

# Digging deeper into OncoSpan's NGS data

---

## Authors

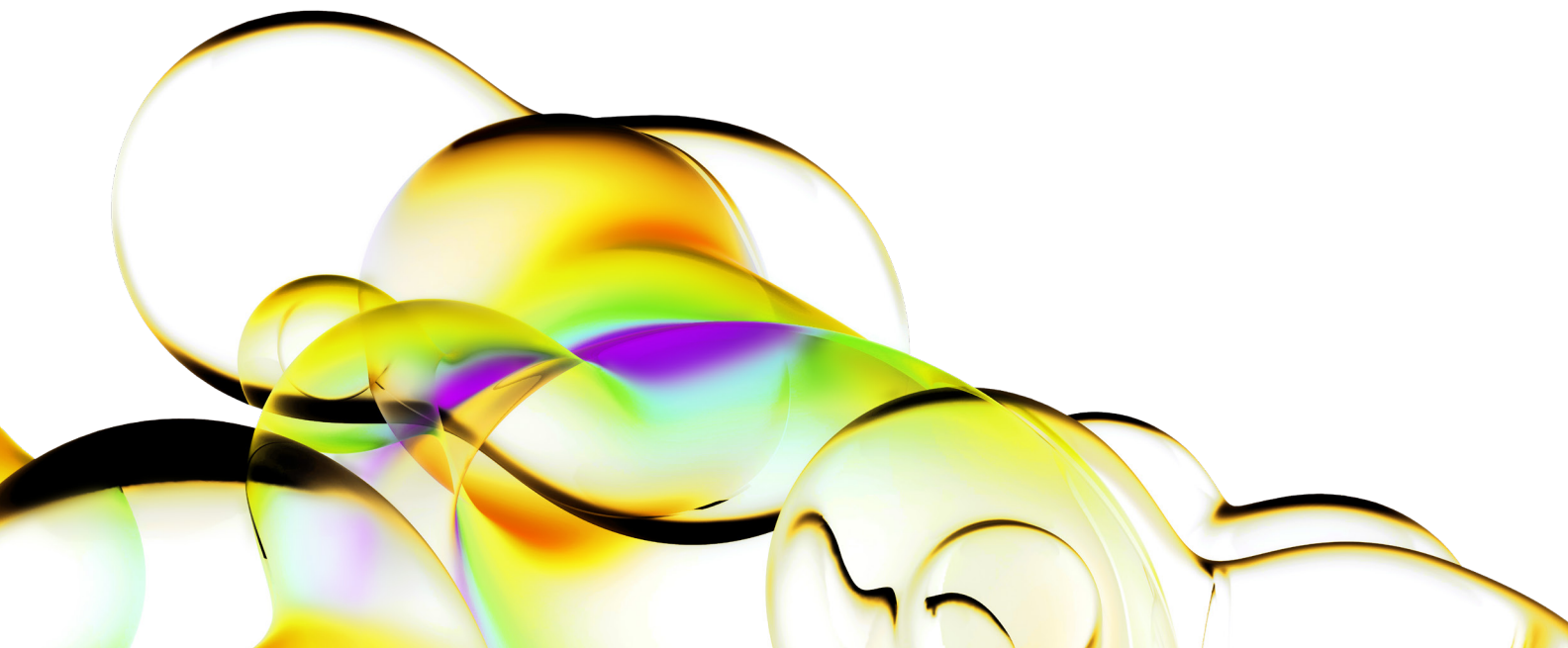
Bernice Freeman,  
Jesse Stombaugh &  
Janine Ostick  
Revvity

## 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 OncoSpan reference standards are cell line-derived, to provide a close representation of the genomic complexity found in tumor samples. These reference standards are commutable (performing similarly to a 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, Revvity 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.

Table 1: File types provided for the OncoSpan product range

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

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, and average ddPCR allele frequencies, if applicable.

## What informatics tools are used for 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.

Table 2: Coverage depth for OncoSpan products

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

\*NGS provider maximum coverage for cfDNA

## How was product validation performed?

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 highconfidence 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.

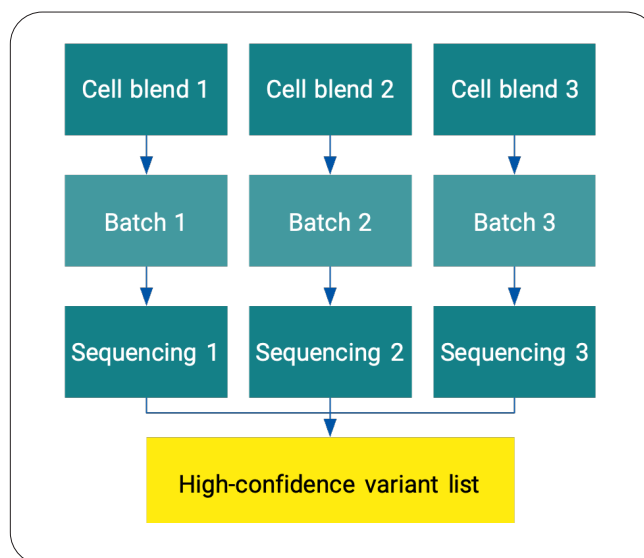


Figure 1: Validation process flow for OncoSpan products

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.

## References

4. 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
5. 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
6. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26. doi:10.1038/nbt.1754
7. 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
8. 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

Want to learn more about how Revvity's Mimix™ 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 [here](#).

The Revvity logo is displayed in a lowercase, sans-serif font. It is positioned in the lower right quadrant of the page, above a large yellow wavy graphic that spans the bottom of the page. The logo consists of the word "revvity" in a clean, modern typeface.