Development and Validation Procedures for Improving SearchLight DNA™, a Canine Cancer Genomic Diagnostic Panel

Manisha Warrier[1], Shukmei Wong[1*], Sharda Sakthikumar[1], Zeeshan Ahmed[1], Salvatore Facista[1], Natalie Duran[1], Jessica Molnar[1], Jonathan Adkins[1], Nidhi Patel[1], Esther Chon[1], Derick Whitley[1], Giulia Siravegna[1], Kathryn Banovich[1], David Haworth[1], Guannan Wang[1], Zhanyang Zhu[1], William PD Hendricks[1]*

[1]*The Translational Genomics Research Institute (TGen) and Vidium Animal Health, a Subsidiary of TGen*

INTRODUCTION

SearchLight DNA™ v1.0 is an extensively validated canine cancer genomic diagnostic panel that identifies mutations with biomarker associations in 120 genes. The content is based on ongoing curation of canine and human cancer literature. The panel has now been expanded to capture new mutations and biomarker evidence, supported by technical improvements in the bioinformatics workflow. SearchLight DNA™ v1.2 contains additional probes that can capture 86 new hotspot mutations across 25 genes relevant to canine cancer, including one new gene PLCG1. Improvements to copy number variant calling were achieved with the addition of 32 new target regions to improve coverage in AKT1 and SMO, a copy number baseline generated using 14 additional normal samples, and a tool version upgrade. A variant database and web interface were also developed to improve clinical annotation and reporting processes. Here we describe the development and validation procedures for this new and improved hybrid-capture panel that is capable of identifying a broader range of genomic variation across 500kb and has the potential to elucidate key biomarkers with diagnostic, prognostic, or therapeutic significance in canine cancers.

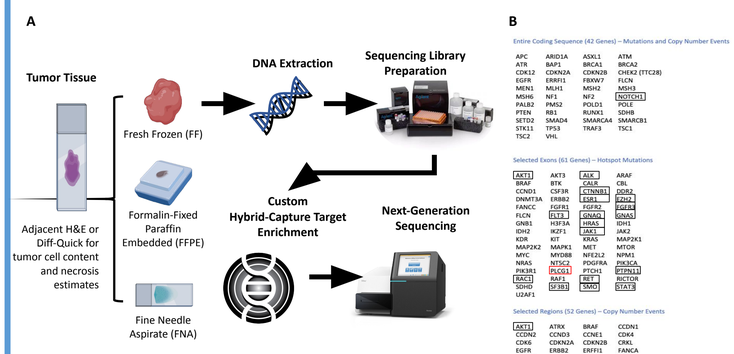
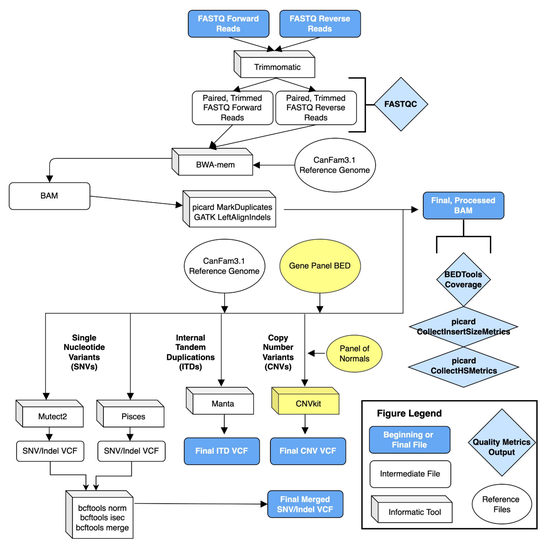


Figure 1. SearchLight DNA™ v1.2 Workflow and Content. SearchLight DNA™ v1.2 is a tumor-only, NGS, hybrid-capture, canine gene panel covering > 500 kb across 120 genes associated with canine or human cancer. (**A**) Wet laboratory workflow for sequencing tumor tissue (FF, FFPE, FNA). (**B**) Genes and mutation types evaluated by SearchLight DNA™ v1.2. Black boxes indicate genes for which new probes were added. Red box indicates the new gene added to v1.2.

METHODS



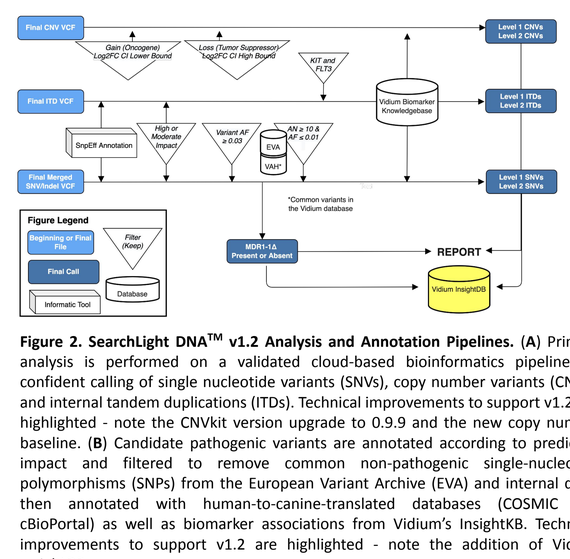


Figure 2. SearchLight DNA™ v1.2 Analysis and Annotation Pipelines. (**A**) Primary analysis is performed on a validated cloud-based bioinformatics pipeline for confident calling of single nucleotide variants (SNVs), copy number variants (CNVs), and internal tandem duplications (ITDs). Technical improvements to support v1.2 are highlighted - note the CNVkit version upgrade to 0.9.9 and the new copy number baseline. (**B**) Candidate pathogenic variants are annotated according to predicted impact and filtered to remove common non-pathogenic single-nucleotide polymorphisms (SNPs) from the European Variant Archive (EVA) and internal data, then annotated with human-to-canine-translated databases (COSMIC and cBioPortal) as well as biomarker associations from Vidium's InsightKB. Technical improvements to support v1.2 are highlighted - note the addition of Vidium InsightDB.

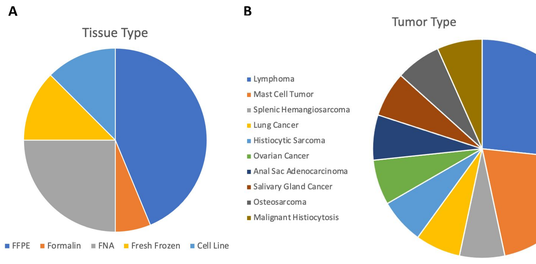


Figure 3. SearchLight DNA™ v1.2 Sample Selection. Validation samples were selected to account for maximum variation in tumor and tissue type. The sample set included 14 clinical samples sequenced previously on v1.0, and two cell lines. Tissue types included FFPE, FNA, and fresh frozen tissue (A). All validation samples, based on v1.0 results, harbor a variety of mutations, including single-nucleotide variants (SNVs), copy number alterations (CNVs), and internal tandem duplications (ITDs). Tumor types are shown in (B).

Table 1. SearchLight DNA™ v1.2 Validation Requirements. QC thresholds for run-level and sample-level metrics are shown below, along with minimum requirements for variant concordance.

Run-level Metrics	Sample-level Metrics	Variant Concordance
% PF >= 70%	Mean target coverage >= 200x	SNV concordance >= 90%
% Q30 >= 80%	Fraction bases on-target/unique >= 38%	CNV concordance % at gene-level, case-by-case
	Fraction bases on-target >= 100x >= 89%	ITD concordance 100%
	Coefficient of variation <= 0.6	Ability to call new hotspot mutations

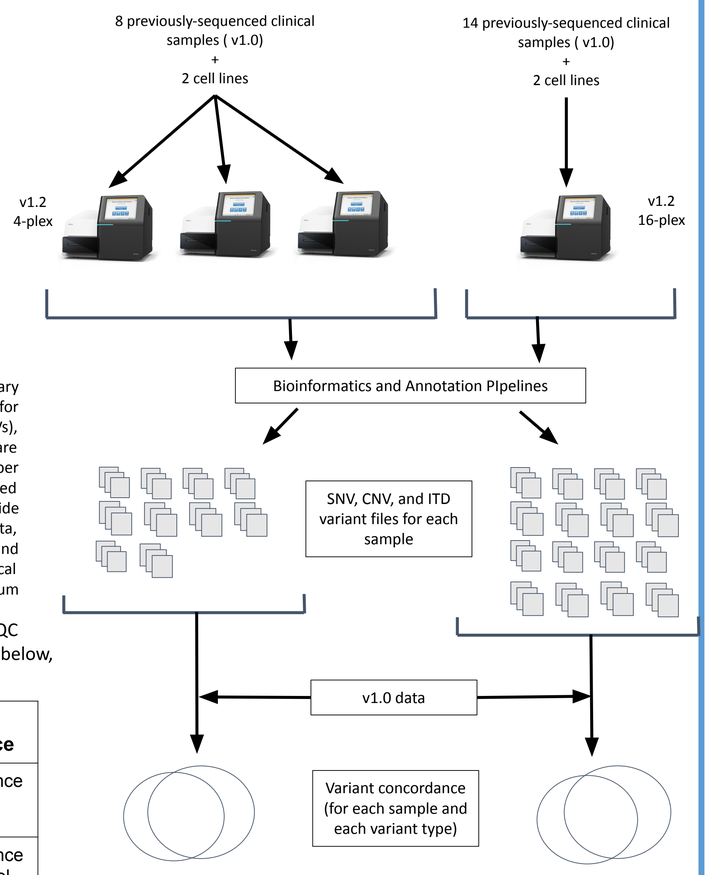


Figure 4. SearchLight DNA™ v1.2 Validation Design. Sequencing runs were designed using the above approach. Three 4-plex runs and one 16-plex run were performed using the validation sample set. 8 clinical samples were sequenced in both the 4-plex and 16-plex runs, and provided replicate data for the reproducibility analysis. 6 additional clinical samples were sequenced in 16-plex alone. All filtered variants identified with v1.2 were compared with those obtained from the initial v1.0 panel in a variant concordance analysis. In addition, v1.2 4-plex data was compared with v1.2 16-plex data for concordance between replicates of the 8 clinical samples sequenced twice (not shown in figure).

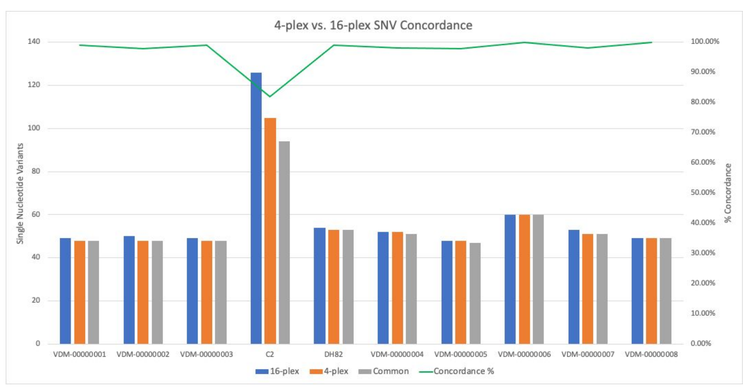
RESULTS

All new regions added to SearchLight DNA™ v1.2 met or exceeded coverage thresholds. Targeted hotspot mutations in PTPN11, FLT3, ESR1, KMT2D, and MEN1 were identified in the validation dataset. Small variant concordance in common regions between v1.0 and v1.2 was >95% and repeatability across sequencing runs was >95% for all clinical samples in our validation set. Internal tandem duplications in KIT were also identified using the v1.2 workflow. Copy number variant concordance between panel versions was ~86% and repeatability was >93%. We confirmed the absence of copy number events in all normal samples used to generate the CNV baseline, indicating high specificity. Bioinformatics workflow improvements and the addition of the Vidium InsightDB interface have greatly enhanced quality control and clinical reporting processes, ensuring fast turnaround times and high-quality, insightful data.

Table 2. SearchLight DNA™ v1.2 Run-Level QC. All run-level QC metrics passed QC thresholds. Table shows %PF and %Q30 across both 4-plex and 16-plex validation runs. On average, %PF and %Q30 were 94% and 95%, respectively.

Run	Plexing	% PF	% Q30
Run 1	4	92	94
Run 2	4	94	95
Run 3	4	93	94
Run 1	16	96	96

Figure 7. SearchLight DNA™ v1.2 4-plex and 16-plex SNV Concordance. We performed SearchLight DNA™ v1.2 sequencing of 14 clinical samples previously sequenced on v1.0. (**A**) SNV variant concordance between clinical samples sequenced in 4-plex and the v1.0 datasets for the same clinical samples. (**B**) SNV variant concordance between clinical samples sequenced in 16-plex and the v1.0 datasets for the same clinical samples. (**C**) SNV variant concordance between 4-plex and 16-plex data for the validation samples (clinical samples and DH82 cell line) in reproducibility analysis.



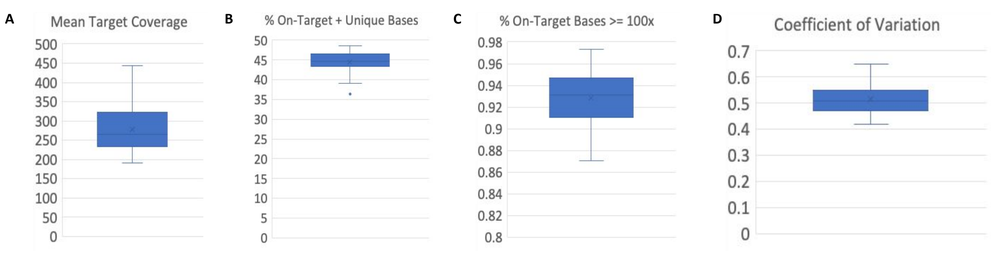


Figure 5. SearchLight DNA™ v1.2 Sample-Level QC. Average mean target coverage across the 4-plex and 16-plex validation datasets was 277x (A). Average on-target and unique bases covered was found to be ~45% (B), and average on-target bases covered with 100x or more was ~93% (C). Average CV was approximately 0.51 (D).

Table 3. SearchLight DNA™ v1.2 4-plex and 16-plex ITD Variants. We performed SearchLight DNA™ v1.2 sequencing of 14 clinical samples previously sequenced on v1.0. Data shown here are from the two samples known to have KIT ITDs, VDM-00000009 and C2 cell line. Results are obtained from Manta, the structural variant caller. Samples that did not have either the KIT or FLT3 ITD based on v1.0 also did not have those ITDs based on the v1.2 workflow. All ITDs were confirmed by visual inspection in IGV.

Sample	Plexing	v1.2 ITD Coordinates (Manta)
VDM-00000009	16	chr13:47178577-47178650
C2	16	chr13:47178578-47178623
C2	4	run1 - chr13:47178578-47178623 run2 - chr13:47178578-47178623 run3 - chr13:47178578-47178623

Table 4. SearchLight DNA™ v1.2 4-plex and 16-plex CNV Concordance. We performed SearchLight DNA™ v1.2 sequencing of 14 clinical samples previously sequenced on v1.0. Data shown below include the CNV variant concordance between clinical samples sequenced in 4-plex and the v1.0 datasets for the same clinical samples, CNV variant concordance between clinical samples sequenced in 16-plex and the v1.0 datasets for the same clinical samples, and CNV variant concordance between 4-plex and 16-plex data for the validation samples in our reproducibility analysis. Weighted concordance takes into account the size of the "truth" set of copy number events, where samples with 0 events have low weight and those with more events have a higher weight.

Analysis	Average Concordance %	Weighted Concordance %
4-plex vs. v1.0	92.96%	90.83%
16-plex vs. v1.0	81.56%	88.54%
4-plex vs. 16-plex	95.96%	94.68%

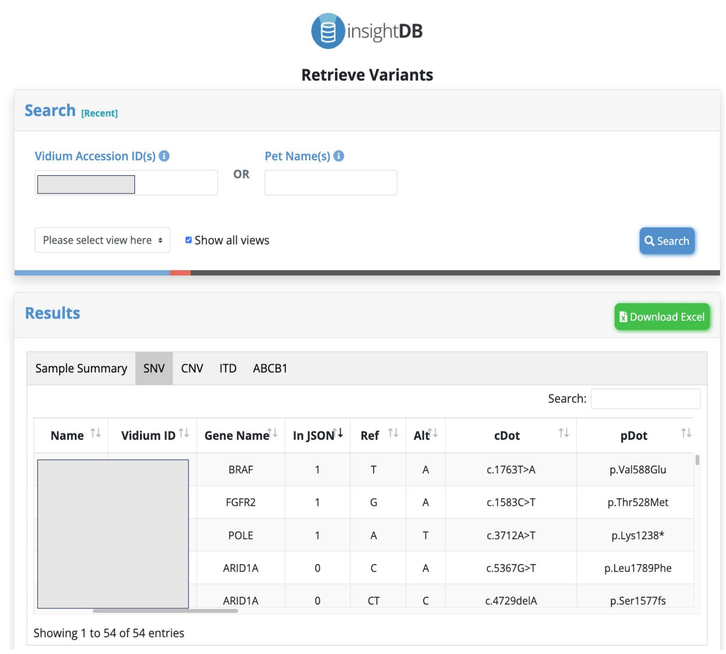


Figure 6. Vidium InsightDB. Pictured above is a screen capture of the custom web interface and database designed by Vidium that houses clinical sample metadata and genomic information used for reporting at the sample level and for biomarker discovery at the cohort level. It is able to display sample information such as sample name, accession ID, gender, breed, presumptive diagnosis, etc. Variant information for each variant type include gene name, location of the mutation, functional impact and effect, frequency, etc. Such a database will not only improve clinical reporting procedures but will also facilitate cohort analyses across hundreds of clinical cases, allowing us to expand our knowledge and understanding of canine cancers.

CONCLUSIONS

- SearchLight DNA™ v1.2 spans 500kb of sequence space across important canine cancer genes. Relative to v1.0, it contains additional probes capable of capturing 86 new hotspot mutations and 32 new regions to improve copy number calling, allowing us to identify genomic variants that may be highly relevant to canine cancer diagnosis, prognosis, and treatment.
- For SearchLight DNA™ v1.2 validation, we achieved our target metrics including: >95% average SNV concordance and >85% average CNV concordance across validation samples compared to v1.0 datasets, as well as reproducibility >95% for SNVs and >93% for CNVs across v1.2 replicate data.
- The ability to version and validate genomic diagnostic panels such as SearchLight DNA™ as performed here is critical for ensuring high quality while enabling incorporation of the latest emerging biomarker data.

FUNDING & ACKNOWLEDGEMENTS

We thank all veterinarians and pet parents who pursued SearchLight DNA for genomic analysis. *Financial disclosures: This study was funded by Vidium Animal Health (VAH), a TGen subsidiary. All authors are employees of TGen and Vidium Animal Health.*