Saliva samples collected and stabilized with Oragene®•DNA are a reliable source of DNA for next generation sequencing

Rafal M. Iwasiow
DNA Genotek Inc., Ottawa, Canada

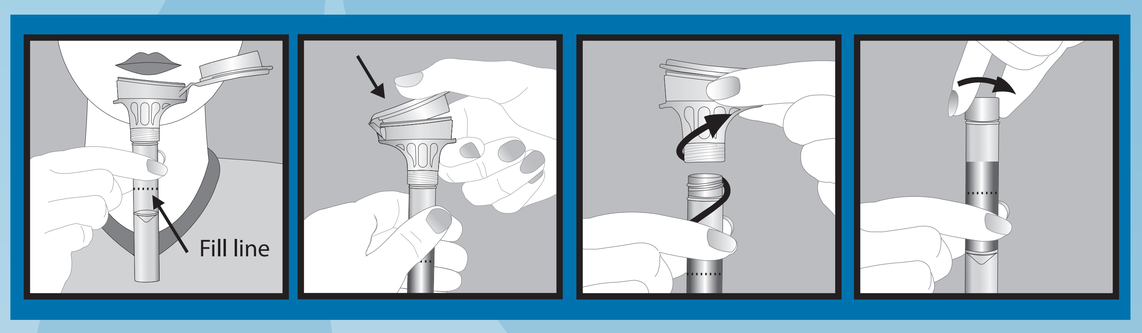
Introduction

In recent years we have observed significant advancements in next generation sequencing, decreasing overall costs and greater acceptance of targeted capture technologies. Together these factors are enabling the use of next generation sequencing for clinical applications. It is quickly becoming more practical and economical to process large numbers of samples for the purpose of genetic analysis. Such analysis requires the collection of reliable, high quality samples. Saliva collected with Oragene•DNA provides a non-invasive alternative method to blood samples for collecting large amounts of high quality genomic DNA that is suitable for next generation sequencing applications. Oragene•DNA devices contain a stabilizing reagent that ensures the sample is of high quality and allows long term storage at ambient temperature. In this study we investigated DNA extracted from saliva collected using Oragene•DNA for use with Agilent SureSelect Human All Exon Kit and paired-end sequencing on the Illumina® Genome Analyzer II.

To evaluate the performance of saliva samples collected using Oragene•DNA we compared the data against DNA from blood samples collected from the same individuals. Using the Agilent SureSelect Human All Exon Kit to capture regions totaling 38Mb = 1.22% of human genomic regions we demonstrate the ability to enrich either blood or saliva DNA samples to investigate genetic variations using next generation sequencing technologies.

Materials and methods

Collection of saliva using Oragene•DNA

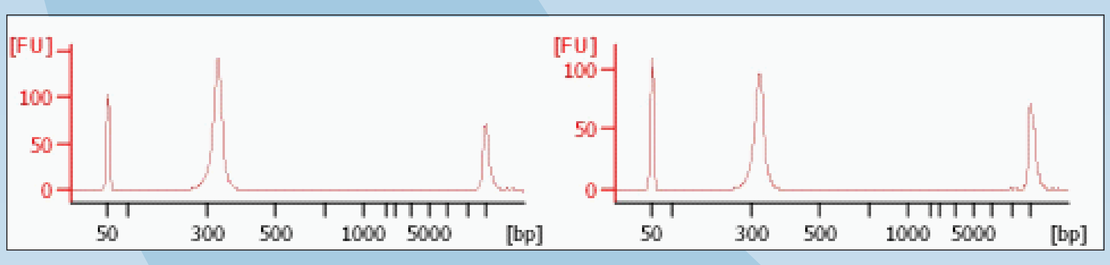


- Saliva was collected according to the instructions provided in the kit.
- Saliva samples were collected from 3 donors.
- Oragene•DNA/saliva samples were purified according to DNA Genotek protocol PD-PR-006.
- DNA was quantified using the Invitrogen Picogreen Quanti-it kit.

- Blood samples were collected from the same 3 donors that donated saliva.
- 8 mL of whole blood was collected using EDTA tubes.
- Blood samples were centrifuged and the buffy coat was collected.
- DNA from buffy coat was purified using the Qiagen Blood mini kit.

- DNA samples were enriched for the human exome using the Agilent SureSelect All Exon kit.
- Samples were sequenced using paired-end sequencing on the Illumina Genome Analyzer II.
- One lane per sample was used.

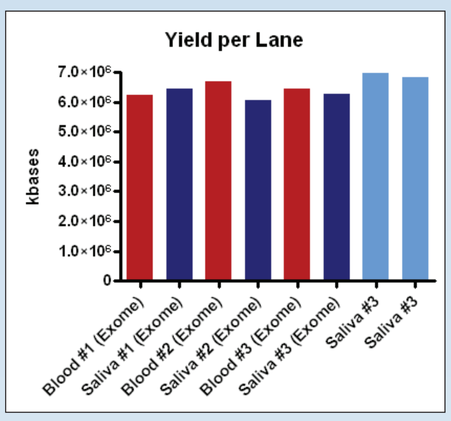
Bioanalyzer traces of Exome-enriched libraries

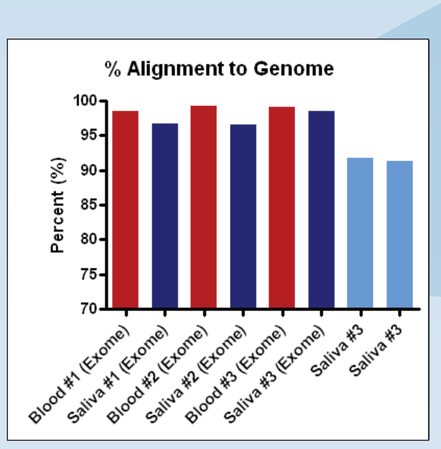


Saliva #1 Blood #1

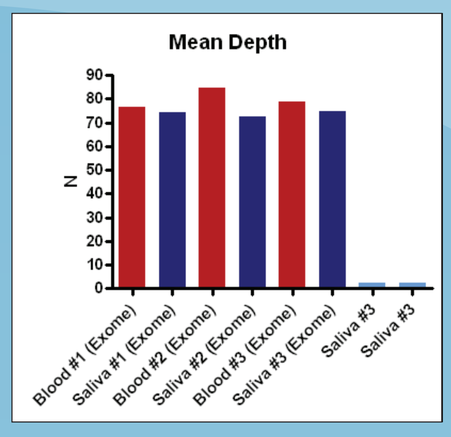
Results

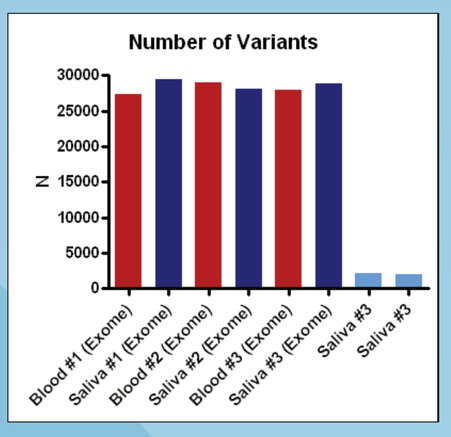
All samples generated in excess of 6 x 10^9 bases per lane on the Illumina Genome Analyzer II. Statistically there was no difference in total bases sequenced between saliva and blood samples or between whole genome or exome enriched samples. For exome enriched blood and saliva samples in excess of 96.6% of all bases aligned to the human genome. In comparison, saliva samples not enriched using the Agilent Sureselect kit were observed to have 91.5% of bases aligned to the human genome.



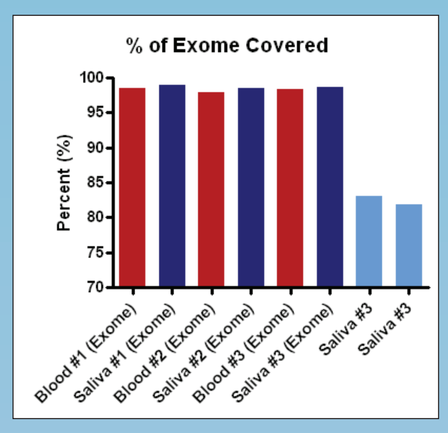


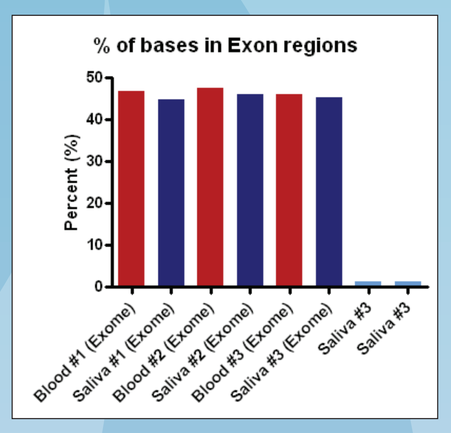
Using a single lane on the Illumina Genome Analyzer II flow cell we were able to achieve a mean depth of 77x for both saliva and blood samples that were exome-enriched. In contrast samples that were not enriched only had a mean depth of 2.4x. Exome-enriched saliva and blood samples identified a similar mean number of variants: 28,738 for saliva and 28,067 for blood.





Exome-enriched saliva and blood samples achieved 98.5% coverage of the exon regions targeted by the Agilent SureSelect All Exon Kit. On average 46% of sequenced bases were located in the exon regions. In contrast, samples which were not enriched only had 82.4% coverage of the exon regions with much lower depth.





Both saliva and blood samples can be used to investigate genetic variations using next generation sequencing technologies. Here we demonstrate that using DNA from saliva or blood we successfully queried SNPs of interest and observed a perfect correlation between the sample types. We queried the following previously reported SNPS: rs2075650 (APOE)[1], rs10513025 (SEMA5A)[2], rs2736990 (SNCA)[3,4], and rs356220 (SNCA)[4,5].

rsID	Gene	Chr	Pos	Donor 1		Donor 2		Donor 3	
				Blood	Saliva	Blood	Saliva	Blood	Saliva
rs356220	SNCA	4	90860363	A	A	A	A	A	A
rs2075650	APOE	19	50087459	R	R	A	A	A	A
rs2736990	SNCA	4	90897564	C	C	C	C	C	C
rs10513025	SEMA5A	5	9676622	T	T	T	T	T	T

Conclusions

Saliva is a reliable source of DNA for use with the Agilent SureSelect Human All Exon Kit and sequencing using paired-end methodology on the Illumina Genome Analyzer II as demonstrated by the coverage (98.7%), depth (74x), and genome aligment (97.3%). Additionally, samples that were not enriched had 91.5% of their sequence align to the reference human genome indicating that saliva can also be used for whole genome sequencing.

Donor	Enrichment	Sample Source	Lane Yield (kbases)	% Align Genome	Coverage	Variants	Mean Depth	kbases Not In Exon	kbases In Exon	Proportion in Exon
A	Agilent SureSelect All Exon	Blood	6,217,788	98.45%	98.50%	27,278	76.4	3,253,014	2,845,003	46.65%
		Saliva	6,434,248	96.74%	98.90%	29,444	74.38	3,416,661	2,781,056	44.87%
B		Blood	6,676,662	99.20%	97.94%	28,983	84.78	3,461,969	3,139,072	47.55%
		Saliva	6,072,430	96.61%	98.55%	27,996	72.3	3,147,312	2,693,723	46.12%
C		Blood	6,447,860	99.08%	98.39%	27,941	78.85	3,433,914	2,932,943	46.07%
		Saliva	6,268,330	98.46%	98.61%	28,773	74.72	3,359,432	2,785,456	45.33%
	N/A	Saliva	6,982,002	91.68%	83.02%	2,161	2.444	6,308,938	76,708	1.20%
	N/A	Saliva	6,818,024	91.23%	81.78%	1,893	2.385	6,129,287	73,742	1.19%

Lane Yield (kbases)	The number of bases provided by the PF reads
% Align Genome	The % of PF reads that aligned to the reference
Coverage	The % of the Exome covered by sequencing reads
Variants	The number of variants (non-reference bases) identified for that sample, after passing quality thresholds
Mean Depth	The average depth for all bases covered in the Exome
Bases Not In Exon	The number of bases that aligned to the genome but not within a direct exon boundary
Bases In Exon	The number of bases that fall within an exact exon boundary
Proportion in Exon	The % of bases aligned within an exon versus all bases aligned

References

(1) Sudha Seshadri et. al. (2010). Genome-wide Analysis of Genetic Loci Associated With Alzheimer Disease. *JAMA*. Vol. 303, No. 18, 1832-1840.
(2) Lauren A. Weiss et. al. (2009). A genome-wide linkage and association scan reveals novel loci for autism. *Nature*. Vol. 461, 802-808.
(3) Javier Simon-Sanchez et. al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genetics*. Vol. 41, No. 12, 1308-1312.
(4) The UK Oarkinson's Disease Consortium and The Wellcome Trust Case Control Consortium 2. (2011). Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21. *Human Molecular Genetics*. Vol. 20, No. 2, 345-353.
(5) Taye H Hamza et. al. (2010). Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature Genetics*. Vol. 42, No. 9, 781-785.

DNA genotek *Superior samples Proven performance*