



Introducing SNP & Variation Suite 8

March 11, 2014

Gabe Rudy
VP Product Development



- SVS 7 has been adopted by hundreds of client organizations worldwide.
- GenomeBrowse has 2,500 registered users.



Andreas Scherer, PhD, President & CEO



Researchers

- Human genome
- Plant DNA
- Animal DNA



Clinicians

- Also use genomic information to:
- Diagnose diseases
- Identify best treatment options for patients



Students

- Currently need to be a “computer science” whiz before they can begin analyzing data
- SVS reduces the initial hurdle to enter the field



- Sequencing technology becoming cheaper
- Scalability is key
- Sample sizes are increasing
- More data per sample
- **Improvement of:**
 - Data management
 - Data import and manipulation capabilities





Company maturity:

- Service engagements
- Deeply ingrained in the community
- Hosting webcasts
- Website becomes knowledge hub

our 2 snps...
A BLOG BY GOLDEN HELIX

Home Authors @gabinformatics

← First Golden Helix Research Competition a Great Success!

The Next Chapter of Golden Helix: Our Launch of SVS 8

Posted on March 10, 2014 by goldenadmin

Today is a big day for us. Today we are launching [JV1] a major release of our flagship product, SNP & Variation Suite (SVS), to the general public. SVS 8 is a substantial improvement over the previous release in a number of dimensions (see detailed discussion on our [What's New](#) page).

We've come a long way.

Over five years ago, in November 2008, we introduced SVS 7 as a powerful tool to conduct next-generation sequencing and GWAS studies. Since then, we have been adopted by hundreds of client organization worldwide. SVS is being used by leading research organizations in the US, Canada, Latin America, Asia, Australia, Africa, and Europe.

About one and a half years ago, we launched the first version of our free, standalone genome browser. This tool was designed to help researchers to view large sequencing files alongside public annotation databases in a fluid and intuitive way. Over 2,500 researchers in our field are using [GenomeBrowse](#) today.

As customers have seen the intuitive way GenomeBrowse allows them to navigate through extensive databases and large amount of data, we have been asked to bring the same visualization look and feel to SVS (which is the biggest change for SVS 8). This becomes increasingly important as SVS is adopted outside of our traditional audience.

Our primary users are researchers who analyze human, plant, or animal DNA to study diseases and traits. Our software is also used in clinical settings where scientists are working hand-in-hand with their clinical counterparts to use genomic information to diagnose diseases and identify best treatment options for their patients. In addition, we see a third group of users becoming increasingly important: undergrads and graduate students who are learning how to conduct analytics work. Here, the ease-of-use of our platform reduces the initial hurdle to enter the field. It eliminates the need for a student to become a biostatistics whiz before they can begin analyzing data and focusing on the science behind it.

Another major topic that we addressed in this release is scalability. As sequencing technology has

Search

About
Welcome to the Our 2 SNPs...[®] blog by Golden Helix, a leading bioinformatics company. This blog seeks to inform our customers and the genetic research community as a whole on the latest in analysis methods, best practices, and the future of the industry.

Please, don't be shy! We want to hear your two SNPs as well! Also, if there is a specific topic that you would like to hear more about, please contact us at info@goldenhelix.com. For more information about Golden Helix, visit our [main site](#).

Follow...
Your Email Address
Subscribe

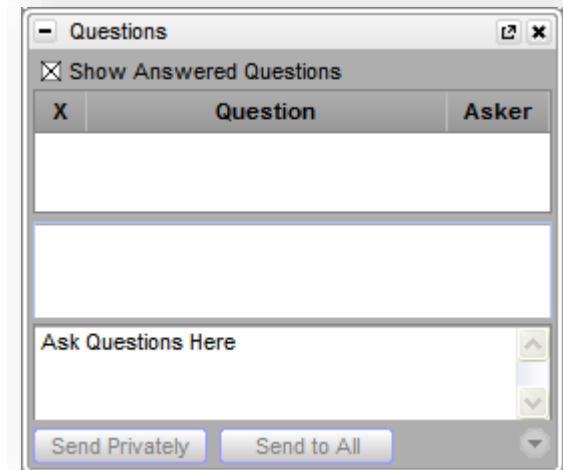
Categories

- About GHI (7)
- Add-on scripts & data repository (12)
- Assessment of new methods (5)
- Best practices in genetic analysis (16)
- Big picture (21)
- Bioinformatic support (9)
- Clinical genetics (6)
- Customer success (18)
- General statistical genetics principles (8)
- How to's and advanced workflows (14)
- News, events, & announcements (44)
- Paper review (2)



Questions during the presentation

Use the Questions pane in your GoToWebinar window





1 History

2 New Annotation Infrastructure

3 GenomeBrowse Grows in Its Utility

4 What's New in GWAS Analysis

5 Conclusion

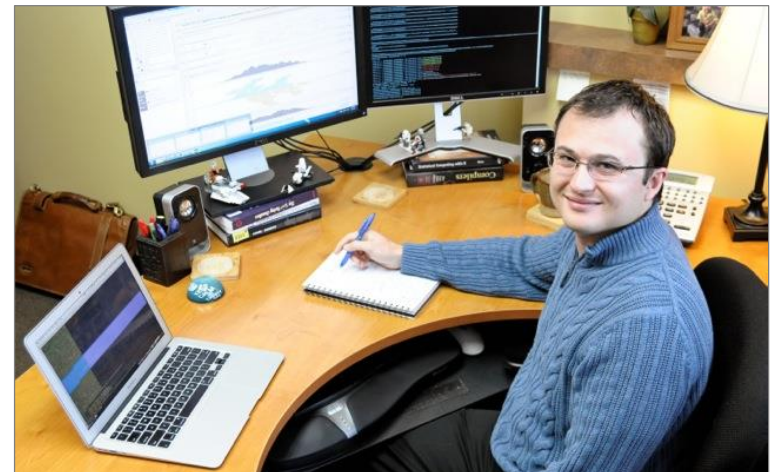


■ Golden Helix

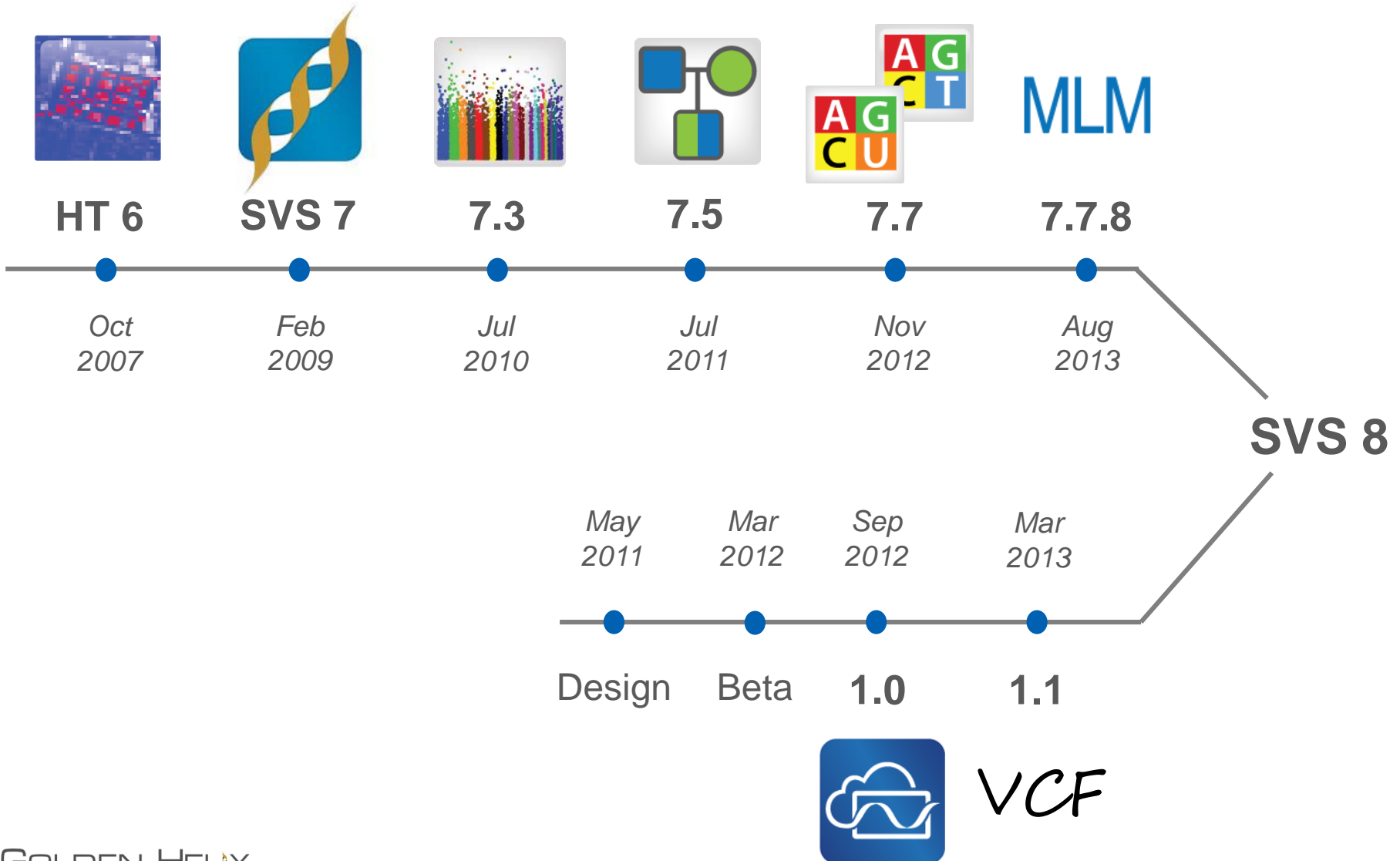
- Founded in 1998
- Genetic association software
- Analytic services
- Hundreds of users worldwide
- Over 800 customer citations in scientific journals

■ Products I Build with My Team

- **SNP & Variation Suite (SVS)**
 - SNP, CNV, NGS tertiary analysis
 - Import and deal with all flavors of upstream data
- **GenomeBrowse**
 - Visualization of everything with genomic coordinates. All standardized file formats.
- **RNA-Seq Pipeline**
 - Expression profiling bioinformatics



Historical Context



Integrated Product Solution





1 History

2 **New Annotation Infrastructure**

3 GenomeBrowse Grows in Its Utility

4 What's New in GWAS Analysis

5 Conclusion



- Annotations are the cornerstone of DNA analysis and interpretation
- **Our rebuilt infrastructure results in:**
 - New **file format** that provides compressed files with richer usage and indexed search
 - New **Convert** wizard
 - Update Annotation and Filtering **workflows** to utilize these sources
 - Redesigned **data repository** with the future in mind





- Amazing in-place compression with column data storage techniques
- Scales to hundreds of millions of records without slowing down
- Integrated documentation
- Indexed field searching

@gabeinformatics
AN "OUR 2 SNPs..." BLOG BY GOLDEN HELIX

Home Authors @gabeinformatics

← CALL FOR ABSTRACTS: Dell Ultrabook Laptop and SVS Giveaway → Back to Basics: Importing/Exporting Data in Imputation Program Data Formats with SVS →

All I Want for Christmas Is a New File Format for Genomics

Posted on December 16, 2013 by Gabe Rudy

[t](#) [f](#) [s](#)

'Tis the season of quiet, productive hours. I've been spending a lot of mine thinking about file formats. Actually I've been spending mine implementing a new one, but more on that later.

File formats are amazingly important in big data science. In genomics, it is hard not to be awed by how successful the BAM file format is.

I thought one of the most tweetable moments at ASHG 2013 was when [Jeffrey Reid](#) from BCM Human Genome Sequencing Center (HGSC) talked about how they offloaded to the cloud (via DNAnexus) 2.4 million hours of compute time to perform the alignment and variant calling on ~4k genomes and ~12k exomes.

In the process, they produced roughly half a petabyte of BAM files (well mostly BAM files, VCFs are an order of magnitude smaller, but part of the output mix).

I'd speculate that [Heng Li's](#) binary file format for storing alignments of short reads to a reference genome is responsible for more bytes of data being stored on the cloud (and maybe in general) than any other file format in the mere 4 years since it was invented.

But really, the genius of the format was not in the clever and extensible encoding of the output of alignment algorithms (the CIGAR string and key-value pair "tag" field have held up remarkably well through years of innovation and dozens of tools), but in the one-to-one relationship it shared with its text-based counterpart, the SAM file.

Wanted: Grey Beard Unix Hookers for Big Data Bolonoe

Text files have been, and will continue to be, the dominant medium of data scientists -- and for good reasons. Even a basic familiarity with the [Unix toolset](#) allows for an immense amount of data management and processing tasks to take place. And to take place efficiently. I dare you to write a program that is [faster](#) at searching text files than [grep](#).

In fact, any bioinformatic pipeline worth its salt is probably heavily utilizing [grep](#), [awk](#), [cut](#), [sed](#), [sort](#), [xargs](#) and, yes, the rusty edged [Swiss army knife](#) that is [perl](#). We've worked on a couple here at Golden Helix. You can find the [perl-based examples](#) [here](#) and the [perl-based examples](#) [here](#).

About
Welcome to the Our 2 SNPs... blog by Golden Helix, a leading bioinformatics company. This blog seeks to inform our customers and the genetic research community as a whole on the latest in analysis methods, best practices, and the future of the industry.

Please, don't be shy! We want to hear your two SNPs as well! Also, if there is a specific topic that you would like to hear more about, please contact us at info@goldenhelix.com. For more information about Golden Helix, visit our [main site](#).

Follow...
Your Email Address

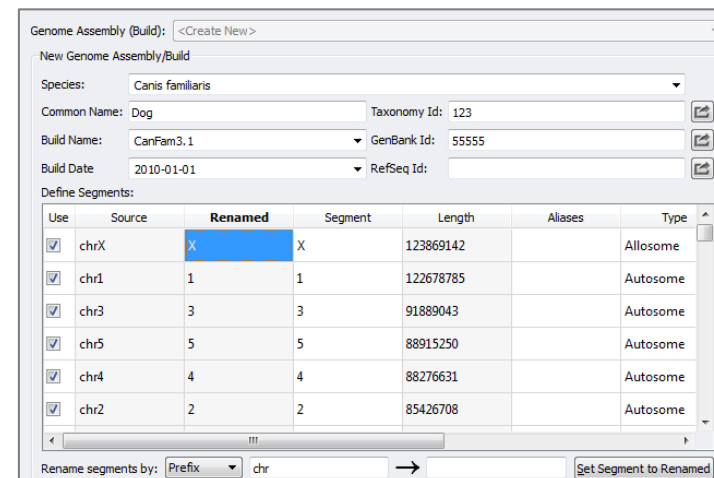
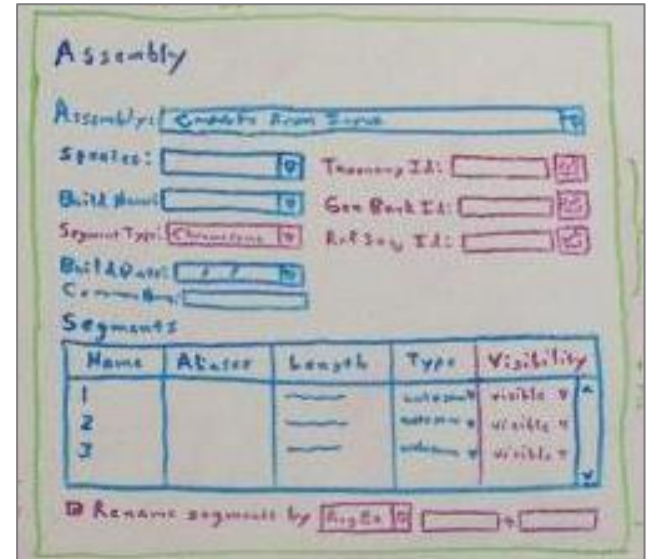
Categories

- About GH (7)
- Add-on scripts & data repository (12)
- Assessment of new methods (5)
- Best practices in genetic analysis (16)
- Big picture (21)
- Bioinformatic support (9)
- Clinical genetics (6)
- Customer success (18)
- General statistical genetics principles (6)
- How to's and advanced workflows (14)
- News, events, & announcements (44)
- Paper review (2)
- Personal genomics (7)
- Plant & animal (6)
- Technology review (25)

Convert Wizard



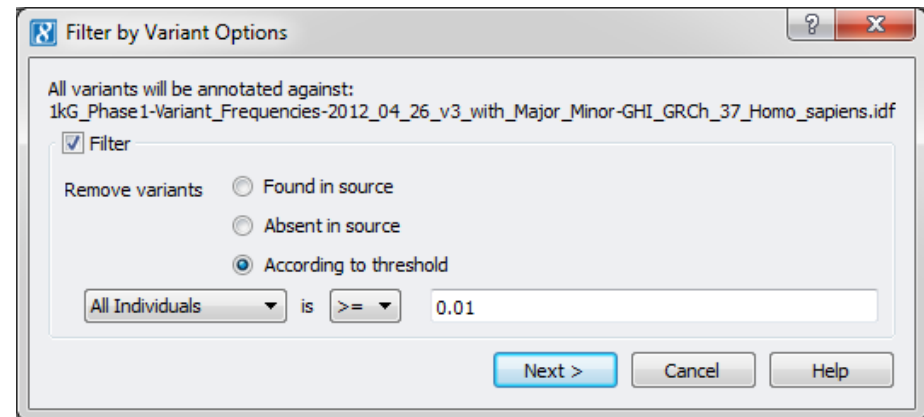
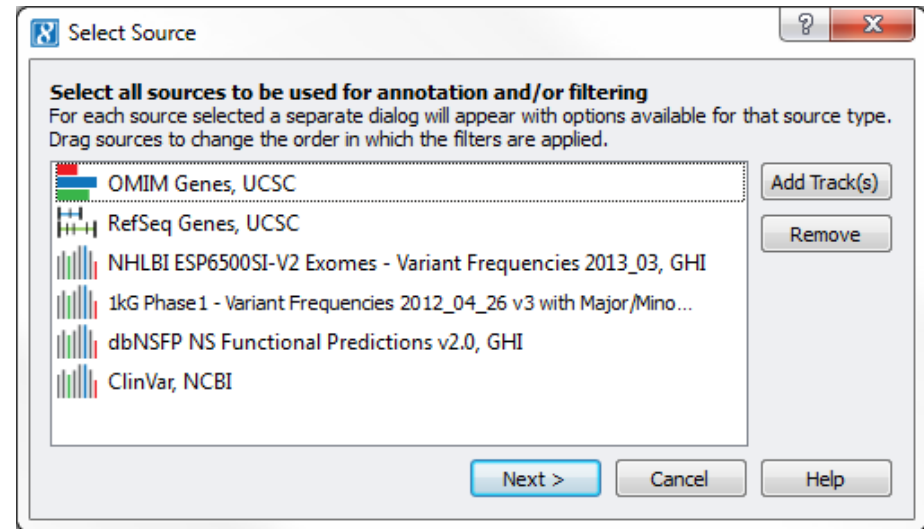
- Supports new file formats – every type of file we have seen, we can convert
- Preview your data as you go
- Document your data source in the wizard or afterwards
- Curate your own genomes
- Easy to share your results – just one file



Annotation Workflow

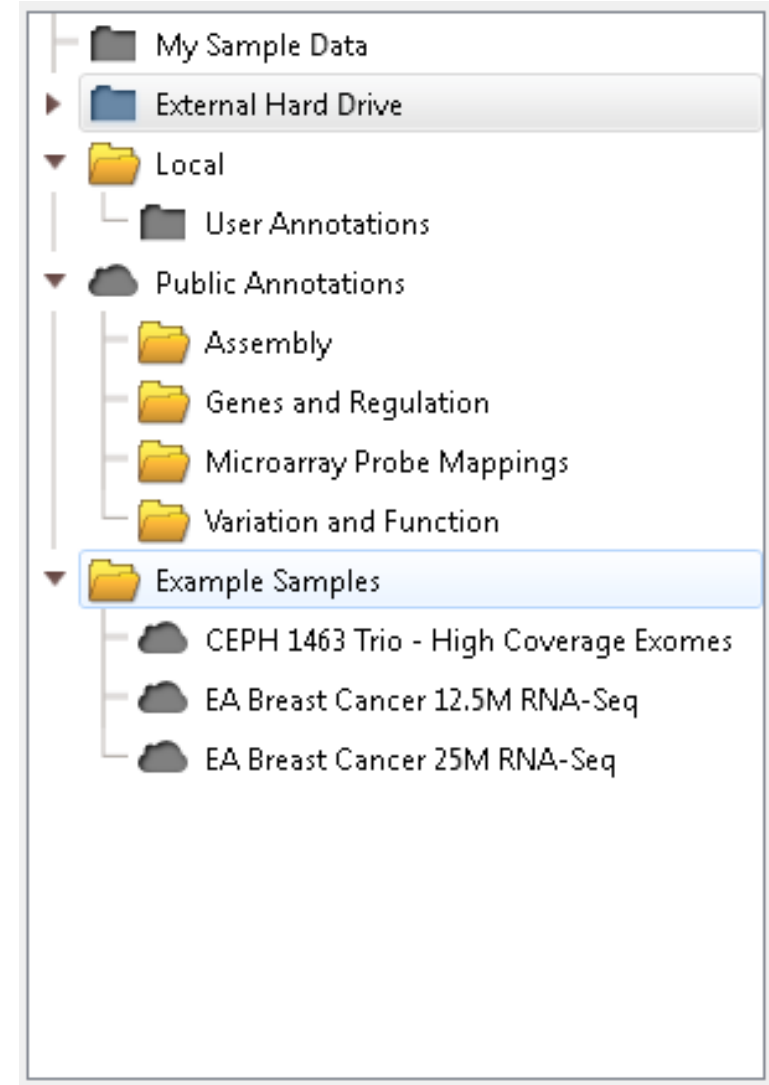


- Pick all your annotations up front, be guided through choices on each one
- Annotate smaller datasets over the network
- Chain your filters





- Have hierarchy and most important use versioning
- We never delete a previously published data source. Access any version at any time.
- Fast and easy streaming, or download using our new downloader
- Documentation built in
- Historical repository converted to TSF, but your existing files will continue to work.





[Demo]



1 History

2 New Annotation Infrastructure

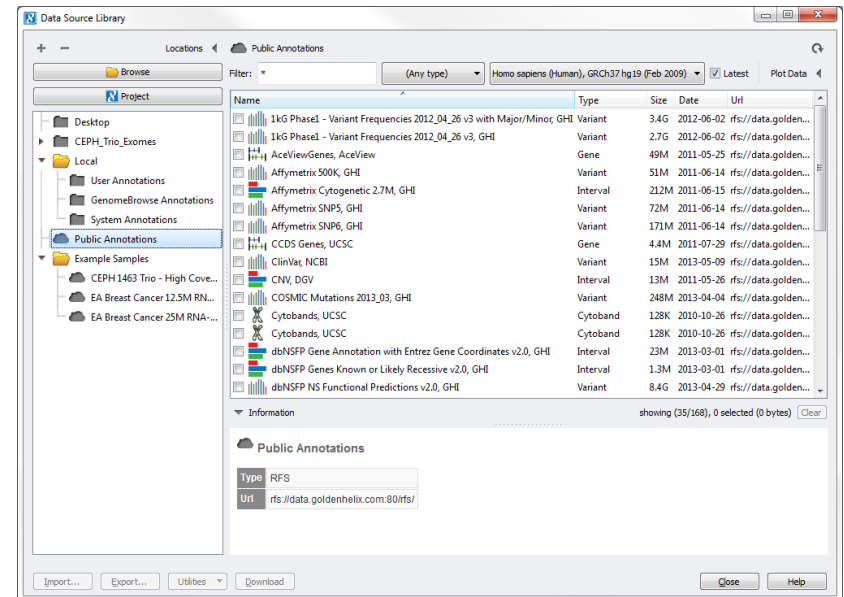
3 **GenomeBrowse Grows in Its Utility**

4 What's New in GWAS Analysis

5 Conclusion



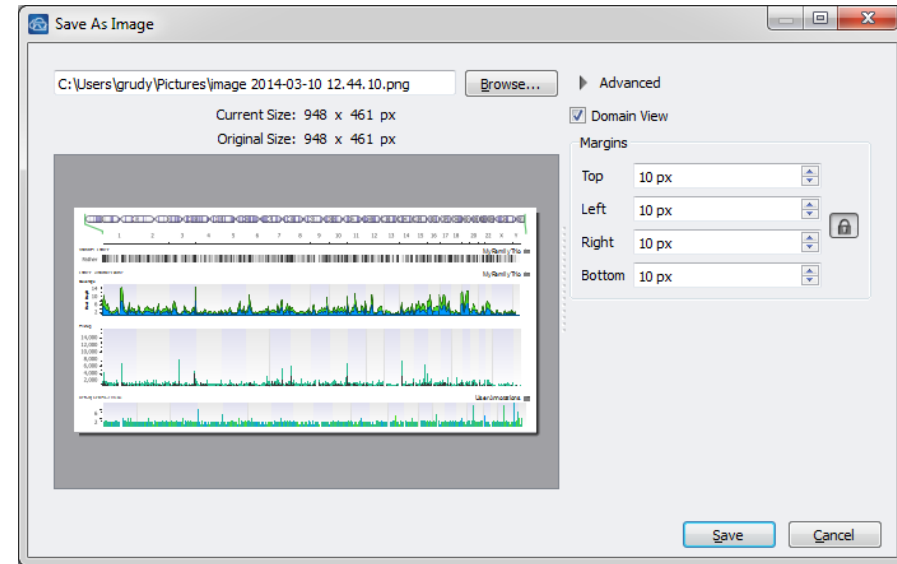
- Re-imagined GenomeBrowse's data management interface
 - Provides integrated preview of documentation
 - Easily add permanent folders to be in your library (with ability to be recursive)
 - Browse any folder to quickly find a plottable source



Save as Image



- Have had people tell us that they love GenomeBrowse as its renderings are better than publication quality
- Needed a flexible way to export that beautiful rendering
- We didn't want to cut corners, and the use cases are quite complex
- Have rewritten half of GenomeBrowse to be in a flexible architecture for all the plot rendering to be laid out and rendered to devices other than the screen.



- For this release, we have a great start, but have more to come



[Demo]



1 History

2 New Annotation Infrastructure

3 GenomeBrowse Grows in Its Utility

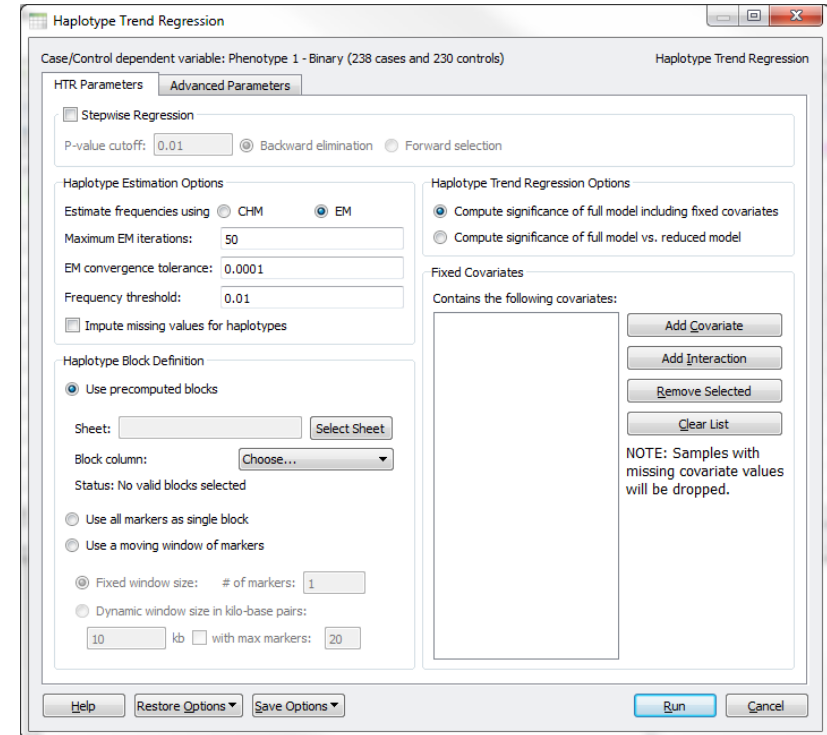
4 **What's New in GWAS Analysis**

5 Conclusion

Haplotype Trend Regression



- Haplotype Trend Regression one of our most requested features from our HelixTree line that wasn't supported in SVS 7
- Now SVS 8 has HTR with more:
 - Marker blocks
 - Covariates
 - Can use binary or continue dependent variable





[Demo]



- **Fst by Marker**
- **DNA-Seq Additions**
 - Classify by Inheritance Pattern
 - Calculate Alt Read Ratio
 - Score Variants by Dominant Model
 - Filter based on VCF Quality Metrics

GOLDEN HELIX
Accelerating the Quest for Significance

Products Services Community Resources Support Company

Search

QUICK LINKS

- Request Evaluation
- Request Pricing
- Package Selection Guide
- What's New
- Email or Call
1-888-569-4629

PRODUCTS

- SNP & Variation Suite
- SNP Analysis Package
- DNA-Seq Analysis Package
- RNA-Seq Analysis Package
- CNV Analysis Package
- Power Seat
- PBAT Analysis
- Viewer
- GenomeBrowse

OUR 2 SNPs...[®] GHI Blog

The New Human Genome Reference and Clinical Grade Annotations
[Read more](#)

What's New in SNP & Variation Suite 8

Welcome to Version 8 of SNP & Variation Suite (SVS), an integrated collection of user-friendly, yet powerful analytic tools for managing, analyzing, and visualizing multifaceted genomic and phenotypic data. Version 8 of SVS improves on many of the things you love about Version 7 and offers a plethora of new features that truly earn it the distinction of a version upgrade.

Taking Visualization to the Next Level

In September 2012, the Golden Helix GenomeBrowse[®] visualization tool was introduced to the genomic community as an intuitive way to view DNA-seq and RNA-seq pile-up and coverage data. As a free, stand-alone product, GenomeBrowse quickly gained popularity and today boasts over 2,500 users.

Since then, users of SNP & Variation Suite who also wanted to see visualizations in GenomeBrowse had to manage with a clunky copy-and-paste workflow to get data from one product to the other. Additionally, while SVS did have a genome browser, it was not able to provide the seamless user experience that GenomeBrowse could.

Not anymore.

Version 8 of SNP & Variation Suite takes visualization to a whole new level with the integration of GenomeBrowse for viewing any data in genomic space inside of SVS. Import your data once and have it at your fingertips for filtering, quality assurance, analysis, and then visualization with just one click. If you haven't yet tried GenomeBrowse, expect a fluid interface with easy-to-use controls such as zooming via the mouse scroll wheel as well as powerful options for changing the view like the ability to set the y-axis.

And when you are ready to publish, save your GenomeBrowse window as an image in a variety of file formats including PNG, JPG, and TIF.

Improving the Way Data Is Handled

With the release of SVS Version 8, annotation tracks will now be in a completely re-envisioned file format: TSF (replacing the old IDF file format). While the details are pretty cool, what you really need to know is that TSF gives the user more efficient data storage (at times up to 80% smaller file sizes) while providing an expanded number of field types and integrated field and source level documentation.

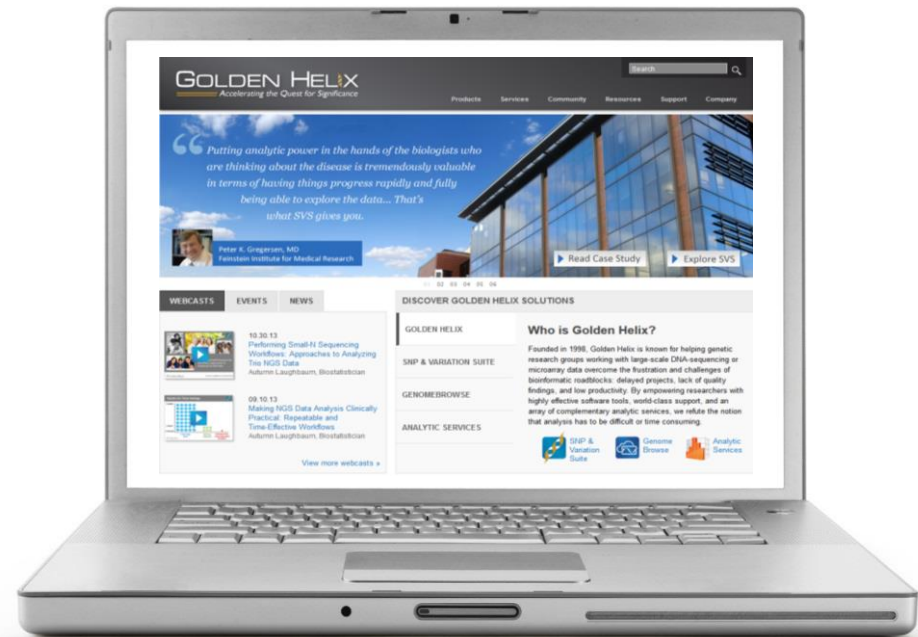
The conversion to TSF file format gave us the opportunity to create an all new data conversion wizard that can accept, churn, and spit out almost anything you throw at it. The benefit? The ability to make a custom annotation track in no time flat. (Our in-house data curation team loves it.)

Sign up for updates & info:
Name:
Email:



Updates are always free for customers

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com





Questions?

Use the Questions pane in your GoToWebinar window

