Cancer Analysis Service

The most accurate data for comprehensive cancer research studies.

Highlights

- **Robust Combined Somatic Variant Calling**
  Best method for analysis of non-homogeneous tumor samples
- **Most Widely Adopted and Published NGS Technology**
  Cited in over 3,700 publications*
- **Most Interpretable Data**
  Data output in universal format for seamless integration with third-party analysis software
- **Lowest Sample Input Requirement**
  Only 500 ng sample needed, conserving precious samples for follow-on studies

Highly Accurate Data Advances Cancer Research

Researchers need high-confidence genomic data to decipher the mechanisms of cancer. The experts in the Illumina Genome Network (IGN) provide a Cancer Analysis Service that meets this need and more. Requiring the lowest sample input (500 ng) of any sequencing service, IGN delivers accurate, easily interpretable, and highly actionable whole-genome sequencing (WGS) data of tumor/normal samples.

IGN Experts Leverage Proven Technologies

IGN consists of genomic groups at leading institutions worldwide, all well-published experts in Illumina technologies, offering sequencing services to individual researchers. Using proven technologies and methods from start to finish, the IGN Cancer Analysis Service generates high-quality WGS data using industry-leading Illumina HiSeq® systems powered by TruSeq® technology. Data analysis includes IGN's optimized combined calling method, ensuring the most accurate somatic mutation calls for non-homogenous tumor samples. Delivered in industry-standard format, data sets are ready for visualization and downstream analysis using commercial and open-source third-party software. In addition, data can be seamlessly integrated with follow-on genotyping, RNA-Seq, and methylation analysis studies to obtain a clearer, more comprehensive understanding of cancer mechanisms.

Small Sample Size Yields Big Results

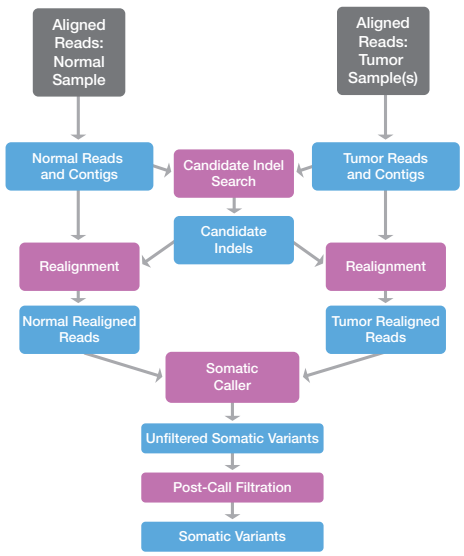
IGN's Cancer Analysis Service has the lowest sample input requirement (500 ng) of any cancer sequencing service. A cost-effective solution for small and large cohort studies, the service delivers at least 40× average post-alignment coverage of the normal genome and at least 80× average post-alignment coverage of the tumor genome. These IGN-recommended coverage levels have been optimized to deliver highly accurate somatic calls**. Per sample data analysis (read mapping and variant calling of SNVs, indels, SVs, and CNVs) is performed for tumor and normal samples using Illumina's Isaac† Alignment and Variant Calling workflow and ancillary software. Cancer Analysis results are available as an add-on analysis, in conjunction with standard WGS services, for the detection of somatic SNVs and small indels using the combined calling analysis method[1]. The results can also be used for the detection of somatic copy-number aberrations, and structural variants such as large indels (> 50 bp), duplications, translocations, and inversions.

Ensuring Accurate Somatic Mutation Call Rates

Cancer samples can possess significant heterogeneity at the genetic and histological levels. Most primary tumor biopsies are likely to contain multiple sub-clones of cancer cells, as well as normal cell contamination. Competing service offerings make variant calls for the tumor sample against a standard reference and again for the normal sample against the reference, subtracting the results to identify the somatic calls. This approach does not properly model the complexities of multiple cancer sub-clones or normal sample contamination.

IGN's Bayesian combined-calling method eliminates this problem by modeling normal samples as mixtures of diploid germline variation and noise, and tumor samples as mixtures of normal cells and somatic

Figure 1: Somatic SNV and Small Indel Identification Workflow



Somatic SNVs and small indels (≤ 50 bp) are identified by a Bayesian combined calling method where both the tumor reads and the normal reads are considered.

*Publications from introduction of technology in 2006 to May 2013

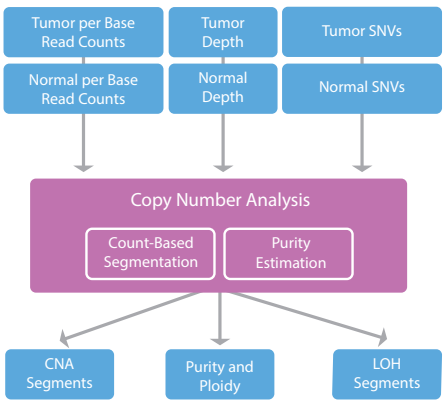
**Additional coverage levels are available

† Illumina Sequence Analysis and Comparison software

variation. By more accurately modeling real-life tumor samples, IGN's combined-calling method recovers 97% of known SNVs. Here's how it works:

Once the sequencing reads from each genome are properly mapped to the reference, the IGN somatic caller combines tumor and normal sample data sets together (combined calling) to better model the varying levels of tumor purities present (Figure 1). The combined analysis takes into account the somatic variation and noise that can occur at any allele frequency ratio, and reports somatic SNVs and small indels as output. Structural variants (SVs) are obtained by comparing the SVs identified in the tumor and the SVs in the matched normal. Somatic copy number aberrations (CNAs), as well as tumor purity and ploidy estimates, are derived from a count-based method to perform read depth estimation for tumor and matched normal samples[2] (Figure 2).

Figure 2: Copy Number Aberration Workflow



CNAs, tumor purity, and ploidy estimates are derived from a count-based method to perform read depth estimation for tumor and matched normal samples.

The advantage of the combined calling method can be best seen in the results of a melanoma cell-line study (Figures 3). By mixing a tumor sample data set with fractions of a normal sample data set, a range of test data sets was created ranging from 40–100% tumor purity. The data were analyzed using IGN's combined calling method, with the results demonstrating its superior sensitivity (> 97.00%) and specificity (> 98.5%) for calling somatic variants and true variants, even when the tumor purity is as low as 40%.

Research-Ready Data Set for Downstream Analysis

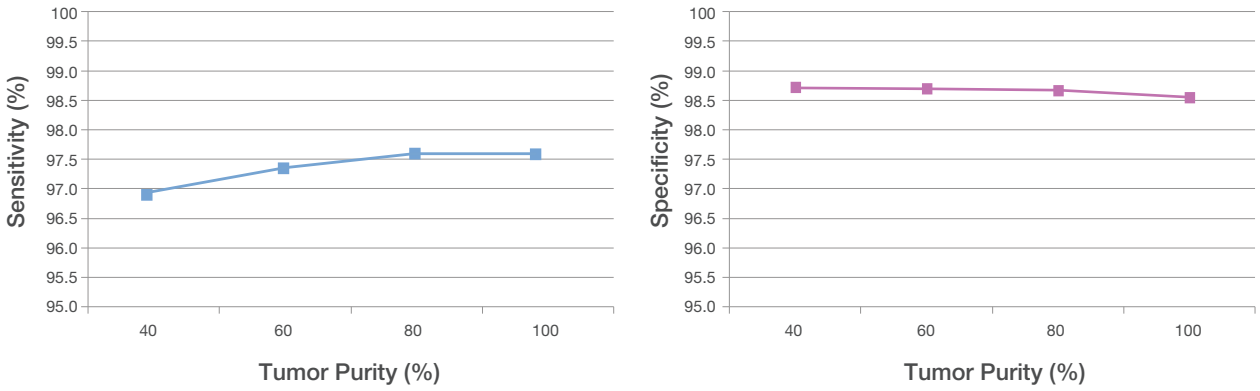
Data sets are provided in the most widely adopted formats (BAM and VCF) for streamlined analysis, and include called variants for the tumor and normal genomes against the reference, as well as somatic mutations for the tumor genome against the paired normal genome (Table 1). Illumina has partnered with leading data analysis and visualization companies, including Knome and Diagnomics, to provide optimized software tools for downstream filtering and interrogation of IGN data sets that will deliver biologially relevant and actionable results.

Follow-on projects, such as genotyping, RNA-Seq, and methylation studies, can be performed by IGN or Certified Service Provider (CSPro®) partners using Illumina systems and products, enabling a clearer understanding of cancer and the biological pathways it impacts.

Cancer Analysis Service without Compromise

Powered by TruSeq technology and utilizing an optimized combined-calling method, the IGN Cancer Analysis Service provides researchers with easily interpretable tumor/normal sequencing results. With the lowest sample input volume requirement in the service industry, researchers can sequence the most tumor samples, while conserving them for later follow-on studies. In all, the IGN Cancer Analysis Service provides researchers with the confidence to take the next step in their research. Learn more at www.illumina.com/genomenetwork.

Figure 3: IGN's Combined Calling Method Enhances the Ability to Call True Somatic Variants



Even at low tumor purity, sensitivity (ability to call variants) and specificity (ability to call true variants) for somatic SNVs remain high.

References

1. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, et al. (2012) Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 28: 1811–1817.
2. Ivakhno S, Royce T, Cox AJ, Evers DJ, Cheetham K, et al. (2010) CNAseg—a novel framework for identification of copy number changes in cancer from second-generation sequencing data, Bioinformatics 26:3051–3058.

Table 1: IGN Cancer Analysis Service Details

Sample Input Requirements	Normal Sample	Tumor Sample
Input DNA	0.5–1 µg	0.5–1 µg
Minimum Average Autosomal Fold Coverage	40×*	80×*
Percent of Non-N Reference Coverage	> 90%	> 90%
Minimum Number of SNVs Detected	3 million	3 million
Genotype Concordance	only with 1 µg service	only with 1 µg service

		WGS Data	
Data Deliverables	**Somatic Variants**	**Normal Sample**	**Tumor Sample**
Variants (VCF format)			
Single-Nucleotide Variations (SNVs)	✓	✓	✓
Indels (1–50 bp)	✓	✓	✓
Structural Variants (VCF Format)			
Large Deletions	✓	✓	✓
Large Insertions	✓	✓	✓
Inversions	✓	✓	✓
Segmental Duplications	✓	✓	✓
Translocation	✓	✓	✓
Interchromosomal	✓	✓	✓
Intrachromosomal	✓	✓	✓
Reads (Archival BAM)			
Aligned and Unaligned Reads with Compressed Quality Scores	N/A	✓	✓
Annotations/Other Deliverables			
dbSNP Comparisons	N/A	✓	✓
Gene and Functional Impact Annotations	✓	✓	✓
Disease Association	✓	✓	✓
Allele Frequencies	N/A	✓	✓
Concordance with Genotyping Arrays	N/A	✓	✓
Raw .idat Files for Genotyping Arrays	✓	✓	✓
Summary Report	✓	✓	✓

*Recommended coverage, other coverages available.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 770-2012-005 Current as of 16 April 2014

illumina®