Video Article

Collection and Extraction of Saliva DNA for Next Generation Sequencing

Michael R. Goode[1], Soo Yeon Cheong[1], Ning Li[1], William C. Ray[1,2], Christopher W. Bartlett[1,3]

[1]Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital

[2]The Interdisciplinary Graduate Program in Biophysics, The Ohio State University

[3]Department of Pediatrics, The Ohio State University

Correspondence to: Christopher W. Bartlett at christopher.bartlett@nationwidechildrens.org

Abstract

The preferred source of DNA in human genetics research is blood, or cell lines derived from blood, as these sources yield large quantities of high quality DNA. However, DNA extraction from saliva can yield high quality DNA with little to no degradation/fragmentation that is suitable for a variety of DNA assays without the expense of a phlebotomist and can even be acquired through the mail. However, at present, no saliva DNA collection/extraction protocols for next generation sequencing have been presented in the literature. This protocol optimizes parameters of saliva collection/storage and DNA extraction to be of sufficient quality and quantity for DNA assays with the highest standards, including microarray genotyping and next generation sequencing.

Video Link

The video component of this article can be found at http://www.jove.com/video/51697/

Introduction

Obtaining high quality DNA for human genetic studies is essential in the disease gene discovery process. Blood, though requiring an invasive procedure and also being more expensive than saliva collection, is favored for creating immortalized cell lines as an infinite source of DNA, or iPSCs for functional studies, and sometimes blood DNA is used when cell lines are not available. However, obtaining blood requires a trained phlebotomist and blood has a shorter half-life than saliva[1]. DNA from saliva is less expensive and easier to obtain, since it can be collected and sent through the mail without the need for a phlebotomist, thereby increasing potential subject pools well beyond the catchment area of hospitals and laboratories[2]. Study enrollment may be improved when subjects have the option of giving a saliva sample instead of blood[3, 4]. Concerns about the quantity and quality of DNA from saliva may have limited its widespread use despite numerous studies recent studies showing the suitability of whole saliva, with an average of 4.3×10^5 cells per milliliter, for DNA testing over the older buccal swabs methods that did not obtain significant amounts of saliva[2, 3, 4, 5, 6]. While a modest literature exists showing the suitability of whole saliva derived DNA for genotyping applications including microarray-based methods[8, 9, 10], no studies have examined next generation sequencing (NGS). The goal for optimizing this whole saliva DNA extraction protocol was to maximize quantity and quality for genetics applications in a cost effective way that is easily implemented in laboratories with common reagents and consumables.

DNA extraction from saliva requires several procedures: 1) collection and storage, 2) cell lysis, 3) RNase treatment, 4) protein precipitation, 5) ethanol precipitation, 6) DNA rehydration. The DNA Stabilization Buffer solution, described previously[2], functions adequately without alteration. No attempt to optimize the RNase treatment and DNA rehydration steps was made. For each remaining step, several variables that could affect yield were identified. Each variable was manipulated individually and improvement in yield and quality was assessed statistically. For variables that were shown to improve yield and/or DNA quality, the optimal values were included in the final protocol.

Protocol

NOTE: Prior to providing saliva samples all subjects gave informed consent conforming to the guidelines for treatment of human subjects at Nationwide Children's Hospital.

1. Saliva Collection and Storage

1. Prior to saliva collection, ensure that the subject's mouth is free of food or other foreign substances by having the subject rinse their mouth with water and avoiding eating or drinking for 30 min before collecting the sample.
2. Open a 15 ml centrifuge tube with 2.5 ml of DNA stabilization buffer[2] making sure to avoid touching the inside of the cap or tube, and have the subject spit 2.5 ml of saliva into the buffer solution. Note: Collecting more than 2.5 ml of saliva can lead to sample degradation from an

insufficient ratio of sample to DNA stabilization buffer. Collecting too little saliva will reduce expected yields from the protocol. To evaluate collection volumes, use the numbered gradients on the side of the tube.

3. Replace cap and mix by inversion until the mixture is homogenized. Vigorous shaking is not necessary. Store the samples at RT for short-term storage or 4 °C for long-term storage (>3 months).

2. Initial Preparations and Cell Lysis

1. Prior to starting the extraction, heat a water bath to 37 °C, and prepare an ice bucket. Three 15 ml conical centrifuge tubes will be needed for each extracted sample. The three tubes will be used to hold the cell and protein pellet, the final extracted gDNA, and the isopropanol and ethanol supernatants.
2. Retrieve samples from storage, and invert samples several times then vortex at medium speed for 15 sec.
3. Dispense 2.5 ml of sample into a clean 15 ml centrifuge tube, and add 5 ml of Cell Lysis Solution. Mix the sample 50 times by inversion, and incubate at RT for 30 min.

3. RNA Removal

1. Add 40 µl of RNase A Solution at 100 mg/ml, and incubate at 37 °C for 15 min.
2. Remove the sample from the 37 °C water bath and cool on ice for 3 min.
    1. After the RNase A incubation, increase the temperature of the water bath to 65 °C for the DNA rehydration step of the protocol.

4. Protein and Lipid Removal

1. Add 50 µl of Proteinase K Solution at 20 mg/ml, mix several times by inversion, and incubate at RT for a minimum of 30 min. Note: This is a possible pausing point for the protocol. After the addition of the Proteinase K Solution, the sample can be stored at 4 °C until the extraction can be completed. Storage at 4 °C for up to 24 hr was not shown to have a significant effect on the extraction yields or DNA quality. Long-term storage at this stage has not been evaluated.
2. Add 1.7 ml of Protein Precipitation Solution, vortex vigorously for 20 sec at high speed, and place on ice for 10 min.
3. Once the samples have cooled on ice for 10 min, centrifuge for 10 min at 3,000 x g and 4 °C. The precipitated proteins must form a tight pellet to continue. If the pellet is not tight or the solution is still cloudy, the samples can be cooled on ice for 5 min more and centrifugation repeated. The samples must be kept on ice to ensure a tight pellet.

5. Isolation and Purification of gDNA

1. Into a clean 15 ml centrifuge tube, pipet 5 ml of Isopropanol and 8 µl of pure Glycogen Solution at 20 mg/ml.
2. Pour the supernatant containing the gDNA from step 4.3 into the tube containing the Isopropanol and Glycogen Solution, leaving behind the precipitated protein pellet. Once the supernatant has been added, gently mix the sample 50 times by inversion and centrifuge for 30 min at 3,000 x g and 4 °C.
3. Pour the supernatant slowly into a clean 15 ml tube. After removal of the supernatant, add 1 ml of 70% ethanol to wash the pellet by slowly rocking and gently moving the ethanol over the precipitated pellet several times. Retain the ethanol in the tube.
4. After the initial wash, centrifuge the sample for 1 min at 2,000 x g and 20 °C. This centrifugation step can be done at either 4 °C or 20 °C. No significant effect of temperature has been shown for this step.
5. Following the initial wash and centrifugation of the pellet, slowly pour the ethanol wash from the tube and discard, then perform a second wash by repeating steps 5.3 and 5.4.
6. After the removal of the supernatant from the second wash, allow the pellet to air dry for 15 min.
    1. If the sample has not completely dried, air dry for another 15 min.

6. Rehydration of gDNA

1. Once the sample has dried, add 300 µl of Tris-EDTA to rehydrate the dried gDNA pellet.
2. Vortex the sample for 5 sec at medium speed and place in a 65 °C hot water bath for 1 hr.
3. Remove the samples from the water bath and incubate O/N at RT.
    NOTE: All products and reagents used are listed in the Materials Table, as well as **Table 4**.

Representative Results

To determine optimal parameters for DNA extraction a series of paired DNA extractions was performed. A single saliva sample was split and each portion tested with one of two possible values for a given variable. At least eight replicates of each paired test were performed (*e.g.*, a single saliva sample was aliquoted to test extraction both with and without initial 50 °C incubation). Optimization was based on four standard metrics: total DNA yield, the 260/280 value, the 260/230 value, and visual inspection of electrophoresed DNA to assess fragmentation. Not all possible combinations of the variables were assessed statistical interactions (N=169 combinations), opting instead to assess the marginal effect of each variable individually. Effects were tested using a multi-way repeated-measures ANOVA and estimated effects were derived from the equivalent regression equation. All significant effects are summarized in **Table 1**, shown as average change in yield (ng/µl per ml of saliva input) or DNA quality (260/280 and 260/230).

Cell lysis (step 2) was optimized by assessing: 1) the presence/absence of a 50 °C incubation (1 hr) prior to cell lysis to ensure that Proteinase K degradation and cell lysis mediated by the storage buffer went to completion, 2) presence/absence of a homogenization by vortexing step (medium speed, 15 sec), and 3) lysis solution incubation time (5 versus 30 min). The 30 min cell lysis incubation increased yield by an average of 3.5% ($p<.01$) but no other cell lysis variable had a significant effect on yield. Vortexing decreased the 260/280 ratio by a statistically significant ($p<.001$) but practically small 0.03.

Protein precipitation (step 4) is preceded by Proteinase K digestion to disrupt amino acid chains, improving protein precipitation efficiency and releasing captured DNA. The amount of Proteinase K was varied ten-fold. Centrifugation temperature was reduced from 20 °C to 4 °C. Increasing the amount of Proteinase K caused a statistically significant decrease in yield (8.7%) and also slightly improved both the 260/280 and 260/230 ratios.

Ethanol precipitation (step 5) was the last stage of the protocol examined. The amount of glycogen carrier (0, 8 µl) was varied, as was the total centrifugation time (5 vs. 30 min[11]). Only centrifugation time significantly affected yield, with an average increase of 290%. The longer spin also decreased the 260/280 ratio slightly (0.05). No significant effect of glycogen on yield was observed during the experiments; though the total quantity of DNA in these extractions was sufficiently large that glycogen would not typically be used. Despite the lack of effect in these samples, it is still recommend to use glycogen to minimize the risk of reduced yields whenever saliva input volume is lower than given here or if there is any other reason to believe yield will be low.

Visual inspection of the representative DNA samples (**Figure 1**) indicated that the extracted DNA was not greatly fragmented for any saliva DNA extraction procedure, but rather showed an appropriate high molecular weight band without the smearing indicative of degraded DNA. After RNase A digestion, the protocol produced an average 260/280 of 1.74.
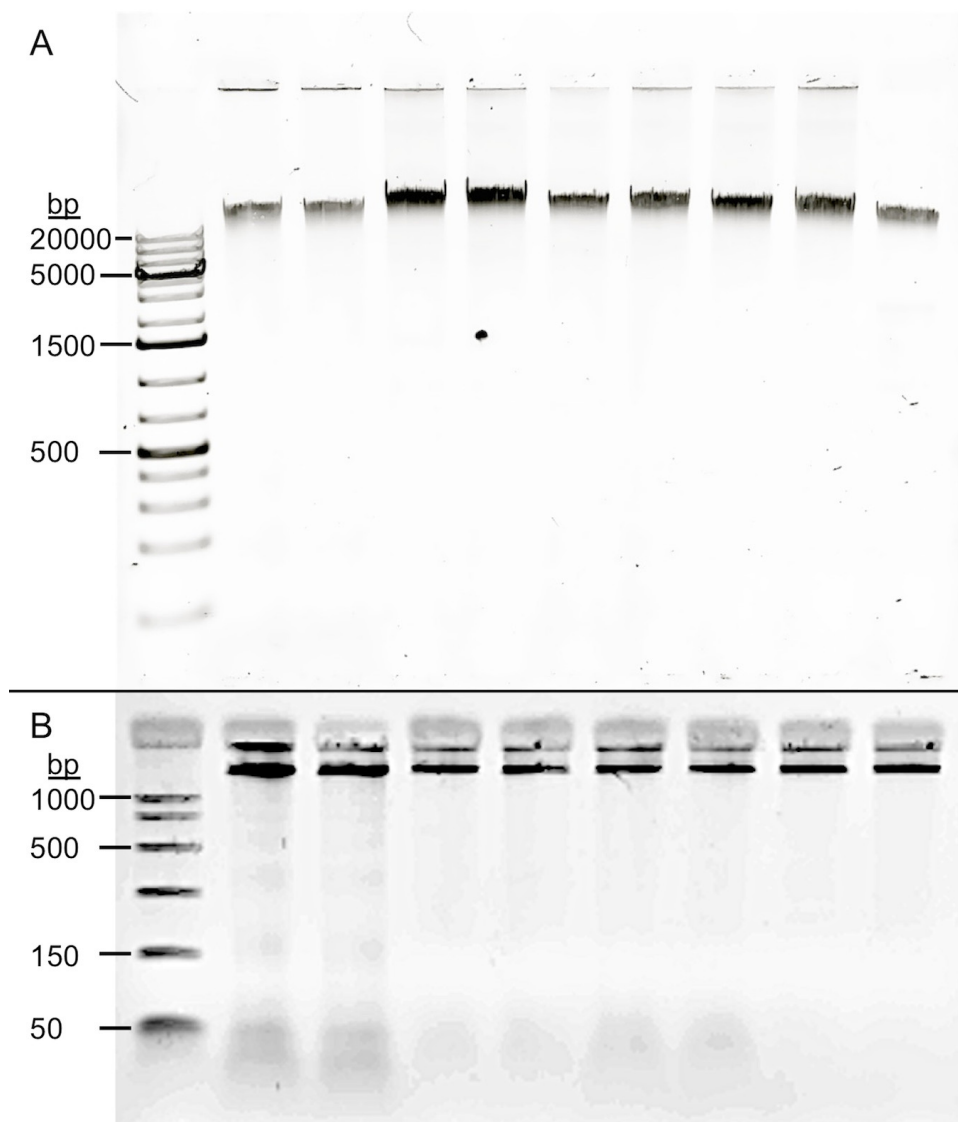
Figure 1. Quality of saliva derived DNA. Four extraction procedures were applied to the same saliva collection. (**A**) Samples were electrophoresed on a 0.8% agarose gel (250 ng DNA). All variations of the saliva DNA extraction protocols result in high molecular weight (>20 kb) DNA, with no evidence of degradation. Lane: 1 DNA ladder, 2 & 3 Oragene prepIT L2P Protocol samples, 4 & 5 Gentra Puregene Body Fluids Protocol, 6 & 7 the optimized protocol without RNA removal step, 8 & 9 the optimized protocol with the RNA remove step. Lanes 2-7 are directly analogous protocols on the same saliva samples. Lanes 8 & 9 show that the RNA removal step does not introduce DNA degradation. (**B**) Samples were electrophoresed on a 2% agarose gel (150 ng DNA). A slight RNA peak is observable near the bottom of the gel in lanes 2 through 7 (conventions as above). Lanes 8 & 9 show the effectiveness of the RNA removal step.

The RNA Removal Step (step 3 with RNase A) is critical for accurate quantification of DNA. During testing, consistently high RNA content was observed, as determined by the ratio of double stranded DNA to RNA measured by a Qubit 2.0 Fluorometer. On average, nucleic acid content from samples without RNase A treatment consisted of 46.6% (±0.4) RNA. Samples that underwent the RNA Removal Step read as "<20 ng/ml", which is the lowest possible reading for the Qubit's RNA detection.

The DNA obtained through this optimized protocol was of sufficient quality for high throughput sequencing when the additional RNase A step was applied. To attain targeted resequencing data, a custom Agilent SureSelect Target Enrichment kit was applied to 24 samples, targeting 2.6 Mb of sequence. High throughput sequencing was conducted on 12 barcoded (indexed) samples per lane. Sequence reads were BWA-aligned to the hg19 reference genome[12], then application of GATK[13] base quality score recalibration, indel realignment, duplicate removal, SNP discovery and genotyping simultaneously across all 24 samples was performed using the best practice hard filtering parameter values[14]. All 24 samples yielded high quality NGS data (**Table 2**). Of reads that passed Illumina's standard filters and had Q>20, 91.4% aligned to the sequence enrichment target regions, providing an average on-target coverage depth of >30x coverage at Q>100, well within the necessary limits for rare SNP discovery in each sample. The average strand balance was 49.9%. Comparing variant calls with Illumina microarray genotypes yielded a concordance of 98.9%.

Candidate Variable	ng/µl	260/280	260/230
Vortex	n.s.	-0.03***	n.s.

x30 min Cell Lysis Incubation	3.5%**	n.s.	n.s.
Proteinase K x10	-8.7%*	-0.05***	-0.03***
30 min spin	290.2%***	-0.05***	n.s.
Glycogen	n.s.	n.s.	-0.37***

Table 1. Effect Size of Optimized Variable on Quantity/Quality Metrics. All effect sizes are in the units listed in the column header. p-values from ANOVA: *$p<.05$; **$p<.01$; ***$p<.001$; n.s. not significant

Quality Metric	Value
Target length	1,708 kb
Target Covered > 30x	85.22%
SNPs in dbSNP	92.55%
Array agreement	98.99%

Table 2. Quality of High Throughput Sequence from Target Enrichment.

New Optimized Protocol	Item	Distributer	Catalog #	Purchasing Unit	Cost/Unit	Cost/Collection
	15 ml Centrifuge Tubes	Fisher	12-565-268	500 Tubes	$233.50	$3.2690
	Cell Lysis Solution	Qiagen	158908	1 L	$401.00	$3.2080
	Proteinase K	Sigma	P6556	1 g	$713.00	$1.5686
	Protein Precipitation Solution	Qiagen	158912	350 ml	$350.00	$2.7200
	Isopropanol	Fisher	A416-4	Case of 4 x 4 L	$486.71	$0.2434
	Glycogen Solution (20 mg/ml)	EZ-BioResearch	S1003	1 ml	$51.00	$0.8160
	70% Ethanol	Fisher	04-355-305	Case of 4 x 1 gal.	$123.03	$0.0325
	Tris-EDTA (TE)	Fisher	BP2473-1	1 L	$68.67	$0.0412
	NaCl	Fisher	AC194090010	1 kg	$34.65	$0.000004
	Tris HCl	Fisher	BP1757-100	100 ml	$57.84	$0.0116
	EDTA(0.5 M) Solution	Fisher	03-500-506	100 ml	$33.60	$0.0134
	Sodium Dodecyl Sulfate	Fisher	BP166-100	100 g	$59.95	$0.0060
					Total Cost	**$11.93**
Oragene	Item	Distributer	Catalog #	Purchasing Unit	Cost/Unit	Cost/Collection
	15 ml Centrifuge Tubes	Fisher	12-565-268	500 Tubes	$233.50	$0.9340
	100% Ethanol	Fisher	BP2818-100	100 ml	$34.29	$1.6459
	70% Ethanol	Fisher	04-355-305	Case of 4 x 1 gal.	$123.03	$0.0081
	Tris-EDTA (TE)	Fisher	BP2473-1	1 L	$68.67	$0.0687
	1.5 ml tube	Genesee	22-281A	500 Tubes	$22.85	$0.0457
	Oragene Collection KIT	Oragene	OG-500	1	$25.00	$25.00
					Total Cost	**$27.70**
Puregene	Item	Distributer	Catalog #	Purchasing Unit	Cost/Unit	Cost/Collection
	15 ml Centrifuge Tubes	Fisher	12-565-268	500 Tubes	$233.50	$0.9340
	Cell Lysis Solution	Qiagen	158908	1 L	$401.00	$4.0100
	Proteinase K	Qiagen	158918	650 ml	$73.10	$7.8723

Protein Precipitation Solution	Qiagen	158912	350 ml	$350.00	$4.0000
Isopropanol	Fisher	A4164	Case of 4 x 4 L	$486.71	$0.3650
Glycogen Solution	Qiagen	158930	500 ml	$64.30	$2.5720
70% Ethanol	Fisher	04-355-305	Case of 4 x 1 gal.	$123.03	$0.0975
DNA Hydration Solution	Qiagen	158914	100 ml	$69.80	$0.1396
NaCl	Fisher	AC19409-0010	1 kg	$34.65	$0.000004
Tris HCl	Fisher	BP1757-100	100 ml	$57.84	$0.0116
EDTA(0.5 M) Solution	Fisher	03-500-506	100 ml	$33.60	$0.0134
Sodium Dodecyl Sulfate	Fisher	BP166-100	100 g	$59.95	$0.0060
				Total Cost	**$19.99**

Table 3. Cost comparison of the optimized protocol to extract DNA from 2 ml of whole saliva with other commercially available protocols. All reagents and consumables required for DNA extraction have been assessed using standard list prices available on the internet as of September 30, 2013. Note that this protocol is written to extract DNA from 1.25 ml of whole saliva (*i.e.*, 2.5 ml of saliva and buffer, see step 2.3) as this value represents half of the total volume in the saliva collection tube. For the cost comparison, 2 ml was chosen as this is the amount associated with a common commercially available kit (Oragene).

DNA Stabilization Buffer (250 ml)		
Component	Volume (ml)	[Final]
1 M NaCl	1.461 g	0.1 M
Tris HCl	2.5	0.01 M
0.5M EDTA	5	0.01 M
10% SDS	12.5	0.014 M
Proteinase K Solution (20 mg/ml)	2.5	6.92×10^{-6} M
ddH$_2$O	227.5	
Proteinase K Solution (20 mg/ml)		
Component	Volume (ml)	[Final]
Proteinase K powder	500 mg	6.92×10^{-4} M
ddH$_2$O	25	
70% Ethanol (500 ml)		
Component	Volume (ml)	[Final]
Ethanol 95%	368.5	12.63 M
ddH$_2$O	131.5	

Table 4. Recipes for Reagents.

Discussion

The present procedure is an optimized DNA extraction protocol that has considerably improved yield of high molecular weight DNA compared to standard methods, without compromising DNA quality. The critical step with the biggest effect on yield the most was step 5.2, which includes a longer centrifugation step during ethanol precipitation than any published protocol reviewed here, except one that was not widely distributed[11]. No changes in DNA quality associated with this longer centrifugation were detected, indicating that most of the available DNA from whole saliva collection is not degraded and high molecular weight.

Whole saliva collection has limitations in sample quality such as the potential for foreign contaminants, which needs to be minimized at the collection stage, and the presence of excessive protein in the sample that can be a sign of an underlying infection. Large amounts of protein or foreign contaminants can remain in the final extracted DNA thereby making quantification inaccurate. If there are residual proteins or contaminants after rehydration (step 6) are suspected, a sample clean up can be performed by starting at the protein precipitation step with reagent volumes scaled to reflect the sample input volume. Another limitation of the protocol is the length of time required to perform the steps. Multiple samples can be run in parallel; however, it is recommended that no more than 24 parallel extractions be run simultaneously. This is

particularly important for the protein precipitation step, where the sample must remain cold to ensure a tight pellet and running more than 24 samples may allow pellets time to re-dissolve.

The protocol presented here is the most cost effective method considered (see **Table 3**). While this protocol does use reagents from the Puregene extraction kit, a smaller volume than recommended in the Puregene protocol is used without compromising extraction yield and it is this reduction in reagents that drives the cost savings relative to the protocol. Note that the calculated cost for extraction uses the list price for each item and does not reflect any discounts. The cost per extraction with the optimized protocol can be reduced further with bulk orders or discounts through company sales representatives.

Evidence has also been provided that this protocol is suitable for use with next generation sequencing. The data obtained provides further evidence of the utility of saliva samples for human genetics research in many diseases where blood is not routinely available. While blood and cell line derived DNA continues to be the preferred source of genetic material for testing, whole saliva collection is a viable alternative when such sources are not available, when patient enrollment is affected by their collection or phlebotomy is not available or impractical.

Disclosures

The authors have nothing to disclose.

Acknowledgements

References

1. Quinque, D., Kittler, R., Kayser, M., Stoneking, M, Nasidze, I. Evaluation of saliva as a source of human DNA for population and association studies. *Analytical Biochemistry.* **353**, 272-277 (2006).
2. Min, J.L., *et al.* High mircosatellite and SNP genotyping success rates established in a large number of genomic DNA samples extracted from mouth swabs and genotypes. *Twin Research and Human Genetics.* **9**, 501-506 (2006).
3. Dlugos, D.J., Scattergood, T.M., Ferraro, T.N., Berrettinni, W.H., Buono, R.J. Recruitment rates and fear of phlebotomy in pediatric patients in a genetic study of epilepsy. *Epilepsy & Behavior.* **6**, 444-446 (2005).
4. Etter, J.F., Neidhart, E., Bertand, S., Malafosse, A., Bertrand, D. Collecting saliva by mail for genetic and cotinine analyses in participants recruited through the internet. *European Journal of Epidemiology.* **20**, 833-838 (2005).
5. Hansen, T.V., Simonsen, M.K., Nielsen, F.C., Hundrup, Y.A. Collection of blood, saliva, and buccal cell samples in a pilot study on the danish nurse cohort: Comparison of the response rate and quality of genomic DNA. *Cancer Epidemiol Biomarkers Prev.* **16**, 2072-2076 (2007).
6. Van Schie, R.C.A.A., Wilson, M.E. Saliva: A convenient source of DNA for analysis of bi-allelic polymorphisms of fcγ receptor iia (cd32) and fcγ receptor iiib (cd16). *Journal of Immunological Methods.* **208**, 91-101 (1997).
7. Dawes, C. Estimates, from salivary analyses, of the turnover time of oral mucosal epithelium in humans and the number of bacteria in an edentulous mouth. *Archives of Oral Biology.* **48**, 329-336 (2003).
8. Bahlo, M., *et al.* Saliva-derived DNA performs well in large-scale, high-density single-nucleotide polymorphism microarray studies. *Cancer Epidemiol Biomarkers Prev.* **19**, 794-798 (2010).
9. Hu, Y., *et al.* Genotyping performance between saliva and blood-derived genomic dnas on the dmet array: A comparison. *PLoS ONE* **7** (3), e33968 (2012).
10. Simmons, T.R., *et al.* Increasing genotype-phenotype model determinism: Application to bivariate reading/language traits and epistatic interactions in language-impaired families. *Human Heredity.* **70**, 232-244 (2010).
11. Zeugin, J.A., Hartley, J.L. Ethanol precipitation of DNA. *Focus.* **7**, 1-2, (1985).
12. Li, H., Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics.* **25**, 1754-60 (2009).
13. McKenna, A., *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-303 (2010).
14. DePristo, M., *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics.* **43**, 491-498 (2011).