

## Digging deeper into OncoSpan's NGS data

Bernice Freeman, Jesse Stombaugh & Janine Ostick  
Horizon, a PerkinElmer company

### Introduction

The OncoSpan product range offers the world's largest oncology reference standard set with a companion high coverage, batch specific in silico NGS data set. Available in cell-free DNA (cfDNA), formalin fixed paraffin embedded (FFPE) curl and genomic DNA (gDNA) formats, these products are accompanied by FASTQ, BAM and VCF files for in depth analysis of exome sequencing data.

All of Horizon's OncoSpan reference standards are cell line-derived, to provide a close representation of the genomic complexity found in patient tumor samples. These reference standards are commutable (performing similarly to a patient genome sample during sequencing and analysis). Derivation from cell lines means a reliable source of reproducible, high quality reference standards is available for use in your end-to-end process assay control.

Here we explore deep sequencing information provided with the OncoSpan product range.

### What files are provided with OncoSpan Products?

For our OncoSpan customers, Horizon provides raw FASTQ, BAM and VCF files for each product. The README.txt file contains the MD5 for each file, so you can confirm the files downloaded are not corrupt. A list of files provided for each product batch can be found in Table 1.

File Type	OncoSpan		
	gDNA	FFPE	cfDNA
Raw FASTQ	✓	✓	✓
Genome build GRCh37/hg19 specific BAM	✓	✓	✓
Genome build GRCh38/hg38 specific BAM	✓	✓	✓
VCF	✓	✓	✓

Table 1: File types provided for the OncoSpan product range

BAM-formatted files contain the recalibrated, deduplicated, aligned reads for a specific batch of OncoSpan.

For each genome build shown in Table 1 above, we provide 3 VCF-formatted files:

- 1) high-confidence list of 25 ddPCR-validated variants;
- 2) high-confidence list of >380 NGS-validated variants which includes 25 ddPCR-validated;
- 3) other variants for the exome (the output from the variant caller used for that specific batch) which is subject to batch-to-batch variability.

Each VCF file contains a list of variants and the batch specific allele frequency and read depth. For each product, we also provide a Microsoft Excel spreadsheet, which lists out all the high-confidence variants, their average allele frequencies and read depths across multiple batches, COSMIC (v85-v90)/dbSNP (v151) annotations and average ddPCR allele frequencies, if applicable.

### What informatics tools are used for Horizon's variant calling?

Historically some of our early OncoSpan gDNA batches were processed through the GATK Best Practices Pipeline by our sequencing provider, and the variant caller used was either GATK HaplotypeCaller or LoFreq. For current batches of gDNA (Batch 44799 onwards) and all batches of the FFPE and cfDNA products, raw reads are run through the Illumina DRAGEN Somatic Pipeline, which performs mapping, sorting and variant calling. Using the same pipeline allows for consistency and comparability across all three OncoSpan products.

### How are these files generated?

All OncoSpan samples were sequenced on the Illumina platform using the Agilent SureSelect Human All Exon V6 kit. Coverage depth is shown in Table 2 below.

OncoSpan Product type	Coverage depth
gDNA	500x
FFPE	500x
cfDNA	*120x

Table 2: Coverage depth for OncoSpan products

\*NGS provider maximum coverage for cfDNA

### How did Horizon perform product validation?

For OncoSpan FFPE and cfDNA, three batches of each product were sequenced by using a combination of whole exome sequencing (WES) and ddPCR validation to produce a list of high confidence variants. Only variants observed in all 3 batches (and the original OncoSpan high-confidence list) were considered for the high-confidence variant list as shown in Figure 1 below. The initial gDNA products were validated under a more complex process, combining WES and ddPCR validation.

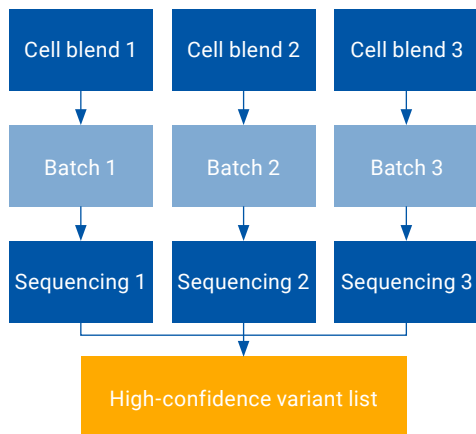


Figure 1: Validation process flow for OncoSpan products

The average NGS allele frequency and read depth, along with the average ddPCR allele frequency across all 3 batches can be found in the Excel file, which can be downloaded from the product pages. It should be noted that only OncoSpan gDNA NGS files from Batch 44799 onwards have data from three batches to provide an equally robust high-confidence variant list.

Want to learn more about how Horizon's Reference material can aid in identifying true positives? Learn more on how to get the most from your NGS data in our blog post here.

Want to get started using OncoSpan products for NGS sequencing controls? Take a look at our OncoSpan Product range

### Written by Bernice Freeman, Jesse Stombaugh & Janine Ostick



Bernice is a Senior Scientist in the Diagnostic's team at Horizon Discovery, UK. She obtained her First-class BSc in Human Biosciences at Coventry University and has been working within the Diagnostics team for over 6 years, developing and improving Horizon's Oncology reference standard. Previously, she worked in Phase II human clinic trials at hVIVO, investigating various targeted drugs to respiratory viruses.



Jesse is a Senior Bioinformatics Scientist in the Bioinformatics team at Horizon Discovery, US. He obtained his PhD from Bowling Green State University, where he focused on RNA 3D structure prediction. At Horizon, he specializes in developing and scaling software solutions for processing and analyzing NGS data. These solutions are used for developing new products and assessing the quality of our current products.



Janine is a Scientific Support Specialist in the Technical Support team at Horizon Discovery, UK. She obtained her PhD from the University of Cambridge, focussing on iPSC derived neurons and organoids to study mutations in MAPT. At Horizon, she assists customers with their scientific questions about Horizon's products.

### References

1. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 2010;26(17):2204-2207. doi:10.1093/bioinformatics/btq351
2. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178-192. doi:10.1093/bib/bbs017
3. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754
4. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033
5. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinforma*. 2014;2014.11.12.1-11.12.34. doi:10.1002/0471250953.bi1112s47

### For more information

To find the contact information in your country for your technology of interest, please visit us at [horizondiscovery.com/contact-us](http://horizondiscovery.com/contact-us)

Horizon Discovery, 8100 Cambridge Research Park, Waterbeach, Cambridge, CB25 9TL, United Kingdom  
WF691800

©2022 Horizon Discovery Group Ltd and its affiliated companies. All rights reserved.

PerkinElmer is a trademark of PerkinElmer, Inc., registered in the United States and other countries. Horizon Discovery is a trademark mark of Horizon Discovery Ltd. Dharmacon is a trademark of Dharmacon, Inc. All PerkinElmer, Inc., Horizon Discovery Ltd and Dharmacon, Inc. trademarks are used with permission. Other product names and brand names may be trademarks or registered trademarks of their respective owners.

**horizon**  
a PerkinElmer company