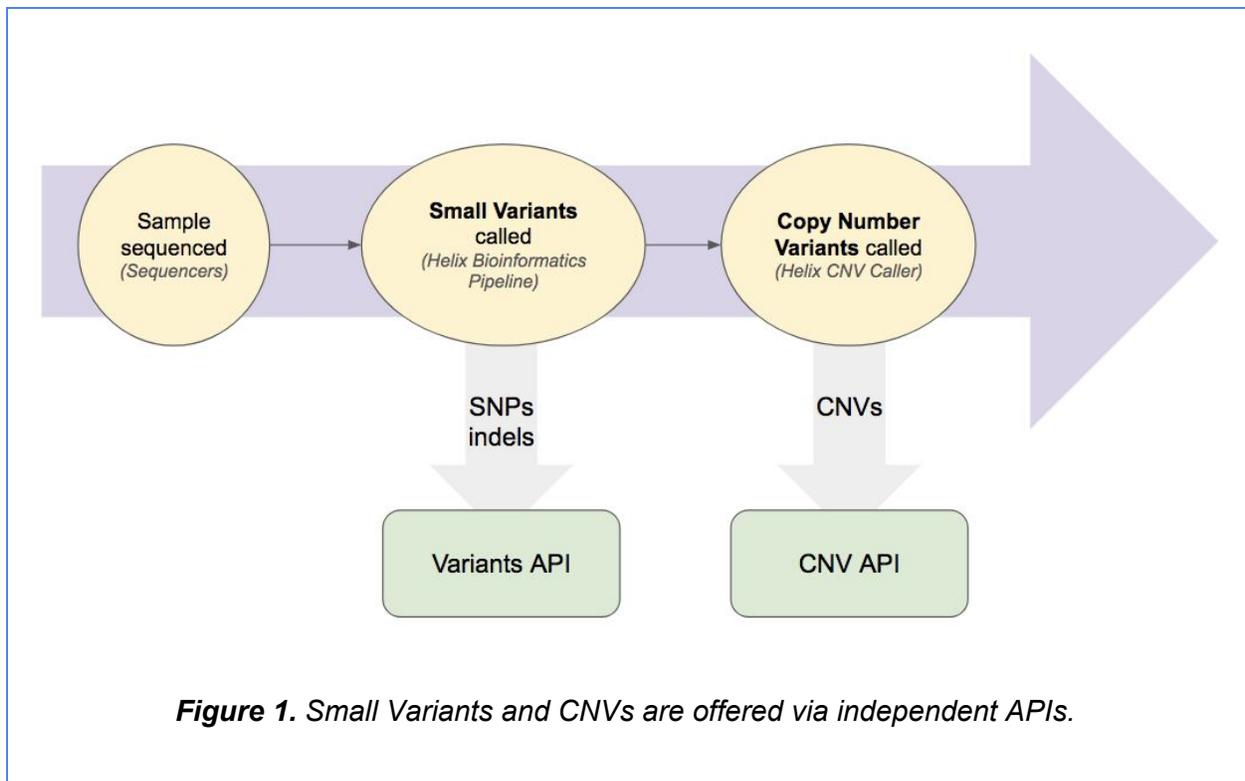


Helix Copy Number Variations White Paper

April 4, 2019

The Helix CNV Caller delivers clinically validated copy number variations (CNVs) across the Helix Exome+, an assay that queries the full human exome as well as hundreds of thousands of non-coding targets in Helix's CLIA-certified and CAP-accredited production laboratory. These results, delivered through a separate API endpoint from our small variants endpoint, offer an expansion on the type of genetic variation that can be detected, studied, and reported upon in the Helix model (*Figure 1*).

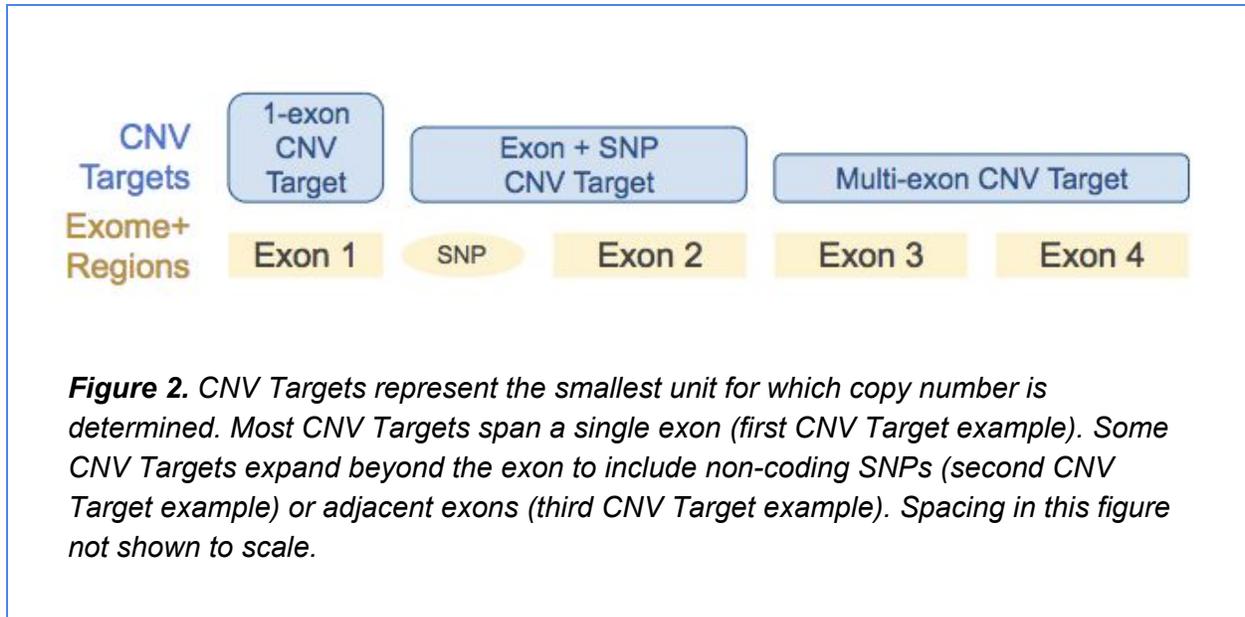


Detection of Exome-wide CNVs

CNV Targets are the Units Used for CNV Measurement

Prior to identifying CNVs, the Exome+ is segmented into hundreds of thousands of CNV Targets, which are predefined genomic regions representing the smallest unit for which copy number is determined. The majority of CNV Targets represent a single exon, but due to proximity in space, a small subset of CNV Targets span multiple exons or include nearby

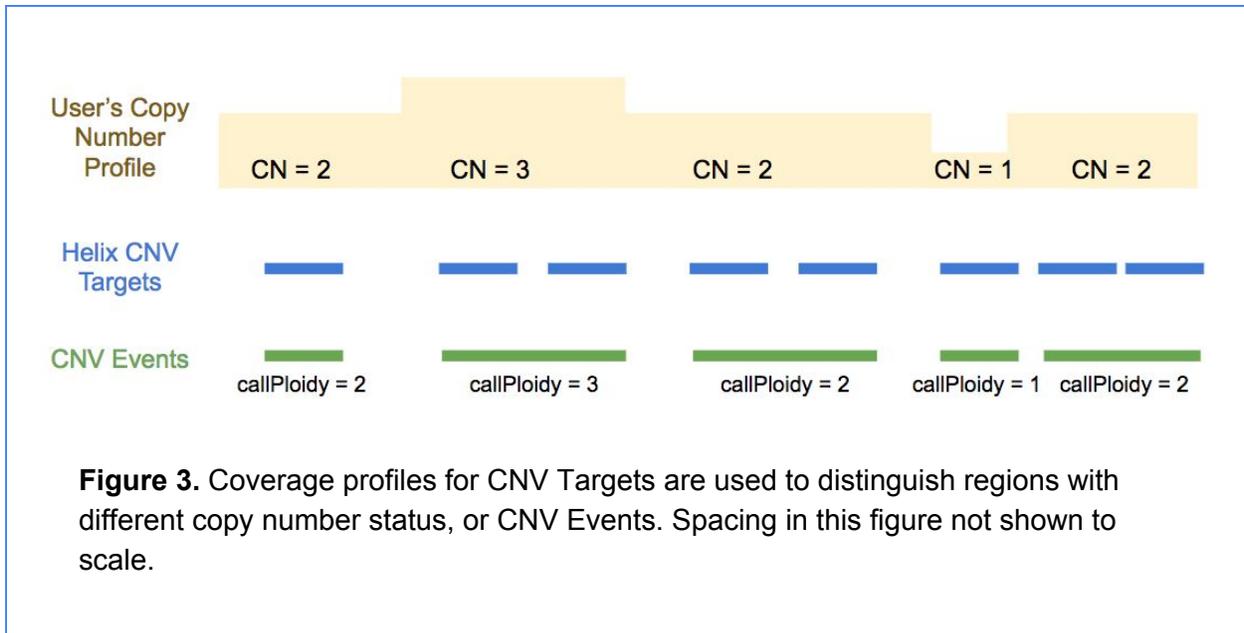
non-coding SNPs (*Figure 2*). The distribution of CNV Targets are defined per version of the Exome+, leading to a slight difference in CNV Targets between Exome+ v1 (retired in November 2018) and Exome+ v2 (currently live in production).



CNV Targets were created based on the distribution of Exome+ regions and then assessed to determine if they qualify for inclusion in CNV analysis. CNV Targets that did not meet the quality requirements were blacklisted, resulting in their exclusion prior to CNV analysis. These quality checks included: (1) a check to remove CNV Targets with a high no-call rate; (2) a check to remove CNV Targets that cover very common CNVs; and (3) a check to remove CNV Targets where the underlying sequence is not unique within the genome.

Determining Copy Number of CNV Targets

Copy number is determined by organizing CNV Targets into CNV Events, where a CNV Event is comprised of adjacent targets sharing the same ploidy. Ploidy is determined by comparing read depth per CNV Target for each individual against the expected profile of read pair counts as determined by a baseline of many individuals sequenced through the same processes at the same time. A tuned Hidden Markov Model (HMM) takes into account the expected dispersion for a CNV Target and the likelihood of transitions to new copy number states between CNV Targets to inform the start and end of each CNV Event as well as the copy number status for that CNV Event (output as **callPloidy**, see *Figure 3*).



CNVs are identified as CNV events where *callPloidy* does not equal *expectedPloidy*.

Helix CNV Caller Performance Metrics

Performance metrics across the full Exome+ were evaluated using data generated using both Exome+ v1 and Exome+ v2, as described in *Table 1* and in the sections below.

	Exome+ v1	Exome+ v2
Sensitivity	97.7% SE: 1.0% - 194 replicates across - 54 samples supporting - 69 documented CNVs	97.9% SE:1.4% - 235 replicates across - 53 samples supporting - 68 documented CNVs
Specificity	99.994% SE: 0.0005% - 8 replicates of NA12878	99.998% SE:0.0001% - 61 replicates of NA12878
Call Rate	99.99% SE:0.0003% - 351 samples	99.99% SE:0.0003% - 351 samples

Table 1. Sensitivity, Specificity, and Call Rate of the Helix CNV Caller for Exome+ v1 and Exome+ v2. For each metric, the estimate and standard error (SE) are reported, as well as a description of the samples used to generate these results.

CNV Sensitivity

In order to determine the sensitivity of the Helix CNV Caller, a wide range of samples with documented CNVs ranging from single-exon events to larger cytogenomic events were run through our standard production processes.

For Exome+ v1:

- All CNVs spanning two CNV Targets or more were detected successfully.

For Exome+ v2:

- All missed CNVs spanned a single CNV Target except one two-target CNV was missed in one replicate (but was detected in other replicates).

CNV Specificity

Relying on eight replicates of NA12878 for Exome+ v1 and 61 replicates of NA12878 for Exome+ v2, specificity was determined by first collecting all documented deletions across the Exome+ based on a map of CNVs published in Mills et al.¹. From there, all CNVs detected in the NA12878 replicates were compared to the known deletions and any novel CNVs were considered False Positives (FPs). Confirmation was not applied to the presumed FPs. Since the gold standard used includes only deletions, it remains possible that some duplications counted as FPs could in fact be TPs, leading to an underestimation of exome-wide CNV specificity.

CNV Target Call Rate

The Helix Exome+ is represented by hundreds of thousands of CNV Targets whose sensitivity and specificity are estimated above. In order to make a call on copy number, we require a quality score *callQuality* ≥ 20 . Where *callQuality* < 20 , the CNV Target is no-called, reducing the CNV Call Rate.

CNV Event Call Rate

The results of the CNV Caller described above apply to the entire Exome+, yet most partners will be most interested in CNVs in regions of specific interest to their product, such as genes involved in predisposition to cancer or cardiovascular disease.

To understand the frequency of CNVs identified across a relevant section of the exome, we counted the number of CNV Events identified across the ACMG59² genes, using 2,000 female and 2,000 male production samples that pass CNV QC on Exome+ v2 (*Table 2*). Currently, confirmation is expected for CNVs delivered in a clinical setting, and so these results can

provide an indication to how common confirmation will be expected for an ACMG59-like product.

ACMG59 CNV Results, across 4,000 samples	
Count of CNV events	22 events
Count of samples with CNV events	21 samples
% of samples with 0 CNV events	99.45%
% of samples with 1 or 2 CNV events	0.55%

Table 2. Characteristics of CNV Events Spanning the ACMG59 genes. Description of the frequency of CNV events identified across 4,000 samples (half male, half female) overlapping the ACMG59 genes.

Helix CNV Caller Limitations

The Helix CNV Caller has the following limitations:

- Non-unique regions such as PMS2, exons 12-15, are outside of the reportable range.
- Events smaller than a CNV Target are likely to be missed, else they are reported as if they represent the full CNV Target.
- In the case of whole chromosome aneuploidy or large but partial chromosome aneuploidy, the entire chromosome is excluded from analysis. An exception is that CNVs will continue to be called in the presence of Trisomy 21.
- If sex is not inferred by the Helix Bioinformatics Pipeline, which can occur in the presence of sex chromosome aneuploidy, then chrX is excluded from analysis.
- ChrY and chrM are outside of the reportable range.
- Mosaic events and structural variations such as inversions and translocations are outside of the Helix CNV Analytical Range.
- CNV results must be confirmed by a diagnostic laboratory prior to making any medical decisions or taking any medical actions.
- Individuals who are symptomatic, pregnant (if performing carrier testing), or who have a family history should be directed toward comprehensive diagnostic testing in lieu of using a Helix-supported screen. Examples of such a family history include:
 - Onset of disease at an earlier age than population average, and/or
 - Family history of the same disease multiple times in multiple relatives (e.g., multiple relatives diagnosed with an arrhythmia), and/or

- Personal and/or family history suggestive of a syndrome (e.g., colon and uterine cancer in the same side of the family can indicate the family is at an increased odds to have Lynch syndrome), and/or
- Personal and/or family history of a diagnosis of a rare disease.

References

1. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, et al. (2011) Mapping copy number variation by population-scale genome sequencing. *Nature* 470: 59–65. <https://www.nature.com/articles/nature09708>
2. NCBI Genome Remapping Service. <http://www.ncbi.nlm.nih.gov/genome/tools/remap>
3. ACMG 59. <https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/>