Clinical grade whole genome sequencing data from various samples types: dried blood spot, whole blood, saliva and tumor tissue

Christin Collins[1], C. Alexander Valencia[1], Gerald Irzyk[1], Zeqiang Ma[1], Edward Szekeres[1], Zdenek Markovic[1], Suresh Shenoy[1], Dhara Shah[1], Yang Wang[1], Alice Tanner[1], Madhuri Hegde[1,2]
[1]PerkinElmer Genomics, [2]Emory University

ABSTRACT

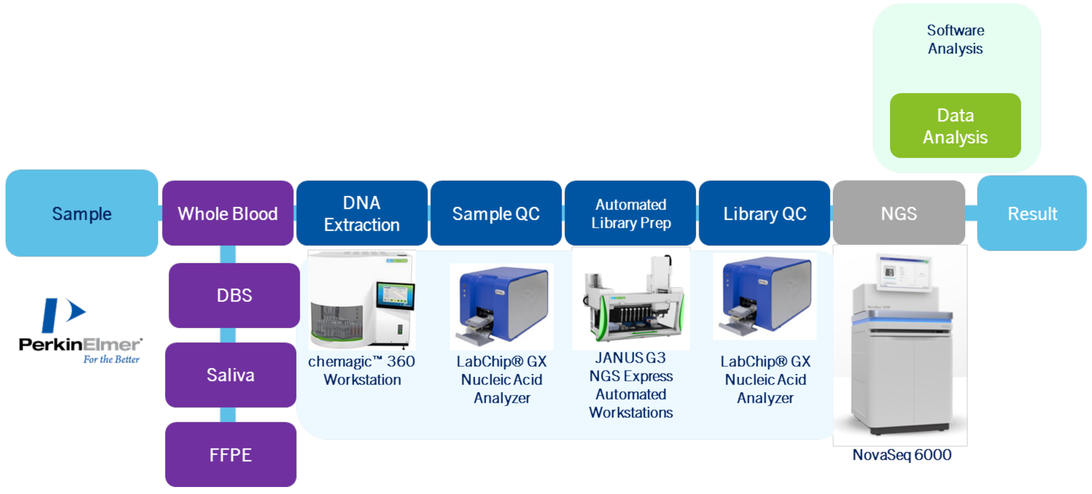
Whole genome sequencing (WGS) is becoming increasingly attractive as an alternative to other molecular methods, due to its broader coverage and decreasing cost as well as higher sensitivity for the detection of variants. Providing sample type options, dried blood spots (DBS), saliva, whole blood (WB, EDTA) and tumor to patients may increase the participation in genetic testing. The objective of this study was to assess whether a variety of sample types provide clinical grade whole genome data by comparing the global metrics and performance parameters. DNA was extracted from DBS, Saliva (Oragene® Dx TM), and whole blood using a chemagic 360 instrument and the average yield was 2 µg/600 µL for saliva and 2 µg/200 µL for whole blood (WB). DNA from tumor was extracted using the QIAmp DNA FFPE Tissue Kit with a yield of 637 ng from 10 curls. For DBS, two 4.75 mm2 punches were used for DNA extraction, which gave a yield of 329 ng ± 187 ng. In the experimental design, we sequenced the whole genome without DNA amplification using the KAPA library kit and 100 ng of input gDNA from three healthy adult individuals that provided DBS, saliva, and WB, along with control DNA (NA12878) and a tumor sample, on the NovaSeq 6000 using S2 flowcells. The data were aligned to the reference GRCh Build 37 (hg19) and variants were called using the Edico Dragen analysis pipeline. The average coverage was 41.16X, 58.67X, 43.83X and 44X for DBS, saliva, WB and tumor, respectively. The coverage of the genome at >10X (DBS 90.65X, saliva 91.04X, WB 90.90X and tumor 90.00X), total number of SNPs (DBS 4,048,341, saliva 4,041,393, WB 4,025,813 and tumor 4,161,595), total number indels (DBS 912,534, saliva 934,077, WB 914,920 and tumor 945,591), number of heterozygous (DBS 3,148,678, saliva 3,152,630, WB 3,120,487 and tumor 3,389,454) and homozygous calls (DBS 1,812,196, saliva 1,822,840, WB 1,820,246 and tumor 1,717,732) were similar for the four sample types. Based on the WGS of NA12878 compared to genome in a bottle (GIAB), the accuracy was calculated to be 98.90% by including all SNPs (3,195,810). The precision for the four sample types was also very close, specifically, 99.3% for DBS, 99.4% saliva, 99.4% for WB and 97.0% for tumor. The sensitivity, defined as the low coverage regions in genes associated with disease, values for DBS, saliva and WB were alike, namely, 99.07% (585 low coverage exons), 99.30% (444 low coverage exons) and 99.17% (533 low coverage exons), respectively. Similarly, the specificity values were comparable for the three sample types (DBS 99.30%, saliva 99.40% and WB 99.37%). This evidence supports that properly stored DBS, saliva, blood and tumor were all satisfactory sources of DNA for WGS studies using Illumina NovaSeq technology. The viability of DBS as a source of quality DNA for opens up the possibility that next-generation sequencing (single genes, gene panels, exome and genome) can play a role in newborn sequencing.

INTRODUCTION

- Dried blood spot samples (DBS) have been collected and stored for decades as part of newborn screening programs worldwide
- Newborn Screening Biobanks are of immense value in medical studies, for example, to examine the genetics of various disorders
- Due to the low yield of gDNA from DBS, previous studies employed whole genome amplification (WGA) prior to whole genome sequencing (WGS) and whole exome sequencing (WES)
- **CHALLENGE**: Introduction of de novo variants due to the amplification process
- **APPROACH:** PCR-free WGS was performed by using the KAPA library construction kit and the latest sequencing platform, Illumina NovaSeq 6000 from a number of samples types, namely, whole blood versus DBS, saliva and FFPE

METHODS

WGS sequencing workflow



RESULTS

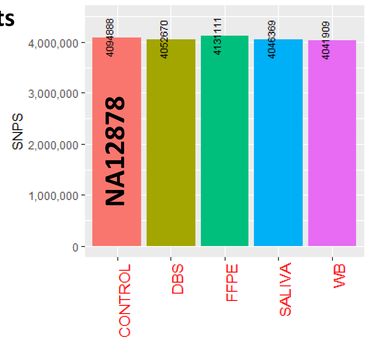
Similar: coverage % at 10X/20X, # of heterozygous/homozygous calls, # of SNPs across sample type; Variable: number of INDELs

Coverage % at 10X/20X, Number (#) of heterozygous and homozygous variants

	Sample Name	Average Coverage	Coverage > 10X (%)	Coverage > 20X (%)	Heterozygous	Homozygous
Sample 1	DBS	44.26	90.5	89.76	3,179,137	1,792,890
	SALIVA	54.1	90.7	90.24	3,173,874	1,799,573
	WB	48.45	90.68	90.12	3,165,648	1,797,590
	Average	**48.94**	**90.63**	**90.04**	**3,172,886**	**1,796,684**
Sample 2	DBS	39.28	90.73	86.32	3,116,181	1,822,634
	SALIVA	65.33	91.29	90.62	3,151,239	1,831,045
	WB	40.81	91	87.05	3,089,832	1,827,476
	Average	**48.47**	**91.01**	**88.00**	**3,119,084**	**1,827,052**
Sample 3	DBS	39.95	90.72	86.47	3,150,717	1,821,066
	SALIVA	56.59	91.13	88.76	3,132,777	1,837,904
	WB	42.22	91.03	87.58	3,105,983	1,835,674
	Average	**46.25**	**90.96**	**87.60**	**3,129,826**	**1,831,548**
Sample 4	FFPE	43.88	90.42	86.66	3,389,454	1,717,732
	WB	30.12	90.46	87.73	3,206,045	1,742,090
	Average	**37.00**	**90.44**	**87.20**	**3,297,750**	**1,729,911**

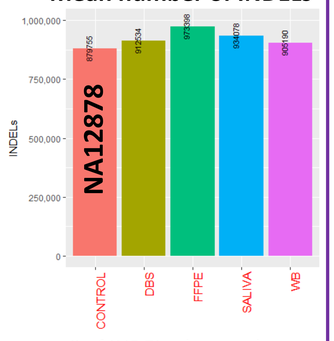
- Similar coverage % and number of heterozygous/homozygous variants within the same individual across sample types

Mean number of SNPs



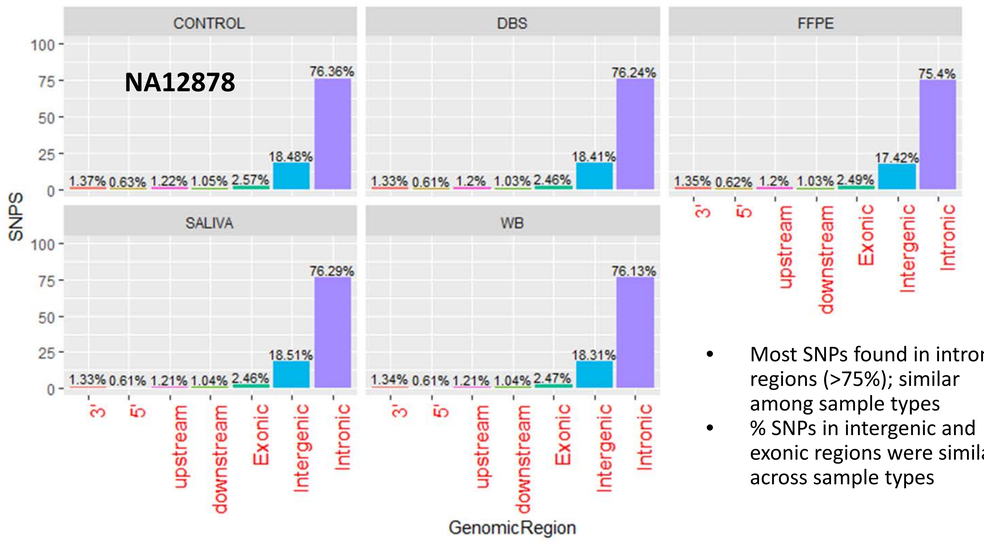
- # of SNPs similar across various sample types
- # of SNPs from all sample types comparable to control (NA12878) sample

Mean number of INDELs



- # of INDELs detected was variable across sample types
- Accuracy of INDELs is lower
- FFPE showed highest # of INDELs

Similar genomic region SNP distribution by sample type



- Most SNPs found in intronic regions (>75%); similar among sample types
- % SNPs in intergenic and exonic regions were similar across sample types

DISCUSSION/CONCLUSIONS

- Global statistics were similar across sample types
  - Coverage % at 10X/20X, # of heterozygous/homozygous variants, # of SNPs/INDELs across sample type
  - Similar SNP distribution by sample type
  - INDELs were shown to be more variable – INDEL calls are less accurate
- Evidence supports that properly stored DBS, saliva, blood and tumor were all satisfactory sources of DNA for WGS studies using Illumina NovaSeq technology
- The viability of DBS as a source of quality DNA opens up the possibility that next-generation sequencing (single genes, gene panels, exomes and genomes) can play a role in newborn sequencing