Targeted Next–Generation Sequencing for Canine Genotyping using the Fluidigm® D3™ Assay Design Engine with the Access Array™ System

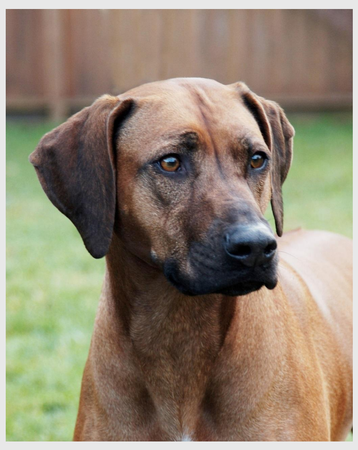
Min Lin[1*], Mark Neff[2], Alison Ruhe[3], Jiashun Zheng[4], Nico Tuason[1], Michael Lin[1], Janet Ziegle[1], Honrado Lopez[1] and Julian Walker[1]

[1]Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080
[2]Van Andel Research Institute, 333 Bostwick Ave NE, Grand Rapids, MI 49503
[3]ProjectDog, 636 San Pablo Ave, Albany, CA 94706
[4]University of California, San Francisco, CA 94080

Introduction

For proof-of-concept, Fluidigm® D3™ assay design engine and Access Array™ system were implemented for targeted next-generation sequencing for non-human genotyping identification to achieve a rapid and more economical workflow. We assayed 48 variants mined from the genome of a dog affected with adult onset deafness. A prior genome-wide association study (GWAS) had mapped the causative gene to a ~4 Mb interval. The top 48 single-nucleotide polymorphisms (SNPs) from within this interval were converted to assays suitable for NGS analysis. The Fluidigm® D3™ assay design engine was applied to generate target-specific primers, and the Fluidigm® Access Array™ System provided a fast and automated target-enrichment sample preparation. All 48 variants within the 48 samples were successfully genotyped in a single sequencing run. 24 heterozygous variants were excluded because the inherited deafness is a Mendelian recessive defect. Among the remaining 24 variants, three showed low "carrier" rates among a population sampling, consistent with a recent and private mutation in the breed isolate. Our results demonstrate this approach is faster (<three weeks) and more cost effective (<$0.30/dog/variant) than current platforms for SNP genotyping by NGS.

Study Background



Rhodesian Ridgeback

- **Inherited deafness**: An adult onset deafness segregates in the Rhodesian Ridgeback breed with an apparent autosomal recessive mode. Affected dogs completely lose the ability to hear between six months and one year of age. Male dogs appear to lose their hearing earlier than female dogs, though there does not appear to be a sex bias of affected status.

- **Prior GWAS:** A previous case-control association study (26:230) mapped the cause of the defect to a single locus spanning 4 Mb. There were no canine orthologs of previously-implicated human deafness genes found in the mapped interval.

- **Prior whole genome sequencing:** The whole genome of an affected dog was sequenced at 32X coverage using multiple lanes on an Illumina HiSeq. Conventional NGS pipelines for alignment and variant calling on the Amazon Elastic Compute Cloud identified a total of 10,351 sequence variants within the mapped interval.

Conclusions

- This study demonstrates the utility of a simplified, targeted resequencing method for custom SNP genotyping. The results of this experiment, with 48 SNPs typed on 48 dogs, excluded half the variants as being causal and prioritized three SNPs as being strongest candidates.

- 95.0% of NGS reads aligned to the canine reference, and 99.9% of those mapped to the targeted amplicon sites, demonstrating the high quality of the NGS library.

- One of the advantages of genotyping by using the Fluidigm® Access Array™ System is SNP discovery. In this study, in addition to the targeted SNPs, we identified 56 additional SNPs in our 48 target regions.

- The Fluidigm® systems leveraged here provided rapid and cost-effective NGS library preparation, agnostic to the sample type and scale of genetic study by NGS. High read coverage per variant assayed suggests that a greater depth of pooling may be obtained in the future, further increasing the efficiency of this approach and decreasing the cost.

Results

Fluidigm® Systems and Workflow

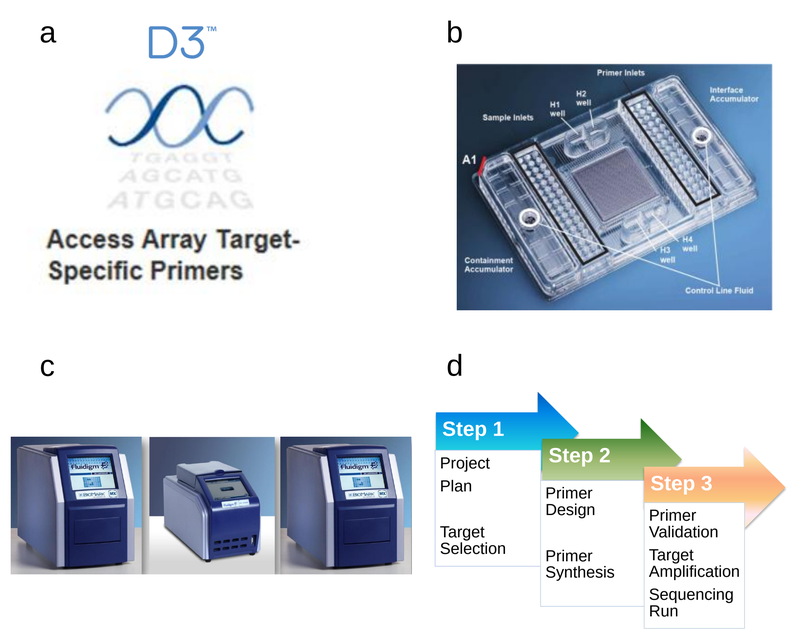


Figure 1: Workflow

a) The Fluidigm® D3™ assay design engine provides target-specific primers to prepare sequencing-ready libraries for NGS. **b)** The 48.48 Access Array™ integrated fluidic circuit (IFC) is a microfluidic chip that systematically combines 48 sample inputs with 48 primer inputs to enable all possible 2,304 combinations of samples and primer pools of up to 10 primer pairs. Once samples and primers are loaded and combined, the chip undergoes thermal cycling to amplify regions of interest from the samples. Upon completion of PCR, the resulting amplicons recovered on a per-sample basis in the IFC are ready for sequencing. **c)** The Access Array™ System is comprised of two IFC controllers: Pre-PCR for loading samples and primers and Post-PCR for harvesting PCR products, and a thermal cycler for PCR reaction. **d)** High primer success rate (95% for ≤65% GC content) designed by the D3™ assay design engine , with simple and automatic workflow, ensure a fast turnaround time from project planning to NGS data.

Primer Validation and Sample QC

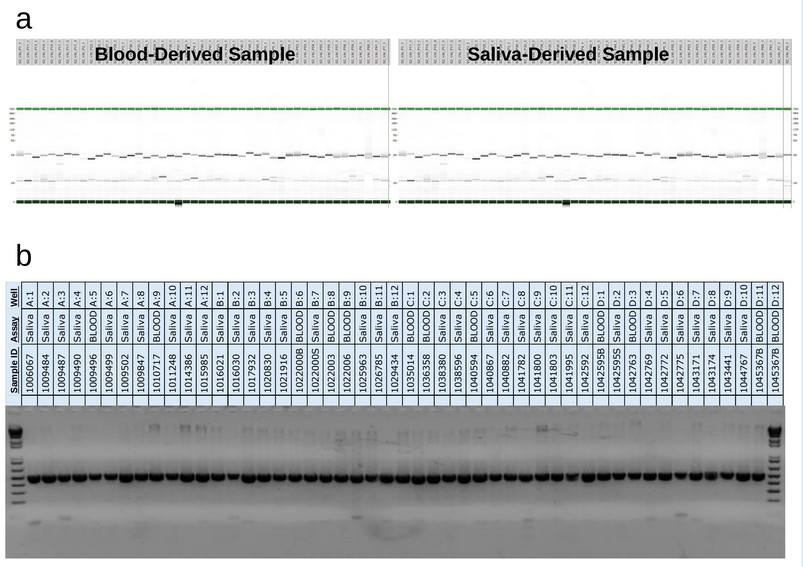
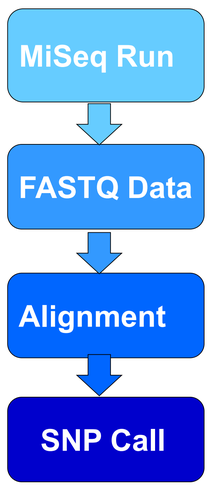


Figure 2: Gel images for primer validation and sample qualification

a) 48 primers were validated with a control blood-derived sample and a saliva-derived sample to ensure the two types of samples can generate sufficient PCR products with the correct size targets. Low yield PCR product or unspecific product may result in low quality library and low coverage for variants. **b)** All samples were screened by Nanodrop and run on 1.5% agarose gel to identify high purity (data not shown here), and the final assessment applied a 353 bp assay to confirm intact DNA based on the target amplicon size of about 300 bp. The high signal bands from 48 samples indicate the intact genomic DNA derived from blood and saliva. Degraded genomic DNA may prevent efficient amplification of the target region, resulting in low yield reads.

Data Analysis Workflow



The FASTQ files were generated by MiSeq Reporter with demultiplexing for each sample and the FASTQ files are suitable for the secondary analysis. The alignment was performed using SNAP (snap.cs.berkeley.edu) with canFam3 reference genome. SNP calling was done using the GATK (HaplotypeCaller), SNPs were merged into a single VCF file.

Data Analysis Workflow

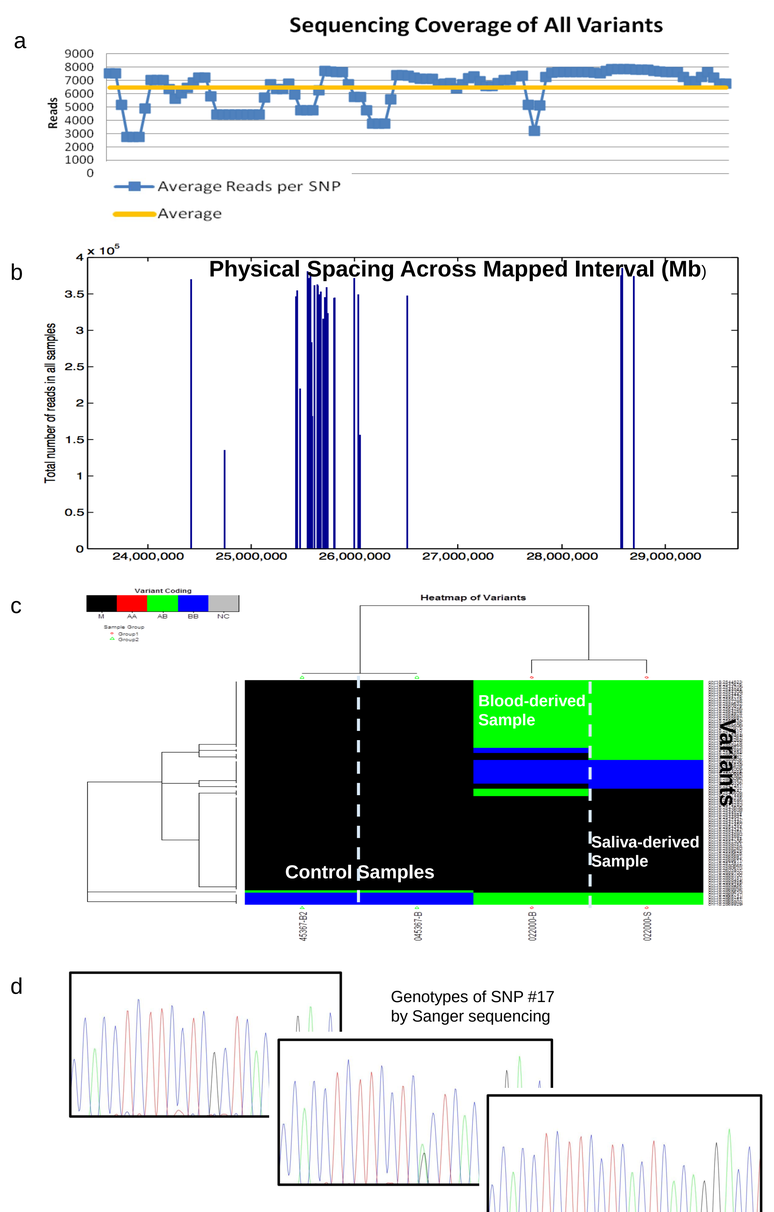


Figure 3: Variants calling

a) Each targeted variant (48) was well covered with an average of 6.5k reads per sample. High read depth per variant means future studies can explore deeper pooling of individuals and/or variants without loss of call rates. Additional variants (56) were also newly discovered. These polymorphisms were located in sequences adjacent to targeted variants (same amplicon). The additional polymorphisms were also well covered and reliably scored, further supporting fine mapping of the locus. **b)** The read coverage of variants across the mapped interval was consistent. The total number of reads per variant across all 48 dogs differed less than fourfold, suggesting that the number of markers assayed could be increased. **c)** Two pairs of identical samples tested showed high replication of genotype calls. Replication was perfect when both gDNAs were derived from blood. Replication was nearly perfect when comparing blood-derived and saliva-derived gDNAs. This result indicates that the approach was robust and DNA template-tolerant. **d)** A subset of genotypes were scored in common between the NGS (n = 4,784 genotypes total) and Sanger sequencing (n = 288). Genotypes typed by both approaches were perfectly concordant.

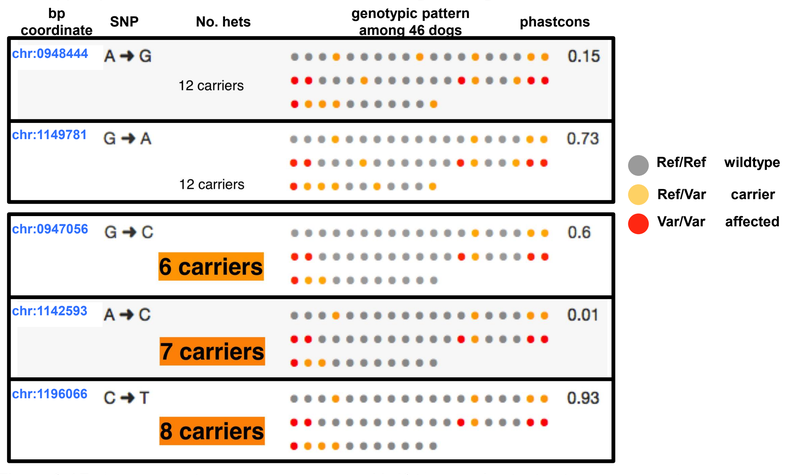
Marker Selection

30 variants were selected for possible causality

10,351	total variants discovered in the affected case genome
4,928	variants homozygous, as expected for a Mendelian recessive defect
79	variants unique to the Rhodesian Ridgeback breed, compared to 30 other dog genomes analyzed (no other breeds are known to exhibit a simple adult onset deafness)
30	variants tied to non-zero phastcons scores in the reference genome sequence, as expected for base pair coordinates of functional effect
0	variants deemed to be non-synonymous coding changes

Analysis of Genotype Data

Genotypic pattern of 5/48 variants on 46 dogs using a custom-variant browser



Five of 45 dogs genotyped were cases. Half the variants tested were not homozygous in these dogs, thereby excluding the variants as being causal. Of 24 remaining SNPs, most suggested a dozen carriers from among a random cohort of 38 dogs. Three Variants were less frequent, each identifying six, seven or eight carriers individually. These three variants will be able to be screened against a larger population sampling.

Fluidigm Corporation
7000 Shoreline Court, Suite 100 • South San Francisco, CA 94080
Toll-free: 1.866.FLUIDLINE
www.fluidigm.com