Title

Whole exome and whole genome sequencing with dried blood spot DNA without whole genome amplification

Journal

Human Mutation, 39(1)

ISSN

1059-7794

Authors

Bassaganyas, Laia
Freedman, George
Vaka, Dedeepya
et al.

Publication Date

2018

DOI

10.1002/humu.23356

Peer reviewed

HHS Public Access

Author manuscript

Hum Mutat. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

Hum Mutat. 2018 January ; 39(1): 167–171. doi:10.1002/humu.23356.

Whole exome and whole genome sequencing with dried blood spot DNA without whole genome amplification

Laia Bassaganyas[1,*], **George Freedman**[2,*], **Dedeepya Vaka**[3], **Eunice Wan**[3], **Richard Lao**[3], **Flavia Chen**[3], **Mark Kvale**[3], **Robert J. Currier**[4], **Jennifer M. Puck**[2,3], and **Pui-Yan Kwok**[1,3,4]

[1]Cardiovascular Research Institute, University of California, San Francisco, CA

[2]Department of Pediatrics, University of California, San Francisco, CA

[3]Institute for Human Genetics, University of California, San Francisco, CA

[4]Genetic Disease Screening Program, California Department of Public Health, Richmond, CA

[5]Department of Dermatology, University of California, San Francisco, CA

Abstract

Newborn screening for rare conditions is performed in all 50 states in the USA. We have partnered with the California Department of Public Health Genetic Disease Laboratory to determine whether sufficient DNA can be extracted from archived dried blood spots for next generation sequencing in the hopes that next generation sequencing can play a role in newborn screening. We optimized the DNA extraction and sequencing library preparation protocols for residual infant dried blood spots archived over 20 years ago and successfully obtained acceptable whole exome and whole genome sequencing data. This sequencing study using dried blood spot DNA without whole genome amplification prior to sequencing library preparation provides evidence that properly stored residual newborn dried blood spots are a satisfactory source of DNA for genetic studies.

Keywords

Newborn screening; Dried blood spot DNA preparation; Whole exome sequencing; WES; Whole genome sequencing. WGS

Introduction

Newborn screening (NBS) for rare conditions that are not clinically apparent, but that require immediate treatment, is an essential public health program in all 50 states in the USA. The falling cost of next generation sequencing provides an opportunity to ask whether DNA sequencing might become a cost-effective tool to be integrated into newborn screening. To test whether it is possible, and if so whether advantageous, to perform whole exome sequencing (WES) or whole genome sequencing (WGS) as part of NBS, we have partnered with the California Department of Public Health (CDPH) Genetic Disease

Corresponding author information: Pui-Yan Kwok, MD, PhD, University of California, San Francisco, 555 Mission Bay Blvd South, MC-3118, San Francisco, CA 94143, pui.kwok@ucsf.edu, Phone: 415-514-3802, Fax: 415-514-1172.

*Co-first authors

Laboratory to sequence DNA from archived dried blood spots (DBS) of individuals as a feasibility pilot.

California began statewide newborn screening in 1980, collecting DBS specimens from essentially all newborns since that time. After the completion of newborn screening, residual specimens were archived in freezers at −20°C in containers with desiccant. Between 1980 and 1982, the California NBS program was managed without computerized records linking individual identifiers with the residual DBS specimens. Consequently, those specimens had no associated patient information at all and were of only limited value for research. The California Biobank Program removed these specimens from the inventory, making a small number of them available to this project as anonymous specimens to evaluate protocols for DNA extraction and sequencing from DBS.

DBS specimens have been made available for approved research projects through the California Biobank Program. Additional DBS resources may be found in the Virtual Repository of Dried Blood Spots (VRDBS) maintained by the Newborn Screening Translational Research Network (NBSTRN). As a reflection of whole-population screening, archived residual infant DBS specimens represent an unparalleled resource for investigation of conditions affecting the newborn, particularly rare genetic disorders, such as the inborn errors of metabolism.

While DNA extracted from DBS has been successfully sequenced on next-generation sequencers in the past for WES and WGS, previous protocols required whole genome amplification [Hollegaard et al., 2013; Poulsen et al., 2016]. To test the hypothesis that properly stored DBSs, even ones archived decades ago, can yield DNA of sufficient quantity and quality for next-generation sequencing, we conducted three sets of studies to prepare DBS DNA for next-generation sequencing without whole genome amplification. In the first study, we established that acceptable WES data could be obtained without whole genome amplification, although the sequencing library had smaller than expected insert size. In the second study, we optimized the sequencing library preparation protocol to produce large insert size for WES sequencing. In the third study, we extended the protocol for WGS. Our finding shows that sufficient amounts of DNA can be extracted from old DBSs to produce acceptable WES and WGS data.

Materials and Methods

DNA extraction

After completion of mandated state newborn screening, residual DBS on Grade 903 filter paper were archived (stacked together without any protective barriers) and stored desiccated at −20°C. After obtaining a waiver of consent from the CDPH IRB for study of de-identified cases and controls, samples were located, brought to room temperature, and punched to remove two 3.2-mm discs, each corresponding to approximately 3 μL of blood. The discs were deposited into wells of 96-well deep plates (Corning 3960, Corning, NY, USA). A clean blank filter card was punched twice between each sample punch to minimize cross-contamination. Each plate included an internal positive control (punch from a DBS prepared from anonymous cord blood) and a negative punch from a blank filter card.

The DBS discs were digested in 200 μL detergent solution with proteinase K (0.5 mg/mL) for 16 h at 65°C with 240 rpm shaking, after which the plates were spun at 370×g for 5 min. Genomic DNA was isolated using an AutoGenprep 965 robot (Autogen, Inc., Holliston, MA, USA) with reagents supplied by the company. The standard tissue DNA extraction protocol was followed with an extra 70% ethanol pellet wash. Briefly, punches were incubated in 50 μL RNA digestion solution and organic extraction with 180 μL phenol/chloroform, with agitation followed by phase separation by centrifugation (4000 rpm, 20 min). The resulting aqueous phase, 390 μL, was aspirated into fresh 96-well plates for DNA precipitation with 230 μL isopropanol. After centrifugation, DNA pellets were washed twice with 450 μL 70% ethanol. Pellets were air dried for 60 min and then hydrated overnight at RT in 50 μL of 10% TE buffer. DNA from the duplicate punches was pooled to yield 100 μL, of which 5 μL was run on a 0.8% agarose TAE gel for 2 h. DNA size, integrity and quantity were estimated by densitometry with a ChemiDoc Touch imaging system (Bio-Rad, Hercules, CA, USA) using Image Lab software.

DNA shearing

DNA shearing was done on the Covaris S2 Sonicator (Woburn, MA, USA) according to the manufacturer's instructions. All samples were sheared using the C1 program (intensity-5; duty cycle-10%; cycles per burst-200) at various treatment times to achieve optimal insert size. For insert size ~250 bp, the treatment time of 90s (8 WES samples 01-02 to 01-09) or 120s (10 WGS samples M037, M068, M69, M75, M089, M097, M157, M179, U036, U040) was used. For insert size of ~450 bp (4 WES samples DB1a, DB1b, DB2a, DB2b), the treatment time of 30s was used.

Sequencing library preparation

For WES and WGS library preparation, an average of 55 ng and 5 ng of input DNA were used, respectively. The sequencing libraries were made using the KAPA HTP or Hyper Library Preparation Kit according to manufacturer's instructions (Kapa Biosystems, Wilmington, MA, USA). For libraries optimized for larger inserts sizes, the ligation products were treated with the KAPA dual-SPRI size selection washing procedure. Samples DB1a and DB2a were subjected to the 0.4×–0.6× wash, while DB2a and DB2b were subjected to the 0.5×–0.7× wash.

Whole exome capture

For WES experiments, the pooled libraries were incubated for 72 hours with the capture probes (SeqCap EZ Human Exome Kit v3.0, Roche Nimblegen, Madison, WI, USA) according to the manufacturer's instructions.

DNA sequencing

WES sequencing for the initial 8 samples (01-02 to 01-09) was performed on the HiSeq 2500 (Illumina, La Jolla, CA, USA) with the paired-end 100 protocol, while the remaining 4 WES samples and the 10 WGS samples were sequenced on the HiSeq 4000, using the paired-end 150 protocol.

DNA sequencing QC analysis

Raw image files were analyzed and converted to base calls by real-time analysis using the default settings recommended by the manufacturer for data from the HiSeq 2500 or Hiseq 4000. Real-time analysis output base call files (*.bcl) were converted to FASTQ files with consensus assessment of sequence and variation (CASAVA pipeline, version 1.8, Illumina and version 2.17) using computer servers of the UCSF Institute for Human Genetics (IHG). CASAVA also de-multiplexed the data to obtain FASTQ files for individual samples. Initial sample and library quality control (QC) metrics produced from the Illumina SAV viewer .bin were evaluated using an in-house custom script which parses the bin files. FastQC (version 0.11.5, Babraham Institute, Cambridge, UK; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) reports were generated and manually inspected to ensure cycle uniformity and run quality. Blue Collar Bioinformatics (Bcbio version 1.0.0a0-914e368; https://github.com/chapmanb/bcbio-nextgen) was used to run the QC pipeline. Bcbio-nextgen provides best-practice pipelines for automated analysis of high throughput sequencing data (http://bcbio-nextgen.readthedocs.io/en/latest/index.html). Processing of all data followed Genome Analysis Tool Kit (GATK Unified GenoTyper version 2015.1.1-3.4.46-0-ga8e1d99, Broad Institute, Cambridge, MA, USA) best practices (https://software.broadinstitute.org/gatk/best-practices/bp_3step.php?case=GermShortWGS, https://gatkforums.broadinstitute.org/gatk/discussion/2804/howto-call-variants-with-the-unifiedgenotyper). Specifically, Burrows-Wheeler Aligner (BWA version 0.7.15) [Li & Durbin, 2010] software was used to map reads in paired-end mode to the University of California, Santa Cruz (UCSC) canonical-ordering build of the human reference genome (hg19 version, Genome Reference Consortium GRCh37, version 26). The BAM files obtained were used for the subsequent steps in the analysis pipeline. Picard (version 2.5.0, SourceForge, Mountain View, CA, USA; https://broadinstitute.github.io/picard/) was used to sort mapped reads into coordinate order and to ensure all mate-pair information was properly updated. Picard was also used to mark duplicate reads and low-quality reads, defined as reads with low mapping quality score (MAQ, SourceForge; http://samtools.sourceforge.net/), that were unpaired or unmapped, or that failed a platform/vendor quality check. The GATK tool [McKenna *et al.*, 2010] was then used for recalibration of the base quality score and for insertion/deletion (INDEL) realignment, with Picard used again for marking duplicates, to obtain analysis-ready BAM files. Multiple QC programs (Samtools version 1.3.1, Picard, Qualimap version 2.2; http://qualimap.bioinfo.cipf.es/) were run through bcbio-nextgen on the BAM files to obtain information about uniquely mapped reads, % of duplicates, % of reads aligned to the target, mean target coverage, zero coverage regions, and mean insert sizes.

Results

High molecular weight DNA was recovered from 100% of the samples. DNA extraction yields from a representative selection of DBS were highly variable but normally distributed (Figure 1), with mean recovery from 2 punches of 370 ng/μL (Range 150–920 ng/μL, SD 190 ng/μL).

The DNA samples extracted from dried blood spots were used in 3 sets of pilot experiments. In the first experiment, we aimed to determine if old, archived, but properly stored DBS could yield DNA samples of sufficient quantity and quality for next-generation sequencing without whole genome amplification. Eight DNA samples extracted from DBS obtained in 1980 were sheared by sonication to produce DNA fragments ~200 bp in length, according to the Covaris protocol; however, actual DNA length was observed to be only ~150 bp. Shearing time was subsequently reduced to account for presumed single strand DNA nicking in the archived DBS, given the lack of evidence of degraded DNA in double strand DNA gels. The sheared DNA fragments were repaired, ligated to adaptors, and amplified to produce the sequencing library before incubation with whole exome capture probes to yield the WES library. The samples were sequenced in 4-plex (4 samples per lane) on the Illumina HiSeq 2500 sequencer in high output mode using the paired-end 100 bp protocol. The results are shown in Table 1 (samples 01–02 to 01–09). The average coverage of the targeted bases was 69× (range 56× to 87×), with <0.01% of the target region not covered and 96.1% (range 95.3% – 98.1%) of the target region with ≥10× coverage. This protocol has been used to generate WES data for ~1500 DBS DNA samples to be examined for gene mutations corresponding to their associated tandem mass spectrometry results for NBS metabolic disorders (unpublished data). The average coverage of target bases for the first 188 out of 1500 samples sequenced was 60× with only 2.6% of the target bases not covered (data not shown). Heterozygous variants are identified readily at this level of overall coverage.

In the second experiment, we aimed to produce longer DNA fragments to take advantage of the new paired-end 150 bp sequencing protocol offered on the HiSeq 4000 sequencer. Four DNA samples (DB1a, DB1b, DB2a, and DB2b), also extracted from DBS obtained in 1980, were sheared for an even shorter time of 30s to produce >400 bp fragments. After sample library preparation and exome capture, the samples were sequenced on the HiSeq 4000 sequencer (4-plex, paired-end 150 bp protocol). As shown in Table 1, the average coverage of the targeted bases is 91× (range 74× to 114×), with <0.01% of the target region not covered and 97.2% (range 97.0% – 97.4%) of the target region with ≥10× coverage. Alternative read fraction test was used to assess the degree of contamination found in the samples and no evidence of cross contamination was seen in the archival DBS-derived DNA.

In the third experiment, we explored the feasibility of WGS with DBS derived DNA samples. Ten residual DBS DNA samples (archived from 1996–2013) used in a WES study were sheared and prepared for WGS. Despite the low amount input DNA for shearing (~5 ng) the resultant libraries were sequenced on the HiSeq 4000 sequencer (paired-end 150 bp protocol, one sample per lane) to evaluate the lower limit of input DNA for WGS. Because the insert size was ~250 bp, and the starting DNA amounts were suboptimal, the mean exome target coverage was low (20×, range 11× – 28×). Nevertheless, 95.3% of the exome target region had ≥10× mean coverage for 6 of the samples that achieved ≥20× mean exome target coverage. The results are promising but far from perfect. We anticipate that with the protocol that yields sequencing libraries with larger inserts and sequencing to a higher depth (30× mean exome target coverage), the data quality will be more acceptable.

Discussion

Dried blood spots stored in repositories around the world are potential resources for genetic and genomic analysis. This is to our knowledge the first study to use DBS DNA without whole genome amplification prior to WES and WGS sequencing library preparation. Our success was likely due to 4 factors. First, the California DBS were preserved at −20° C under desiccation, conditions favoring preservation of intact DNA. Second, although DNA quantity recovered from DBS was highly variable [Lawson et al., 2016; Hall et al., 2015], the automated DNA extraction protocol on average yielded more high-molecular-weight DNA than has been produced by manual DNA extraction protocols. Third, newer sequencing library and exome capture preparation kits require only small amounts of input DNA. Fourth, our shorter DNA shearing time compensated for the shorter than expected fragment size observed, possibly attributable to single strand nicking of DNA stored in the form of DBS.

In our study input DNA amount affected the proportion of duplicate reads, and, by extension, exome target coverage. However, even very low amounts of genomic DNA, when subjected to WGS without incurring the losses attendant to exome capture, could generate acceptable exome coverage. Indeed, the two WGS samples with >25× exome target coverage yielded the same level of exome coverage (>96% of exome targets with at least 10× coverage) as WES samples with >60× exome target coverage. While our results are promising, it is important that any future WGS study generates at least 30× overall coverage of exome target regions in order to support heterozygous variant calls with confidence.

Based on these pilot study results, we have performed WES on a larger set of ~1500 DBS DNA samples. Analysis for concordance with metabolic screening in the CDPH newborn screening program is ongoing.

Our pilot study described here demonstrates that appropriately stored NBS DBS DNA is a good source of material for genomic studies in diverse populations. It is remarkable that no sign of cross contamination is detected even though the DBS cards are kept with any barriers between them. Among the 18 million DBS specimens in the California Biobank repository, there is substantial representation of many ethnic groups, as well as an increasing representation of individuals with backgrounds of mixed ethnicities. Although all of the disorders that are targets for NBS are rare, the ethnic distribution of disorders varies greatly [Feuchtbaum et al., 2012]. Appropriate assessment of DNA variants in the context of NBS for rare disorders needs to be based on a deep understanding of the differences of variant frequencies in populations of different ethnicities. As an appropriate model, recent work on cystic fibrosis examined the fifty most common CFTR variants by ethnicity, and found that 20 of the most common variants in Hispanic CF patients were not shared with other ethnicities [Feuchtbaum et al., 2012]. This restriction of common variants to one ethnicity is reproduced in Asian and African American populations as well [Schrijver et al., 2016]. The ability to produce high quality exome and genome sequences from newborn DBS, when achieved, offers an unparalleled opportunity to advance our understanding of rare disease associated and polymorphic variants in non-northern European populations, allowing for

improved interpretation of DNA based screening and diagnosis in populations of diverse ancestry.

Acknowledgments

Funding Acknowledgement and Disclaimer: Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award U19HD077627. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Feuchtbaum L, Carter J, Dowray S, Currier RJ, Lorey F. Birth prevalence of disorders detectable by newborn screening by race/ethnicity. Genet Med. 2012 Nov; 14(11):937–45. [PubMed: 22766612]

Hall, Elizabeth M., Flores, Sharon R., De Jesús, Víctor R. Influence of Hematocrit and Total-Spot Volume on Performance Characteristics of Dried Blood Spots for Newborn Screening. Int J Neonatal Screen. 2015 Aug 21.1:69–78. [PubMed: 28868497]

Hollegaard MV, Grauholm J, Nielsen R, Grove J, Mandrup S, Hougaard DM. Archived neonatal dried blood spot samples can be used for accurate whole genome and exome-targeted next-generation sequencing. Mol Genet Metab. 2013 Sep-Oct;110(1–2):65–72. [PubMed: 23830478]

Lawson AJ, Bernstone L, Hall SK. Newborn screening blood spot analysis in the UK: influence of spot size, punch location and haematocrit. J Med Screen. 2016 Mar; 23(1):7–16. [PubMed: 26113437]

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Mar 1; 26(5):589–95. [PubMed: 20080505]

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep; 20(9):1297–303. [PubMed: 20644199]

Poulsen JB, Lescai F, Grove J, Bækvad-Hansen M, Christiansen M, Hagen CM, Maller J, Stevens C, Li S, Li Q, Sun J, Wang J, Nordentoft M, Werge TM, Mortensen PB, Børglum AD, Daly M, Hougaard DM, Bybjerg-Grauholm J, Hollegaard MV. High-Quality Exome Sequencing of Whole-Genome Amplified Neonatal Dried Blood Spot DNA. PLoS One. 2016 Apr 18.11(4):e0153253. [PubMed: 27089011]

Schrijver I, Pique L, Graham S, Pearl M, Cherry A, Kharrazi M. The Spectrum of CFTR Variants in Nonwhite Cystic Fibrosis Patients: Implications for Molecular Diagnostic Testing. J Mol Diagn. 2016 Jan; 18(1):39–50. [PubMed: 26708955]

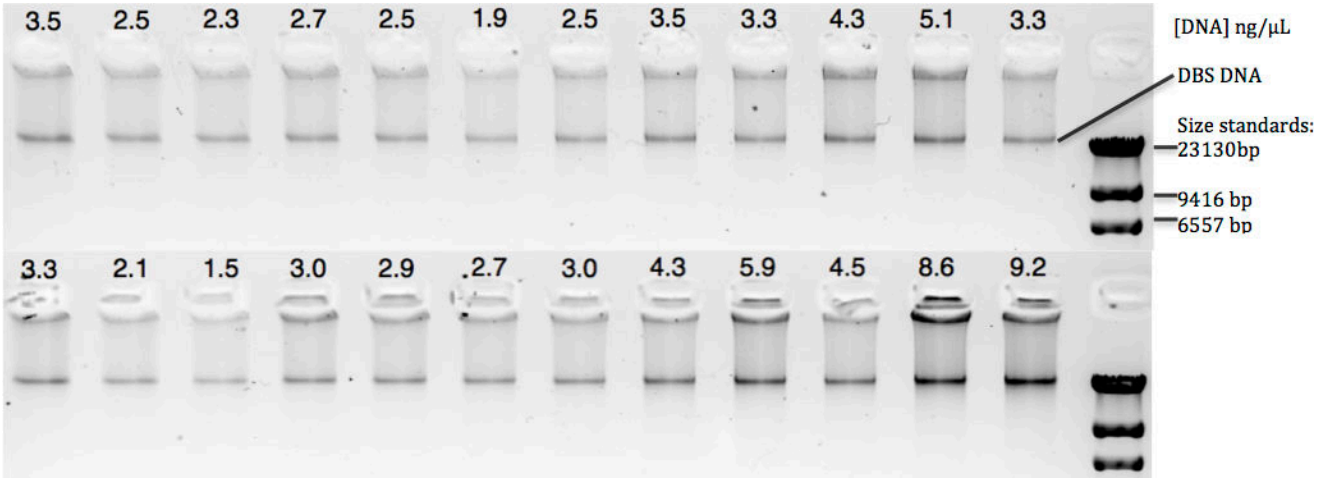

Figure 1.

Agarose gel image of DNA extracted from 24 representative DBS samples with a DBS size standard (right lane). The sample DNA concentrations, estimated by densitometry, are shown above each lane. Tight DNA bands of high molecular weight were observed for every sample, with no observable DNA degradation.

Table 1

Quality control data of next generation sequencing experiments with DNA extracted from dried blood spots

Sample #	Expt. Type/#	HiSeq Model	Protocol	Mean Coverage on Target	% Exome Target with No coverage	% Exome Target with 10× Coverage	Average Insert Size (bp)
01-02	WES/1	2500	PE100	60×	0.01%	95.30%	218
01-03	WES/1	2500	PE100	70×	0.02%	95.80%	218
01-04	WES/1	2500	PE100	56×	0.01%	95.30%	213
01-05	WES/1	2500	PE100	76×	0.01%	96.00%	219
01-06	WES/1	2500	PE100	56×	0.02%	95.50%	215
01-07	WES/1	2500	PE100	87×	0.01%	96.30%	221
01-08	WES/1	2500	PE100	73×	0.01%	96.20%	222
01-09	WES/1	2500	PE100	75×	0.01%	98.10%	212
DB1a	WES/2	4000	PE150	74×	0.01%	97.10%	481
DB1b	WES/2	4000	PE150	92×	0.01%	97.20%	423
DB2a	WES/2	4000	PE150	84×	0.01%	97.00%	460
DB2b	WES/2	4000	PE150	114×	0.01%	97.40%	417
M037	WGS/3	4000	PE150	12×	1.90%	60.60%	251
M068	WGS/3	4000	PE150	11×	1.90%	65.00%	251
M069	WGS/3	4000	PE150	16×	1.60%	79.40%	269
M075	WGS/3	4000	PE150	17×	1.60%	88.90%	248
M089	WGS/3	4000	PE150	21×	1.50%	94.30%	267
M091	WGS/3	4000	PE150	20×	1.50%	93.50%	267
M157	WGS/3	4000	PE150	23×	1.50%	95.20%	248
M179	WGS/3	4000	PE150	23×	1.70%	95.80%	260
U036	WGS/3	4000	PE150	26×	1.40%	96.10%	298
U040	WGS/3	4000	PE150	28×	1.60%	96.70%	265