DNASTAR®

DNASTAR Lasergene®

Rapid Assembly and Analysis of Clinical Sequencing Data on a Desktop Computer: Using DNASTAR Software to Identify Potential Disease-Causing Mutations

Thomas Schwei[1], Timothy Durfee PhD[1], Katherine Maxfield[1], Matthew Keyser[1], Daniel Nash[1], Schuyler Baldwin[1], Richard Nelson PhD[1], Kenneth Dullea[1], John Schroeder[1], Pavel Pinkas PhD[1], Guy Plunkett III PhD[1,2], Frederick Blattner PhD[1,2,3]

[1] DNASTAR, Inc., Madison, Wisconsin, USA
[2] University of Wisconsin, Department of Genetics, Madison, Wisconsin, USA
[3] Scarab Genomics LLC, Madison, Wisconsin, USA

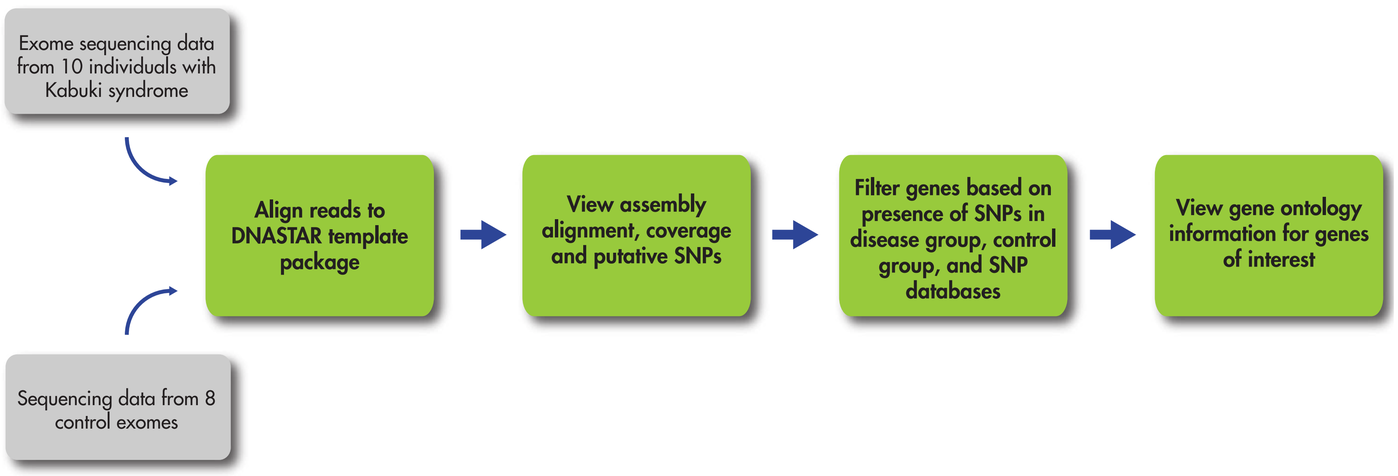
Abstract

DNASTAR offers an integrated suite of software for assembling and analyzing sequence data from all major next-generation sequencing platforms. The software supports key workflows on a desktop computer, including analysis of clinical sequencing data.

In this project, exome data from 10 individuals with Kabuki syndrome and 8 control individuals are aligned against the entire human genome using DNASTAR's proprietary assembly software. After assembly, interactive views within the software allow for visual inspection of coverage and read alignment along the genome; the ability to pool and filter SNP data from all samples; and integration of data from dbSNP, GERP and COSMIC databases, and gene ontology information from the GO Consortium. In this case, the integrated SNP filtering is used to rapidly identify 1) a small set of genes that contain novel non-synonymous mutations in most of the Kabuki affected individuals and none of the control exomes and 2) genes that contain novel loss-of-function mutations in any number of the Kabuki affected individuals. The results of each of these searches corroborate a previous study of the same data, pointing to mutations in the gene MLL2 as a likely factor in Kabuki syndrome.

The integration of all these tools into one software package facilitates fast, comprehensive analysis, helping scientists move quickly from raw next-gen sequencing data to results showing the genetic and genomic impact of mutations. By using innovative algorithms within the software, scientists can have all of the assembly and analysis capabilities available to them on their desktop computer, including support for large data sets generated by next-gen sequencing instruments.

DNASTAR Clinical Sequence Analysis Pipeline



View Alignment, Coverage and SNPs

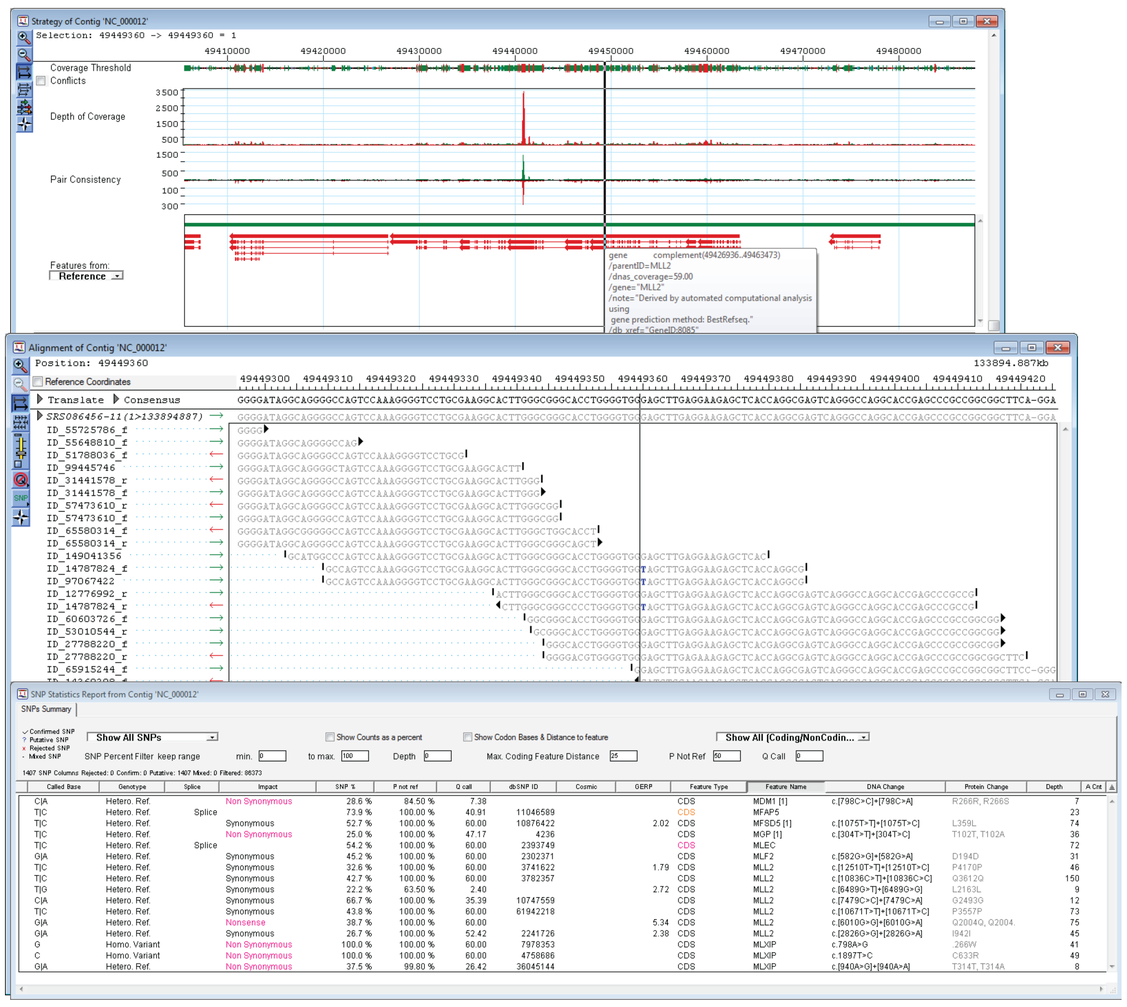


Figure 1. Here, one of the assemblies from the Kabuki samples is shown in SeqMan Pro. The Strategy View (top) shows depth of coverage along the genome; the Alignment View (middle) shows the individual reads and putative SNPs; and the interactive SNP report (bottom) can be used to navigate to SNPs in the other views and filter SNPs based on depth of coverage, SNP percentage, distance to the closest coding feature, and SNP impact.

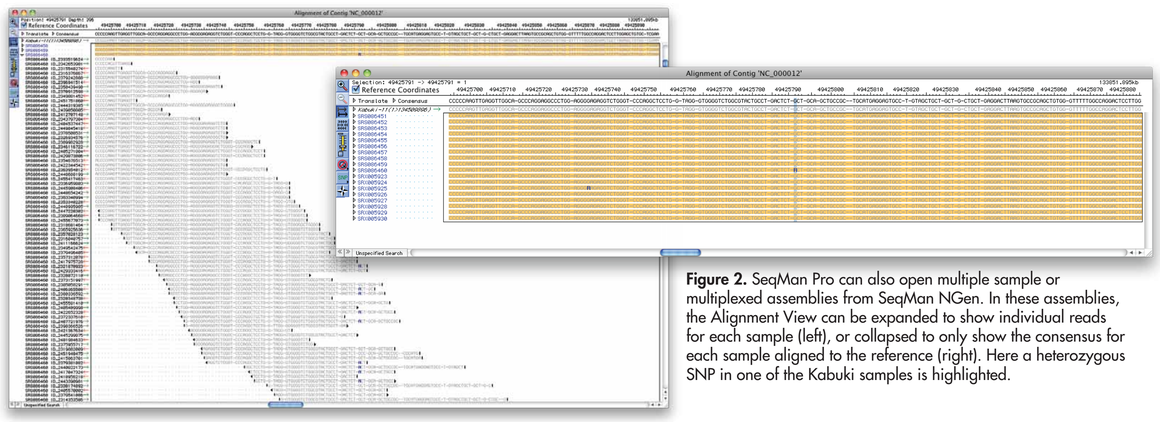


Figure 2. SeqMan Pro can also open multiple sample or multiplexed assemblies from SeqMan NGen. In these assemblies, the Alignment View can be expanded to show individual reads for each sample (left), or collapsed to only show the consensus for each sample aligned to the reference (right). Here a heterozygous SNP in one of the Kabuki samples is highlighted.

Filter Genes Based on SNP Data

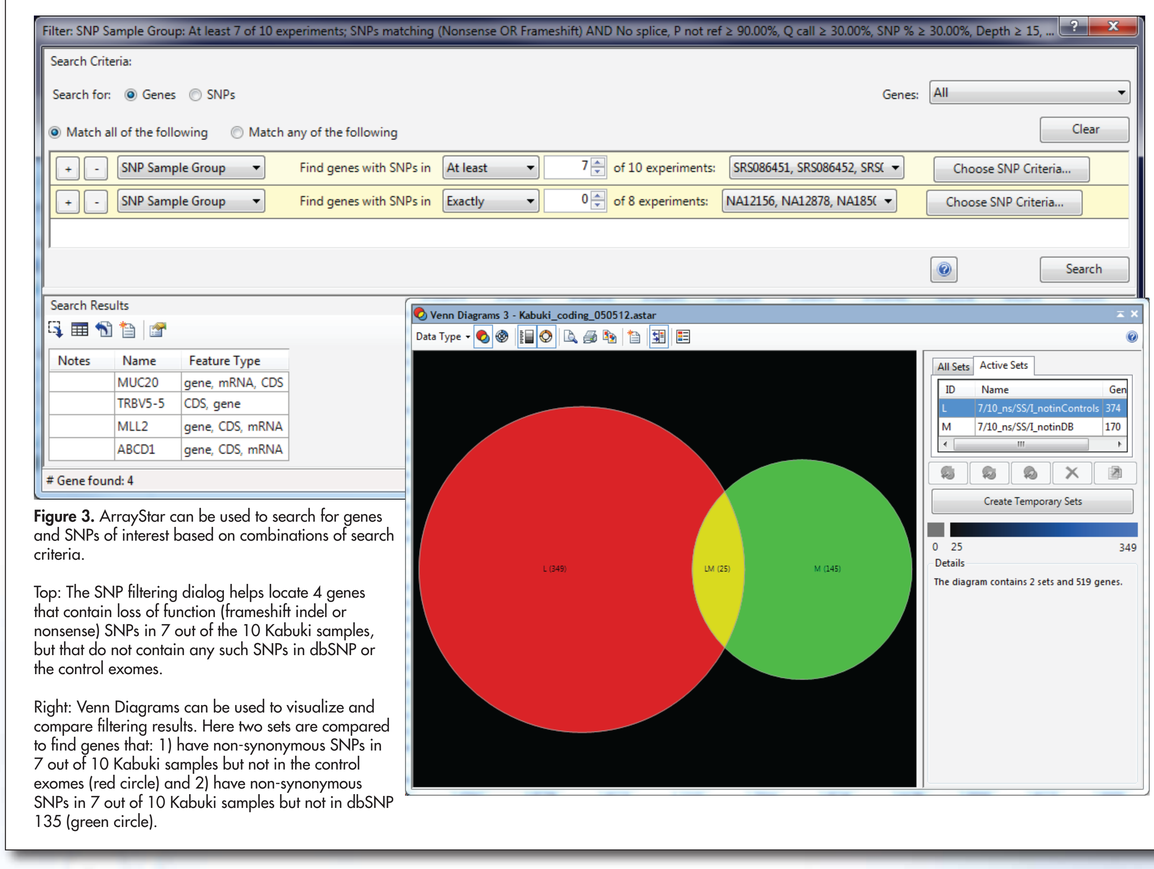


Figure 3. ArrayStar can be used to search for genes and SNPs of interest based on combinations of search criteria.

Top: The SNP filtering dialog helps locate 4 genes that contain loss of function (frameshift indel or nonsense) SNPs in 7 out of the 10 Kabuki samples, but that do not contain any such SNPs in dbSNP or the control exomes.

Right: Venn Diagrams can be used to visualize and compare filtering results. Here two sets are compared to find genes that: 1) have non-synonymous SNPs in 7 out of 10 Kabuki samples but not in the control exomes (red circle) and 2) have non-synonymous SNPs in 7 out of 10 Kabuki samples but not in dbSNP 135 (green circle).

View Gene Ontology

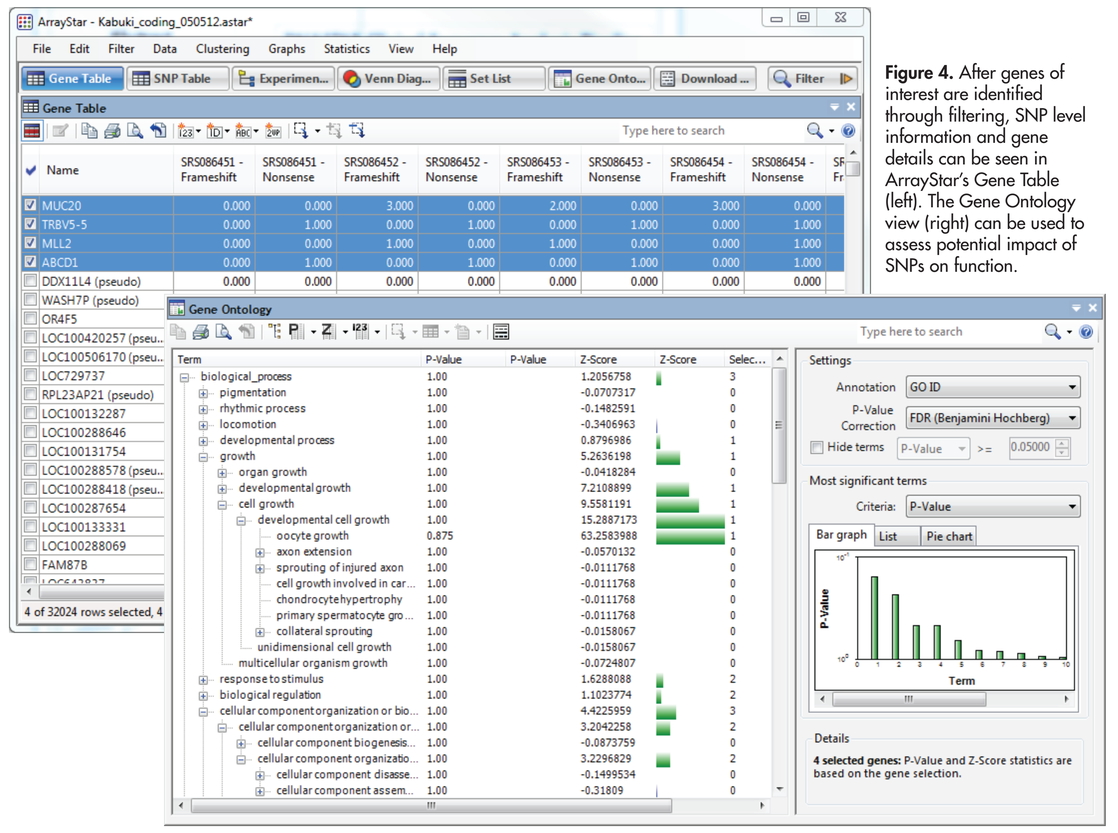


Figure 4. After genes of interest are identified through filtering, SNP level information and gene details can be seen in ArrayStar's Gene Table (left). The Gene Ontology view (right) can be used to assess potential impact of SNPs on function.

References

The Next Generation Mendelian Genetics project was provided by NIH grant 1RC2 HG005608-01 to Drs. Debbie Nickerson, Jay Shendure, Michael Bamshad, and Wendy Raskind, and research on Kabuki Syndrome by 5RO1-HD48895 to Michael Bamshad. The Kabuki syndrome datasets used for the analyses described in this poster were obtained from the database of Genotype and Phenotype (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000295.v1.p1.

The eight human exome samples used for the analyses described in this poster were obtained from the Sequence Read Archive found at http://trace.ncbi.nlm.nih.gov/Traces/sra/ under the SRA study SRP000910.