplification (RPA)	Recombinase, single stranded binding protein, DNA polymerase, forward primer, reverse primer	37–42 °C	>1 h	Fluorescence, antigenic-tag (antibody)	Performed in the presence of PCR inhibitors. Fast and sensitive	Inhibited by detergents (SDS and CTAB). Non-specific/high background signal
Nucleic acid sequence-based amplification (NASBA)	RNase H, reverse transcriptase, T7 DNA-dependent RNA polymerase, forward primer with T7 promoter sequence, reverse primer	41 °C	>2 h	Fluorescence	More sensitive and less time-consuming	Non-specific reactions/false positives
Strand-displacement amplification (SDA)	DNA polymerase, restriction endonuclease, primers, dCTP, dTTP, dGTP, dATPα	37–49 °C	>2 h	Fluorescence	High specificity. Detection of large RNA molecules	Non-specific reaction/high background signal
Rolling circle amplification (RCA)	DNA ligase, DNA polymerase, primer, padlock probe	30–37 °C	>1.5 h	Fluorescence	High specificity	False negatives and false positives

* Advantages and disadvantages in comparison with RT-qPCR methods.

4. Non-Conventional Methods

4.1. Genome Sequencing

Unprecedentedly massive genome sequencing has been undertaken with SARS-CoV-2 strains. The total number of sequenced genomes is approximately a million at present [30]. As the number of genomes increases rapidly, the World Health Organization (WHO) has provided guidelines for the genome sequencing of SARS-CoV-2 [159]. According to this guideline, the genome sequencing of SARS-CoV-2 can be used for understanding the emergence of SARS-CoV-2, understanding the biology of SARS-CoV-2, improving diagnostics and therapeutics, investigating virus transmission and spread, and inferring epidemiological parameters [159]. Based on the accumulative genome sequences of SARS-CoV-2, the emergence of the variants of concern [43,44,48,49,57,62], the origin of SARS-CoV-2 [2,31], and the mutation frequency of RT-qPCR primer/probe sites [30] can be known to humanity. Currently, various whole genome sequencing methods of the virus are being developed [160–166]. The sequencing methods of viruses can be categorized into the metagenomics approach and target enrichment based methods. In the metagenomics approach, the viral genome can be extracted from clinical samples and the extracted nucleic acid are sequenced. This can also be done with cultured viruses. These approaches have a clear advantage over target enrichment based methods. The metagenomics approach can be used even if there is no information of the pathogen or there are novel pathogens that were not previously known. However, a high proportion of host cell genetic materials can be found, which should be removed or reduced for sequencing. The removal or depletion methods of the host genetic materials vary by the type of sample or the virus [167–174]. Due to the nature of the metagenomics approaches, the clinical samples should ideally have a high titer of the pathogens. The metagenomics approach also can be done with cultured pathogens. However, the isolation and culturing of the pathogen are very time-consuming and labor-intensive work. In some cases, the isolation and culturing of some pathogens are not possible or are very difficult [175,176]. Alternatively, target enrichment methods can be used for the genome sequencing. The genetic materials of specific pathogens can be enriched through hybrid capture probes [177,178]. The sequences of the probes are complementary to the genome sequences of specific pathogens and these target enrichments effectively remove not-target sequences and increase the proportion of the target sequences. One of the advantages of hybrid capture approaches is the tolerance of sequence mismatch, allowing the capture of divergent variants. However, the hybrid capture approaches are relatively more expensive and complicated than other approaches. Another group of target specific enrichment approaches is the amplicon based approaches. The amplicon-based approaches are mainly dependent on PCR reactions. The PCR reaction can selectively enrich the genome of the target pathogens in the presence of non-target nucleic acids just like host genetic materials. Due to the nature of PCR, the amplicon based approaches are relatively more inexpensive, sensitive, and specific than other approaches. The WHO guidelines suggest that the complete genome sequencing can be done from the sample with Ct values of up to 30 and the partial genome sequencing can be done from the sample with Ct values of 30–35, although the Ct value can vary with various factors [159]. Currently, the most widely used primer panels for SARS-CoV-2 are ARTIC network amplicon sets [179]. At least three commercially available SARS-CoV-2 primer panels (CleanPlex SARS-CoV-2 Panel; Paragon genomics, QIAseq SARS-CoV-2 Primer Panel; Qiagen, and NEBNext ARTIC SARS-CoV-2 Library Prep Kit; NEB) are based on ARTIC network amplicon sets. However, there are also limitations. The design of primers requires prior knowledge of full sequence information of the genome. In addition, the primer intolerance to the sequence mismatch hinders the genome sequencing of the variants. The amplicon appears that the amplicon based approaches can be applied only to previously well-known pathogens. Alternatively, the genomic materials can be amplified sequence-independently [163,180,181]. Single primer isothermal amplification (SPIA) can amplify the genomic materials in a sequence-independent manner. As SPIA can amplify the genomic materials, prior knowledge of the pathogens is not required for a target enrichment approach. However, removing

non-target genomic materials is mandatory for sequencing based SPIA as SPIA can also amplify non-target genomic material. Due to these characteristics, a high proportion of target genomic materials in the samples is crucial for successful SPIA based sequencing. SPIA based sequencing with low viral input showed very low coverage compared to other methods [180].

For the genome sequencing of pathogens, various sequencing technology can be used. While the conventional Sanger sequencing can still be used for viral genome sequencing [182], most SARS-CoV-2 genome sequencing is done with NGS sequencing technology. Currently, the most widely used NGS sequencing technology is the sequencing platforms of Illumina. Although the sequence length of individual reads is relatively short (paired-end 150 bp), the throughput and the accuracy of the individual reads are outstanding. Ion Torrent is another short reads sequencing platform technology, where the length of individual reads is 400 bp or 600 bp. The running time of the Ion Torrent sequencer is shorter than that of the Illumina sequencer. Long read alternatives are also available. The lengths of individual reads from PacBio and Oxford Nanopore Technology sequencers are tens of kilo base pairs or more. The individual reads from these long read sequences can cover most of the viral genome. However, the throughput of the long read sequencers is relatively lower than that of short read sequencers such as Illumina sequencing platforms. Furthermore, the accuracy of the individual reads is relatively lower than that of Illumina sequencing platforms. The sequencing platforms of Oxford Nanopore Technology maximize the benefits and drawbacks of the long read sequencer. The maximum length of the individual read is recorded up to a megabase scale [183]. However, due to the relative low accuracy of the individual reads, the WHO guidelines do not recommend these for SARS-CoV-2 genome sequencing unless the sequencing is replicated [159,184]. The coverage and depth of viral genome sequencing can be varied by the number of samples in single runs. Generally, most multiplex sequencing library kits for NGS support up to 384 samples per single runs. However, the production scale sequencers of Illumina (Hiseq, NextSeq, and NovaSeq) generate massive reads for small genome of viruses even with multiplex libraries. Though short individual reads require relatively more depth for high coverage, massive generation of sequence reads and low sequencing error rates of individual reads can compensate for short individual read length. Due to the massive sequence reads generation of the Illumina sequencer, the metagenomics approach of viral genome sequencing is practically available to only Illumina sequencer or similar platforms. Even if most of the sequence reads are non-target sequence reads (host genetic materials, contamination, etc.), a high quality genome assembly of the virus can be produced from the small remaining fraction of the target sequences. The long read sequencers such as sequencers by Pacbio and Oxford Nanopore Technology can generate very long individual reads that can cover most of the viral genome. However, due to the relatively low yield of total sequence reads and low accuracy of individual reads, the long read sequencers are not adequate for the metagenomics approach of viral genome sequencing. Instead, the long sequencers are more suitable for the amplicon-based approach. The target specific amplification can overcome the drawbacks of long read sequencers such as the low yield of total sequence reads and improving the accuracy of individual reads. Moreover, long sequencers can use long amplicons unlike short read sequencers. The schematic procedure of genome sequencing is shown in Figure 3.

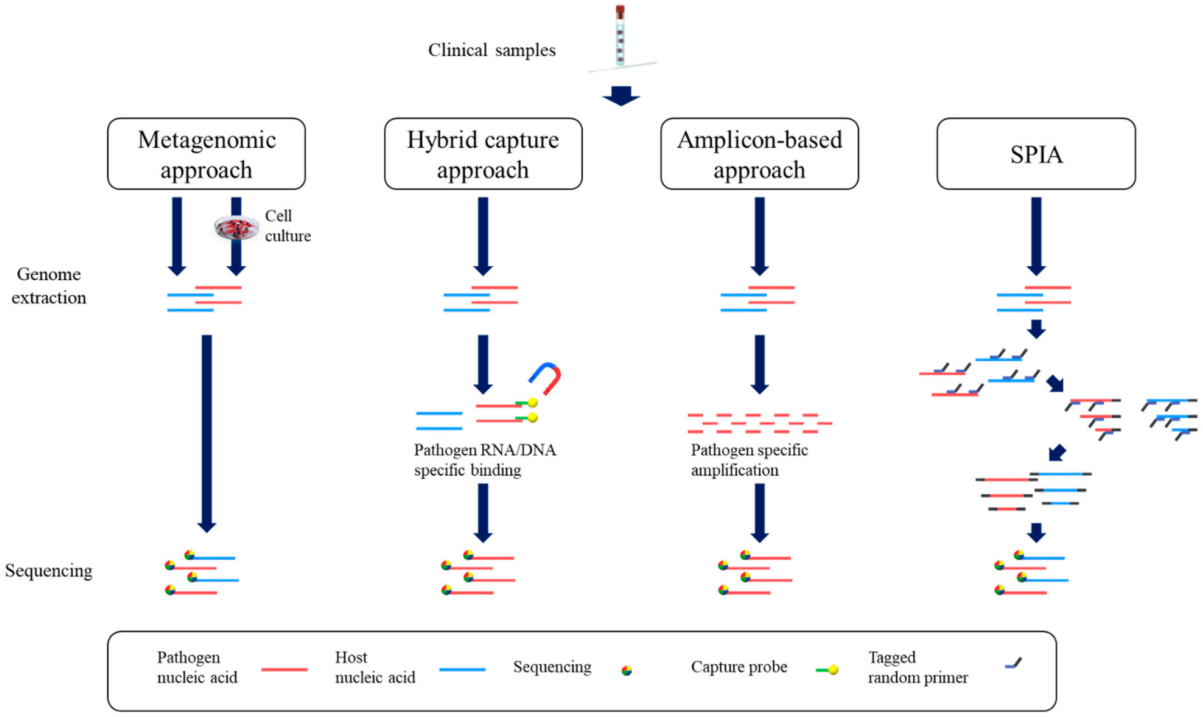


Figure 3. Schematic procedure of genome sequencing.

4.2. CRISPR Based COVID-19 Detection

The clustered regularly interspaced short palindromic repeats (CRISPR)-Cas technique has been repurposed for diagnostics and is one of the widely used nucleic acid detection methods [185,186]. Many types of Cas proteins have been developed to create highly accurate and sensitive diagnostic methods [187]. Cas9 has been widely used for genome editing while DNA-targeting Cas12 (also known as Cpf1 or C2c1) effectors and RNA-targeting Cas 13a are more suitable for disease diagnosis [188].

Compared to conventional diagnostic methods such as RT-qPCR, CRISPR-based approaches can quickly provide rapid, visual, highly sensitive, and specific detection due to the collateral cleavage of a reporter dye in the presence of a target [189].

Numerous CRISPR-Cas detection systems have been developed. For example, techniques of a CRISPR–Cas12-based assay have been developed for the detection of SARS-CoV-2 from patient sample RNA, called SARS-CoV-2 DNA Endonuclease-Targeted CRISPR Trans Reporter (DETECTR) [190]. This assay includes simultaneous reverse transcription and isothermal amplification using loop-mediated amplification (RT–LAMP) [191]. An RNA targeting Cas13a dependent platform [156], the SHERLOCK (Specific High Sensitivity Enzymatic Reporter UnLOCKing) technique, offers a simplified test and has a limit of detection of 100 copies of the viral genome [192,193].

Recently, a lyophilized CRISPR-Cas12 assay for SARS-CoV-2 detection (Lyo-CRISPR SARS-CoV-2 kit) has been developed based on reverse transcription (RT), isothermal amplification, and CRISPR-Cas12 reaction [194]. The schematic procedure of CRISPR detection systems is shown in Figure 1D.

4.3. Nanotechnology Based Methods

Nanotechnology has already proven its value through its diagnostic, vaccine, and therapeutic applications that have expanded into clinical applications [195]. Scientists have shown that nucleic acid detection using nanomaterials for viral infectious diseases now have various advantages in the diagnostic field [196]. Moreover, nanomaterials are powerful tools for the diagnosis, prevention, and treatment of COVID-19 [197].

Magnetic nanoparticles have been used in RT-qPCR diagnosis for the extraction of viral RNA from SARS-CoV-2. This method merges the lysis and binding steps, and the poly (amino ester) with carboxyl groups (PC)-coated magnetic nanoparticles (pcMNPs)-RNA complexes can be directly introduced into RT-qPCR reactions [198]. In addition, a test has been developed to diagnose SARS-CoV-2 that can rapidly detect the virus. The test is performed using gold nanoparticles to detect specific proteins such as nucleocapsid phosphoprotein. In the presence of gold nanoparticles, the test is positive upon the color of the liquid reagent changing from purple to blue [199]. Nanoparticles that can be applied in the RT-qPCR diagnosis are shown in Figure 1E.

5. Resources and Standardization for SARS-CoV-2 Nucleic Acid Testing

5.1. International Activity for Standardization

The Consultative Committee on the Quantity of Material (CCQM) is responsible for developing and documenting the equivalence of national standards such as certified reference materials (CRMs)/reference materials (RMs) and reference methods for biological and chemical measurements [200]. It advises the International Committee for Weights and Measures (Comité international des poids et mesures, CIPM) on matters related to biological and chemical measurements including advice on the International Bureau of Weights and Measures' (Bureau international des poids et mesures, BIPM) scientific program activities [200]. One of the responsibilities of the CCQM is to contribute to the development of a worldwide admitted system of national measurement standards, reference methods, and facilities for biological and chemical measurements [201]. Several National Metrology Institutes (NMIs) and expert laboratories from many countries have performed highly sensitive and accurate measurements of the amount of the SARS-CoV-2 viral RNA tested using reverse transcription-digital PCR (RT-dPCR) [202]. The capability to accurately measure the amount of COVID-19 causing viral nucleic acid with equivalence globally will remarkably improve diagnostic testing confidence and support countries in effectually tackling the pandemic situation [202]. Moreover, the biological, chemical, and physical measurement ability of NMIs are enabling industry and laboratories to effectively and quickly face the COVID-19 challenge [203].

NMIs support quality assurance by developing and providing CRM/RM for COVID-19 and ensures equivalent pathology testing, while minimizing false negative and positive test outcomes.

Non-NMIs including both public and private sectors have developed standards and reference materials for SARS-CoV-2 RNA. The National Institute for Biological Standards and Control (NIBSC) distributed the 1st WHO International Standards for SARS-CoV-2 RNA in 2020, which is used for the standardization of nucleic acid amplification technique (NAT)-based diagnostic assays. The U.S. Food and Drug Administration developed the SARS-CoV-2 Reference Panel in 2020 to precisely compare the performance of NAT-based assays for SARS-CoV-2 detection. Non-profit organizations such as American Type Culture Collection (ATCC) and private industries such as LGC SeraCare have also developed and distributed reference materials for NAT-based SARS-CoV-2 diagnosis.

5.2. Reference Materials

A RM can be defined as a "material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process" [204]. According to the general requirements for the competence of reference material producers (ISO 17034:2016), a CRM is defined as a "reference material characterized by a metrologically valid procedure for one or more specified properties, accompanied by a reference material certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability" [204]. Generally, major producers of RMs are NMIs [205]. However, RMs of SARS-CoV-2 are produced by not only NMIs, but also by commercial institutes and culture collections. The first RM of SARS-CoV-2 was produced by the National Institute of

Metrology of China (NIMC). The RM was synthetic RNA based on the genome sequence of Wuhan-Hu-1, the first identified SARS-CoV-2 strain. The RM of NIMC was quantified with the ddPCR method and contained N, E, and partial genes of RdRp genes that can cover the WHO announced in-house assay for SARS-CoV-2 [91]. Other NMIs such as the Joint Research Center in Europe (JRC), Korea Research Institute of Standards and Science (KRISS), National Measurement Institute of Australia (NMIA), National Institute of Standards and Technology (NIST), and The National Metrology Institute of Turkey (UME) have also produced RMs for SARS-CoV-2 nucleic acid testing. NIBSC also produced a RM for SARS-CoV-2 nucleic acid testing and it has been used as WHO reference materials. Type culture collection also produced RMs. ATCC and BEI resources produced RMs with inactivated SARS-CoV-2 strain (USA-WA1/2020). Not only public institutes but also commercial institutes have produced RMs. Seracare, which is a subsidiary of LGC, produced RMs in the early phase of the pandemic. The RMs were SARS-CoV-2 RNA covered with viral proteins, which can be used for the WHO announced in-house assays. Twist Bioscience produced RMs with various SARS-CoV-2 strains including emerging variants such B.1.1.7, B.1.351, and P.1.

The majority of these RMs were quantitated with dPCR. The comparison of qPCR and dPCR methods showed that dPCR can quantitate specific genes regardless of primer sequences [72], indicating that the dPCR assay is a suitable method for the measurement of RMs. The reference materials of SARS-CoV-2 nucleic acid testing are listed in Tables 4 and S1.

Table 4. Reference materials and resources for SARS-CoV-2 nucleic acid testing.

Institute	Type	Numbers
ATCC	heat-inactivated	3
	Synthetic RNA	5
Bio-Rad	Synthetic RNA	1
JRC	Synthetic RNA	1
KRISS	Synthetic RNA	1
	Virus Like Particle	1
NIBSC	Heat-inactivated	1
NMIA	Inactivated	1
NIMC	Synthetic RNA	1
NIST	Synthetic RNA	1
Randox Qnostics	Heat-inactivated	3
Seracare	Virus Like Particle	4
Thermo Scientific	Inactivated	1
	Genomic RNA	1
	Synthetic RNA	2
Twist Bioscience	Synthetic RNA	20
UME	Synthetic RNA	2
ZeptoMetrix	Chemical-inactivated	1

ATCC; American Type Culture Collection, JRC; Joint Research Center, KRISS; Korea Research Institute of Standards and Science, NIBSC; National Institute for Biological Standards and Control, NMIA; National Measurement Institute Australia, NIMC; National Institute of Metrology of China, NIST; National Institute of Standards and Technology, UME; The National Metrology Institute of Turkey.

5.3. Other Resources

The genomic sequences of SARS-CoV-2 are mainly deposited in GISAID databases and are also available at the NCBI GenBank [27,30]. The metadata of strains such as

collection date, patient age, gender, sequencing methods, etc. are also available in GISAID. There are also other specialized databases. Virus Pathogen Resources (ViPR) provides detailed genome sequence information and analysis tools [206]. Nextstrain provides real-time tracking evolution and spreading of SARS-CoV-2 [28]. Nextstrain visualizes tracking of SARS-CoV-2 lineages by integrating geographic information and sequence information. CoV-GLUE is a mutation dedicated database for SARS-CoV-2 [207]. CoV-GLUE summarizes the mutations of amino acid replacements, insertions, and deletion.

Although more than a million SARS-CoV-2 genome sequences are publicly available, live SARS-CoV-2 sources are relatively limited. As SARS-CoV-2 is regarded as risk group 3 and handling of live SARS-CoV-2 requires biosafety level 3 laboratories, only a few culture collections distribute SARS-CoV-2 strains. BEI resources have been established by the National Institute of Allergy and Infectious Diseases (NIAID) and provide live SARS-CoV-2 strains and derivatives of SARS-CoV-2. The National Culture Collection for Pathogens of Korea (NCCP) also provides live viruses and derivatives from SARS-CoV-2 isolated in Korea. European Virus-Archive Global (EVAg) is a network between 25 laboratories including 16 EU member state institutions and nine non-EU institutions. It also provides isolated strains and derivatives from the EU and related countries. Training courses such as SARS-CoV-2 diagnostic training are also available at EVAg.

6. Conclusions

Currently, representative diagnostic methods of SARS-CoV-2 are RT-qPCR assays. Most countries use the RT-qPCR assay as a primary method for diagnostics. Though alternative methods are available, their sensitivity, specificity, or costs are not comparable to RT-qPCR assays. However, RT-qPCR assays require relatively expensive instruments and highly trained personnel. These requirements restrict the expansion of diagnostics capacity in some countries. To overcome these drawbacks, future diagnostics methods should be inexpensive and simple, which can be used for point-of-care testing. For example, if inexpensive whole viral genome sequencing methods are developed in the future, they will be capable of replacing RT-qPCR assays as a standard disease diagnostic test. The whole genome sequencing of the virus can provide more information including variant information of the viruses. The reference materials for the diagnostics are also important to assess newly developed diagnostic methods.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms22116150/s1, Table S1: Extended table of Reference materials and resources for SARS-CoV-2 nucleic acid testing.

Author Contributions: Conceptualization, S.K.; Investigation, H.M.Y., I.-H.K. and S.K.; Writing—original draft preparation, H.M.Y., I.-H.K. and S.K.; Writing—review and editing H.M.Y., I.-H.K. and S.K.; Funding acquisition, H.M.Y. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Council of Science and Technology grant by the Ministry of Science and ICT (Grant No. CRC-16-01-KRICT). This work was also supported by the "Establishment of measurement standards for microbiology", grant number GP2021-0003-08, funded by the Korea Research Institute of Standards and Science.

Acknowledgments: The professional illustrations were done by Scapiens.

Conflicts of Interest: Seil Kim and Hee-Min Yoo are producers of KRISS SARS-CoV-2 reference materials.

Abbreviations

COVID-19	Coronavirus disease 2019
PCR	polymerase chain reaction
RT-PCR	Reverse transcription PCR
RT-qPCR	Reverse transcription quantitative PCR
RT-dPCR	Reverse transcription digital PCR
RT-ddPCR	Reverse transcription droplet digital PCR
RT-LAMP	Reverse transcription loop-mediated isothermal amplification
RT-RPA	Reverse transcription recombinase polymerase amplification
NGS	Next-generation sequencing
NMI	National Metrology Institutes
RM	Reference materials
SARS-CoV-2	Severe acute respiratory syndrome coronavirus-2
GISAID	Global Initiative on Sharing All Influenza Data
PANGO lineage	Phylogenetic Assignment of Named Global Outbreak LINeages
CDC	US Centers for Disease Control and Prevention
PHE	Public Health England
VUI	Variant Under Investigation
VOC	Variants Of Concern
BAL	Bronchoalveolar lavage
ET	Endotracheal aspirate
FBB	Fibrobronchoscope brush biopsy
Cq	Quantification cycle
Ct	Threshold cycle
LoD	Limit of detection
LoQ	Limit of quantification
FNR	False-negative rates
FPR	False-positive rates
NSLB	Nucleic acid stabilization and lysis buffer
VTM	Viral transport medium
RdRp	RNA-dependent RNA polymerase
E	Envelope
N	Nucleocapsid
cDNA	Complementary DNA
LAMP	Loop-mediated isothermal amplification
RPA	Recombinase polymerase amplification
NASBA	Nucleic acid sequence-based amplification
SDA	Strand-displacement amplification
RCA	Rolling circle amplification
FIP	Forward inner primer
BIP	Backward inner primer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
WHO	World Health Organization
SPIA	Single primer isothermal amplification
DETECTR	DNA Endonuclease-Targeted CRISPR Trans Reporter
SHERLOCK	Specific High Sensitivity Enzymatic Reporter UnLOCKing
RT	Reverse transcription
CCQM	Consultative Committee on the Quantity of Material
CRMs	Certified reference materials
CIPM	Comité international des poids et mesures
BIPM	Bureau international des poids et mesures
NMI	National Metrology Institutes
NIBSC	National Institute for Biological Standards and Control
NAT	Nucleic acid amplification technique
ATCC	American Type Culture Collection

NIMC	National Institute of Metrology of China
JRC	Joint Research Centre in Europe
KRISS	Korea Research Institute of Standards and Science
NMIA	National Measurement Institute of Australia
NIST	National Institute of Standards and Technology
UME	The National Metrology Institute of Turkey
ViPR	Virus Pathogen Resources
NIAID	National Institute of Allergy and Infectious Diseases
NCCP	National Culture Collection for Pathogens of Korea
EVAg	European Virus-Archive Global

References

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef] [PubMed]
2. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270–273. [CrossRef] [PubMed]
3. Van Der Hoek, L.; Pyrc, K.; Jebbink, M.F.; Vermeulen-Oost, W.; Berkhout, R.J.M.; Wolthers, K.C.; Wertheim-Van Dillen, P.M.E.; Kaandorp, J.; Spaargaren, J.; Berkhout, B. Identification of a New Human Coronavirus. *Nat. Med.* **2004**, *10*, 368–373. [CrossRef]
4. Vabret, A.; Mourez, T.; Gouarin, S.; Petitjean, J.; Freymuth, F. An Outbreak of Coronavirus OC43 Respiratory Infection in Normandy, France. *Clin. Infect. Dis.* **2003**, *36*, 985–989. [CrossRef] [PubMed]
5. Vijgen, L.; Keyaerts, E.; Moës, E.; Thoelen, I.; Wollants, E.; Lemey, P.; Vandamme, A.-M.; Van Ranst, M. Complete Genomic Sequence of Human Coronavirus OC43: Molecular Clock Analysis Suggests a Relatively Recent Zoonotic Coronavirus Transmission Event. *J. Virol.* **2005**, *79*, 1595–1604. [CrossRef] [PubMed]
6. Owczarek, K.; Szczepanski, A.; Milewska, A.; Baster, Z.; Rajfur, Z.; Sarna, M.; Pyrc, K. Early Events during Human Coronavirus OC43 Entry to the Cell. *Sci. Rep.* **2018**, *8*, 7124. [CrossRef] [PubMed]
7. Pene, F.; Merlat, A.; Vabret, A.; Rozenberg, F.; Buzyn, A.; Dreyfus, F.; Cariou, A.; Freymuth, F.; Lebon, P. Coronavirus 229E-Related Pneumonia in Immunocompromised Patients. *Clin. Infect. Dis.* **2003**, *37*, 929–932. [CrossRef] [PubMed]
8. Hendley, J.O.; Fishburne, H.B.; Gwaltney, J.M. Coronavirus Infections in Working Adults. Eight-Year Study with 229 E and OC 43. *Am. Rev. Respir. Dis.* **1972**, *105*, 805–811. [PubMed]
9. Woo, P.C.Y.; Lau, S.K.P.; Chu, C.; Chan, K.; Tsoi, H.; Huang, Y.; Wong, B.H.L.; Poon, R.W.S.; Cai, J.J.; Luk, W.; et al. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. *J. Virol.* **2005**, *79*, 884–895. [CrossRef]
10. Lau, S.K.P.; Woo, P.C.Y.; Yip, C.C.Y.; Tse, H.; Tsoi, H.W.; Cheng, V.C.C.; Lee, P.; Tang, B.S.F.; Cheung, C.H.Y.; Lee, R.A.; et al. Coronavirus HKU1 and Other Coronavirus Infections in Hong Kong. *J. Clin. Microbiol.* **2006**, *44*, 2063–2071. [CrossRef]
11. Nitsche, A.; Schweiger, B.; Ellerbrok, H.; Niedrig, M.; Pauli, G. SARS Coronavirus Detection. *Emerg. Infect. Dis.* **2004**, *10*, 1300–1303. [CrossRef]
12. Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H.-R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R.A.M.; et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1967–1976. [CrossRef]
13. Poutanen, S.M.; Low, D.E.; Henry, B.; Finkelstein, S.; Rose, D.; Green, K.; Tellier, R.; Draker, R.; Adachi, D.; Ayers, M.; et al. Identification of Severe Acute Respiratory Syndrome in Canada. *N. Engl. J. Med.* **2003**, *348*, 1995–2005. [CrossRef]
14. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **2003**, *348*, 1953–1966. [CrossRef]
15. Peiris, J.S.M.; Lai, S.T.; Poon, L.L.M.; Guan, Y.; Yam, L.Y.C.; Lim, W.; Nicholls, J.; Yee, W.K.S.; Yan, W.W.; Cheung, M.T.; et al. Coronavirus as a Possible Cause of Severe Acute Respiratory Syndrome. *Lancet* **2003**, *361*, 1319–1325. [CrossRef]
16. Killerby, M.E.; Biggs, H.M.; Midgley, C.M.; Gerber, S.I.; Watson, J.T. Middle East Respiratory Syndrome Coronavirus Transmission. *Emerg. Infect. Dis.* **2020**, *26*, 191–198. [CrossRef]
17. Zaki, A.M.; van Boheemen, S.; Bestebroer, T.M.; Osterhaus, A.D.M.E.; Fouchier, R.A.M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **2012**, *367*, 1814–1820. [CrossRef]
18. Al Johani, S.; Hajeer, A.H. MERS-CoV Diagnosis: An Update. *J. Infect. Public Health* **2016**, *9*, 216–219. [CrossRef]
19. Seddiq, N.; Al-Qahtani, M.; Al-Tawfiq, J.A.; Bukamal, N. First Confirmed Case of Middle East Respiratory Syndrome Coronavirus Infection in the Kingdom of Bahrain: In a Saudi Gentleman after Cardiac Bypass Surgery. *Case Rep. Infect. Dis.* **2017**, *2017*, 1–4. [CrossRef]
20. Zheng, F.; Liao, C.; Fan, Q.H.; Chen, H.B.; Zhao, X.G.; Xie, Z.G.; Li, X.L.; Chen, C.X.; Lu, X.X.; Liu, Z.S.; et al. Clinical Characteristics of Children with Coronavirus Disease 2019 in Hubei, China. *Curr. Med. Sci.* **2020**, *40*, 275–280. [CrossRef]
21. Tay, M.Z.; Poh, C.M.; Rénia, L.; MacAry, P.A.; Ng, L.F.P. The Trinity of COVID-19: Immunity, Inflammation and Intervention. *Nat. Rev. Immunol.* **2020**, *20*, 363–374. [CrossRef] [PubMed]
22. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. Available online: https://coronavirus.jhu.edu/map.html (accessed on 31 May 2021).

23. Toyoshima, Y.; Nemoto, K.; Matsumoto, S.; Nakamura, Y.; Kiyotani, K. SARS-CoV-2 Genomic Variations Associated with Mortality Rate of COVID-19. *J. Hum. Genet.* **2020**, *65*, 1075–1082. [CrossRef] [PubMed]
24. Vandenberg, O.; Martiny, D.; Rochas, O.; van Belkum, A.; Kozlakidis, Z. Considerations for Diagnostic COVID-19 Tests. *Nat. Rev. Microbiol.* **2020**, *19*, 171–183. [CrossRef] [PubMed]
25. Dong, Y.; Dai, T.; Wei, Y.; Zhang, L.; Zheng, M.; Zhou, F. A Systematic Review of SARS-CoV-2 Vaccine Candidates. *Signal Transduct. Target. Ther.* **2020**, *5*, 1–14. [CrossRef]
26. Elbe, S.; Buckland-Merrett, G. Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health. *Glob. Challenges* **2017**, *1*, 33–46. [CrossRef]
27. Agarwala, R.; Barrett, T.; Beck, J.; Benson, D.A.; Bollin, C.; Bolton, E.; Bourexis, D.; Brister, J.R.; Bryant, S.H.; Canese, K.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13.
28. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. NextStrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* **2018**, *34*, 4121–4123. [CrossRef]
29. Rambaut, A.; Holmes, E.C.; O'Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403–1407. [CrossRef]
30. GISAID Clade and Lineage Nomenclature Aids in Genomic Epidemiology Studies of Active HCoV-19 Viruses. Available online: https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/ (accessed on 31 May 2021).
31. Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; et al. On the Origin and Continuing Evolution of SARS-CoV-2. *Natl. Sci. Rev.* **2020**, *7*, 1012–1023. [CrossRef]
32. Hodcroft, E.B.; Hadfield, J.; Neher, R.A.; Bedford, T. Year-Letter Genetic Clade Naming for SARS-CoV-2 on Nextstain.Org. Available online: https://nextstrain.org/blog/2020-06-02-SARSCoV2-clade-naming (accessed on 31 May 2021).
33. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [CrossRef]
34. Xia, S.; Lan, Q.; Su, S.; Wang, X.; Xu, W.; Liu, Z.; Zhu, Y.; Wang, Q.; Lu, L.; Jiang, S. The Role of Furin Cleavage Site in SARS-CoV-2 Spike Protein-Mediated Membrane Fusion in the Presence or Absence of Trypsin. *Signal Transduct. Target. Ther.* **2020**, *5*, 92. [CrossRef]
35. Hakim, M.S. SARS-CoV-2, Covid-19, and the Debunking of Conspiracy Theories. *Rev. Med. Virol.* **2021**, e2222. [CrossRef]
36. Zhou, H.; Chen, X.; Hu, T.; Li, J.; Song, H.; Liu, Y.; Wang, P.; Liu, D.; Yang, J.; Holmes, E.C.; et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr. Biol.* **2020**, *30*, 2196–2203.e3. [CrossRef]
37. Lam, T.T.Y.; Jia, N.; Zhang, Y.W.; Shum, M.H.H.; Jiang, J.F.; Zhu, H.C.; Tong, Y.G.; Shi, Y.X.; Ni, X.B.; Liao, Y.S.; et al. Identifying SARS-CoV-2-Related Coronaviruses in Malayan Pangolins. *Nature* **2020**, *583*, 282–285. [CrossRef]
38. Zhang, T.; Wu, Q.; Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr. Biol.* **2020**, *30*, 1346–1351.e2. [CrossRef]
39. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.e19. [CrossRef]
40. Volz, E.; Hill, V.; McCrone, J.T.; Price, A.; Jorgensen, D.; O'Toole, Á.; Southgate, J.; Johnson, R.; Jackson, B.; Nascimento, F.F.; et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **2021**, *184*, 64–75.e11. [CrossRef]
41. Hou, Y.J.; Chiba, S.; Halfmann, P.; Ehre, C.; Kuroda, M.; Dinnon, K.H.; Leist, S.R.; Schäfer, A.; Nakajima, N.; Takahashi, K.; et al. SARS-CoV-2 D614G Variant Exhibits Efficient Replication Ex Vivo and Transmission in Vivo. *Science* **2021**, *370*, 1464–1468.
42. PHE SARS-CoV-2 Variants of Concern and Variants under Investigation in England. Available online: https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-cases-data (accessed on 31 May 2021).
43. CDC SARS-CoV-2 Variant Classifications and Definitions. Available online: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html (accessed on 31 May 2021).
44. Davies, N.G.; Abbott, S.; Barnard, R.C.; Jarvis, C.I.; Kucharski, A.J.; Munday, J.D.; Pearson, C.A.B.; Russell, T.W.; Tully, D.C.; Washburne, A.D.; et al. Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England. *Science* **2021**, *372*, eabg3055. [CrossRef]
45. Ferasin, L.; Fritz, M.; Ferasin, H.; Legros, V.; Leroy, E.M. Myocarditis in Naturally Infected Pets with the British Variant of COVID-19. *bioRxiv* **2021**. [CrossRef]
46. Fujino, T.; Nomoto, H.; Kutsuna, S.; Ujiie, M.; Suzuki, T.; Sato, R.; Fujimoto, T.; Kuroda, M.; Wakita, T.; Ohmagari, N. Novel SARS-CoV-2 Variant Identified in Travelers from Brazil to Japan. *Emerg. Infect. Dis.* **2021**, *27*, 1243–1245. [CrossRef] [PubMed]
47. Sabino, E.C.; Buss, L.F.; Carvalho, M.P.S.; Prete, C.A.; Crispim, M.A.E.; Fraiji, N.A.; Pereira, R.H.M.; Parag, K.V.; da Silva Peixoto, P.; Kraemer, M.U.G.; et al. Resurgence of COVID-19 in Manaus, Brazil, despite High Seroprevalence. *Lancet* **2021**, *397*, 452–455. [CrossRef]

48. Naveca, F.; Souza, V.; Corado, A.; Nascimento, F.; Silva, G.; Costa, Á.; Duarte, D.; Pessoa, K.; Mejía, M.; Brandão, M.; et al. COVID-19 Epidemic in the Brazilian State of Amazonas Was Driven by Long-Term Persistence of Endemic SARS-CoV-2 Lineages and the Recent Emergence of the New Variant of Concern P.1. *Res. Sq.* **2021**. [CrossRef]
49. Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.d.S.; Mishra, S.; Crispim, M.A.E.; Sales, F.C.S.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [CrossRef]
50. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215–220. [CrossRef]
51. Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural Basis of Receptor Recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [CrossRef] [PubMed]
52. Starr, T.N.; Greaney, A.J.; Hilton, S.K.; Ellis, D.; Crawford, K.H.D.; Dingens, A.S.; Navarro, M.J.; Bowen, J.E.; Tortorici, M.A.; Walls, A.C.; et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182*, 1295–1310.e20. [CrossRef]
53. Tian, F.; Tong, B.; Sun, L.; Shi, S.; Zheng, B.; Wang, Z.; Dong, X.; Zheng, P. Mutation N501Y in RBD of Spike Protein Strengthens the Inter-Action between COVID-19 and Its Receptor ACE2. *bioRxiv* **2021**. [CrossRef]
54. FDA Fact Sheet for Health Care Providers Emergency Use Authorization (EUA) of Bamlanivimab and Etesevimab. Available online: https://www.fda.gov/media/145802/download (accessed on 31 May 2021).
55. Wang, P.; Nair, M.S.; Liu, L.; Iketani, S.; Luo, Y.; Guo, Y.; Wang, M.; Yu, J.; Zhang, B.; Kwong, P.D.; et al. Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *Nature* **2021**, *593*, 130–135. [CrossRef]
56. FDA Fact Sheet for Health Care Providers Emergency Use Authorization (EUA) of Regen-Covtm (Casirivimab with Imdevimab). Available online: https://www.fda.gov/media/145611/download (accessed on 31 May 2021).
57. Wang, P.; Casner, R.G.; Nair, M.S.; Wang, M.; Yu, J.; Cerutti, G.; Liu, L.; Kwong, P.D.; Huang, Y.; Shapiro, L.; et al. Increased Resistance of SARS-CoV-2 Variant P.1 to Antibody Neutralization. *Cell Host Microbe* **2021**, *29*, 747–751.e4. [CrossRef]
58. Edara, V.V.; Floyd, K.; Lai, L.; Gardner, M.; Hudson, W.; Piantadosi, A.; Waggoner, J.J.; Babiker, A.; Ahmed, R.; Xie, X.; et al. Infection and MRNA-1273 Vaccine Antibodies Neutralize SARS-CoV-2 UK Variant. *medRxiv* **2021**. [CrossRef]
59. Wu, K.; Werner, A.P.; Moliva, J.I.; Koch, M.; Choi, A.; Stewart-Jones, G.B.E.; Bennett, H.; Boyoglu-Barnum, S.; Shi, W.; Graham, B.S.; et al. mRNA-1273 Vaccine Induces Neutralizing Antibodies against Spike Mutants from Global SARS-CoV-2 Variants. *bioRxiv* **2021**. [CrossRef]
60. Collier, D.A.; De Marco, A.; Ferreira, I.A.T.M.; Meng, B.; Datir, R.; Walls, A.C.; Kemp S, S.A.; Bassi, J.; Pinto, D.; Fregni, C.S.; et al. SARS-CoV-2 B.1.1.7 Sensitivity to MRNA Vaccine-Elicited, Convalescent and Monoclonal Antibodies. *medRxiv* **2021**. [CrossRef]
61. Altmann, D.M.; Boyton, R.J.; Beale, R. Immunity to SARS-CoV-2 Variants of Concern. *Science* **2021**, *371*, 1103–1104. [CrossRef]
62. Lule Bugembe, D.; VTPhan, M.; Ssewanyana, I.; Semanda, P.; Dhaala, B.; Nabadda, S.; Niamh, Á.O.; Rambaut, A.; Kaleebu, P.; Cotten, M. A SARS-CoV-2 Lineage A Variant (A.23.1) with Altered Spike Has Emerged and Is 1 Dominating the Current Uganda Epidemic 2. *medRxiv* **2021**. [CrossRef]
63. Deng, X.; Garcia-knight, M.A.; Khalid, M.M.; Servellita, V.; Morris, M.K.; Sotomayor-gonzález, A.; Glasner, D.R.; Reyes, R.; Gliwa, A.S.; Reddy, N.P.; et al. Transmission, Infectivity, and Neutralization of a Spike L452R SARS-CoV-2 Variant. *Cell* **2021**, in press. [CrossRef]
64. Yadav, P.D.; Sapkal, G.N.; Abraham, P.; Ella, R.; Deshpande, G.; Patil, D.Y.; Nyayanit, D.A.; Gupta, N.; Sahay, R.R.; Shete, A.M.; et al. Neutralization of Variant under Investigation B.1.617 with Sera of BBV152 Vaccinees. *Clin. Infect. Dis.* **2021**, in press. [CrossRef]
65. Rahbari, R.; Moradi, N.; Abdi, M. RRT-PCR for SARS-CoV-2: Analytical Considerations. *Clin. Chim. Acta* **2021**, *516*, 1–7. [CrossRef]
66. Eftekhari, A.; Alipour, M.; Chodari, L.; Dizaj, S.M.; Ardalan, M.R.; Samiei, M.; Sharifi, S.; Vahed, S.Z.; Huseynova, I.; Khalilov, R.; et al. A Comprehensive Review of Detection Methods for SARS-CoV-2. *Microorganisms* **2021**, *9*, 232. [CrossRef]
67. Sharma, K.; Aggarwala, P.; Gandhi, D.; Mathias, A.; Singh, P.; Sharma, S.; Negi, S.S.; Bhargava, A.; Das, P.; Gaikwad, U.; et al. Comparative Analysis of Various Clinical Specimens in Detection of SARS-CoV-2 Using rRT-PCR in New and Follow up Cases of COVID-19 Infection: Quest for the Best Choice. *PLoS ONE* **2021**, *16*, e0249408. [CrossRef]
68. Tong, Y.; Bao, A.; Chen, H.; Huang, J.; Lv, Z.; Feng, L.; Cheng, Y.; Wang, Y.; Bai, L.; Rao, W.; et al. Necessity for Detection of SARS-CoV-2 RNA in Multiple Types of Specimens for the Discharge of the Patients with COVID-19. *J. Transl. Med.* **2020**, *18*, 1–8. [CrossRef]
69. Udvardi, M.K.; Czechowski, T.; Scheible, W.R. Eleven Golden Rules of Quantitative RT-PCR. *Plant Cell* **2008**, *20*, 1736–1737. [CrossRef]
70. Larionov, A.; Krause, A.; Miller, W.R. A Standard Curve Based Method for Relative Real Time PCR Data Processing. *BMC Bioinform.* **2005**, *6*, 1–16. [CrossRef]
71. Sherina, V.; McMurray, H.R.; Powers, W.; Land, H.; Love, T.M.T.; McCall, M.N. Statistical Approaches to Decreasing the Discrepancy of Non-detects in qPCR Data. *bioRxiv* **2017**. [CrossRef]
72. Park, C.; Lee, J.; Hassan, Z.U.; Ku, K.B.; Kim, S.J.; Kim, H.G.; Park, E.C.; Park, G.S.; Park, D.; Baek, S.H.; et al. Comparison of Digital PCR and Quantitative PCR with Various SARS-CoV-2 Primer-Probe Sets. *J. Microbiol. Biotechnol.* **2021**, *31*, 358–367. [CrossRef]
73. Han, M.S.; Byun, J.H.; Cho, Y.; Rim, J.H. RT-PCR for SARS-CoV-2: Quantitative versus Qualitative. *Lancet Infect. Dis.* **2021**, *21*, 165. [CrossRef]

74. Wölfel, R.; Corman, V.M.; Guggemos, W.; Seilmaier, M.; Zange, S.; Müller, M.A.; Niemeyer, D.; Jones, T.C.; Vollmar, P.; Rothe, C.; et al. Virological Assessment of Hospitalized Patients with COVID-2019. *Nature* **2020**, *581*, 465–469. [CrossRef]
75. Pan, Y.; Zhang, D.; Yang, P.; Poon, L.L.M.; Wang, Q. Viral Load of SARS-CoV-2 in Clinical Samples. *Lancet Infect. Dis.* **2020**, *20*, 411–412. [CrossRef]
76. Arnaout, R.; Lee, R.A.; Lee, G.R.; Callahan, C.; Cheng, A.; Yen, C.F.; Smith, K.P.; Arora, R.; Kirby, J.E. The Limit of Detection Matters: The Case for Benchmarking Severe Acute Respiratory Syndrome Coronavirus 2 Testing. *Clin. Infect. Dis.* **2021**, in press. [CrossRef] [PubMed]
77. Platts, A.E.; Johnson, G.D.; Linnemann, A.K.; Krawetz, S.A. Real-Time PCR Quantification Using a Variable Reaction Efficiency Model. *Anal. Biochem.* **2008**, *380*, 315–322. [CrossRef] [PubMed]
78. Lievens, A.; Van Aelst, S.; Van Den Bulcke, M.; Goetghebeur, E. Enhanced Analysis of Real-Time PCR Data by Using a Variable Efficiency Model: FPK-PCR. *Nucleic Acids Res.* **2012**, *40*, e10. [CrossRef]
79. Forootan, A.; Sjöback, R.; Björkman, J.; Sjögreen, B.; Linz, L.; Kubista, M. Methods to Determine Limit of Detection and Limit of Quantification in Quantitative Real-Time PCR (QPCR). *Biomol. Detect. Quantif.* **2017**, *12*, 1–6. [CrossRef]
80. Wikramaratna, P.S.; Paton, R.S.; Ghafari, M.; Lourenço, J. Estimating the False-Negative Test Probability of SARSCoV- 2 by RT-PCR. *Eurosurveillance* **2020**, *25*, 2000568. [CrossRef]
81. Xiao, A.T.; Tong, Y.X.; Zhang, S. False Negative of RT-PCR and Prolonged Nucleic Acid Conversion in COVID-19: Rather than Recurrence. *J. Med. Virol.* **2020**, *92*, 1755–1756. [CrossRef]
82. Erster, O.; Shkedi, O.; Benedek, G.; Zilber, E.; Varkovitzky, I.; Shirazi, R.; Shorka, D.O.; Cohen, Y.; Bar, T.; Yechieli, R.; et al. Improved Sensitivity, Safety, and Rapidity of COVID-19 Tests by Replacing Viral Storage Solution with Lysis Buffer. *PLoS ONE* **2021**, *16*, e0249149. [CrossRef]
83. Tsang, N.N.Y.; So, H.C.; Ng, K.Y.; Cowling, B.J.; Leung, G.M.; Ip, D.K.M. Diagnostic Performance of Different Sampling Approaches for SARS-CoV-2 RT-PCR Testing: A Systematic Review and Meta-Analysis. *Lancet Infect. Dis.* **2021**, in press. [CrossRef]
84. Karligkiotis, A.; Arosio, A.; Castelnuovo, P. How to Obtain a Nasopharyngeal Swab Specimen. *N. Engl. J. Med.* **2020**, *383*, e14.
85. Pinninti, S.; Trieu, C.; Pati, S.K.; Latting, M.; Cooper, J.; Seleme, M.C.; Boppana, S.; Arora, N.; Britt, W.J.; Boppana, S.B. Comparing Nasopharyngeal and Midturbinate Nasal Swab Testing for the Identification of Severe Acute Respiratory Syndrome Coronavirus 2. *Clin. Infect. Dis.* **2021**, *72*, 1253–1255. [CrossRef]
86. Mohammadi, A.; Esmaeilzadeh, E.; Li, Y.; Bosch, R.J.; Li, J.Z. SARS-CoV-2 Detection in Different Respiratory Sites: A Systematic Review and Meta-Analysis. *EBioMedicine* **2020**, *59*, 102903. [CrossRef]
87. Fani, M.; Teimoori, A.; Ghafari, S. Comparison of the COVID-2019 (SARS-CoV-2) Pathogenesis with SARS-CoV and MERS-CoV Infections. *Future Virol.* **2020**, *15*, 317–323. [CrossRef]
88. Kim, J.M.; Kim, H.M.; Lee, E.J.; Jo, H.J.; Yoon, Y.; Lee, N.J.; Son, J.; Lee, Y.J.; Kim, M.S.; Lee, Y.P.; et al. Detection and Isolation of SARS-CoV-2 in Serum, Urine, and Stool Specimens of COVID-19 Patients from the Republic of Korea. *Osong Public Heal. Res. Perspect.* **2020**, *11*, 112–117. [CrossRef]
89. Wang, W.; Xu, Y.; Gao, R.; Al, E. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* **2020**, *323*, 1843–1844. [CrossRef]
90. Vogels, C.B.F.; Brito, A.F.; Wyllie, A.L.; Fauver, J.R.; Ott, I.M.; Kalinich, C.C.; Petrone, M.E.; Casanovas-Massana, A.; Catherine Muenker, M.; Moore, A.J.; et al. Analytical Sensitivity and Efficiency Comparisons of SARS-CoV-2 RT–QPCR Primer–Probe Sets. *Nat. Microbiol.* **2020**, *5*, 1299–1305. [CrossRef]
91. Jung, Y.; Park, G.S.; Moon, J.H.; Ku, K.; Beak, S.H.; Lee, C.S.; Kim, S.; Park, E.C.; Park, D.; Lee, J.H.; et al. Comparative Analysis of Primer-Probe Sets for RT-qPCR of COVID-19 Causative Virus (SARS-CoV-2). *ACS Infect. Dis.* **2020**, *6*, 2513–2523. [CrossRef]
92. Pillonel, T.; Scherz, V.; Jaton, K.; Greub, G.; Bertelli, C. Letter to the Editor: SARS-CoV-2 Detection by Real-Time RT-PCR. *Eurosurveillance* **2020**, *25*, 1–2. [CrossRef]
93. Klungthong, C.; Chinnawirotpisan, P.; Hussem, K.; Phonpakobsin, T.; Manasatienkij, W.; Ajariyakhajorn, C.; Rungrojcharoenkit, K.; Gibbons, R.V.; Jarman, R.G. The Impact of Primer and Probe-Template Mismatches on the Sensitivity of Pandemic Influenza A/H1N1/2009 Virus Detection by Real-Time RT-PCR. *J. Clin. Virol.* **2010**, *48*, 91–95. [CrossRef]
94. Nayar, G.; Seabolt, E.E.; Kunitomi, M.; Agarwal, A.; Beck, K.L.; Mukherjee, V.; Kaufman, J.H. Analysis and Forecasting of Global Real Time RT-PCR Primers and Probes for SARS-CoV-2. *Sci. Rep.* **2021**, *11*, 8988. [CrossRef]
95. van Kasteren, P.B.; van der Veer, B.; van den Brink, S.; Wijsman, L.; de Jonge, J.; van den Brandt, A.; Molenkamp, R.; Reusken, C.B.E.M.; Meijer, A. Comparison of Seven Commercial RT-PCR Diagnostic Kits for COVID-19. *J. Clin. Virol.* **2020**, *128*, 104412. [CrossRef]
96. Simas, A.M.; Crott, J.W.; Sedore, C.; Rohrbach, A.; Monaco, A.P.; Gabriel, S.B.; Lennon, N.; Blumenstiel, B.; Genco, C.A. Pooling for SARS-CoV2 Surveillance: Validation and Strategy for Implementation in K-12 Schools. *medRxiv* **2020**. [CrossRef]
97. Sun, Q.; Li, J.; Ren, H.; Pastor, L.; Loginova, Y.; Madej, R.; Taylor, K.; Wong, J.K.; Zhang, Z.; Zhang, A.; et al. Saliva as a Testing Specimen with or without Pooling for SARS-CoV-2 Detection by Multiplex RT-PCR Test. *PLoS ONE* **2021**, *16*, e0243183. [CrossRef]
98. de Salazar, A.; Aguilera, A.; Trastoy, R.; Fuentes, A.; Alados, J.C.; Causse, M.; Galán, J.C.; Moreno, A.; Trigo, M.; Pérez-Ruiz, M.; et al. Sample Pooling for SARS-CoV-2 RT-PCR Screening. *Clin. Microbiol. Infect.* **2020**, *26*, 1687.e1–1687.e5. [CrossRef] [PubMed]
99. Barak, N.; Ben-Ami, R.; Sido, T.; Perri, A.; Shtoyer, A.; Rivkin, M.; Licht, T.; Peretz, A.; Magenheim, J.; Fogel, I.; et al. Lessons from Applied Large-Scale Pooling of 133,816 SARS-CoV-2 RT-PCR Tests. *Sci. Transl. Med.* **2021**, *13*, eabf2823. [CrossRef] [PubMed]

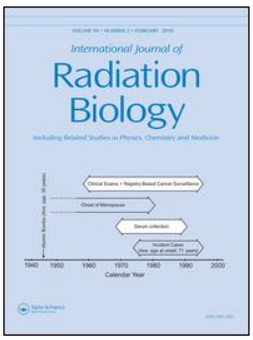
100. Cleary, B.; Hay, J.A.; Blumenstiel, B.; Harden, M.; Cipicchio, M.; Bezney, J.; Simonton, B.; Hong, D.; Senghore, M.; Sesay, A.K.; et al. Using Viral Load and Epidemic Dynamics to Optimize Pooled Testing in Resource-Constrained Settings. *Sci. Transl. Med.* **2021**, *13*, eabf1568. [CrossRef] [PubMed]
101. Kim, S.Y.; Lee, J.; Sung, H.; Lee, H.; Han, M.G.; Yoo, C.K.; Lee, S.W.; Hong, K.H. Pooling Upper Respiratory Specimens for Rapid Mass Screening of COVID-19 by Real-Time RT-PCR. *Emerg. Infect. Dis.* **2020**, *26*, 2469–2472. [CrossRef]
102. Wacker, M.J.; Godard, M.P. Analysis of One-Step and Two- Step Real-Time RT-PCR Using Superscript III. *J. Biomol. Tech.* **2005**, *16*, 266–271.
103. Dharavath, B.; Yadav, N.; Desai, S.; Sunder, R.; Mishra, R.; Ketkar, M.; Bhanshe, P.; Gupta, A.; Redhu, A.K.; Patkar, N.; et al. A One-Step, One-Tube Real-Time RT-PCR Based Assay with an Automated Analysis for Detection of SARS-CoV-2. *Heliyon* **2020**, *6*, e04405. [CrossRef]
104. Yu, C.Y.; Chan, K.G.; Yean, C.Y.; Ang, G.Y. Nucleic Acid-Based Diagnostic Tests for the Detection SARS-CoV-2: An Update. *Diagnostics* **2021**, *11*, 53. [CrossRef]
105. Miranda, J.P.; Osorio, J.; Videla, M.; Angel, G.; Camponovo, R.; Henríquez-Henríquez, M. Analytical and Clinical Validation for RT-qPCR Detection of SARS-CoV-2 Without RNA Extraction. *Front. Med.* **2020**, *7*, 567572. [CrossRef]
106. Das Mukhopadhyay, C.; Sharma, P.; Sinha, K.; Rajarshi, K. Recent Trends in Analytical and Digital Techniques for the Detection of the SARS-Cov-2. *Biophys. Chem.* **2021**, *270*, 106538. [CrossRef]
107. Park, M.; Won, J.; Choi, B.Y.; Lee, C.J. Optimization of Primer Sets and Detection Protocols for SARS-CoV-2 of Coronavirus Disease 2019 (COVID-19) Using PCR and Real-Time PCR. *Exp. Mol. Med.* **2020**, *52*, 963–977. [CrossRef]
108. Chen, C.J.; Hsieh, L.L.; Lin, S.K.; Wang, C.F.; Huang, Y.H.; Lin, S.Y.; Lu, P.L. Optimization of the CDC Protocol of Molecular Diagnosis of COVID-19 for Timely Diagnosis. *Diagnostics* **2020**, *10*, 333. [CrossRef]
109. Bustin, S.A.; Nolan, T. RT-qPCR Testing of SARS-COV-2: A Primer. *Int. J. Mol. Sci.* **2020**, *21*, 3004. [CrossRef]
110. Wyllie, A.L.; Fournier, J.; Casanovas-Massana, A.; Campbell, M.; Tokuyama, M.; Vijayakumar, P.; Warren, J.L.; Geng, B.; Muenker, M.C.; Moore, A.J.; et al. Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. *N. Engl. J. Med.* **2020**, *383*, 1283–1286. [CrossRef]
111. Takeuchi, Y.; Furuchi, M.; Kamimoto, A.; Honda, K.; Matsumura, H.; Kobayashi, R. Saliva-Based PCR Tests for SARS-CoV-2 Detection. *J. Oral Sci.* **2020**, *62*, 350–351. [CrossRef]
112. Gallo, G.; La Torre, M.; Pietroletti, R.; Bianco, F.; Altomare, D.F.; Pucciarelli, S.; Gagliardi, G.; Perinotti, R. Italian Society of Colorectal Surgery Recommendations for Good Clinical Practice in Colorectal Surgery during the Novel Coronavirus Pandemic. *Tech. Coloproctol.* **2020**, *24*, 501–505. [CrossRef]
113. Mizumoto, K.; Kagaya, K.; Zarebski, A.; Chowell, G. Estimating the Asymptomatic Proportion of Coronavirus Disease 2019 (COVID-19) Cases on Board the Diamond Princess Cruise Ship, Yokohama, Japan, 2020. *Eurosurveillance* **2020**, *25*, 2000180. [CrossRef]
114. Zhu, H.; Zhang, H.; Xu, Y.; Laššáková, S.; Korabečná, M.; Neužil, P. PCR Past, Present and Future. *Biotechniques* **2020**, *69*, 317–325. [CrossRef]
115. Kojabad, A.A.; Farzanehpour, M.; Galeh, H.E.G.; Dorostkar, R.; Jafarpour, A.; Bolandian, M.; Nodooshan, M.M. Droplet Digital PCR of Viral DNA/RNA, Current Progress, Challenges, and Future Perspectives. *J. Med. Virol.* **2021**, *93*, 4182–4197. [CrossRef]
116. Sedlak, R.H.; Jerome, K.R. Viral Diagnostics in the Era of Digital Polymerase Chain Reaction. *Diagn. Microbiol. Infect. Dis.* **2013**, *75*, 1–4. [CrossRef]
117. Salipante, S.J.; Jerome, K.R. Digital PCR—An Emerging Technology with Broad Applications in Microbiology. *Clin. Chem.* **2020**, *66*, 117–123. [CrossRef]
118. Quan, P.L.; Sauzade, M.; Brouzes, E. DPCR: A Technology Review. *Sensors* **2018**, *18*, 1271. [CrossRef] [PubMed]
119. Kuypers, J.; Jerome, K.R. Applications of Digital PCR for Clinical Microbiology. *J. Clin. Microbiol.* **2017**, *55*, 1621–1628. [CrossRef] [PubMed]
120. Alteri, C.; Cento, V.; Antonello, M.; Colagrossi, L.; Merli, M.; Ughi, N.; Renica, S.; Matarazzo, E.; Ruscio, F.D.; Tartaglione, L.; et al. Detection and Quantification of SARS-CoV-2 by Droplet Digital PCR in Real-Time PCR Negative Nasopharyngeal Swabs from Suspected COVID-19 Patients. *PLoS ONE* **2020**, *15*. [CrossRef] [PubMed]
121. Falzone, L.; Musso, N.; Gattuso, G.; Bongiorno, D.; Palermo, C.I.; Scalia, G.; Libra, M.; Stefani, S. Sensitivity Assessment of Droplet Digital PCR for SARS-CoV-2 Detection. *Int. J. Mol. Med.* **2020**, *46*, 957–964. [CrossRef]
122. Suo, T.; Liu, X.; Feng, J.; Guo, M.; Hu, W.; Guo, D.; Ullah, H.; Yang, Y.; Zhang, Q.; Wang, X.; et al. DdPCR: A More Accurate Tool for SARS-CoV-2 Detection in Low Viral Load Specimens. *Emerg. Microbes Infect.* **2020**, *9*, 1259–1268. [CrossRef]
123. Dang, Y.; Liu, N.; Tan, C.; Feng, Y.; Yuan, X.; Fan, D.; Peng, Y.; Jin, R.; Guo, Y.; Lou, J. Comparison of Qualitative and Quantitative Analyses of COVID-19 Clinical Samples. *Clin. Chim. Acta* **2020**, *510*, 613–616. [CrossRef]
124. Liu, X.; Feng, J.; Zhang, Q.; Guo, D.; Zhang, L.; Suo, T.; Hu, W.; Guo, M.; Wang, X.; Huang, Z.; et al. Analytical Comparisons of SARS-COV-2 Detection by qRT-PCR and ddPCR with Multiple Primer/Probe Sets. *Emerg. Microbes Infect.* **2020**, *9*, 1175–1179. [CrossRef]
125. Kevadiya, B.D.; Machhi, J.; Herskovitz, J.; Oleynikov, M.D.; Blomberg, W.R.; Bajwa, N.; Soni, D.; Das, S.; Hasan, M.; Patel, M.; et al. Diagnostics for SARS-CoV-2 Infections. *Nat. Mater.* **2021**, *20*, 593–605. [CrossRef]

126. Gatti, M.; Calandri, M.; Biondo, A.; Geninatti, C.; Piatti, C.; Ruggirello, I.; Santonocito, A.; Varello, S.; Bergamasco, L.; Bironzo, P.; et al. Emergency Room Comprehensive Assessment of Demographic, Radiological, Laboratory and Clinical Data of Patients with COVID-19: Determination of Its Prognostic Value for in-Hospital Mortality. *Intern. Emerg. Med.* **2021**, in press. [CrossRef]
127. Mostafa, H.H.; Hardick, J.; Morehead, E.; Miller, J.A.; Gaydos, C.A.; Manabe, Y.C. Comparison of the Analytical Sensitivity of Seven Commonly Used Commercial SARS-CoV-2 Automated Molecular Assays. *J. Clin. Virol.* **2020**, *130*, 104578. [CrossRef]
128. D'Cruz, R.J.; Currier, A.W.; Sampson, V.B. Laboratory Testing Methods for Novel Severe Acute Respiratory Syndrome-Coronavirus-2 (SARS-CoV-2). *Front. Cell Dev. Biol.* **2020**, *8*, 1–11. [CrossRef]
129. Ganguli, A.; Mostafa, A.; Berger, J.; Aydin, M.Y.; Sun, F.; Stewart de Ramirez, S.A.; Valera, E.; Cunningham, B.T.; King, W.P.; Bashir, R. Rapid Isothermal Amplification and Portable Detection System for SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 22727–22735. [CrossRef]
130. Carter, L.J.; Garner, L.V.; Smoot, J.W.; Li, Y.; Zhou, Q.; Saveson, C.J.; Sasso, J.M.; Gregg, A.C.; Soares, D.J.; Beskid, T.R.; et al. Assay Techniques and Test Development for COVID-19 Diagnosis. *ACS Cent. Sci.* **2020**, *6*, 591–605. [CrossRef]
131. Bodulev, O.L.; Sakharov, I.Y. Isothermal Nucleic Acid Amplification Techniques and Their Use in Bioanalysis. *Biochemistry* **2020**, *85*, 147–166. [CrossRef]
132. Udugama, B.; Kadhiresan, P.; Kozlowski, H.N.; Malekjahani, A.; Osborne, M.; Li, V.Y.C.; Chen, H.; Mubareka, S.; Gubbay, J.B.; Chan, W.C.W. Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* **2020**, *14*, 3822–3835. [CrossRef]
133. Mautner, L.; Baillie, C.K.; Herold, H.M.; Volkwein, W.; Guertler, P.; Eberle, U.; Ackermann, N.; Sing, A.; Pavlovic, M.; Goerlich, O.; et al. Rapid Point-of-Care Detection of SARS-CoV-2 Using Reverse Transcription Loop-Mediated Isothermal Amplification (RT-LAMP). *Virol. J.* **2020**, *17*, 1–14. [CrossRef]
134. Qian, J.; Boswell, S.A.; Chidley, C.; Lu, Z.X.; Pettit, M.E.; Gaudio, B.L.; Fajnzylber, J.M.; Ingram, R.T.; Ward, R.H.; Li, J.Z.; et al. An Enhanced Isothermal Amplification Assay for Viral Detection. *Nat. Commun.* **2020**, *11*, 1–10. [CrossRef]
135. Chaibun, T.; Puenpa, J.; Ngamdee, T.; Boonapatcharoen, N.; Athamanolap, P.; O'Mullane, A.P.; Vongpunsawad, S.; Poovorawan, Y.; Lee, S.Y.; Lertanantawong, B. Rapid Electrochemical Detection of Coronavirus SARS-CoV-2. *Nat. Commun.* **2021**, *12*, 1–10. [CrossRef]
136. Zhang, C.; Zheng, T.; Fan, H.; Zhang, T.; Han, D. Aligner-Mediated Cleavage-Based Isothermal Amplification for SARS-CoV-2 RNA Detection. *ACS Appl. Bio Mater.* **2021**, *4*, 3805–3810. [CrossRef]
137. Xing, W.; Liu, Y.; Wang, H.; Li, S.; Lin, Y.; Chen, L.; Zhao, Y.; Chao, S.; Huang, X.; Ge, S.; et al. A High-Throughput, Multi-Index Isothermal Amplification Platform for Rapid Detection of 19 Types of Common Respiratory Viruses Including SARS-CoV-2. *Engineering* **2020**, *6*, 1130–1140. [CrossRef]
138. Notomi, T.; Okayama, H.; Masubuchi, H.; Yonekawa, T.; Watanabe, K.; Amino, N.; Hase, T. Loop-Mediated Isothermal Amplification of DNA. *Nucleic Acids Res.* **2000**, *28*, E63. [CrossRef] [PubMed]
139. Tomita, N.; Mori, Y.; Kanda, H.; Notomi, T. Loop-Mediated Isothermal Amplification (LAMP) of Gene Sequences and Simple Visual Detection of Products. *Nat. Protoc.* **2008**, *3*, 877–882. [CrossRef] [PubMed]
140. Mori, Y.; Kitao, M.; Tomita, N.; Notomi, T. Real-Time Turbidimetry of LAMP Reaction for Quantifying Template DNA. *J. Biochem. Biophys. Methods* **2004**, *59*, 145–157. [CrossRef] [PubMed]
141. Yu, L.; Wu, S.; Hao, X.; Dong, X.; Mao, L.; Pelechano, V.; Chen, W.-H.; Yin, X. Rapid Detection of COVID-19 Coronavirus Using a Reverse Transcriptional Loop-Mediated Isothermal Amplification (RT-LAMP) Diagnostic Platform. *Clin. Chem.* **2020**, *66*, 975–977. [CrossRef] [PubMed]
142. Dao Thi, V.L.; Herbst, K.; Boerner, K.; Meurer, M.; Kremer, L.P.M.; Kirrmaier, D.; Freistaedter, A.; Papagiannidis, D.; Galmozzi, C.; Stanifer, M.L.; et al. A Colorimetric RT-LAMP Assay and LAMP-Sequencing for Detecting SARS-CoV-2 RNA in Clinical Samples. *Sci. Transl. Med.* **2020**, *12*, eabc7075. [CrossRef] [PubMed]
143. Lu, R.; Wu, X.; Wan, Z.; Li, Y.; Jin, X.; Zhang, C. A Novel Reverse Transcription Loop-Mediated Isothermal Amplification Method for Rapid Detection of Sars-Cov-2. *Int. J. Mol. Sci.* **2020**, *21*, 2826. [CrossRef]
144. Park, G.S.; Ku, K.; Baek, S.H.; Kim, S.J.; Kim, S.I.; Kim, B.T.; Maeng, J.S. Development of Reverse Transcription Loop-Mediated Isothermal Amplification Assays Targeting Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *J. Mol. Diagnostics* **2020**, *22*, 729–735. [CrossRef]
145. Baek, Y.H.; Um, J.; Antigua, K.J.C.; Park, J.H.; Kim, Y.; Oh, S.; Kim, Y.I.; Choi, W.S.; Kim, S.G.; Jeong, J.H.; et al. Development of a Reverse Transcription-Loop-Mediated Isothermal Amplification as a Rapid Early-Detection Method for Novel SARS-CoV-2. *Emerg. Microbes Infect.* **2020**, *9*, 998–1007. [CrossRef]
146. Wei, S.; Kohl, E.; Djandji, A.; Morgan, S.; Whittier, S.; Mansukhani, M.; Hod, E.; D'Alton, M.; Suh, Y.; Williams, Z. Direct Diagnostic Testing of SARS-CoV-2 without the Need for Prior RNA Extraction. *Sci. Rep.* **2021**, *11*, 1–6. [CrossRef]
147. Lalli, M.A.; Langmade, J.S.; Chen, X.; Fronick, C.C.; Sawyer, C.S.; Burcea, L.C.; Wilkinson, M.N.; Fulton, R.S.; Heinz, M.; Buchser, W.J.; et al. Rapid and Extraction-Free Detection of SARS-CoV-2 from Saliva by Colorimetric Reverse-Transcription Loop-Mediated Isothermal Amplification. *Clin. Chem.* **2021**, *67*, 415–424. [CrossRef]
148. Pang, B.; Xu, J.; Liu, Y.; Peng, H.; Feng, W.; Cao, Y.; Wu, J.; Xiao, H.; Pabbaraju, K.; Tipples, G.; et al. Isothermal Amplification and Ambient Visualization in a Single Tube for the Detection of SARS-CoV-2 Using Loop-Mediated Amplification and CRISPR Technology. *Anal. Chem.* **2020**, *92*, 16204–16212. [CrossRef]

149. Ali, Z.; Aman, R.; Mahas, A.; Rao, G.S.; Tehseen, M.; Marsic, T.; Salunke, R.; Subudhi, A.K.; Hala, S.M.; Hamdan, S.M.; et al. ISCAN: An RT-LAMP-Coupled CRISPR-Cas12 Module for Rapid, Sensitive Detection of SARS-CoV-2. *Virus Res.* **2020**, *288*, 198129. [CrossRef]
150. Zhu, X.; Wang, X.; Han, L.; Chen, T.; Wang, L.; Li, H.; Li, S.; He, L.; Fu, X.; Chen, S.; et al. Multiplex Reverse Transcription Loop-Mediated Isothermal Amplification Combined with Nanoparticle-Based Lateral Flow Biosensor for the Diagnosis of COVID-19. *Biosens. Bioelectron.* **2020**, *166*, 112437. [CrossRef]
151. Rohaim, M.A.; Clayton, E.; Sahin, I.; Vilela, J.; Khalifa, M.E.; Al-natour, M.Q.; Bayoumi, M.; Poirier, A.C.; Branavan, M.; Tharmakulasingam, M.; et al. Artificial Intelligence-Assisted Loop Mediated Isothermal Amplification (AI-LAMP) for Rapid Detection of SARS-CoV-2. *Viruses* **2020**, *12*, 972. [CrossRef]
152. Piepenburg, O.; Williams, C.H.; Stemple, D.L.; Armes, N.A. DNA Detection Using Recombination Proteins. *PLoS Biol.* **2006**, *4*, 1115–1121. [CrossRef]
153. Kersting, S.; Rausch, V.; Bier, F.F.; von Nickisch-Rosenegk, M. Rapid Detection of Plasmodium Falciparum with Isothermal Recombinase Polymerase Amplification and Lateral Flow Analysis. *Malar. J.* **2014**, *13*, 99. [CrossRef]
154. Lobato, I.M.; O'Sullivan, C.K. Recombinase Polymerase Amplification: Basics, Applications and Recent Advances. *TrAC Trends Anal. Chem.* **2018**, *98*, 19–35. [CrossRef]
155. Zhang, W.S.; Pan, J.; Li, F.; Zhu, M.; Xu, M.; Zhu, H.; Yu, Y.; Su, G. Reverse Transcription Recombinase Polymerase Amplification Coupled with CRISPR-Cas12a for Facile and Highly Sensitive Colorimetric SARS-CoV-2 Detection. *Anal. Chem.* **2021**, *93*, 4126–4133. [CrossRef]
156. Patchsung, M.; Jantarug, K.; Pattama, A.; Aphicho, K.; Suraritdechachai, S.; Meesawat, P.; Sappakhaw, K.; Leelahakorn, N.; Ruenkam, T.; Wongsatit, T.; et al. Clinical Validation of a Cas13-Based Assay for the Detection of SARS-CoV-2 RNA. *Nat. Biomed. Eng.* **2020**, *4*, 1140–1149. [CrossRef]
157. Arizti-Sanz, J.; Freije, C.A.; Stanton, A.C.; Petros, B.A.; Boehm, C.K.; Siddiqui, S.; Shaw, B.M.; Adams, G.; Kosoko-Thoroddsen, T.S.F.; Kemball, M.E.; et al. Streamlined Inactivation, Amplification, and Cas13-Based Detection of SARS-CoV-2. *Nat. Commun.* **2020**, *11*, 5921. [CrossRef]
158. Kim, H.S.; Abbas, N.; Shin, S. A Rapid Diagnosis of SARS-CoV-2 Using DNA Hydrogel Formation on Microfluidic Pores. *Biosens. Bioelectron.* **2021**, *177*, 113005. [CrossRef]
159. WHO. *Genomic Sequencing of SARS-CoV-2*; WHO: Geneva, Switzerland, 2021; ISBN 9789240018440.
160. Quick, J.; Loman, N.J.; Duraffour, S.; Simpson, J.T.; Severi, E.; Cowley, L.; Bore, J.A.; Koundouno, R.; Dudas, G.; Mikhail, A.; et al. Real-Time, Portable Genome Sequencing for Ebola Surveillance. *Nature* **2016**, *530*, 228–232. [CrossRef]
161. Hoenen, T.; Groseth, A.; Rosenke, K.; Fischer, R.J.; Hoenen, A.; Judson, S.D.; Martellaro, C.; Falzarano, D.; Marzi, A.; Squires, R.B.; et al. Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg. Infect. Dis.* **2016**, *22*, 331–334. [CrossRef]
162. Houldcroft, C.J.; Beale, M.A.; Breuer, J. Clinical and Biological Insights from Viral Genome Sequencing. *Nat. Rev. Microbiol.* **2017**, *15*, 183–192. [CrossRef]
163. Myrmel, M.; Oma, V.; Khatri, M.; Hansen, H.H.; Stokstad, M.; Berg, M.; Blomström, A.-L. Single Primer Isothermal Amplification (SPIA) Combined with next Generation Sequencing Provides Complete Bovine Coronavirus Genome Coverage and Higher Sequence Depth Compared to Sequence-Independent Single Primer Amplification (SISPA). *PLoS ONE* **2017**, *12*, e0187780. [CrossRef]
164. Reyes, G.R.; Kim, J.P. Sequence-Independent, Single-Primer Amplification (SISPA) of Complex DNA Populations. *Mol. Cell. Probes* **1991**, *5*, 473–481. [CrossRef]
165. Maurier, F.; Beury, D.; Fléchon, L.; Varré, J.S.; Touzet, H.; Goffard, A.; Hot, D.; Caboche, S. A Complete Protocol for Whole-Genome Sequencing of Virus from Clinical Samples: Application to Coronavirus OC43. *Virology* **2019**, *531*, 141–148. [CrossRef] [PubMed]
166. Kim, K.W.; Choi, S.; Shin, S.K.; Lee, I.; Ku, K.B.; Kim, S.J.; Kim, S.; Yi, H. A Half-Day Genome Sequencing Protocol for Middle East Respiratory Syndrome Coronavirus. *Front. Microbiol.* **2021**, *12*, 1–8. [CrossRef] [PubMed]
167. Hall, R.J.; Wang, J.; Todd, A.K.; Bissielo, A.B.; Yen, S.; Strydom, H.; Moore, N.E.; Ren, X.; Huang, Q.S.; Carter, P.E.; et al. Evaluation of Rapid and Simple Techniques for the Enrichment of Viruses Prior to Metagenomic Virus Discovery. *J. Virol. Methods* **2014**, *195*, 194–204. [CrossRef] [PubMed]
168. Deng, L.; Silins, R.; Castro-Mejía, J.L.; Kot, W.; Jessen, L.; Thorsen, J.; Shah, S.; Stokholm, J.; Bisgaard, H.; Moineau, S.; et al. A Protocol for Extraction of Infective Viromes Suitable for Metagenomics Sequencing from Low Volume Fecal Samples. *Viruses* **2019**, *11*, 667. [CrossRef] [PubMed]
169. Sathiamoorthy, S.; Malott, R.J.; Gisonni-Lex, L.; Ng, S.H.S. Selection and Evaluation of an Efficient Method for the Recovery of Viral Nucleic Acid Extraction from Complex Biologicals. *Npj Vaccines* **2018**, *3*, 1–6.
170. Tan, C.C.S.; Maurer-Stroh, S.; Wan, Y.; Sessions, O.M.; de Sessions, P.F. A Novel Method for the Capture-Based Purification of Whole Viral Native RNA Genomes. *AMB Express* **2019**, *9*, 45. [CrossRef]
171. Manso, C.F.; Bibby, D.F.; Mbisa, J.L. Efficient and Unbiased Metagenomic Recovery of RNA Virus Genomes from Human Plasma Samples. *Sci. Rep.* **2017**, *7*, 1–13. [CrossRef]
172. Zhang, D.; Lou, X.; Yan, H.; Pan, J.; Mao, H.; Tang, H.; Shu, Y.; Zhao, Y.; Liu, L.; Li, J.; et al. Metagenomic Analysis of Viral Nucleic Acid Extraction Methods in Respiratory Clinical Samples 11 Medical and Health Sciences 1108 Medical Microbiology. *BMC Genomics* **2018**, *19*, 1–8.

173. Temmam, S.; Monteil-Bouchard, S.; Robert, C.; Pascalis, H.; Michelle, C.; Jardot, P.; Charrel, R.; Raoult, D.; Desnues, C. Host-Associated Metagenomics: A Guide to Generating Infectious RNA Viromes. *PLoS ONE* **2015**, *10*, e0139810. [CrossRef]
174. Ponce-Rojas, J.C.; Costello, M.S.; Proctor, D.A.; Kosik, K.S.; Wilson, M.Z.; Arias, C.; Acosta-Alvear, D. A Fast and Accessible Method for the Isolation of RNA, DNA, and Protein to Facilitate the Detection of SARS-CoV-2. *J. Clin. Microbiol.* **2020**, *59*, e02403-20.
175. Jones, M.K.; Grau, K.R.; Costantini, V.; Kolawole, A.O.; de Graaf, M.; Freiden, P.; Graves, C.L.; Koopmans, M.; Wallet, S.M.; Tibbetts, S.A.; et al. Human Norovirus Culture in B Cells. *Nat. Protoc.* **2015**, *10*, 1939–1947. [CrossRef]
176. Pyrc, K.; Sims, A.C.; Dijkman, R.; Jebbink, M.; Long, C.; Deming, D.; Donaldson, E.; Vabret, A.; Baric, R.; van der Hoek, L.; et al. Culturing the Unculturable: Human Coronavirus HKU1 Infects, Replicates, and Produces Progeny Virions in Human Ciliated Airway Epithelial Cell Cultures. *J. Virol.* **2010**, *84*, 11255–11263. [CrossRef]
177. Charre, C.; Ginevra, C.; Sabatier, M.; Regue, H.; Destras, G.; Brun, S.; Burfin, G.; Scholtes, C.; Morfin, F.; Valette, M.; et al. Evaluation of NGS-Based Approaches for SARS-CoV- 2 Whole Genome Characterisation. *Virus Evol.* **2020**, *6*, 1–8. [CrossRef]
178. Gaudin, M.; Desnues, C. Hybrid Capture-Based next Generation Sequencing and Its Application to Human Infectious Diseases. *Front. Microbiol.* **2018**, *9*, 1–9. [CrossRef]
179. Tyson, J.R.; James, P.; Stoddart, D.; Sparks, N.; Wickenhagen, A.; Hall, G.; Choi, J.H.; Lapointe, H.; Kamelian, K.; Smith, A.D.; et al. Improvements to the ARTIC Multiplex PCR Method for SARS-CoV-2 Genome Sequencing Using Nanopore. *bioRxiv* **2020**. [CrossRef]
180. Liu, T.; Chen, Z.; Chen, W.; Chen, X.; Hosseini, M.; Yang, Z.; Li, J.; Ho, D.; Turay, D.; Gheorghe, C.; et al. A Benchmarking Study of SARS-CoV-2 Whole-Genome Sequencing Protocols Using COVID-19 Patient Samples. *bioRxiv* **2020**. [CrossRef]
181. Maurano, M.T.; Ramaswami, S.; Zappile, P.; Dimartino, D.; Boytard, L.; Ribeiro-dos-Santos, A.M.; Vulpescu, N.A.; Westby, G.; Shen, G.; Feng, X.; et al. Sequencing Identifies Multiple Early Introductions of SARS-CoV-2 to the New York City Region. *Genome Res.* **2020**, *30*, 1781–1788. [CrossRef] [PubMed]
182. Shaibu, J.O.; Onwuamah, C.K.; James, A.B.; Okwuraiwe, A.P.; Amoo, O.S.; Salu, O.B.; Ige, F.A.; Liboro, G.; Odewale, E.; Okoli, L.C.; et al. Full Length Genomic Sanger Sequencing and Phylogenetic Analysis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Nigeria. *PLoS ONE* **2021**, *16*, 1–14. [CrossRef]
183. Payne, A.; Holmes, N.; Rakyan, V.; Loose, M. Bulkvis: A Graphical Viewer for Oxford Nanopore Bulk FAST5 Files. *Bioinformatics* **2019**, *35*, 2193–2198. [CrossRef]
184. Grubaugh, N.D.; Gangavarapu, K.; Quick, J.; Matteson, N.L.; De Jesus, J.G.; Main, B.J.; Tan, A.L.; Paul, L.M.; Brackney, D.E.; Grewal, S.; et al. An Amplicon-Based Sequencing Framework for Accurately Measuring Intrahost Virus Diversity Using PrimalSeq and IVar. *Genome Biol.* **2019**, *20*, 8. [CrossRef]
185. Rezaei, M.; Razavi Bazaz, S.; Zhand, S.; Sayyadi, N.; Jin, D.; Stewart, M.P.; Ebrahimi Warkiani, M. Point of Care Diagnostics in the Age of COVID-19. *Diagnostics* **2020**, *11*, 9. [CrossRef]
186. Wang, M.; Zhang, R.; Li, J. CRISPR/Cas Systems Redefine Nucleic Acid Detection: Principles and Methods. *Biosens. Bioelectron.* **2020**, *165*, 112430. [CrossRef]
187. Bin Moon, S.; Yon Kim, D.; Ko, J.-H.; Kim, Y.-S. Recent Advances in the CRISPR Genome Editing Tool Set. *Exp. Mol. Med.* **2019**, *51*, 1–11. [CrossRef]
188. Jiang, F.; Doudna, J.A. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, *46*, 505–529. [CrossRef]
189. Zhang, W.; Liu, K.; Zhang, P.; Cheng, W.; Li, L.L.; Zhang, F.; Yu, Z.; Li, L.L.; Zhang, X. CRISPR-Based Approaches for Efficient and Accurate Detection of SARS-CoV-2. *Lab. Med.* **2021**, *52*, 116–121. [CrossRef]
190. Broughton, J.P.; Deng, X.; Yu, G.; Fasching, C.L.; Servellita, V.; Singh, J.; Miao, X.; Streithorst, J.A.; Granados, A.; Sotomayor-Gonzalez, A.; et al. CRISPR–Cas12-Based Detection of SARS-CoV-2. *Nat. Biotechnol.* **2020**, *38*, 870–874. [CrossRef]
191. Broughton, J.P.; Deng, X.; Yu, G.; Fasching, C.L.; Singh, J.; Streithorst, J.; Granados, A.; Sotomayor-Gonzalez, A.; Zorn, K.; Gopez, A.; et al. Rapid Detection of 2019 Novel Coronavirus SARS-CoV-2 Using a CRISPR-Based DETECTR Lateral Flow Assay. *medRxiv* **2020**. [CrossRef]
192. Kellner, M.J.; Koob, J.G.; Gootenberg, J.S.; Abudayyeh, O.O.; Zhang, F. SHERLOCK: Nucleic Acid Detection with CRISPR Nucleases. *Nat. Protoc.* **2019**, *14*, 2986–3012. [CrossRef]
193. Joung, J.; Ladha, A.; Saito, M.; Segel, M.; Bruneau, R.; Huang, M.L.W.; Kim, N.G.; Yu, X.; Li, J.; Walker, B.D.; et al. Point-of-Care Testing for COVID-19 Using SHERLOCK Diagnostics. *medRxiv* **2020**. [CrossRef]
194. Kudr, J.; Michalek, P.; Ilieva, L.; Adam, V.; Zitka, O. COVID-19: A Challenge for Electrochemical Biosensors. *Trends Anal. Chem.* **2021**, *136*, 116192. [CrossRef]
195. Varahachalam, S.P.; Lahooti, B.; Chamaneh, M.; Bagchi, S.; Chhibber, T.; Morris, K.; Bolanos, J.F.; Kim, N.Y.; Kaushik, A. Nanomedicine for the SARS-CoV-2: State-of-Theart and Future Prospects. *Int. J. Nanomed.* **2021**, *16*, 539–560. [CrossRef] [PubMed]
196. Cheng, X.; Chen, G.; Rodriguez, W.R. Micro- and Nanotechnology for Viral Detection. *Anal. Bioanal. Chem.* **2009**, *393*, 487–501. [CrossRef]
197. Tang, Z.; Kong, N.; Zhang, X.; Liu, Y.; Hu, P.; Mou, S.; Liljeström, P.; Shi, J.; Tan, W.; Kim, J.S.; et al. A Materials-Science Perspective on Tackling COVID-19. *Nat. Rev. Mater.* **2020**, *5*, 847–860. [CrossRef]
198. Zhao, Z.; Cui, H.; Song, W.; Ru, X.; Zhou, W.; Yu, X. A Simple Magnetic Nanoparticles-Based Viral RNA Extraction Method for Efficient Detection of SARS-CoV-2. *bioRxiv* **2020**. [CrossRef]

199. Moitra, P.; Alafeef, M.; Alafeef, M.; Alafeef, M.; Dighe, K.; Frieman, M.B.; Pan, D.; Pan, D.; Pan, D. Selective Naked-Eye Detection of SARS-CoV-2 Mediated by N Gene Targeted Antisense Oligonucleotide Capped Plasmonic Nanoparticles. *ACS Nano* **2020**, *14*, 7617–7627. [CrossRef] [PubMed]
200. Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology (CCQM). Available online: https://www.bipm.org/en/committees/cc/ccqm (accessed on 31 May 2021).
201. The Asia Pacific Metrology Programme (APMP). Available online: http://www.apmpweb.org/index.html (accessed on 31 May 2021).
202. National Measurement Institutes Demonstrate High Accuracy Reference Measurement System for SARS-CoV-2 Testing. Available online: https://www.bipm.org/en/-/2020-nmi-covid (accessed on 31 May 2021).
203. National Measurement Institute Helping Industry Respond to COVID-19. Available online: https://www.industry.gov.au/news/national-measurement-institute-helping-industry-respond-to-covid-19 (accessed on 31 May 2021).
204. ISO. *ISO 17034:2016 General Requirements for the Competence of Reference Material Producers*; ISO: Geneva, Switzerland, 2016.
205. Wise, S.A. What Is Novel about Certified Reference Materials? *Anal. Bioanal. Chem.* **2018**, *410*, 2045–2049. [CrossRef] [PubMed]
206. Pickett, B.E.; Sadat, E.L.; Zhang, Y.; Noronha, J.M.; Squires, R.B.; Hunt, V.; Liu, M.; Kumar, S.; Zaremba, S.; Gu, Z.; et al. ViPR: An Open Bioinformatics Database and Analysis Resource for Virology Research. *Nucleic Acids Res.* **2012**, *40*, 593–598. [CrossRef]
207. Singer, J.; Gifford, R.; Cotten, M.; Robertson, D. CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. *Preprints* **2020**. [CrossRef]



International Journal of Radiation Biology

ISSN: 0955-3002 (Print) 1362-3095 (Online) Journal homepage: https://www.tandfonline.com/loi/irab20

The Warburg effect: Essential part of metabolic reprogramming and central contributor to cancer progression

Peter Vaupel, Heinz Schmidberger & Arnulf Mayer

To cite this article: Peter Vaupel, Heinz Schmidberger & Arnulf Mayer (2019): The Warburg effect: Essential part of metabolic reprogramming and central contributor to cancer progression, International Journal of Radiation Biology, DOI: 10.1080/09553002.2019.1589653

To link to this article: https://doi.org/10.1080/09553002.2019.1589653

Accepted author version posted online: 01 Mar 2019.

IJRB 60th Anniversary Special Issue

REVIEW

The Warburg effect: Essential part of metabolic reprogramming and central contributor to cancer progression

Peter Vaupel, Heinz Schmidberger and Arnulf Mayer

Department of Radiation Oncology, University Medical Center, Mainz, Germany

Correspondence

Professor Dr. med. Peter Vaupel

Email: vaupel@uni-mainz.de

Department of Radiation Oncology

Tumor Pathophysiology Group

University Medical Center

Langenbeckstrasse 1

56131 Mainz

Germany

Fon: +49 171 1240073

ABSTRACT

Purpose: In the early 1920s, Warburg published experimental data on the enhanced conversion of glucose to pyruvate (followed by lactate formation) even in the presence of abundant oxygen (aerobic glycolysis, Warburg effect). He attributed this metabolic trait to a respiratory injury and considered this a universal metabolic alteration in carcinogenesis. This interpretation of the data was questioned since the early 1950s. Realistic causative mechanisms and consequences of the Warburg effect were described only during the past 15 years and are summarized in this article.

Conclusions: There is clear evidence that mitochondria are not defective in most cancers. Aerobic glycolysis, a key metabolic feature of the Warburg phenotype, is caused by active metabolic reprogramming required to support sustained cancer cell proliferation and malignant progression. This metabolic switch is directed by altered growth factor signaling, hypoxic or normoxic activation of HIF-1α- transcription, oncogene activation or loss-of-function of suppressor genes, and is implemented in the hostile tumor microenvironment. The "selfish" reprogramming includes (a) overexpression of glucose transporters and of key glycolytic enzymes, and an accelerated glycolytic flux with subsequent accumulation and diversion of glycolytic intermediates for cancer biomass synthesis, (b) high-speed ATP production that meets the energy demand, and (c) accumulation of lactate which drives tumor progression and largely contributes to tumor acidosis, which in turn synergistically favors tumor progression and resistance to certain antitumor therapies, and compromises anti-tumor immunity. In all, the Warburg effect is the central contributor to the cancer progression machinery.

KEYWORDS Aerobic glycolysis, Warburg effect, glycolytic phenotype, metabolic reprogramming, cancer metabolism, cancer progression

Introduction

It has been known for 95 years that cancer cells metabolize glucose differently than differentiated cells. In the early 1920s, investigating tumor tissue slices, Otto H. Warburg made the seminal observation that cancer cells predominantly convert glucose to lactate even under conditions of adequate oxygen (O_2) supply, i.e., without exposure to hypoxic conditions (Warburg 1923, Warburg et al., 1924). He attributed this metabolic trait ("aerobic glycolysis", later referred to as "Warburg effect" by E. Racker, 1976) to a mitochondrial "respiratory injury" (mitochondrial dysfunction, mitochondrial defect; Warburg, 1956a) and considered "aerobic glycolysis" as the universal, key metabolic alteration in malignant transformation ("origin of cancer cells", Warburg 1956b).

Warburg's interpretation of his pioneering experimental data ("Warburg hypothesis") was originally questioned by Chance & Castor (1952), Chance (1953) and Chance & Hess (1956, 1959), showing that intact and functional cytochromes detected in most tumor cells clearly speak against a general mitochondrial dysfunction. In addition, Weinhouse (1956, 1976), Aisenberg (1961), Vaupel (1974, 1977), Shapot (1976), Shapot & Vaupel (1976), Zu & Guppy (2004), Kim & Dang (2006), Moreno-Sanchez et al. (2007), Christofk et al. (2008), Potter et al. (2016) and a series of other authors provided substantial evidence that oxidative phosphorylation (OXPHOS) and a normal Krebs cycle (TCA cycle) persist in the vast majority, if not in all malignant tumors, i.e., most cancers exhibit the Warburg effect while retaining mitochondrial respiration (for reviews see e.g., Koppenol et al., 2011,Vaupel & Mayer 2012).

During the past decades, many experimental studies have been published which support or refuse the "original" Warburg effect as a hallmark of cancer. As a result, endless controversies and discussions on Warburg's discovery and between supporters and opponents have attained attention among scientists worldwide (for a review see Vaupel & Mayer, 2012; approx. 20,000

publications u up to December, 2018!). Not noticed by most of the supporters of his original postulate, Warburg himself (1962) has contributed to this confusion conceding that the respiration of cancer cells is in fact not compromised, thus cutting back on his initial "respiration injury" theory. Furthermore, there is evidence that the tumor tissue sections used in Warburg's experiments were thicker than "critical" oxygen diffusion distances, i.e., tissue slices investigated were at least partially hypoxic. Warburg calculated a critical diffusion distance of 470 µm based on invalid O_2 diffusion constants as part of the differential equations for calculation of O_2 diffusion distances (Vaupel 1974, Dang et al., 2008).

Considering contributions by modern genomic and mass spectrometry-based proteomic analyses, and contrary to Warburg's conventional belief, in this review we will demonstrate that (a) varying oxygen availability govern the O_2 consumption rate of cancer cells *in vivo*, clearly speaking against a universal mitochondrial dysfunction and a plethora of studies and communications in favor of the principal role of the Warburg effect in its original concept, and (b) the Warburg effect does not reflect a compensatory increase in glycolytic flux upon permanent impairment of oxidative metabolism in cancer cells which apparently have a great need for ATP. Instead, the Warburg effect describes an aggressive phenotype with "selfish" biosynthetic programs that are required to sustain cell survival, proliferation, migration, invasion, metastasis, and suppression of antitumor immune responses, ultimately leading to disease progression. The respective transcriptional responses are mediated primarily by hypoxia-inducible factors (HIFs), activated PI3K/Akt/mTOR and deactivated LKB1/AMPK pathways, and cMYC, as evidenced during the past twenty years.

Warburg effect: Sequela to defective mitochondrial function ?

For decades, the Warburg effect (aerobic glycolysis) was described as a sequela to damaged mitochondria (model "Damaged mitochondria" according to Zhou

et al., 2017) and as a metabolic signature of cancer cells. Instead, oxygen uptake rates of most malignant tumors *in vivo* is determined by O_2 availability, i.e., blood O_2 concentration x blood flow rate (Vaupel 1974, 1977).

Oxygen uptake rates are a function of blood flow in cancers

O_2 uptake rates strongly depend on the O_2 availability and thus on the efficacy of blood flow as evidenced in a series of isografted or allografted murine tumors (Vaupel 1974, 1977) and various xenografted human tumors (Vaupel et al., 1987, Kallinowski et al., 1989). Flow increases in individual tumors up to 39.5°C upon local hyperthermia was followed by similar rises in the O_2 uptake rates (for a review see Vaupel & Mayer, 2012).

Oxygen uptake rates of isolated tumor cells increase at elevated temperatures

On average, isolated tumor cells (exponentially) increase their O_2 consumption rate upon temperature elevation up to maximum values at approx. 42°C. The respective temperature coefficient Q_{10} (van't Hoff coefficient) obtained is 2.3, thus clearly indicating completely functioning mitochondria (for a review see Vaupel & Mayer, 2012).

Oxygen uptake rates of cancers increase with rising arterial O_2 concentrations

The O_2 concentration in the arterial blood supplying the tumor mass is a function of the arterial O_2 partial pressure (pO_2) and/or the hemoglobin content (cHb). Changing these parameters will thus distinctly influence the oxygen availability and, in turn, the O_2 uptake rate of cancers. Increasing these O_2 transport factors, individually or in combination (e.g., upon arterial hyperoxia and/or correction of tumor-associated or therapy-induced anemia) has led to higher O_2 availability and, thus, to an increase in the O_2 uptake rate, again clearly speaking against a respiration injury in cancer cells as a universal feature (for a review see Vaupel & Mayer 2012).

Although there is unequivocal experimental evidence from *in vivo* studies against Warburg's original notion of defective mitochondria in cancer cells, hypotheses similar to Warburg's "Damaged mitochondria model" are still popular in cancer research based on the discovery of (a) mutations of TCA-cycle enzymes (e.g., Boland et al., 2013), and (b) an increase in ROS in cancer cells due to a drop of the ROS-scavenging enzyme superoxide dismutase-II (SOD-II) in mitochondria (e.g., Sabharwal & Schumacker, 2014).

In an Opinion Article, Denko (2008) has pointed out that the Warburg effect may not be due to increasing glycolysis alone but also by *decreasing mitochondrial activity*. This latter effect is triggered by the HIF-1α transcription factor, which can be activated both under hypoxic and normoxic conditions. HIF is considered a key activator of aerobic and hypoxic (anaerobic) glycolysis (see below). HIF-1α can downregulate mitochondrial function (at 1-2% O_2) by direct activation of PDK1 which in turn inactivates the mitochondrial enzyme pyruvate dehydrogenase (PDH). As a consequence, pyruvate (endproduct of glycolysis) cannot enter the TCA cycle, finally leading to a reduced OXPHOS, cellular oxygen consumption and ROS generation (Denko, 2008). A second HIF-1-induced mechanism is the reduction of the number of mitochondria per cell, due to induction of MXI1, an antagonist of Myc, which regulates mitochondrial biogenesis. A third mechanism, by which HIF-1 controls mitochondrial function is by altering the cytochrome c oxidase activity (Denko, 2008).

Warburg effect: Inefficient means of energy generation?

The Warburg effect, i.e., the enhanced rate of glycolysis and generation of lactate in the presence of adequate O_2 supply and fully functioning mitochondria, has been described as an inefficient means of energy metabolism. In aerobic glycolysis the net production is 2 moles each of ATP and pyruvate *per mole glucose*, whereas the total yield being approx. 36 (-38) ATPs from the

complete oxidation of 1 mole of glucose to CO_2 and H_2O, or rather H^+ and HCO_3^- anions (generation of 2 ATPs in glycolysis + 2 ATPs in TCA cycle + 32 ATPs in electron transport chain and OXPHOS). However, the speed of the cytosolic ATP generation is approx. 100- times (range: 30-200 times) faster than in mitochondria, i.e., aerobic glycolysis is able to supply much more ATP *per unit time* than oxidative metabolism of glucose ("low- yield, but high- speed ATP production", Pfeiffer et al., 2001, Shestov et al., 2014). In fact, the amount of ATP generated over any given time period is comparable when either form of glucose metabolism is utilized as long as abundant glucose is available from the extracellular compartment ("tumor microenvironment", Liberty & Locasale, 2016). In cases of greatly increased ATP demand by cancer cells (e.g., during higher activity of efflux membrane ATPases), aerobic glycolysis can increase rapidly while OXPHOS remains constant due to a *much faster and abundant ATP production* through the Warburg effect (Epstein et al., 2014).

Warburg effect: A central feature of metabolic reprogramming in cancer!

Metabolic reprogramming, as a consequence of metabolic plasticity, has been described as one of the hallmarks of cancer (Hanahan & Weinberg, 2011). This trait relates to -*inter alia*- elevations in glycolytic and glutaminolytic fluxes. Acceleration of the glycolytic flux is mainly accomplished by *activation of the PI3K/Akt/mTOR- pathway and of HIF-1α* ("master regulators") upon binding of growth factors totheir surface receptor , and upon activation of HIF-1 by hypoxic stress (Semenza, 2010), or (constitutively) by moderate levels of ROS, HSP90, mutation of tumor suppressors (e.g., PTEN, p53, VHL) or activation of oncogenes (e.g., cMYC, Ras, Raf), respectively. Acceleration of the glycolytic flux can also be achieved by *decrease of AMPK signaling*, as well as *cMyc-driven* (a) upregulation of the glucose transporters GLUT-1 and GLUT-3, thus substantially facilitating glucose transport into the cytoplasm, and (b) overexpression of the enzymes catalyzing the initial steps of the glycolytic pathway in more than 70% of all human cancer types, i.e., hexokinase (HK) and

phosphofructokinase (PFK, Marin-Hernandez et al., 2009, Tran et al., 2016). In all, the alterations mentioned above are implemented in the hostile tumor microenvironment, contributing to the development of the Warburg phenotype.

In addition to providing ATP, aerobic glycolysis *generates antioxidant moieties* (2 NADHs per mole glucose) and allows for the *diversion (shuttling) of glycolytic intermediates* ("carbon sources") into several *biosynthetic pathways* (generation of nucleotides and NADPH -via the pentose phosphate pathway-, lipids and nonessential amino acids) thus supporting biosynthetic programs as long as an adequate glucose supply is maintained. (The role of glutaminolysis providing nitrogens for biosynthesis of nucleotides and amino acids in cancer cells is discussed below.) These biosynthetic pathways are facilitated through a decreased activity of the enzyme pyruvate kinase, which catalyzes the final step of aerobic glycolysis converting phosphoenolpyruvate to pyruvate. In contrast to normal tissues, in most (if not all) human cancers tested to date, tyrosine kinase receptor stimulation and signaling (as well as induction by HIF-1α) leads to the expression of the nearly inactive dimeric M2-isoform of pyruvate kinase (PK-M2, Mazurek & Eigenbrodt 2003, Mazurek et al. 2005), This limits the phosphoenolpyruvate-to-pyruvate conversion and thus leads to an accumulation of upstream glycolytic intermediates ("glucose carbons") and herewith shuttling into the biosynthetic pathways mentioned above (model "Cell proliferation requirement" according to Zhou et al., 2017).

Aerobic glycolysis can minimize the production of reactive oxygen species (ROS; Lu et al., 2015) which, at lower concentrations, can promote increased survival of cancer cells (Sporn & Liby, 2012) as long as an adequate glucose supply is maintained ("tumors act as glucose traps", Shapot, 1980). High glycolytic rates can lead to glucose depletion in the tumor microenvironment (TME) below physiological levels (TME glucose concentrations in human cancers: < 2.5 mM, Vaupel, 1992; glucose concentration of the interstitial fluid in experimental rat tumors: < 1 mM, Gullino et al., 1964).

All in all, the Warburg effect, as a key feature of *"selfish" metabolic reprogramming of cancer cells*, enables/supports rapid and continuous proliferation and growth, maintenance of the cancer stem cell state, survival, and immune system evasion rather than highly efficient ATP production (Fig. 1). This possibility was never anticipated by Warburg, his supporters and contemporaries (e.g., DeBerardinis et al., 2008, Vander Heiden et al., 2009, Cairns et al., 2011, Mihaylova & Shaw, 2011, Ward & Thompson, 2012, Soga, 2013, Amoedo et al., 2013, Phan et al., 2014, Liberty & Locasale, 2016, Bose & Le, 2018, Kato et al., 2018).

Warburg effect: Release of lactate, a key oncometabolite, signaling molecule, and fuel for normoxic cancer cells

High lactate levels in the TME (up to 45 mM) result from *upregulation of aerobic glycolysis*, HIF-1α-driven *glycolysis* as an adaptation in hypoxic/anoxic tumor regions and Myc-driven *glutaminolysis*, another key feature of metabolic reprogramming in cancers, to a certain extent, contribute to lactate accumulation (see below).

pO_2 declines faster with distance from tumor microvessels than do glucose concentrations (Vaupel, 2004). Therefore, anaerobic glycolysis significantly contributes to lactate formation with increasing proportion (a) with enlarging distance from vessels, and (b) with declining pO_2 values from the arterial (inlet) to the venous (outlet) end of the tumor vessel (see Figs.2 and 3). Finally, high lactate levels may also be the consequence of insufficient waste drainage in poorly vascularized tumor areas (Mayer & Vaupel, 2013).

In the 1970s and 1980s the "glycolytic rate" (GR), i.e., the fraction of utilized glucose which is eliminated as lactate via the venous blood, was examined in "tissue-isolated" experimental tumors *in vivo*. In animals breathing room air, GR varied considerably from 28% to 65%, depending on the tumor type and volume

(e.g., Vaupel, 1974, Gullino, 1976, Sauer et al., 1982). Upon breathing 100% oxygen, the mean GR significantly decreased from 51% to 42% in isotransplanted tumors in rats (Vaupel, 1974). For human colon carcinomas *in situ* comparable GR values were described (Holm et al., 1995). The individual contribution of lactate production by aerobic vs. anaerobic glycolysis, however, could not be assessed in these latter experiments.

For an accelerated lactate formation in aerobic glycolysis, HIF-1α- and cMyc-activation increase the expression of pyruvate dehydrogenase kinase 1 (PDK-1) and lactate dehydrogenase A (LDH-A), which drives pyruvate into lactate, thus contributing to the aerobic glycolysis phenotype (Liao, 2014, 2017). Increased HIF-directed expression of PDK-1 inhibits pyruvate dehydrogenase (PDH) and, as a consequence, limits pyruvate entry into the mitochondrial matrix and the Krebs cycle, respectively.

In this latter section, the model of "glycolysis as an adaptation to tumor hypoxia" is highlighted, supporting the notion, that glycolysis is preferentially due to the development of hypoxia and may not display the Warburg effect *in vivo* (Zhou et al., 2017).

Lactate, a master oncometabolite and signaling molecule

Lactate probably is the only oncometabolite involved and necessary in nearly all main sequelae for malignant progression. A series of these pathophysiological conditions results after binding of the *ligand lactate* to the G-protein coupled receptor 81 (GPR81-receptor) and *activation of the PI3K/Akt/mTOR pathway* (San-Millan & Brooks, 2017). Sequelae *inter alia* include sustained angiogenesis, promotion of cell motility, local invasion (partly via hyaluronidase production by cancer-associated fibroblasts) and distant metastasis, resistance to apoptosis, promotion of a stem cell phenotype, suppression of anti-tumor immune responses, and may mediate *radioresistance* by virtue of its antioxidant properties (for a review see Mayer & Vaupel, 2013). In addition, lactate

indirectly stabilizes HIF-1α through inhibition of prolyl-hydroxylase (PHD; Pavlova & Thompson, 2016, Sormendi & Wielockx, 2018).

Lactate/lactic acid causes acidification of the tumor microenvironment

The export of lactate$^-$ anions and H^+ through hypoxia-/HIF-1α-driven MCT4-symporters into the extracellular space lead to a substantial *acidification*/*acidosis* of the tumor microenvironment (TMEwith pH-values< 6.8). (Lact-)Acidosis is a detrimental trait since it drives *inter alia* genetic instability and mutagenesis, triggers resistance to apoptosis, promotes stem cell phenotypes (Lee & Kim, 2016), cancer cell motility, clonal selection of cells with high invasive capacity, upregulates MMP-2- , MMP-9- and cathepsin activities, induces epithelial-to-mesenchymal transition (EMT) and promotes metastasis via enhanced ROS formation (Thews & Riemann, 2019). Acidosis stabilizes HIF-1α, induces vigorous angiogenesis via VEGF-overexpression and triggers autophagy. Furthermore, it can lead to radioresistance, impedes anti-tumor immunity, diminishes the effectiveness of basic chemotherapeutic drugs and may contribute to drug resistance via mTOR-signaling (Pavlova & Thompson, 2016, Corbet & Feron, 2017).

In cancer (as well as in normal tissues) protons can also act as ligands. The respective pH_e sensing and signaling in malignant tumors has been described by Damaghi et al. (2013).

Noteworthy to know: (a) The frequent (but necessarily evolving) co-existence of tumor hypoxia, lactate accumulation and acidosis *in vivo* often complicates the interpretation of their individual roles in tumor progression (Vaupel et al., 2004). (b) Other pathomechanisms involved in and factors contributing to the acidification of the extracellular space in tumor are ATP- hydrolysis, ketogenesis and CO_2 production (which is subsequently hydrated into bicarbonate$^-$ and H^+), by the Krebs cycle, the pentose phosphate pathway (see

above), and -to a certain extent- *glutaminolysis*, another key element of "selfish" metabolic reprogramming in cancers. Glutamine is the second principal growth supporting substrate for cancer cells, providing nitrogens required for the synthesis of nucleotides and nonessential amino acids (Li & Le, 2018). Glutamine serves as an energy source in normoxic cancer cells, regulated by cMyc and HIF-1α. It is taken up into cancer cells through ASC2- transporters and then converted to glutamate and α-ketoglutarate. The subsequent reaction steps in the Krebs cycle (TCA cycle) lead to malate. The conversion of malate to lactate is catalyzed by malic enzyme (ME) and lactate dehydrogenase A (LDH A), both overexpressed in malignant tumors. Glutaminolysis drives the *de novo* synthesis of glutathione, the conversion of malate to pyruvate the production of NADPH, both molecules providing reducing power (Cairns et al. 2011).

Lactate fuels normoxic cancer cells

Upon export through MTC4-symporters into the extracellular space and accumulation in the TME, lactate - by contrast- can be imported by normoxic cancer (and stromal) cells via MTC1-symporters and utilized to fuel oxidative phosphorylation (OXPHOS), thus sparing the uptake of glucose which is -as described above- avidly consumed by hypoxic cells (Sonveaux et al., 2008, 2012, Semenza 2008, Romero-Garcia et al., 2016),

Conclusions

The Warburg effect (as we know it today) describes the enhanced rate of glycolysis (i.e., predominant conversion of glucose to pyruvate) and generation of lactate in the presence of abundant oxygen and fully functioning mitochondria. Switching to aerobic glycolysis is the main feature of metabolic reprogramming, one of the hallmarks of cancer. It constitutes an early and rapid adaptation of cancer cells upon hypoxia, oncogenic gain-of-function and loss-of-

function of suppressor genes in order to meet biosynthetic and bioenergetic demands in growing tumors and malignant progression.

Historical perspective versus the modern concept

Warburg's landmark observations that cancer cells predominantly convert large amounts of glucose to lactate even under conditions of adequate O_2 supply are still acknowledged 95 years later. However, his dogma relying on his original interpretation of the experimental data ("aerobic glycolysis in cancer cells is due to a mitochondrial damage" and "aerobic glycolysis is the universal metabolic alteration in malignant transformation") is no longer tenable (Vaupel & Mayer, 2017).

Advanced developments in molecular biology and high-throughput molecular analyses have revealed, that the selection for high rates of aerobic glycolysis, which is a prerequisite for unlimited growth, is due to an accumulation of signaling pathways which are (a) altered by gene mutations or changes in gene expressions, and (b) often affected by the tumor microenvironment (Cairns et al., 2011) rather than being caused by mitochondrial dysfunctions. In metabolic reprogramming processes responsible for the Warburg effect, these altered signaling pathways act both independently or in concert with each other (Soga, 2013). Our new understanding of cancer metabolism has advanced significantly in the past two decades and is being used for the development of novel diagnostic tools and innovative targeted therapies (Vander Heiden, 2011, Olivares et al., 2015, Velazquez et al., 2016).

Disclosure Statement

The authors report no potential conflict of interest.

Notes on contributors

Peter Vaupel is a well recognized specialist and Professor (Emeritus) of Tumor Pathophysiology. Until 2008 he was Director and Chairman of the Institute of Physiology and Pathophysiology, University of Mainz (Germany). He currently is a member of the Tumor Pathophysiology Group in the Department of Radiation Oncology at the University Medical Center in Mainz. Since 1996 he acts as a Full Member of the Academy of Science in Mainz. His research interests include the tumor oxygenation status, pathogenesis and detection of tumor hypoxia, hypoxia-induced malignant progression, tumor blood flow and microcirculation, tumor oxygen and glucose uptake, tumor pH distribution and tumor bioenergetics. He has published more than 400 original research articles and review papers.

Heinz Schmidberger is Professor of Radiation Oncology and Chair of the Department of Radiation Oncology at the University Medical Center in Mainz. His clinical interests are gastrointestinal cancer, urogenital cancer and lymphoma. In translational research he has been working on DNA repair as a factor of radiation resistance in tumors and a factor for the development of late effects after radiotherapy. He has published 160 original articles and approximately 30 review papers.

Arnulf Mayer is a Senior Lecturer in the Department of Radiation Oncology at the University Medical Center in Mainz. He is trained in Radiation Oncology and Physiology/Pathophysiology. His clinical focus is the multimodal treatment of head and neck cancer, brain tumors, and the utilization of PET in treatment planning. In translational research, he is investigating mechanisms of therapy resistance mediated by the tumor microenvironment using single-cell based image-analytical methods on multi-channel fluorescence specimens of complete tumor sections (“virtual slides”).

Acknowledgement

Figure 3 represents an immunostain from an unpublished study supported by a grant from Varian Medical Systems. Distances between hypoxic cells and microvessels have been calculated in QuPath (https://qupath.github.io), Bankhead et al., 2017)using a script written by Dr. Peter Bankhead, University of Edinburgh, UK.

References

Aisenberg AC. 1961. The glycolysis and respiration of tumors. New York, London: Academic Press

Amoedo ND, Valencia JP, Rodrigues MF, Galina A, Rumjanek FD. 2013. How does the metabolism of tumour cells differ from that of normal cells? Biosc Rep. 33: e00080.

Boland ML, Chourasia AH, Macleod KF. 2013. Mitochondrial dysfunction in cancer. Front Oncol. 3: 292.

Bose S, Le A. 2018. Glucose metabolism in cancer. Adv Exp Med Biol. 1063: 3-12.

Cairns RA, Harris IS, Mak TW. 2011. Regulation of cancer cell metabolism. Nat. Rev.Cancer 11: 85-95

Chance B. 1953. Dynamics of respiratory pigments of ascites tumor cells. Trans NY Acad Sci. 16: 74-75.

Chance B, Hess B. 1956. On the control of metabolism in ascites tumor cell suspensions. Ann NY Acad Sci. 63: 1008-1016.

Chance B, Hess B. 1959. Metabolic control mechanisms. J biol Chem. 234: 2416-2427.

Chance B, Castor LN. 1952. Some patterns of respiratory pigments of ascites tumors in mice. Science 116: 200-202.

Christofk HR, Vander Heiden MG, Harris, MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL Cantley LO. 2008. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature 452:230-233.

Corbet C, Feron O. 2017. Tumour acidosis: From the passenger to the driver's seat. Nat Rev Cancer 17: 577- 593.

Damaghi M, Tafreshi NK, Lloyd MC, Sprung R, Estrelle V, Wijtkowiak JW, Morse DL, Koomen JM, Bui MM, Gatenby RA, Gillies RJ. 2015. Chronic acidosis in the tumour microenvironment selects for overexpression of LAMP2 in the plasma membrane. Nat Commun. 6: 8752.

Dang CV, Kim JW, Gao P, Yustein J. 2008. The interplay between MYC and HIF in cancer. Nat Rev Cancer 8: 51-56.

DeBerardinis RJ, Lum JJ, Hatzivassiliou G Thompson CB. 2008. The biology of cancer: Reprogramming fuels cell growth and proliferation. Cell Metab. 7: 11-20.

Denko, NC. 2008. Hypoxia, HIF 1 and glucose metabolism in the solid tumour. Nat Rev Cancer 8:705-713.

Epstein T, Gatenby RA, Brown JS. 2017. The Warburg effect as an adaptation of cancer cells to rapid fluctuations in energy demand. PLoS One 12(9): e0185085.

Gullino PM. 1976. In vivo utilization of oxygen and glucose by neoplastic tissue. Adv Exp Med Biol. 75: 521-536.

Gullino PM, Clark SH, Grantham FH. 1964. The interstitial fluid of solid tumors. Cancer Res. 24: 780-798.

Hanahan D, Weinberg RA.2011. Hallmarks of cancer: The next generation. Cell 144: 646-674.

Holm E, Hagmüller E, Staedt U, Schlickeiser G, Günther HJ, Leweling H, Tokus M, Kollmar HB. 1995. Substrate balances across colonic carcinomas in humans. Cancer Res. 55: 1373-1378,

Kallinowski, F, Vaupel P, Runkel S, Berg G, Fortmeyer, HP, Baessler KH, Wagner K, Mueller-Klieser W, Walenta S. 1988. Glucose uptake, lactate release, ketone body turnover, metabolic micromilieu, and pH distribution in human breast cancer xenografts in nude rats. Cancer Res. 48: 7264-7272.

Kallinowski F, Schlenger KH, Runkel S, Kloes M, Stohrer M, Okunieff, P, Vaupel P. 1989. Blood flow, metabolism, cellular microenvironment and growth rate of human tumor xenografts. Cancer Res. 49: 3759-3764.

Kato Y, Maeda T, Szuki A, Bala Y. 2018. Cancer metabolism: New insights into classic characteristics. Jpn Dent Sci Rev. 54: 8-12.

Kim JW, Dang CV. 2006. Cancer's molecular sweet tooth and the Warburg effect. Cancer Res. 66: 8927-8930.

Koppenol WH, Bounds, PL, Dang, CV. 2011. Otto Warburg's contributions to current concepts of cancer metabolism. Nat Rev Cancer 11: 325-337.

Lee N, Kim D. 2016. Cancer metabolism: Fueling more than just growth. Mol Cells 39: 847-854.

Li T, Le A. 2018. Glutamine metabolism in cancer. Adv Exp Med Biol. 1063: 13-32.

LiaoY. 2014. Akt and glucose metabolism in cancer. J Metabol Nutr Cancer 1: 80-88.

Liao Y. 2017. Cancer metabolism as we know it today. Genes Dis. 4: 4-6.

Liberty MV, Locasale JW. 2016. The Warburg effect: How does it benefit cancer cells? Trends biochem Sci. 41: 211-218.

Lu J, Tan M, Cai, Q. 2015. The Warburg effect in tumor progression: Mitochondrial oxidative metabolism as an anti-metastasis mechanism. Cancer Lett. 356: 156-164.

Marin-Hernandez A, Gallardo-Perez JC, Ralph SJ, Rodriguez-Enriquez S, Moreno-Sanchez R. 2009. HIF-1alpha modulates energy metabolism in cancer cells by inducing over-expression of specific glycolytic isoforms. Mini Rev Med Chem. 9: 1084-1101.

Mayer A, Vaupel P. 2013. Hypoxia, lactate accumulation and acidosis: Siblings or accomplices driving tumor progression and resistance to therapy? Adv Exp Med Biol. 789: 203-209.

Mazurek S, Eigenbrodt E.2003. The tumor metabolome. Anticancer Res. 23: 1149-1154.

Mazurek S, Boschek CB, Hugo F, Eigenbrodt E. 2005. Pyruvate kinase type M" and its role in tumor growth and spreading. Sem Cancer Biol. 15: 300-308.

Mihaylova MM, Shaw RJ. 2011. The AMP- activated protein kinase (AMPK) signaling pathway coordinates cell growth, autophagy, and metabolism. Nat Cell Biol. 13: 1016-1023.

Moreno-Sanchez R, Rodriguez-Enriquez S, Marin-Hernandez A, Saavedra E. 2007. Energy metabolism of tumor cells. FEBS J. 274 1393-1418.

Olivares O, Däbritz JHM, King A, Gottlieb E, Halsey C. 2015. Research into cancer metabolomics: Towards a clinical metamorphosis. Semin Cell Dev Biol. 43:52-64.

Pavlova NN, Thompson CB. 2016. The emerging hallmarks of cancer metabolism. Cell Metabol. 23: 27-47.

Pfeiffer T, Schuster S, Bonhoeffer S. 2001. Cooperation and competition in the evaluation of ATP-producing pathways. Science 292 504-507.

Phan LM, Yeung S-CJ, Lee M-H. 2014. Cancer metabolic reprogramming: Importance, main features, and potentials for precise targeted anti-cancer therapies. Cancer Biol Med. 11: 1-19.

Potter M, Newport E, Morten KJ. 2016. The Warburg effect: 80 years on. Biochem Soc Trans. 44: 1499-1505.

Racker E. 1976. Why do tumor cells have a high aerobic glycolysis? J Cell Physiol. 89: 697-700.

Romero-Garcia S, Moreno-Altamirano MM, Prado-Garcia H, Sanchez-Garcia FJ. 2016. Lactate contribution to the tumor microenvironment: Mechanisms, effects on immune cells and therapeutic relevance. Front Immunol. 7:52.

Sabharwal SS, Schuhmacker PT. 2014. Mitochondrial ROS in cancer: initiators, amplifiers or an Achilles' heel? Nat Rev Cancer 14: 709-721.

San-Millan I, Brooks GS. 2017. Reexamining cancer metabolism: Lactate production for carcinogenesis could be the purpose and explanation of the Warburg effect. Carcinogenesis 38: 119-133.

Sauer LA, Stayman JW, Dauchy RT. 1982. Amino acid, glucose, and lactic acid utilization in vivo by rat tumors. Cancer Res. 42: 4090-4097.

Semenza GL. 2010. HIF-1: Upstream and downstream of cancer metabolism.
Curr Opin GenetDev. 20: 51-56.

Semenza GL. 2008. Tumor metabolism: Cancer cells give and take lactate. J Clin Invest. 118: 3835-1837.

Shapot VS.1976. Studies on the ability of the tumor respiration in vivo to fulfill its physiological functions. Adv Exp Med Biol. 75: 581-586.

Shapot VS (1980). Biochemical aspects of tumour growth. Moscow: MIR Publishers.

Shapot VS, Vaupel P. 1976. Respiratory gas exchange and metabolism in tumor tissue. Adv Exp Med Biol. 75: 619-620.

Shestov AA, Liu X, Ser Z, Cluntun AA, Hung JP, Huang L, Kim D, Le A, Yellen G, Albeck JG, Locasale JW. 2014. Quantitative determinants of aerobic glycolysis identify flux through the enzyme GADH as a limiting step. Elife.03342.001.

Soga T. (2013). Cancer metabolism: Key players in metabolic reprogramming. Cancer Sci. 104: 275-281.

Sonveaux P, Vegran F, Schroeder T, Wergin MC, Verrax J, Rabbani ZN, De Saedeleer CJ, Kennedy KM, Diepart C, Jordan BF et al. 2008. Targeting lactate-fueled respiration selectively kills hypoxic tumor cells in mice. J Clin Invest.118: 3930-3942.

Sormendi S, Wielockx B. 2018. Hypoxia pathway proteins as central mediators of metabolism in the tumor cells and their microenvironment. Front Immunol. 9: 40.

Sporn MB, Liby KT. 2012. NRF2 and cancer: The good, the bad and the important of context. Nat. Rev Cancer 12: 564-571.

Thews O, Riemann A. 2019. Tumor pH and metastasis: a malignant process beyond hypoxia. Cancer Metastasis Rev. Doi: org/1007/s10555-018-09777-y.

Tran Q, Lee H, Park J, Kim SH, Park J. 2016. Targeting cancer metabolism- Revisiting the Warburg effects. Toxicol Res. 32: 177-193.

Vander Heiden MG. 2011. Targeting cancer metabolism: A therapeutic window opens. Nat. Rev Drug Discov. 10: 671-684.

Vander Heiden MG, Cantley LC, Thompson CB. 2009. Understanding the Warburg effect: The metabolic requirements of cell proliferation. Science 324: 1029-1033.

Vaupel P.1974. Atemgaswechsel und Glucosestoffwechsel von Implantationstumoren (DS-Carcinosarkom) in vivo. Funktionsanalyse biolog Systeme 1: 1-138.

Vaupel P. 1977. Hypoxia in neoplastic tissue. Microvasc Res. 13: 399-408.

Vaupel P. 1992. Physiological properties of malignant tumours. NMR Biomed. 5: 220-225.

Vaupel P. 2004. Tumor microenvironmental physiology and its implications for radiation oncology. Sem Radiat Oncol. 14: 198-206.

Vaupel P, Mayer A. 2012. Availability, not respiratory capacity governs oxygen consumption of solid tumors. Int J Biochem Cell Biol. 44: 1477-1481.

Vaupel P, Mayer A. 2017. Tumor oxygenation status: Facts and fallacies. Adv Exp Med Biol. 977: 91-99.

Vaupel P, Fortmeyer HP, Runkel S, Kallinowski F.1987. Blood flow, oxygen consumption and tissue oxygenation in human breast cancer xenografts in nude rats. Cancer Res. 47: 3496-3503.

Vaupel P, Mayer A, Höckel M. 2004. Tumor hypoxia and malignant progression. Methods Enzymol. 381: 335-354.

Vazquez A, Kamphorst JJ, Markert EK, Schug ZT Tardito S, Gottlieb E. J. 2016. Cancer metabolism at a glance. J Cell Sci.129: 3367-3373.

Warburg O. 1923.Versuche an überlebendem Carcinomgewebe. Biochem Zschr. 142: 317-333.

Warburg O. 1956a. On respiratory impairment in cancer cells. Science 124: 269-270.

Warburg O. 1956 b. On the origin of cancer cells. Science 123: 309-314.

Warburg O. 1962. In: New methods of cell physiology. Applied to cancer, photosynthesis, and mechanisms of X-ray action. Developed 1945-1961. Pp. 631-632 (ed. Warburg, O.) Interscience Publishers, New York.

Warburg O, Posener K, Negelein E. 1924. Über den Stoffwechsel derCarcinomzelle. Biochem Zschr. 152: 319-344.

Ward PS, Thompson CB. 2012. Metabolic reprogramming: A cancer hallmark even Warburg did not anticipate. Cancer Cell 21: 297-308.

Weinhouse S. 1956. On respiratory impairment in cancer cells. Science 124: 267-269.

Weinhouse S. 1976. The Warburg hypothesis fifty years later. Zschr Krebsforsch. 87: 115-126.

Zu XL, Guppy YM. 2004. Cancer metabolism: Facts, fantasy, and fiction. Biochem Biophys Res Commun. 313: 459-465

Zhou W, Liotta LA, Petricoin EF. 2017. The Warburg effect and mass spectrometry-based proteomic analysis Cancer Genom Proteom. 14: 211-218.

Figure legends

Figure 1. The Warburg effect (aerobic glycolysis) at a glance: Major aspects and mechanisms leading to malignant metabolic phenotypes and tumor progression, respectively.

Figure 2. Relationship between mean critical diffusion distances for oxygen (O_2, blue curve) and glucose (red curve) in breast cancer tissue (for computation of diffusion distances, *in vivo* data are used; Vaupel, 2004). From these data there is clear indication that the diffusion distances for glucose are approx. 2- times larger than for oxygen. pO_2 declines faster in the tissue than glucose, i.e., the O_2 extraction rate (approx. 44%) is larger than the extraction rate for glucose (about 38%, Vaupel et al., 1987, Kallinowski et al., 1988). Oxidative metabolism and *aerobic glycolysis* are restricted to tissue volumes next to the vessel and the adjacent areas sufficiently supplied with both oxygen and glucose, whereas *anaerobic/hypoxic glycolysis* is expected in hypoxic/anoxic regions, but still exhibiting adequate glucose supply. Critical oxygen supply thus is thought to be the primary substrate limitation.

Figure 3. *Upper panel*: Multichannel staining of a human melanoma specimen showing GLUT-1 (green channel) and tumor microvessels (red channel). DAPI staining of nuclei (blue channel) is used for single cell segmentation in QuPath.
Middle panel: classification of single cells according to the expression intensity of GLUT-1 (strong expression/hypoxia = bright green, weak or absent expression/normoxia = yellow. Vascular endothelial cells are marked in red. *Lower panel:* histogram of distances calculated from hypoxic cells to the nearest endothelial cells. The modal value of the distribution is 110 µm, i.e., a rim of normoxic cells of approximately this thickness surrounds each blood vessel (range of distances: 0-330 µm). Beyond this distance, a zone of hypoxic cells of slightly smaller thickness begins. Distant from the hypoxic zone, necrotic tissue is found. This pattern corresponds qualitatively to the theoretical analysis presented in Fig. 2. Strongly elevated expression of
GLUT-1 is biologically plausible because glucose is still available in hypoxic tissue microregions up to the point where necrosis begins.