Article

Whole-Genome Sequencing Can Identify Clinically Relevant Variants from a Single Sub-Punch of a Dried Blood Spot Specimen

David J. McBride [1,*], Claire Fielding [1,†], Taksina Newington [1,†], Alexandra Vatsiou [1], Harry Fischl [1], Maya Bajracharya [1], Vicki S. Thomson [1], Louise J. Fraser [1], Pauline A. Fujita [2], Jennifer Becq [1], Zoya Kingsbury [1], Mark T. Ross [1], Stuart J. Moat [3,4] and Sian Morgan [5,*]

[1]   Illumina Cambridge Ltd., Cambridge CB21 6DF, UK
[2]   Illumina Inc., San Diego, CA 92122, USA
[3]   Wales Newborn Screening Laboratory, University Hospital of Wales, Cardiff CF14 4XW, UK
[4]   School of Medicine, Cardiff University, Cardiff CF14 4XW, UK
[5]   All Wales Genetics Laboratory, University Hospital of Wales, Cardiff CF14 4XW, UK
[*]   Correspondence: dmcbride@illumina.com (D.J.M.); sian.morgan22@wales.nhs.uk (S.M.)
[†]   These authors contributed equally to this work.

Abstract: The collection of dried blood spots (DBS) facilitates newborn screening for a variety of rare, but very serious conditions in healthcare systems around the world. Sub-punches of varying sizes (1.5–6 mm) can be taken from DBS specimens to use as inputs for a range of biochemical assays. Advances in DNA sequencing workflows allow whole-genome sequencing (WGS) libraries to be generated directly from inputs such as peripheral blood, saliva, and DBS. We compared WGS metrics obtained from libraries generated directly from DBS to those generated from DNA extracted from peripheral blood, the standard input for this type of assay. We explored the flexibility of DBS as an input for WGS by altering the punch number and size as inputs to the assay. We showed that WGS libraries can be successfully generated from a variety of DBS inputs, including a single 3 mm or 6 mm diameter punch, with equivalent data quality observed across a number of key metrics of importance in the detection of gene variants. We observed no difference in the performance of DBS and peripheral-blood-extracted DNA in the detection of likely pathogenic gene variants in samples taken from individuals with cystic fibrosis or phenylketonuria. WGS can be performed directly from DBS and is a powerful method for the rapid discovery of clinically relevant, disease-causing gene variants.

Keywords: whole genome sequencing; newborn screening; dried blood spots; DNA sequencing; cystic fibrosis; phenylketonuria

1. Introduction

Newborn screening (NBS) plays a vital role in healthcare systems for the prompt identification of individuals who may develop one of a set of rare, but severe health conditions [1]. In the UK, for example, dried blood spot (DBS) specimens are routinely collected at 4–5 days of age and are sent to regional screening laboratories for analysis. Sub-punches are taken from the DBS and analyzed for nine conditions "https://www.gov.uk/guidance/newborn-blood-spot-screening-programme-overview" (accessed on 1 September 2023).

Next-generation sequencing (NGS) has transformed genetic testing in the healthcare setting. Over the past 10 years, the technology has enabled clinical laboratories to progress from assays that focus on a single specific variant to assays that simultaneously query all positions in multiple genes or even whole genomes, dramatically changing what information a clinician can expect from a genetic test [2–4]. WGS is the most comprehensive

test available to capture all classes of variants, from small events such as single-nucleotide variants (SNVs) and small indels, to larger changes in genetic material such as copy number changes (CNV) [5,6] and structural variants (SVs) [7]; the last two often originate deep in intronic regions not included on gene panels or whole-exome sequencing (WES). Calling SVs, even from WGS data, can be compromised by the presence of a repetitive or simple sequence involved in the variant, but there has been success in addressing such cases using combinations of callers or custom callers targeted at specific genomic regions [8–10]. The algorithms used to capture variants and distill the data into a clinical format are under continual improvement, with updates to analysis pipelines requiring validation as well as the laboratory workflows. While this rapid expansion of information can seem daunting, the content is incredibly powerful for aiding diagnosis. Indeed, in rare genetic disorders, whole-genome sequencing (WGS) has been established in numerous healthcare systems as a primary test, in a bid to lessen the diagnostic odyssey faced by patients with conditions which are challenging to diagnose [4,11–16]. While the value of such tests (and rapid diagnoses) for those who suffer from these conditions and their families is obvious "https://everylifefoundation.org/burden-study/" (accessed on 30 August 2023), "https://globalgenes.org/wp-content/uploads/2013/04/ShireReport-1.pdf" (accessed on 1 September 2023) [17–22], the impact of applying them more broadly across the population, or even how this might be accomplished, is less clear. The impact and feasibility of applying WGS to newborns are already being explored [12,21,23–28] "http://www.genomicsengland.co.uk/initiatives/newborns" (accessed on 2 August 2023, "https://www.iconseq.org/" (accessed on 1 August 2023), "https://babyscreen.mcri.edu.au/" (accessed on 31 August 2023).

One key consideration that is being explored in these studies is the source of DNA for WGS. Over the years, NGS workflows have evolved to accept smaller amounts of genetic material as input, but advancement has also been made in terms of allowing workflows to proceed directly from raw, unextracted samples such as blood (including DBS) or saliva. The removal of the extraction step from the workflow has the potential to reduce both the cost and the time taken to obtain the sequencing result while maximizing the DNA input to the assay from small amounts of material. Methods proceeding directly from DBS or DNA extracted from DBS have already been tested using both new and historical samples [25,29–34]. The number of sub-punches used by each group varies. In one study, the authors explored the DNA yields from different numbers of sub-punches, but did not compare WGS from these DNAs [29]. Limiting the number of DBS sub-punches is highly desirable: not only is there a considerable burden on the NBS pipeline in preparing repeated punches from the same sample, but different laboratories use different DBS punch configurations; therefore, the amount of material remaining after standard of care testing can vary. The generation of high-quality data from lower inputs would maximize the number of samples that could benefit from WGS testing.

In this study, we confirm that high-quality WGS data can be generated from libraries derived from DBS and investigate the feasibility of using fewer DBS sub-punches as inputs for detecting clinically relevant variants by WGS.

2. Materials and Methods

2.1. Samples

2.1.1. Consent and Ethics Approval

The study was reviewed by the local research and development department and approved as a service evaluation; formal ethical approval was not required.

2.1.2. Dried Blood Spot Collection

To compare WGS data generated from liquid whole and DBS specimens, we retrieved anonymized residual blood specimens collected into EDTA blood tubes. An aliquot of the liquid blood specimen was taken, and 50 mL aliquots were applied to filter paper collection devices (PerkinElmer 226, Perkin Elmer, Waltham, MA, USA). Liquid whole blood

specimens from the PKU patients were collected during routine monitoring and genetic analysis as part of the KUVAN responsiveness testing pathway [35], and an additional sample was collected from these patients by spotting blood from a finger prick directly onto the blood collection card. The blood spot cards were stored at room temperature for 1 to 6 weeks before processing.

2.1.3. DNA Extraction from Venous Blood

Nucleic acid extraction was undertaken from 2–4 mL of peripheral blood using the Chemagic STAR instrument (Perkin Elmer, Waltham, MA, USA), following the manufacturer's protocol.

2.2. DNA Quantification

DNA samples extracted from blood were quantified on a Gemini XPS Microplate Reader (Molecular Devices, San Jose, CA, USA) using the Quant-iT™ PicoGreen® dsDNA Assay Kit.

2.3. NGS Library Preparation

Sequencing libraries were prepared using the Illumina DNA PCR-Free Prep kit (Illumina Inc., San Diego, CA, USA), which proceeds directly from six 3.2 mm DBS sub-punches when following the standard protocol (Illumina DNA PCR-Free Library Prep Reference Guide, Illumina document #1000000086922 v03). The appropriate number of 3.2 mm or 6 mm sub-punches (6, 3 or 1) was obtained from each DBS using a DBS Puncher® Instrument (Perkin Elmer, Waltham, MA, USA). DNA was released from the sub-punches by shaking at 1000 rpm at 56 °C on a ThermoMixer C (Eppendorf, Hamburg, Germany) in 300 μL of 1× Lysis Master Mix prepared from 10x Modified Lysis Buffer and proteinase K (2 μL), as described in the Illumina Lysis reagent kit protocol (Illumina, Inc., San Diego, CA, USA). After lysis was completed, 135 μL of Illumina Purification Beads were added, and the solution was vortexed briefly and left at room temperature for 5 min. Beads were collected by incubation on a magnetic stand for a further 5 min, then washed for a total of two 30 s washes in 500 μL of 80% EtOH. After the removal of the second wash, beads were air-dried for 5 min before elution in 35 μL of resuspension buffer (RSB). An optional sample QC step was performed by loading 1 μL of lysed material onto the Agilent Genomic DNA ScreenTape System (Agilent, Santa Clara, CA, USA) to assess DNA quality, and 2 μL onto the Qubit™ dsDNA BR Assay Kit (ThermoFisher Scientific, Waltham, MA, USA) to provide an estimate of DNA quantity.

The recovered material from the Illumina lysis kit (30 μL) was fragmented and tagged with NGS adapter sequences by incubating on a thermal cycler with 10 μL PCR-Free Bead-Linked Transposomes (BLT-PF) and 10 μL Tagmentation Buffer 1 (TB1) at 41 °C for 5 min. The reaction was stopped by adding 10 μL of Stop Tagment Buffer 2 (ST2) and incubating for 2 min. Beads were washed with 150 μL of Tagment wash buffer (TWB). Indices were added to adapter sequences using 45 μL of Extension Ligation Mix (ELM) and 5 μL of the appropriate IDT index adapter, then incubating them for 5 min each at 37 °C and 50 °C on a thermal cycler. The beads were, again, washed twice, this time with 75 μL of TWB, before elution from the beads in 45 μL diluted HP3. Libraries were cleaned up and double size selected using sequential incubations (36 μL then 42 μL) with Illumina Purification Beads (plus appropriate 80% EtOH washes). The final libraries were eluted in 22 μL of RSB.

Libraries from the extracted DNA were prepared using the same protocol, with a modification to allow for DNA input (350 ng DNA in 25 μL was added to 15 μL BLT-PF and 10 μL Tagmentation Buffer 1; TB1).

2.4. Sequencing Library Quantification, Pooling, and Sequencing

All libraries were quantified using Kapa qPCR (Merck Life Science UK, Gillingham, UK) on a Roche LightCycler LC480 (Roche Diagnostics Limited, Burgess Hill, UK) and normalized to 0.65 nM. Normalized libraries were pooled at equal volumes and loaded

onto S4 flowcells using Illumina® NovaSeq 6000 Sequencing System Guide XP Workflow. S4 flowcells were sequenced on Novaseq 6000 to generate paired reads of 150 bp.

NovaSeq Control Software version 1.7.5 automatically transferred the base call (*.bcl) files into the BaseSpace Sequencing Hub.

2.5. Data Analysis

Data analysis took place on the BaseSpace Sequencing Hub (BSSH). Data were streamed directly to BSSH from the sequencing instrument. Based on previous experience, we predicted that 130 Gbp of sequencing data would typically be required to achieve an aligned and de-duplicated coverage of 40×. Reads were aligned to Human Reference Genome GRCh38 (alt-aware, with HLAs) using the DragenDNA 3.9.5 analysis workflow. Metrics from the DragenDNA workflow were plotted using R and Python.

Variants which were likely pathogenic were identified by filtering variants called by the DragenDNA workflow using TruSight™ Software Suite (Illumina Inc., San Diego, CA, USA), software that allows rapid analysis and interpretation of the discovered variants. Each experimental condition from each participant was imported as an individual case in the software. Filters were applied in the Interpretation tab of the software to show only PASS variants in or around the gene of interest (i.e., variants occurring in or around *CFTR* for cystic fibrosis and *PAH* for phenylketonuria). We initially searched for two heterozygous variants or one homozygous variant described as pathogenic or likely pathogenic in ClinVar and confirmed the variants discovered by visual inspection by reviewing the aligned reads (bam file) in IGV. For one of the PKU patients, this strategy only revealed a single heterozygous variant. We extended the search to intronic regions of *PAH* where the software had identified a Translocation Breakend variant. Manual review of this variant in IGV confirmed the presence of reads consistent with the presence of a 4.2 Mb inversion.

3. Results

3.1. High-Quality WGS Data Are Achieved from DBS

Library preparation protocols for DNA sequencing can proceed directly from DBS with the inclusion of a simple lysis step, negating the requirement for DNA extraction from blood. DNA sequencing libraries were made initially from both DBS punches (6×3 mm) and extracted DNA (350 ng) using the Illumina DNA PCR-Free library preparation kit following the manufacturer's standard protocol (Supplementary Figure S1). PCR-Free WGS libraries were successfully generated from both sample inputs for all individuals (n = 20) and combined in pools for sequencing. We assessed the relative quality of the genomes obtained by several metrics generated using the DRAGEN (Dynamic Read Analysis for GENomics) analysis pipeline (Illumina Inc., San Diego, CA, USA). The achieved coverage was proportional to the yield of sequencing rather than the input material (Median 44.7×, range 29.0×–66.3×; Supplementary Figure S2). At this level of sequencing coverage, on average, there are enough data to re-sequence each autosomal position more than 40 times; however, the actual depth of coverage at each position varies due to the localized sequence content, with some duplicated regions of the genome causing problems with alignment [36–38]. Sampling a position more than 20 times gives us an excellent chance of seeing any heterozygous variant present at that position. More than 95% of the autosomes and >97.5% of the coding exons were covered by 20 reads or more in the majority of samples (Figure 1a,b, Supplementary Figure S2). Irregularities or biases in the expected sequence representation (as described by sequence coverage variability in regions with different proportions of adenine and thymine bases versus guanine and cytosine bases; GC bias) can be used to identify poor sample or sequencing quality. GC bias curves were plotted for DNA and DBS inputs and show the expected normalized coverage (i.e., a value close to '1') across the majority of the genome when grouped by sequence content (Figure 1c), demonstrating that high-quality WGS data were generated regardless of input.
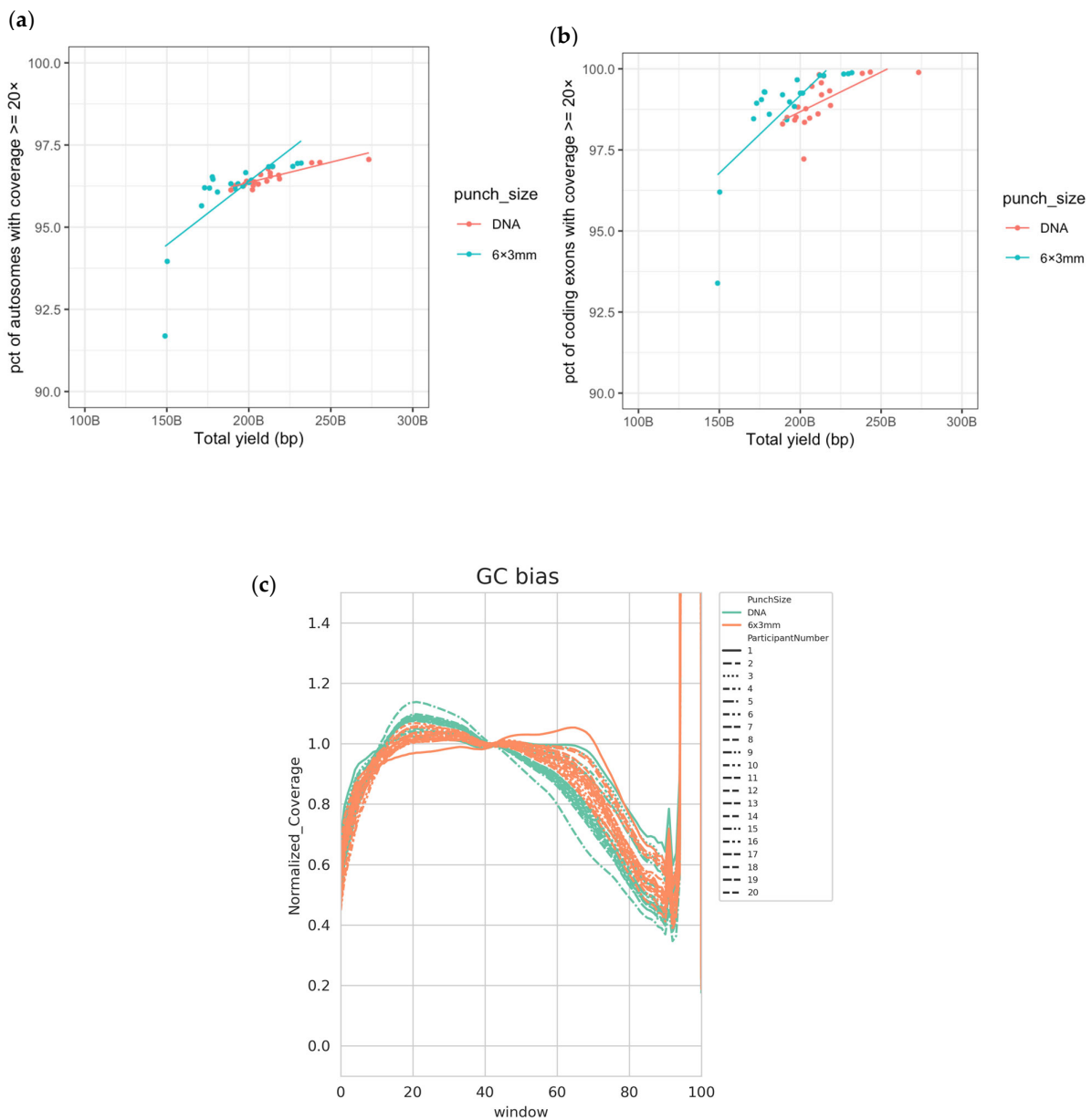
Figure 1. Good-quality WGS data achieved from DBS and extracted DNA. (**a**) Percentage of autosomes with coverage ≥20× vs. total yield sequenced; (**b**) percentage of coding exons with coverage ≥20× vs. total yield sequenced; (**c**) normalized coverage across various windows of GC content in the human genome, grouped by participant. The numbers below the plot indicate the percentage of the human reference genome sequence included in each window of GC content.

3.2. Good-Quality WGS Data Can Be Achieved from a Variety of DBS Punch Configurations

The material collected on DBS cards is normally used as an input for a number of assays. Given that this precious resource can be in short supply, we wanted to investigate whether there was an impact of reducing the number of punches used as input to the WGS assay from the recommended 6 × 3 mm punches (see Methods). Remaining mindful that some laboratories use a 6 mm punch head for processing DBS, we also wanted to explore how well sub-punches of this size performed in the assay. We returned to the DBS cards

taken from participants 9 to 20 to test additional punch configurations (3 mm × 3 mm, 1 mm × 3 mm, and 1 mm × 6 mm) as inputs for comparison.

Altering the punch configuration could result in a change in the amount of input material into WGS library preparation, as there is considerable variation in the volume of blood (9 μL, 3 μL, and 12 μL, respectively [39,40]), and, therefore, the DNA available in each configuration. While all modifications led to some reduction in input material and sequencing library yield, there was sufficient material to generate PCR-Free libraries in each case (Figure 2a). Once again, the coverage which was achieved was proportional to the sequencing yield, with no difference in sequencing quality and only subtle differences in normalized GC content (Figure 2b–d).

While high-quality WGS data were achieved from all input types, we were able to detect subtle differences across the gradient of input types, which might determine the appropriate sub-punch configuration. The fragment length reduced as fewer 3 mm punches were added to the library preparation, consistent with lower DNA input and resulting in shorter library insert sizes for these samples (Figure 2a,e). This is not unexpected given the increased ratio of transposomes to DNA molecules, which would result in a higher degree of fragmentation of the DNA. Insert size (or library fragment length) is an important consideration because sequencing reads are generated from both ends of each fragment of the library in standard (paired-end) Illumina sequencing. Thus, libraries with fragment lengths that are too short can lead to reads with significant overlap, resulting in the generation of redundant information. Short fragment lengths can also result in read pairs that are more challenging to align to the human reference genome sequence, especially those within short, interspersed repeat elements and other sequences that display significant similarity to others present in the genome. While the reduction in library fragment length seen here is not large enough to cause significant impact, more sequencing was required in order to achieve a minimum depth of 20x than for other punch configurations (Figure 2b,c). Library diversity, i.e., the number of unique fragments contained in the library, may therefore be the main driver behind the need to perform additional sequencing of some of the lower-input libraries to achieve the desired level of coverage. Interestingly, the 1 × 6 mm sub-punch inputs yielded libraries of similar fragment sizes to standard 6 × 3 mm despite the lower input, as estimated by the Qubit assay. This may suggest that the Qubit estimation of DNA content in the crude lysate extraction is not as accurate as that of traditionally extracted DNA, but it is interesting to note that the library yield and estimated volume of blood added (via the DBS) were also closer for the 3 × 3 mm and 1 mm × 6 mm sub-punch configurations.

3.3. Identification of Clincally Relevant Variants from DBSs

Three of the participants included in the study (6, 7, and 8) have cystic fibrosis. WGS data were generated from the extracted DNA, and 6 × 3 mm DBS sub-punch inputs from each of these participants and variants were filtered using TruSight™ Software Suite (TSS) to search for high-confidence, likely pathogenic variants. We were able to identify two such variants for each participant. Participant 6 was homozygous for loss of the phenylalanine codon at position 508 of the *CFTR* gene (c.1521_1523del ΔF508), the variant most commonly associated with cystic fibrosis. Participants 7 and 8 were heterozygous for this deletion, but carried additional missense variants (c.3909 C>G Asn1303Lys and c.1624 G > T Gly542Ter, respectively). Likely pathogenic variants in *CFTR* identified by the variant caller were reviewed and confirmed using the Integrative Genomics Viewer (IGV [41,42]; Figure 3). These data demonstrate that clinically relevant variants can be identified from DBS and are in concordance to those variants detected in the peripheral blood.
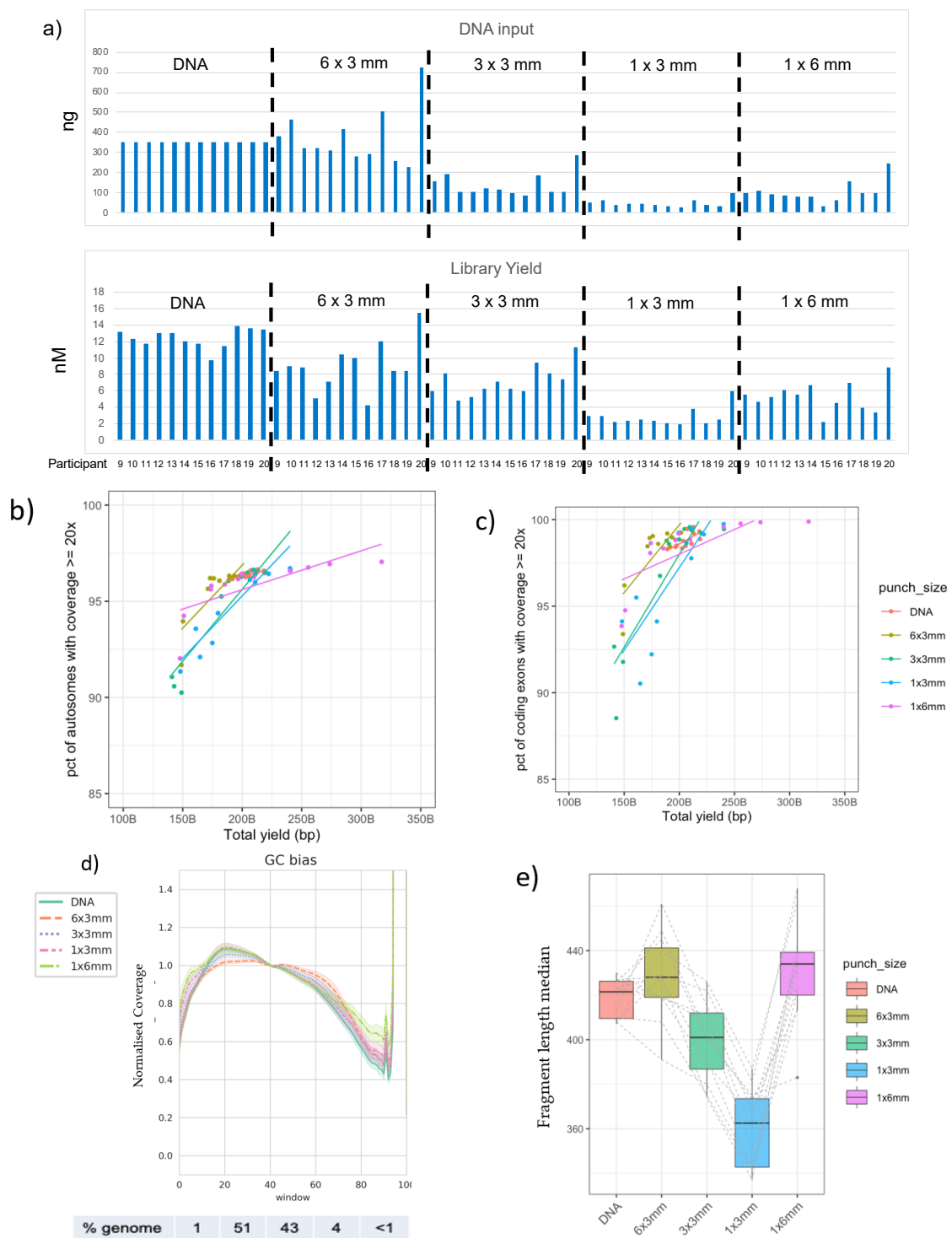
Figure 2. WGS data generated from a variety of DBS punch inputs. (**a**) Qubit values measured post-DBS lysis protocol for participants 9 to 20 are shown above the sequencing library quantifications determined by quantitative PCR (qPCR) for each input type. (**b**) Percentage of autosomes with coverage $\geq 20\times$ vs. total yield sequenced; (**c**) percentage of coding exons with coverage $\geq 20\times$ vs. total yield sequenced; (**d**) normalized coverage across various windows of GC content in the human genome, grouped by sample input type and averaged for the 20 participants. Numbers below the plot indicate the percentages of genomic sequences included in each window of GC content; (**e**) fragment length median for each input type for participants 9 to 20. Ranges of values are summarized by box and whisker plots, with dotted lines link data points from the same participants.
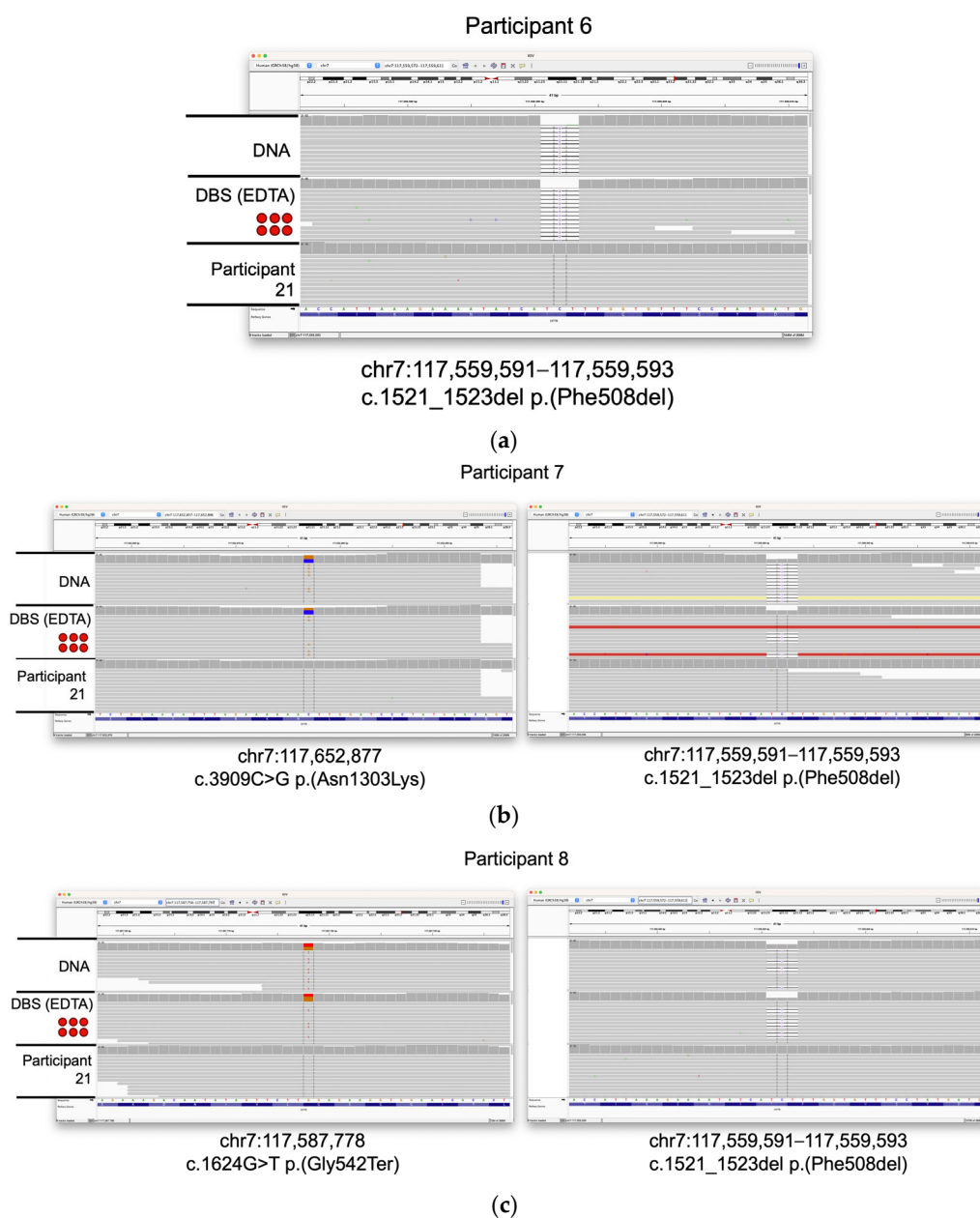
Figure 3. Identification of *CFTR* variants in samples from individuals with cystic fibrosis. IGV screenshots of likely pathogenic variants discovered in three participants with cystic fibrosis. Tracks from DNA and DBS (6 × 3 mm) inputs are shown. Data from participant 21 (extracted DNA) are shown for comparison. (**a**) Homozygous Phe508del variant seen in participant 6. (**b**) Asn1303Lys and Phe508del variants seen in participant 7. (**c**) Gly542Ter and Phe508del variants seen in participant 8.

3.4. Detection of Clinically Relevant Variants from DBS Taken Directly from the Participant

To facilitate direct comparison, all previous DBS were formed using the same EDTA blood specimen that was used for DNA extraction from peripheral blood. To confirm that the blood applied directly to a collection card by capillary blood collection (using a finger prick) would show an equivalent performance, we tested samples from two individuals with phenylketonuria (PKU). DBS were collected by capillary taken directly from the participant as part of routine monitoring, as well as being generated from an EDTA blood draw. The metrics obtained from the libraries generated using both inputs were highly comparable (Figure 4a–c)
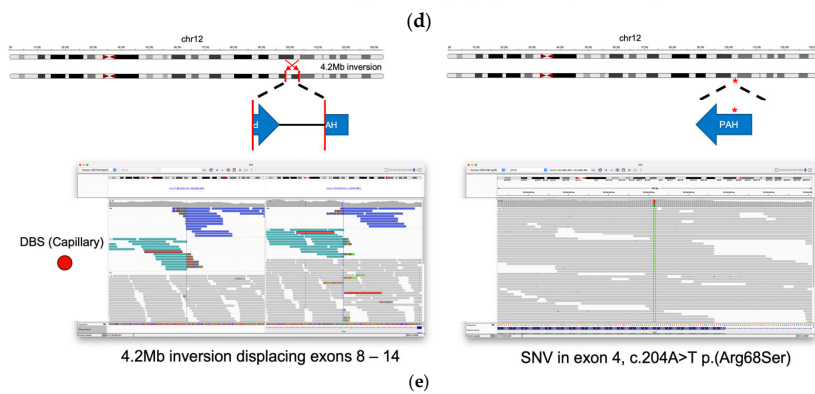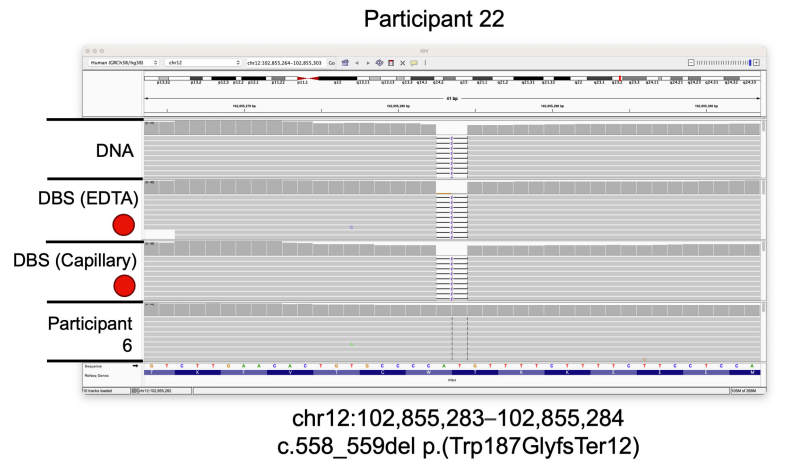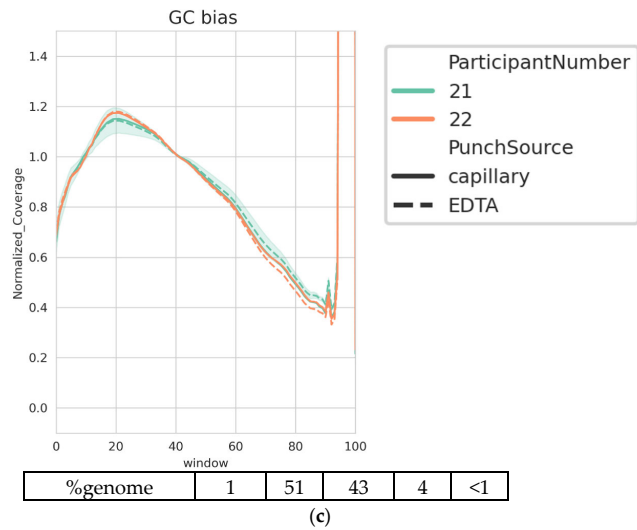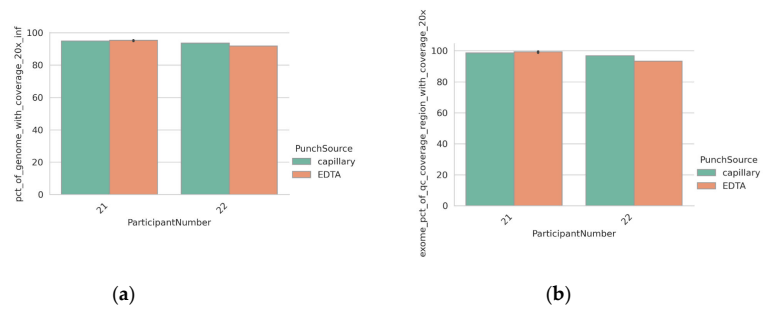
(**a**)

(**b**)

%genome	1	51	43	4	<1

(**c**)

Participant 22

chr12:102,855,283–102,855,284
c.558_559del p.(Trp187GlyfsTer12)

(**d**)

4.2Mb inversion displacing exons 8 – 14

SNV in exon 4, c.204A>T p.(Arg68Ser)

(**e**)

Figure 4. Equivalent performance of a DBS punch (6 mm) and DNA extracted via EDTA blood tubes. (**a**) Percentage of autosomal regions with $\geq 20\times$ coverage for each sample; (**b**) percentage of exome

regions with ≥20× coverage for each sample; (**c**) normalized coverage across various windows of GC content in the human genome, grouped by sample input type; (**d**) IGV screenshot of likely pathogenic gene variants discovered in participant with PKU. Tracks from the DNA and DBS inputs are shown. Data from participant 6 (extracted DNA) are shown for comparison (**e**) Likely pathogenic variants discovered in samples taken from a second participant with PKU. Cartoon representation of the variant location on chromosome 12 and zoomed-in IGV screenshots from the DBS taken directly from the patient. Asterix highlights the presence of the SNV shown in lower panel.

Variants identified from these participants were filtered using TSS in order to identify variants of potential clinical significance. Two variants of interest were identified in each participant regardless of assay input (DBS from patient, DBS via EDTA blood draw, and extracted DNA). One participant carried a homozygous deletion in *PAH*, c.558_559del, which is predicted to result in protein truncation, while the other carried both a heterozygous missense variant in PAF, c.204A > T, and a heterozygous 4.2 Mb inversion displacing exons 8–14 of *PAH* (Figure 4d,e). The identification of the latter demonstrates the power of WGS (from either input) to detect clinically relevant complex structural variants.

4. Discussion

The use of PCR-Free library preparation for WGS is desirable, because PCR amplification steps can introduce coverage biases across the genome and even introduce incorrect bases during amplification [43–45]. In this study, we demonstrate that PCR-Free WGS can be successfully performed directly from a range of DBS inputs (6 × 3 mm, 3 × 3 mm, 1 × 3 mm, 1 × 6 mm) extending beyond those recommended by the manufacturer (6 × 3 mm). The obtained data were highly comparable to those generated using extracted DNA from peripheral blood and were not dramatically affected by the size or number of the sub-punches analyzed. The ability to perform a PCR-Free free library preparation for all sub-punch configurations demonstrates that the 'gold standard' WGS library preparation method is accessible across a range of DBS inputs.

A number of metrics can be used to assess WGS data quality. Here, we focused on depth of coverage because, ultimately, this has the most significant impact on the ability to identify variants in mature NGS pipelines. Our study adds to the growing body of evidence that DBS are a valid input for NGS, offering a real alternative to DNA extracted from blood, where sample collection is challenging [29–33]. Previous studies have investigated the quantity and quality of DNA from DBS before library preparation. The protocol used here is designed to proceed directly from DBS, releasing and lysing cells from the collection card using a simple procedure before starting NGS library construction. While the protocol does not require an assessment of DNA quality, and indeed, QC values may not be comparable to those of DNA extracted by a standard DNA extraction kit, assessment of the material going into the NGS library preparation by gDNA TapeStation suggests that the DNA going into the assay is, in fact, of high quality (Supplementary Figure S3).

While there was no significant impact on data quality when reducing the number of sub-punches, laboratories seeking to do so should be aware that performing optional steps in the protocol may be required in order to achieve a successful WGS workflow. For example, the Illumina DNA library prep protocol can remove the requirement for quantitative PCR of libraries in advance of loading on the sequencing instrument, because the bead linked transposomes are able to normalize the amount of library produced in each reaction. This is desirable when pooling libraries from multiple samples to be run together at the same time on the sequencing instrument to ensure that equal amounts of data are generated from each sample. However, the reduced DNA input from lower numbers of sub-punches may mean that the library normalization properties of the bead-linked transposomes will be lost; thus, assessment of library quantity by quantitative PCR should be maintained to ensure accurate pooling and evenness of coverage. If a single 3 mm punch is used, then additional sequencing may be required in order to achieve the desired

level of coverage, owing to the smaller library fragment sizes. Therefore, the gain of a streamlined DNA punching pipeline may have some impact on cost and turnaround time in the genomic laboratory. The obtained data suggest that a 1×6 mm sub-punch may provide the best combination of simplicity and WGS quality.

The ability to generate PCR-Free whole genomes from DBS is particularly exciting considering that DBS are already collected routinely as part of routine care in newborn screening programs across the world. Material from these DBS is already used as input in a variety of assays, which means that availability of materials could be limited for further testing. The fact that this workflow can proceed from just a single blood sub-punch opens the possibility of performing WGS in parallel to existing tests, adding extra detail to information captured by the current programs or accelerating the identification of genetic changes that would normally require a secondary sample collection procedure (as is the situation today with cystic fibrosis testing). Another challenge of adopting WGS for NBS is the interpretation of the revealed variants. Many of these have unknown significance, especially when studying diverse populations where there is a need for improved reference databases "https://www.ga4gh.org/" (accessed on 31 August 2023), "https://h3africa.org/" (accessed on 31 August 2023) [46–50]. Additionally, recent work using primate sequenced and machine learning may provide valuable information as to whether variants of unknown significance are likely to be pathogenic or benign [51,52]. The individuals in this study had known conditions, allowing us to focus our search for likely pathogenic variants in a limited number of well-described disease genes.

While other NGS assays could proceed from DBS, we believe that the whole-genome sequencing test is the appropriate one in this scenario. Firstly, the non-targeted nature of the WGS assay provides maximum 'future proofing', as clinically relevant genes or regions can be added to reports as they are discovered, without modification to the sample collection or laboratory assay workflows. Secondly, a WGS library is ready for sequencing in a significantly shorter amount of time than one obtained using enrichment strategies, which require additional steps after library preparation. And thirdly, the comprehensive nature of the WGS test means that all classes of variant can be detected, including complex structural variants like the *PAH* inversion discovered in this study.

A WGS test would be a powerful addition to any newborn screening program, potentially providing information on the genetic health of the individual during their lifetime. There are significant challenges in the logistics of sample collection and processing, and the ability to utilize existing testing infrastructures with a proven track record of success is incredibly attractive.

Conflicts of Interest: David J. McBride, Claire Fielding, Taksina Newington, Alexandra Vatsiou, Harry Fischl, Maya Bajracharya, Vicki S. Chambers, Louise Fraser, Pauline A. Fujita, Jennifer Becq, Zoya Kingsbury, and Mark T. Ross are either current or previous full-time employees of Illumina Inc. Stuart J. Moat and Sian Morgan declare no conflicts of interest.

References

1. Therrell, B.L.; Padilla, C.D.; Loeber, J.G.; Kneisser, I.; Saadallah, A.; Borrajo, G.J.; Adams, J. Current status of newborn screening worldwide: 2015. *Semin. Perinatol.* **2015**, *39*, 171–187. [CrossRef] [PubMed]
2. Rehm, H.L.; Bale, S.J.; Bayrak-Toydemir, P.; Berg, J.S.; Brown, K.K.; Deignan, J.L.; Friez, M.J.; Funke, B.H.; Hegde, M.R.; Lyon, E. ACMG clinical laboratory standards for next-generation sequencing. *Anesth. Analg.* **2013**, *15*, 733–747. [CrossRef] [PubMed]
3. Deignan, J.L.; Astbury, C.; Cutting, G.R.; del Gaudio, D.; Gregg, A.R.; Grody, W.W.; Monaghan, K.G.; Richards, S. CFTR variant testing: A technical standard of the American College of Medical Genetics and Genomics (ACMG). *Anesth. Analg.* **2020**, *22*, 1288–1295. [CrossRef] [PubMed]
4. The 100,000 Genomes Project Pilot Investigators. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care—Preliminary Report. *N. Engl. J. Med.* **2021**, *385*, 1868–1880. [CrossRef] [PubMed]
5. Gross, A.M.; Ajay, S.S.; Rajan, V.; Brown, C.; Bluske, K.; Burns, N.J.; Chawla, A.; Coffey, A.J.; Malhotra, A.; Scocchia, A.; et al. Copy-number variants in clinical genome sequencing: Deployment and interpretation for rare and undiagnosed disease. *Anesth. Analg.* **2018**, *21*, 1121–1130. [CrossRef]
6. Whitford, W.; Lehnert, K.; Snell, R.G.; Jacobsen, J.C. Evaluation of the performance of copy number variant prediction tools for the detection of deletions from whole genome sequencing data. *J. Biomed. Inform.* **2019**, *94*, 103174. [CrossRef] [PubMed]
7. Austin-Tse, C.A.; Jobanputra, V.; Perry, D.L.; Bick, D.; Taft, R.J.; Venner, E.; Gibbs, R.A.; Young, T.; Barnett, S.; Belmont, J.W.; et al. Best practices for the interpretation and reporting of clinical whole genome sequencing. *npj Genom. Med.* **2022**, *7*, 27. [CrossRef]
8. Liu, Z.; Roberts, R.; Mercer, T.R.; Xu, J.; Sedlazeck, F.J.; Tong, W. Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol.* **2022**, *23*, 68. [CrossRef]
9. Coutelier, M.; Holtgrewe, M.; Jäger, M.; Flöttman, R.; Mensah, M.A.; Spielmann, M.; Krawitz, P.; Horn, D.; Beule, D.; Mundlos, S. Combining callers improves the detection of copy number variants from whole-genome sequencing. *Eur. J. Hum. Genet.* **2022**, *30*, 178–186. [CrossRef]
10. Dolzhenko, E.; Bennett, M.F.; Richmond, P.A.; Trost, B.; Chen, S.; van Vugt, J.J.F.A.; Nguyen, C.; Narzisi, G.; Gainullin, V.G.; Gross, A.M.; et al. ExpansionHunter Denovo: A computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol.* **2020**, *21*, 102. [CrossRef]
11. Saunders, C.J.; Miller, N.A.; Soden, S.E.; Dinwiddie, D.L.; Noll, A.; Abu Alnadi, N.; Andraws, N.; Patterson, M.L.; Krivohlavek, L.A.; Fellis, J.; et al. Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units. *Sci. Transl. Med.* **2012**, *4*, 154ra135. [CrossRef] [PubMed]
12. Krantz, I.D.; Medne, L.; Weatherly, J.M.; Wild, K.T.; Biswas, S.; Devkota, B.; Hartman, T.; Brunelli, L.; Fishler, K.P.; Abdul-Rahman, O.; et al. Effect of Whole-Genome Sequencing on the Clinical Management of Acutely Ill Infants with Suspected Genetic Disease: A Randomized Clinical Trial. *JAMA Pediatr.* **2021**, *175*, 1218–1226. [PubMed]
13. Clark, M.M.; Hildreth, A.; Batalov, S.; Ding, Y.; Chowdhury, S.; Watkins, K.; Ellsworth, K.; Camp, B.; Kint, C.I.; Yacoubian, C.; et al. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* **2019**, *11*, aat6177. [CrossRef] [PubMed]
14. Kingsmore, S.F.; Cakici, J.A.; Clark, M.M.; Gaughran, M.; Feddock, M.; Batalov, S.; Bainbridge, M.N.; Carroll, J.; Caylor, S.A.; Clarke, C.; et al. A Randomized, Controlled Trial of the Analytic and Diagnostic Performance of Singleton and Trio, Rapid Genome and Exome Sequencing in Ill Infants. *Am. J. Hum. Genet.* **2019**, *105*, 719–733. [CrossRef] [PubMed]
15. Stranneheim, H.; Lagerstedt-Robinson, K.; Magnusson, M.; Kvarnung, M.; Nilsson, D.; Lesko, N.; Engvall, M.; Anderlid, B.-M.; Arnell, H.; Johansson, C.B.; et al. Integration of whole genome sequencing into a healthcare setting: High diagnostic rates across multiple clinical entities in 3219 rare disease patients. *Genome Med.* **2021**, *13*, 40. [CrossRef] [PubMed]
16. Bowdin, S.; Hayeems, R.; Monfared, N.; Cohn, R.; Meyn, M. The SickKids Genome Clinic: Developing and evaluating a pediatric model for individualized genomic medicine. *Clin. Genet.* **2016**, *89*, 10–19. [CrossRef] [PubMed]
17. Willmen, T.; Willmen, L.; Pankow, A.; Ronicke, S.; Gabriel, H.; Wagner, A.D. Rare diseases: Why is a rapid referral to an expert center so important? *BMC Health Serv. Res.* **2023**, *23*, 904. [CrossRef]
18. Bauskis, A.; Strange, C.; Molster, C.; Fisher, C. The diagnostic odyssey: Insights from parents of children living with an undiagnosed condition. *Orphanet J. Rare Dis.* **2022**, *17*, 233. [CrossRef]
19. Farnaes, L.; Hildreth, A.; Sweeney, N.M.; Clark, M.M.; Chowdhury, S.; Nahas, S.; Cakici, J.A.; Benson, W.; Kaplan, R.H.; Kronick, R.; et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *npj Genom. Med.* **2018**, *3*, 10. [CrossRef]
20. Splinter, K.; Adams, D.R.; Bacino, C.A.; Bellen, H.J.; Bernstein, J.A.; Cheatle-Jarvela, A.M.; Eng, C.M.; Esteves, C.; Gahl, W.A.; Hamid, R.; et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *N. Engl. J. Med.* **2018**, *379*, 2131–2139. [CrossRef]

21.   Petrikin, J.E.; Cakici, J.A.; Clark, M.M.; Willig, L.K.; Sweeney, N.M.; Farrow, E.G.; Saunders, C.J.; Thiffault, I.; Miller, N.A.; Zellmer, L.; et al. The NSIGHT1-randomized controlled trial: Rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *npj Genom. Med.* **2018**, *3*, 6. [CrossRef] [PubMed]

22.   Scocchia, A.; Wigby, K.M.; Masser-Frye, D.; Del Campo, M.; Galarreta, C.I.; Thorpe, E.; McEachern, J.; Robinson, K.; Gross, A.; ICSL Interpretation and Reporting Team; et al. Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphology clinic in Mexico. *npj Genom. Med.* **2019**, *4*, 5. [CrossRef] [PubMed]

23.   Kingsmore, S.F.; Consortium, T.B. Dispatches from Biotech beginning BeginNGS: Rapid newborn genome sequencing to end the diagnostic and therapeutic odyssey. *Am. J. Med. Genet. Part C Semin. Med. Genet.* **2022**, *190*, 243–256. [CrossRef] [PubMed]

24.   Kingsmore, S.F.; Smith, L.D.; Kunard, C.M.; Bainbridge, M.; Batalov, S.; Benson, W.; Blincow, E.; Caylor, S.; Chambers, C.; Del Angel, G.; et al. A genome sequencing system for universal newborn screening, diagnosis, and precision medicine for severe genetic diseases. *Am. J. Hum. Genet.* **2022**, *109*, 1605–1619. [CrossRef] [PubMed]

25.   Berg, J.S.; Agrawal, P.B.; Bailey, D.B., Jr.; Beggs, A.H.; Brenner, S.E.; Brower, A.M.; Cakici, J.A.; Ceyhan-Birsoy, O.; Chan, K.; Chen, F.; et al. Newborn Sequencing in Genomic Medicine and Public Health. *Pediatrics* **2017**, *139*, e20162252. [CrossRef] [PubMed]

26.   Seydel, C. Baby's first genome. *Nat. Biotechnol.* **2022**, *40*, 636–640. [CrossRef] [PubMed]

27.   Dimmock, D.; Caylor, S.; Waldman, B.; Benson, W.; Ashburner, C.; Carmichael, J.L.; Carroll, J.; Cham, E.; Chowdhury, S.; Cleary, J.; et al. Project Baby Bear: Rapid precision care incorporating rWGS in 5 California children's hospitals demonstrates improved clinical outcomes and reduced costs of care. *Am. J. Hum. Genet.* **2021**, *108*, 1231–1238. [CrossRef]

28.   Mestek-Boukhibar, L.; Clement, E.; Jones, W.D.; Drury, S.; Ocaka, L.; Gagunashvili, A.; Stabej, P.L.Q.; Bacchelli, C.; Jani, N.; Rahman, S.; et al. Rapid Paediatric Sequencing (RaPS): Comprehensive real-life workflow for rapid diagnosis of critically ill children. *J. Med. Genet.* **2018**, *55*, 721–728. [CrossRef]

29.   Ding, Y.; Owen, M.; Le, J.; Batalov, S.; Chau, K.; Kwon, Y.H.; Van Der Kraan, L.; Bezares-Orin, Z.; Zhu, Z.; Veeraraghavan, N.; et al. Scalable, high quality, whole genome sequencing from archived, newborn, dried blood spots. *npj Genom. Med.* **2023**, *8*, 5. [CrossRef]

30.   Agrawal, P.; Katragadda, S.; Hariharan, A.K.; Raghavendrachar, V.G.; Agarwal, A.; Dayalu, R.; Awasthy, D.; Sharma, S.C.; Sivasamy, Y.K.; Lakshmana, P.; et al. Validation of whole genome sequencing from dried blood spots. *BMC Med. Genom.* **2021**, *14*, 110. [CrossRef]

31.   Poulsen, J.B.; Lescai, F.; Grove, J.; Bækvad-Hansen, M.; Christiansen, M.; Hagen, C.M.; Maller, J.; Stevens, C.; Li, S.; Li, Q.; et al. High-Quality Exome Sequencing of Whole-Genome Amplified Neonatal Dried Blood Spot DNA. *PLoS ONE* **2016**, *11*, e0153253. [CrossRef] [PubMed]

32.   Hollegaard, M.V.; Grauholm, J.; Nielsen, R.; Grove, J.; Mandrup, S.; Hougaard, D.M. Archived neonatal dried blood spot samples can be used for accurate whole genome and exome-targeted next-generation sequencing. *Mol. Genet. Metab.* **2013**, *110*, 65–72. [CrossRef] [PubMed]

33.   Bassaganyas, L.; Freedman, G.; Vaka, D.; Wan, E.; Lao, R.; Chen, F.; Kvale, M.; Currier, R.J.; Puck, J.M.; Kwok, P.-Y. Whole exome and whole genome sequencing with dried blood spot DNA without whole genome amplification. *Hum. Mutat.* **2018**, *39*, 167–171. [CrossRef] [PubMed]

34.   Roman, T.S.; Crowley, S.B.; Roche, M.I.; Foreman, A.K.M.; O'daniel, J.M.; Seifert, B.A.; Lee, K.; Brandt, A.; Gustafson, C.; DeCristo, D.M.; et al. Genomic Sequencing for Newborn Screening: Results of the NC NEXUS Project. *Am. J. Hum. Genet.* **2020**, *107*, 596–611. [CrossRef] [PubMed]

35.   Burton, B.K.; Grange, D.K.; Milanowski, A.; Vockley, G.; Feillet, F.; Crombez, E.A.; Abadie, V.; Harding, C.O.; Cederbaum, S.; Dobbelaere, D.; et al. The response of patients with phenylketonuria and elevated serum phenylalanine to treatment with oral sapropterin dihydrochloride (6R-tetrahydrobiopterin): A phase II, multicentre, open-label, screening study. *J. Inherit. Metab. Dis.* **2007**, *30*, 700–707. [CrossRef] [PubMed]

36.   Ajay, S.S.; Parker, S.C.; Abaan, H.O.; Fajardo, K.V.F.; Margulies, E.H. Accurate and comprehensive sequencing of personal genomes. *Genome Res.* **2011**, *21*, 1498–1505. [CrossRef]

37.   Bentley, D.R.; Balasubramanian, S.; Swerdlow, H.P.; Smith, G.P.; Milton, J.; Brown, C.G.; Hall, K.P.; Evers, D.J.; Barnes, C.L.; Bignell, H.R.; et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **2008**, *456*, 53–59. [CrossRef] [PubMed]

38.   Koboldt, D.C.; Ding, L.; Mardis, E.R.; Wilson, R.K. Challenges of sequencing human genomes. *Brief. Bioinform.* **2010**, *11*, 484–498. [CrossRef]

39.   Moat, S.J.; George, R.S.; Carling, R.S. Use of Dried Blood Spot Specimens to Monitor Patients with Inherited Metabolic Disorders. *Int. J. Neonatal Screen.* **2020**, *6*, 26. [CrossRef]

40.   Moat, S.J.; Zelek, W.M.; Carne, E.; Ponsford, M.J.; Bramhall, K.; Jones, S.; El-Shanawany, T.; Wise, M.P.; Thomas, A.; George, C.; et al. Development of a high-throughput SARS-CoV-2 antibody testing pathway using dried blood spot specimens. *Ann. Clin. Biochem.* **2021**, *58*, 123–131. [CrossRef]

41.   Robinson, J.T.; Thorvaldsdottir, H.; Turner, D.; Mesirov, J.P. igv.js: An embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* **2022**, *39*, btac830. [CrossRef] [PubMed]

42.   Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef] [PubMed]

43. Aird, D.; Ross, M.G.; Chen, W.-S.; Danielsson, M.; Fennell, T.; Russ, C.; Jaffe, D.B.; Nusbaum, C.; Gnirke, A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **2011**, *12*, R18. [CrossRef] [PubMed]

44. Kozarewa, I.; Ning, Z.; Quail, M.A.; Sanders, M.J.; Berriman, M.; Turner, D.J. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* **2009**, *6*, 291–295. [CrossRef] [PubMed]

45. Fang, H.; Wu, Y.; Narzisi, G.; ORawe, J.A.; Barrón, L.T.; Rosenbaum, J.; Ronemus, M.; Iossifov, I.; Schatz, M.C.; Lyon, G.J. Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med.* **2014**, *6*, 89. [CrossRef] [PubMed]

46. Mallick, S.; Li, H.; Lipson, M.; Mathieson, I.; Gymrek, M.; Racimo, F.; Zhao, M.; Chennagiri, N.; Nordenfelt, S.; Tandon, A.; et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **2016**, *538*, 201–206. [CrossRef] [PubMed]

47. Abul-Husn, N.S.; Soper, E.R.; Braganza, G.T.; Rodriguez, J.E.; Zeid, N.; Cullina, S.; Bobo, D.; Moscati, A.; Merkelson, A.; Loos, R.J.F.; et al. Implementing genomic screening in diverse populations. *Genome Med.* **2021**, *13*, 17. [CrossRef]

48. Abul-Husn, N.S.; Team, C.G.; Soper, E.; Odgis, J.A.; Cullina, S.; Bobo, D.; Moscati, A.; Rodriguez, J.E.; Loos, R.; Cho, J.H.; et al. Exome sequencing reveals a high prevalence of BRCA1 and BRCA2 founder variants in a diverse population-based biobank. *Genome Med.* **2019**, *12*, 2. [CrossRef]

49. Brnich, S.E.; Abou Tayoun, A.N.; Couch, F.J.; Cutting, G.R.; Greenblatt, M.S.; Heinen, C.D.; Kanavy, D.M.; Luo, X.; McNulty, S.M.; Starita, L.M.; et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* **2019**, *12*, 3. [CrossRef]

50. Rehm, H.L.; Fowler, D.M. Keeping up with the genomes: Scaling genomic variant interpretation. *Genome Med.* **2019**, *12*, 5. [CrossRef]

51. Gao, H.; Hamp, T.; Ede, J.; Schraiber, J.G.; McRae, J.; Singer-Berk, M.; Yang, Y.; Dietrich, A.S.D.; Fiziev, P.P.; Kuderna, L.F.K.; et al. The landscape of tolerated genetic variation in humans and primates. *Science* **2023**, *380*, eabn8153. [CrossRef]

52. Fiziev, P.P.; McRae, J.; Ulirsch, J.C.; Dron, J.S.; Hamp, T.; Yang, Y.; Wainschtein, P.; Ni, Z.; Schraiber, J.G.; Gao, H.; et al. Rare penetrant mutations confer severe risk of common diseases. *Science* **2023**, *380*, eabo1131. [CrossRef]