NextCode Seeks to Build Business on Proprietary Informatics Infrastructure from DeCode

May 09, 2014
By Uduak Grace Thomas

NEW YORK (GenomeWeb) – Cambridge, Mass-based NextCode Health is seeking to make its bread and butter by providing bioinformatics solutions for research and clinical settings that were originally developed and used internally by DeCode Genetics, now a subsidiary of Amgen, in its genetic studies of the Icelandic population.

Jeff Gulcher, NextCode's president and chief scientific officer, discussed the company's portfolio — which it began marketing last October after obtaining exclusive rights from DeCode to sell the solutions — in a presentation at the Bio-IT World Conference last week. The company's products can be purchased as software-as-a-service and installed on Amazon's cloud infrastructure or on local servers. NextCode offers pay-as-you-go pricing for research and clinical customers who want to analyze a small number of samples at a time — the exact cost differs in both cases. It has a different pricing model for clients that want to run bulk analysis. The company does not disclose the exact amounts it charges.

Underlying NextCode's informatics solutions is a proprietary infrastructure dubbed Genomic Ordered Relational (GOR) architecture that stores large quantities of genomic data in a fashion that makes it easy for informatics solutions running on the front end of the platform to extract information as needed in real time. The GOR architecture uses proprietary methods to store reads according to their position in the human reference genome, Gulcher explained to *BioInform* this week. This provides a "roadmap" that can be used in analyzing sequence data from either BAM or VCF files, he said.

"The tools that we developed to tap into the GOR database … know where to hunt for the data and so it becomes orders of magnitude more efficient to pull out the data [and] to query the database than conventional database infrastructures" allow, Gulcher said.

Furthermore, the GOR platform can manage large quantities of data. At DeCode, it was used to hold data from 350,000 whole genomes, Gulcher noted. This means that users can hold both the raw reads and variant call files in a single location instead of storing files separately and attempting to aggregate data in order to run queries, he said. These files are stored separately in a normalized database within the GOR and combined as needed with annotation information that users pull in during their analysis. That means, Gulcher said, that users do not have to rewrite their BAM or VCF files for each analysis they run which results in both storage and compute usage savings. Also, they still retain access to the raw sequence — as well as coverage information — which, among other things, lets them cross-check their results for false positives, he said.

The base infrastructure also includes a knowledgebase of genomic information that combines data derived from the larger dataset that DeCode gathered in its studies of the Icelandic population — it includes

information on allele frequencies for instance. This information — which supports customers that use the company's solutions for clinical diagnostics — is combined with curated public domain data culled from sources such as the National Center for Biotechnology Information's ClinVar, the Catalogue of Somatic Mutations in Cancer database, and the Online Mendelian Inheritance in Man database. According to the company's website, the database includes more than 1.5 million indels and 6,000 loss-of-function mutations over 4,800 genes.

In one scenario that Gulcher described in his presentation at Bio-IT, the company used its system to identify a genetic mutation that was associated with a rare disorder in two sisters that caused progressive blindness, deafness, and diaphragmatic weakness in two sisters — in this case the sisters had a rare homozygous missense variant in a riboflavin transporter gene paralog which caused their condition.

NextCode's product portfolio includes an internally developed pipeline composed of alignment, mapping, and variant calling algorithms that have been optimized using the DeCode study data. The pipeline includes tools for calling insertions and deletions, predicting the functions of missense mutations, tagging or filtering unstable genomic regions, and also annotating variants with information on phenotype, allele frequency, and more.

The company also markets the NextCode Clinical Sequence Analyzer, (CSA) which provides clinicians with computational tools to identify and report on *de novo* and rare pathogenic mutations that are linked to genetic ailments and also assess disease risk by analyzing genome, exome, and transcriptome data — the company also offers sequencing services for both clinical and research use cases through unnamed partners. The CSA solution includes applications such as NextCode's genome browser which provides visual representations of the data that allow users to compare reads and variants from multiple genomes with reference data. Users can search for clinically relevant mutations in lists of candidate genes and they can also store information on disease-linked genes and mutations for use in future studies.

NextCode also markets Sequence Miner for its more research-oriented clients, mostly in pharma or academia, who are interested in running population-based studies. NextCode Sequence Miner provides these users with statistical tools — as well as the company's genome browser — to run deep queries on their data as part of case-control studies, for example. Users simply group study participants as either cases or controls, define the types of mutations they would like to explore, and the system applies statistical tools and applications, such as the Fisher's Exact test to the data. Also included is the Tumor Sequence Analysis tool, which uses the Mutect and Varscan software — developed by the Broad Institute and Washington University in St. Louis, respectively — to compare sequences from tumor and germline samples.

NextCode's tools are open for customers, meaning that they can customize them internally to fit their needs — it is based on the Unix operating system and SQL programming language — or they can work with the company to tailor the tools to fit their needs, Gulcher said.