Genformatic Genomic Services Overview

Genomic Analysis Services

Genformatic is dedicated to delivering the most accurate genome informatics solutions for genomic research and medicine – providing more accurate variant detection and deeper insights into the functional impact of variants, including health risks, medical implications and other biological information encoded in every genome. The cost of generating high-quality genome data has dropped dramatically over the last decade, the challenge has become how to efficiently and accurately interpret this abundant data. Genformatic has developed innovative bioinformatic solutions for the most important problem in genome sequence analysis – more accurate variant detection, annotation and interpretation. Better tools and methods for decoding, understanding and using genomic information are critical unmet needs of researchers, and Genformatic's algorithms, bioinformatic pipelines, and annotation, visualization and interpretation tools help investigators solve those problems. Genformatic has an array of complete, easy-to-use, analysis pipelines ready to run in the cloud, or available for installation on your local hardware. We have engineered solutions for Whole genomes, Exomes, Methylomes and RNA sequence data. We also offer turnkey service solutions – pass us your data and we'll perform the analysis for you.

Genformatic's tools, analysis pipelines, database content offer complete genome informatics solutions for researchers and clinicians. Genformatic has implemented superior methods for accurately identifying and characterizing important genomic sequence changes, including SNVs, Indels, CNVs and SVs. Our multi-caller variant detection methods increase sensitivity to minimize false negatives, while delivering a statistically rigorous posterior probability metric that provides unprecedented specificity when false positives must be eliminated. Overall our methods assure superior analytic validity and clinical utility. Our informatics solutions are also engineered for efficient computation as well as superior accuracy, minimizing user costs while delivering the highest quality analysis. Our genome informatics solutions are available in the cloud for quick, easy use – the cost is based upon the number of samples and data volume processed., For unlimited applications our analysis solutions are available by enterprise license. Alternatively, we offer a complete service solution if you prefer that we process your data and deliver the results to you. Pass us the sequence data, along with an analysis objective, and we'll quickly and completely analyze your data and deliver a finished report.

Genformatic uses BAYSIC, a novel Bayesian method to improve the specificity of callsets from multiple somatic mutation or variant detection algorithms. First, we produce a BAM file and run multiple callers to produce several VCFs, and then run BAYSIC on the multiple VCFs to generate a single, more accurate BAYSIC call set. Genformatic also employs RUFUS, which finds unique kmers in tumor or proband sequence data. RUFUS defines de novo mutations in malignant tissue versus normal tissue, and germline variants in cases versus controls with unprecedented accuracy.

Beyond its novel methods for improved variant detection, Genformatic has pre-engineered a complete genome sequence analysis pipeline, running standard somatic mutation or germline callers to generate multiple call sets, followed by BAYSIC call set integration for improved accuracy, and finally annotation and interpretation of probable variant effects, including the potential medical implications of detected variants, using a variety of databases.

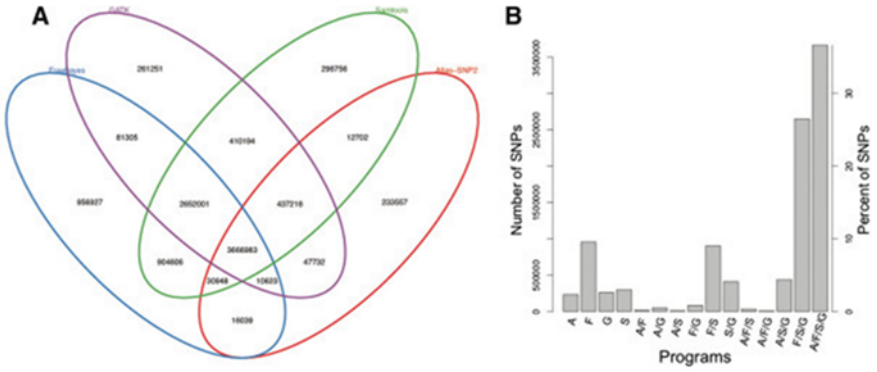
Genformatic also offers Genome Cruiser to visualize, filter and explore the annotated sequence data produced by its pipelines. Genformatic's pipelines and analysis solutions are deployed securely in the cloud and ready to go to aid investigators involved in large projects with many samples, or for application on single genomes, tumor-normal pairs, trios or family data. Genformatic tools and services help scientists understand how sequence changes affect biological functions, impact health risk, or contribute diagnostic and prognostic information or suggest alternative treatment options, including providing information on relevant clinical trials.

Genformatic' s solutions are provided in part by proprietary bioinformatic methods with superior accuracy, and available as stand alone tools, or incorporated into fast, computationally efficient, and easy-to-use analysis pipelines deployed in the cloud and ready for launch. Maximal accuracy and error suppression are critical of course, but the number of errors produced by standard algorithms are alarming, and the downstream consequences of those errors for both genomic research and genomic medicine are severe. A recent scientific paper from Nature Communications assessed the current status of whole genome analysis as follows: "Out of more than 1,000 confirmed somatic SNPs, less than half were unanimously identified by all participating teams. And with insertions and deletions, the concordance was even poorer" [(1)]. The Authors go on to state, "Only one somatic insertion/deletion out of a total of 337 was identified in common by all the parties." The authors conclude that, "We have found that, contrary to common perception, identifying somatic mutations, be they SSMs or SIMs, from WGS data is still a major challenge. Calling mutations with different pipelines on differently prepared sequence read sets resulted in a low level of consensus." Genformatic's innovative methods for improved somatic mutation detection accuracy address this problem.

The authors of this paper recommended several best practices for Whole Genome Sequencing (WGS) analysis including PCR-free library preparation, tumor coverage >100X, control coverage close to tumor coverage (±10%), optimized aligner/variant caller combinations, and running multiple mutation callers. Genformatic developed BAYSIC specifically to deal with the issue of how best to combine the output from multiple variant calling methods to reduce discordance and enhance precision and sensitivity. Using Genformatic's tools we are able to increase mutation sensitivity to 99% compared to 30% for some individual somatic callers and we improve somatic mutation specificity to 99% while individual somatic call sets are littered with false positive calls.

BAYSIC Analysis Strategy

Genformatics has developed a BAYSIC (Bayesian integrated calling) - a fully-Bayesian, unsupervised machine learning (artificial intelligence) method that combines variant calls from many different tools into an integrated call set with improved specificity and sensitivity. BAYSIC represents an innovative bioinformatic method for combining multiple data streams into a new and improved integrated data set. For more information on BAYSIC performance please see our published paper- http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-104.



Genformatic analyzed samples from the 1000 Genomes Project and found many SNPs were present only in one dataset and only 36.8% (3,666,983) of calls were present in all four datasets. (296,756; 956,927; 233,557; 261,251 for SNP detected only by SamTools, FreeBayes, Atlas and GATK, respectively) (Figure A). Only 36.8% (3,666,983) of calls were present in all four sets (Figure B), and only 82.5% (8,222,619) of SNPs were present in two or more sets.

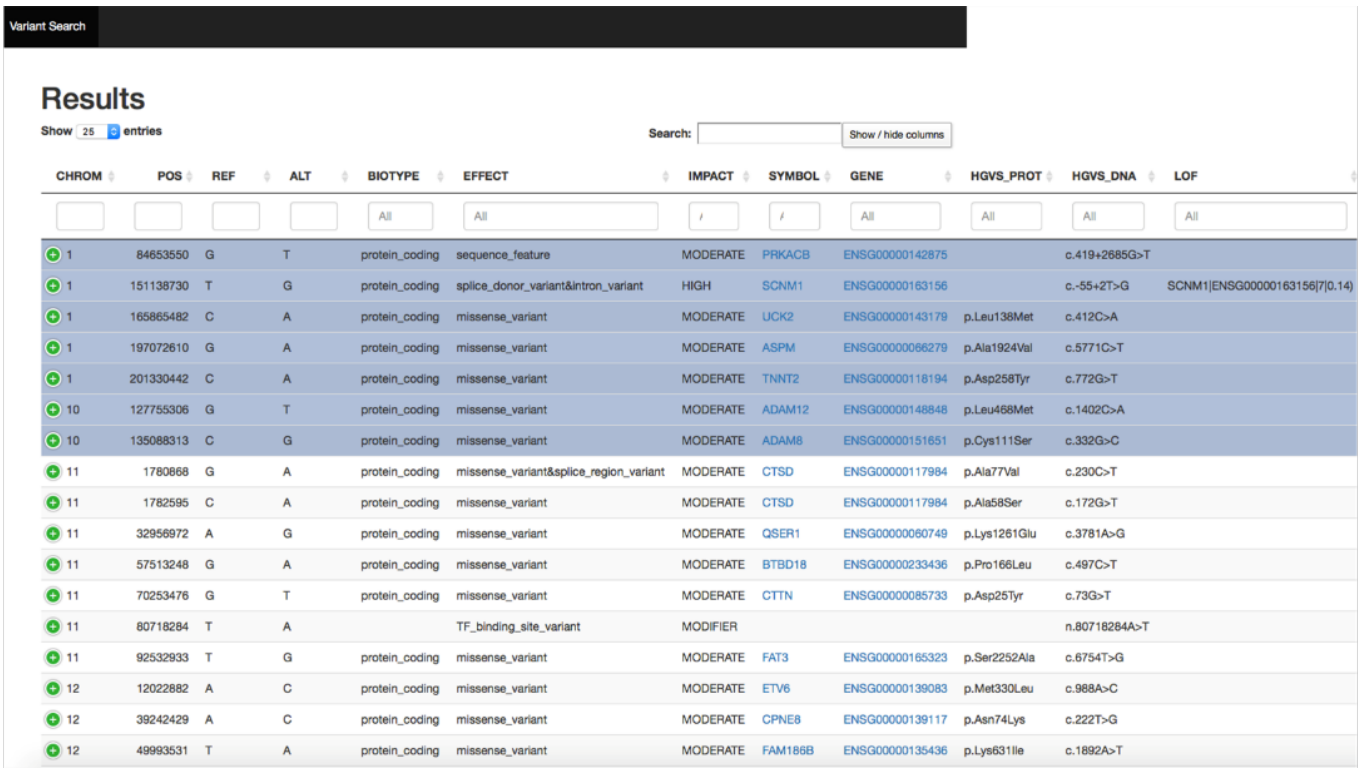
How does it work-

Overview of BAYSIC algorithm. Sets of variant calls produced from one or more programs are the input (in VCF format) to the BAYSIC algorithm. False positive and false negative error rates are estimated, and a posterior probability is assigned to each mutation, allowing the user to generate call sets that emphasize sensitivity or specificity, according to user's tolerance for the risk of false positives or false negatives. Variants whose posterior probability is greater than the cutoff specified by the user are used to generate a set of integrated variant calls. Using BAYSIC, the user can produce VCF files that have higher sensitivity and specificity than individual callers. In clinical samples, BAYSIC identified all of the subsequently confirmed mutations – something that no individual caller could do. In a somatic mutation analysis, the BAYSIC callset also contained a much higher percentage of validated COSMIC SNPS. BAYSIC allows the user to integrate multiple data streams to increase the sensitivity and precision of DNA variation analysis.

Genformatic has also automated and validated RUFUS in its genomic analysis platform. RUFUS (reference-free unique sequence) is another innovative bioinformatic method to avoid errors introduced by sequence read mis-alignment and incorrect read mapping – a pervasive source of error in conventional methods of genomic variant and somatic mutation detection. RUFUS dramatically reduces errors by identification of unique k-mers in sequence read data.

Visualization and Filtering Tools

When variant calling is finished the analyzed data can be viewed in our visualization and filtering tool Genome Cruiser. Genome Cruiser allows the user to filter results by gene, mutation, effect, drug, Database presence and phenotype. Additional information provided include annotations from selected databases, functional impacts, medical prognosis and diagnosis, clinical implications. Genome cruiser is connected a variety of databases including ClinVar, DGIdb, COSMIC, ClinicalTrials.gov, CIVIC, GWAS, Condel, TCGA, dbSNP, SNPeffect, DrugBank, and it can be connected to custom user defined databases.



CHROM	POS	REF	ALT	BIOTYPE	EFFECT	IMPACT	SYMBOL	GENE	HGVS_PROT	HGVS_DNA	LOF
1	84653550	G	T	protein_coding	sequence_feature	MODERATE	PRKACB	ENSG00000142875		c.419+2685G>T	
1	151138730	T	G	protein_coding	splice_donor_variant&intron_variant	HIGH	SCNM1	ENSG00000163156		c.-55+2T>G	SCNM1|ENSG00000163156|7|0.14)
1	165865482	C	A	protein_coding	missense_variant	MODERATE	UCK2	ENSG00000143179	p.Leu138Met	c.412C>A	
1	197072610	G	A	protein_coding	missense_variant	MODERATE	ASPM	ENSG00000066279	p.Ala1924Val	c.5771C>T	
1	201330442	C	A	protein_coding	missense_variant	MODERATE	TNNT2	ENSG00000118194	p.Asp258Tyr	c.772G>T	
10	127755306	G	T	protein_coding	missense_variant	MODERATE	ADAM12	ENSG00000148848	p.Leu468Met	c.1402C>A	
10	135088313	C	G	protein_coding	missense_variant	MODERATE	ADAM8	ENSG00000151651	p.Cys111Ser	c.332G>C	
11	1780868	G	A	protein_coding	missense_variant&splice_region_variant	MODERATE	CTSD	ENSG00000117984	p.Ala77Val	c.230C>T	
11	1782595	C	A	protein_coding	missense_variant	MODERATE	CTSD	ENSG00000117984	p.Ala58Ser	c.172G>T	
11	32956972	A	G	protein_coding	missense_variant	MODERATE	QSER1	ENSG00000060749	p.Lys1261Glu	c.3781A>G	
11	57513248	G	A	protein_coding	missense_variant	MODERATE	BTBD18	ENSG00000233436	p.Pro166Leu	c.497C>T	
11	70253476	G	T	protein_coding	missense_variant	MODERATE	CTTN	ENSG00000085733	p.Asp25Tyr	c.73G>T	
11	80718284	T	A		TF_binding_site_variant	MODIFIER				n.80718284A>T	
11	92532933	T	G	protein_coding	missense_variant	MODERATE	FAT3	ENSG00000165323	p.Ser2252Ala	c.6754T>G	
12	12022882	A	C	protein_coding	missense_variant	MODERATE	ETV6	ENSG00000139083	p.Met330Leu	c.988A>C	
12	39242429	A	C	protein_coding	missense_variant	MODERATE	CPNE8	ENSG00000139117	p.Asn74Lys	c.222T>G	
12	49993531	T	A	protein_coding	missense_variant	MODERATE	FAM186B	ENSG00000135436	p.Lys631Ile	c.1892A>T	

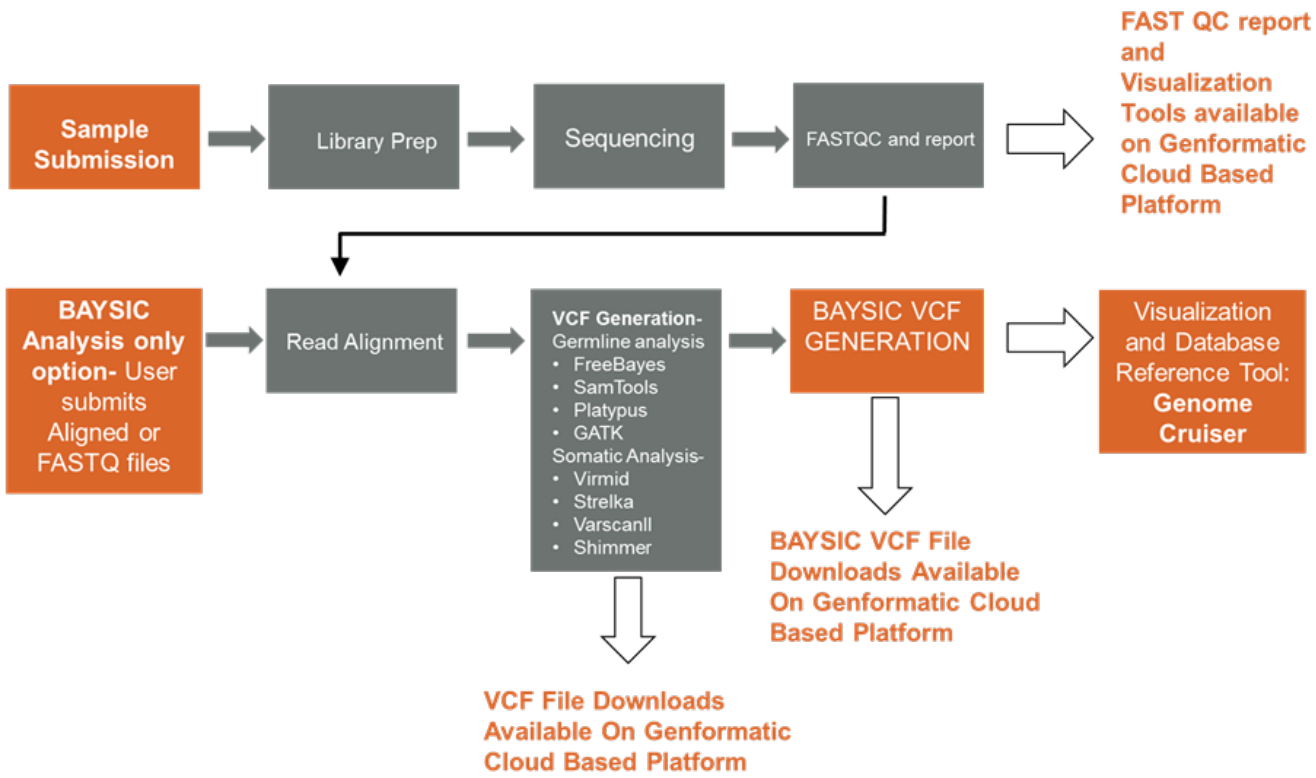
Services Summary

Variant Analysis-Currently Genformatic is offering complete sequencing and analysis services for germline and somatic variant analysis on exomes and whole genomes. We offer tumor/normal (T/N) and tumor/normal/normal (T/N/N) comparisons.

For our DNA variant sequencing and analysis workflows researchers can either send us their DNA sample or raw sequencing data files (FastQ or BAM files). When sequencing is finished raw sequence data will be uploaded to our platform where it will run through our analysis pipelines.

For sequencing applications, if the project requires, we will have the samples sequenced at one of our CAP/CLIA certified sequencing partners or have the samples sent to our research grade sequencing partner .

GENFORMATIC Workflow- Somatic and Germline Sequencing and Analysis services with Visualization tools



Data is returned via our web portal where it can be downloaded or transferred to the researchers own cloud environment. Researcher's analyzed data will be loaded into our cloud platform for viewing and filtering on our Genome Cruiser platform. Deliverables include FAST QC report and FastQ data, BAM alignment files, VCF's generated from Open Source callers such as FreeBayes, SamTools, Strelka, etc. (can be user defined), and our BAYSIC VCF. Analyzed data will also be accessible through our Genome Cruiser tool for visualization, filtering and data annotation. All of our services include 3 months of free storage for raw and analyzed data sets. Our Cloud services platform is run in a HIPAA compliant environment with the highest quality security.

The entire process will be managed with a dedicated project manager who will keep you informed in real time to the status of your project.

Additional Genomic Analysis Services

We offer RNA-seq analysis using the latest cutting edge tools such as QuSage to determine gene set enrichment, EdgeR for differential expression analysis, StringTie and Ballgown for transcript assembly and isoform comparisons. We also have the ability to generate custom informatics analysis for your application.

Genformatic has developed easy to use workflows that enable researchers to quickly and easily understand their analyzed data For RNA seq analysis Genformatic delivers to the researcher, Data QC reports and visualization tools, expression analysis reports and visualization tools, BAM alignment files, and custom user defined reports. Genformatic has experience across a wide breadth of genomic applications including microbiome work and is able to work with you on your custom defined analysis projects.

Contact us to learn more about our genomic analysis services

Genformatics has worked hard to deliver the highest level of accuracy from commercial services providers. Through using tools like BAYSIC and RUFUS researchers receive clear concise analyzed data sets that allow them to quickly get to their variant and/or gene of interest. With prioritization through BAYSIC researchers are quickly able to sort through discordant calls to a high confidence data set. Genformatic services are designed to be easy to use for the non expert genomics researcher and flexible enough to be of value to the expert genomic researcher. Through the use of BAYSIC Genformatic is able to return results that cut through the chemistry variation that is introduced in the sequencing process so the user can have confidence that they don't have to tune their tools to account for data generated on different instruments, chemistry, etc. In somatic sequencing and analysis, Genformatic tools generate data sets that have increased sensitivity for minor allele variants in complex tumor samples with heterogeneity. All of our analysis services can be delivered in a CAP/CLIA and HIPAA compliant environment. For more information on our services and a demonstration of our Genome Cruiser please feel free to contact us at Info@genformatic.com.

References

1) A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing (1). Tyler S. Alioto, Ivo Buchhalter, Sophia Derdak, Barbara Hutter, Matthew D. Eldridge, Eivind Hovig, Lawrence E. Heisler, Timothy A. Beck, Jared T. Simpson, Laurie Tonon, Anne-Sophie Sertier, Ann-Marie

Patch, Natalie Jäger, Philip Ginsbach,Ruben Drews, Nagarajan Paramasivam,Rolf Kabbe, Sasithorn Chotewutmontri, Nicolle Diessl, Christopher Previti et al. Nature Communications 6, Article number: 10001 doi:10.1038/ncomms10001 Received 16 June 2015 Accepted 23 October 2015 Published 09 December 2015“-

2) BAYSIC: a Bayesian method for combining sets of genome variants with improved specificity and sensitivity, Brandi L Cantarel, Daniel Weaver, Nathan McNeill, Jianhua Zhang, Aaron J Mackey and Justin Reese, BMC Bioinformatics201415:104, DOI: 10.1186/1471-2105-15-104© Cantarel et al.; licensee BioMed Central Ltd. 2014, Received: 10 October 2013Accepted: 31 March 2014Published: 12 April 2014