Canopy Biosciences
4320 Forest Park Ave. #303
St. Louis, MO 63108
info@canopybiosciences.com

What is RareSeq?

It is a Next Generation Sequencing Service (NGS) that consists of,

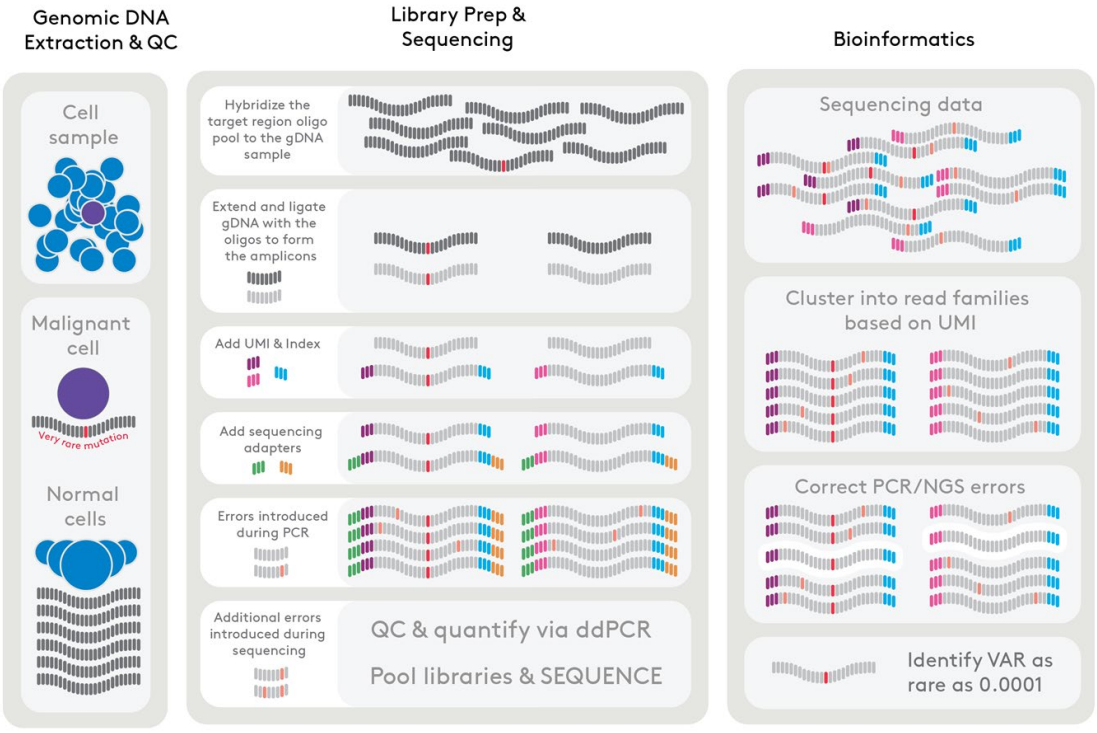
1. Proprietary NGS Library Prep Protocol that includes UMIs
2. Custom Bioinformatics Pipeline

RareSeq combines Illumina® TruSight® Myeloid Sequencing Panel with Error-corrected Sequencing (ECS) to identify rare variant alleles in genes that are frequently mutated in myeloid malignancies. This is an amplicon-based approach with 568 amplicons representing 54 genes. Libraries are generated from gDNA and they represent target regions defined by TruSight® oligos from Illumina®. Sample-specific indices and unique molecular indices (UMI) are added to each library to enable pooling of libraries and clustering of read families. Pooled libraries are loaded into the HiSeq3000 system for sequencing at a depth of > 5000x. We employ our own proprietary bioinformatics pipeline to call variant alleles and it can detect variant allele frequencies as low as 0.0001 with high confidence.

Why is error correction important?

Next-generation sequencing has an error rate of 1-3% and low abundant, rare mutations are often dismissed as errors. With error correction, we can identify these rare mutations and the ability to identify rare variants depends on limit of detection and depth of sequencing.

What is the RareSeq workflow?



What Limits of Detection are offered through RareSeq?

Limit of Detection (LOD)	Error rate	Variant Allele Frequency (VAF)	Sensitivity over standard NGS
1:10,000	0.01%	0.0001	100X
1:5,000	0.02%	0.0002	50X
1:2,000	0.05%	0.0005	10X
1:100 Standard NGS (without a UMI)	1-5%	0.01	Baseline

What can RareSeq detect?

At the LODs offered, our read depths can vary from 2200x to 10,035x. High read depth allows us to detect rare somatic nucleotide variants and indels that are 50 bp or less.

Our waterfall plot summarizes types of mutations at a cohort level for the 54 genes represented in the Illumina® TruSight® Myeloid Sequencing Panel. The sample waterfall plot below shows mutations in 25 of the 54 genes for a cohort of 30 individuals.

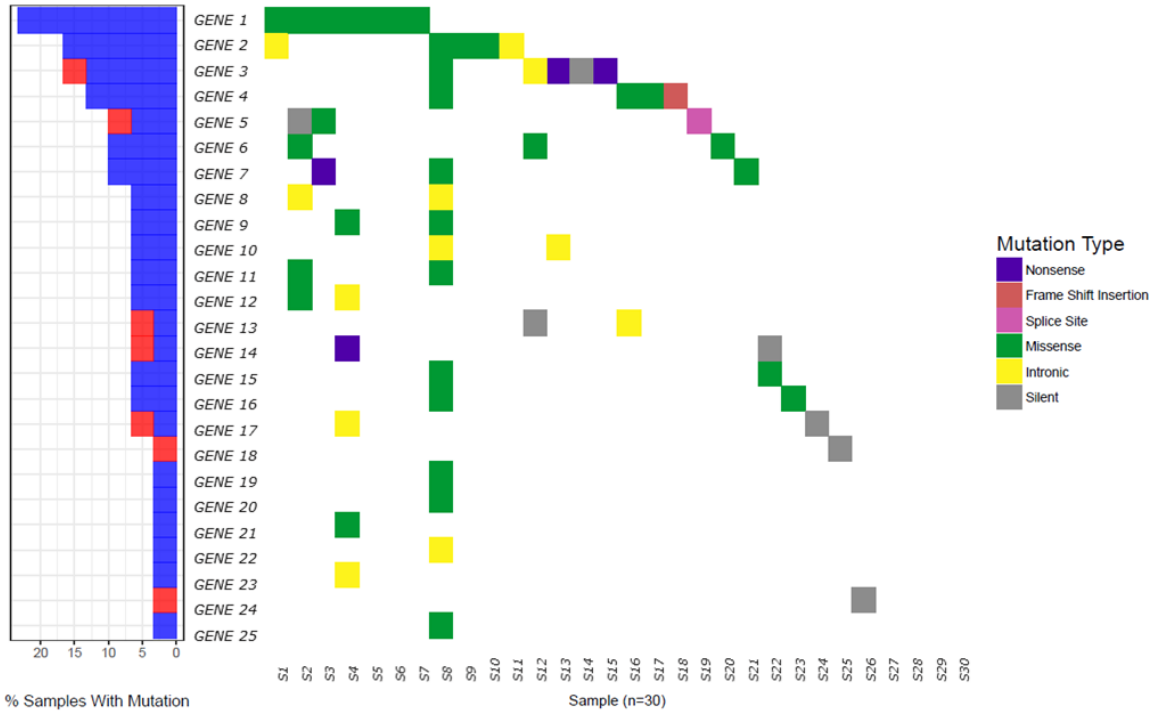


Figure 2. Mutation landscape plot showing different types of mutations in 25 genes of interest.
Columns represent each sample in the cohort and are ordered by the presence of mutations in the most to least frequently mutated gene. The bar graph on the left corresponds to the frequency of mutations for that gene in the entire sample cohort. Nonsense, a premature nonsense or stop codon. Frame shift insertion, insertion causing a shift in the reading frame. Splice site, a mutation affecting splicing. Missense, a single nucleotide change coding for a different nucleotide. Intronic, a nucleotide change in introns. Silent, a change in the nucleotides without a subsequent change in the amino acid or the function of the overall protein.