Genome Trax Review

Codes실 서지혜 _ 2016. 02. 10

INSILICOGEN

www.insilicogen.com

㈜ 인실리코젠

BIOBASE Products

INSILICOGEN
www.insilicogen.com

Human Mutation & Variant Analysis

HGMD®

HGMD® Professional is a unique resource providing comprehensive data on human inherited disease mutations to genetics and genomic research. Its compilation enables quick access to both single mutation queries and advanced search applications. HGMD® is widely used in human genetics research, diagnostics, and personal genomics applications and was an essential tool in analyzing the genomes »

read more

Genome Trax™

Genome Trax™ is a comprehensive compilation of variant knowledge that allows you to quickly and confidently identify pathogenic variants in human whole genome or exome sequences. If you are analyzing whole genome or exome sequencing data — Genome Trax™ is the tool to use. Genome Trax™ makes it easy to upload a complete genome's worth » read more

Pharmacogenomics Data

PGMD™

The PharmacoGenomic Mutation Database The PharmacoGenomic Mutation Database (PGMD™) is a resource for identifying all published genetic variants that have been shown to affect drug response in patients. We have mined the scientific literature for every in vivo patient study that has yielded a significant correlation between genotype and drug response, and offer multiple delivery » read more

Enzymology

BRENDA

BRENDA Professional is the main collection of enzyme functional data available to the scientific community worldwide. The enzymes are classified according to the Enzyme Commission list of enzymes. BRENDA Professional is maintained and developed at the Institute of Biochemistry and Bioinformatics at the Technical University of Braunschweig, Germany. Data on enzyme function are extracted directly » read more

Functional Analysis

PROTEOME™

PROTEOME is a knowledgebase containing data on genes and miRNAs, the diseases they are associated with, the drugs they are targeted by, and the pathways and networks they act within. Based on its extensive compilation of published functional attributes, the powerful ontology search query system allows scientists to quickly find answers to questions relevant to » read more

Gene Regulation Analysis

TRANSFAC®

TRANSFAC® is a unique knowledge-base containing published data on eukaryotic transcription factors and miRNAs, their experimentally-proven binding sites, and regulated genes. The extensive compilation of binding sites provides the most comprehensive data set of transcription factor – gene interactions available. The same data also forms the basis of derived positional weight matrices which can be » read more

ExPlain™

ExPlain™ is a unique data analysis system that combines promoter and pathway analysis tools. Using the power of TRANSFAC®'s transcription factor binding site derived positional weight matrices, ExPlain™ enables you to identify transcription factors affecting gene expression in your microarray and RNA-Seq experiments, as well as predict how they, in combination, can induce observed gene » read more

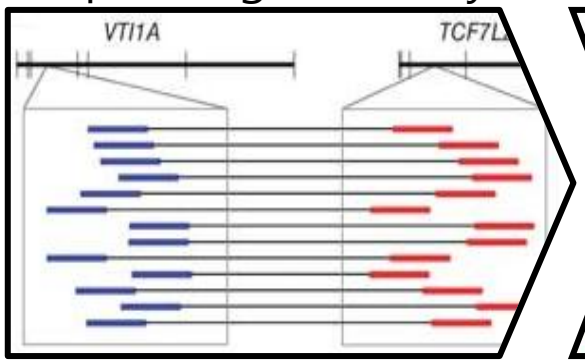
Annotation of Genetic Variants

ANNOVAR™

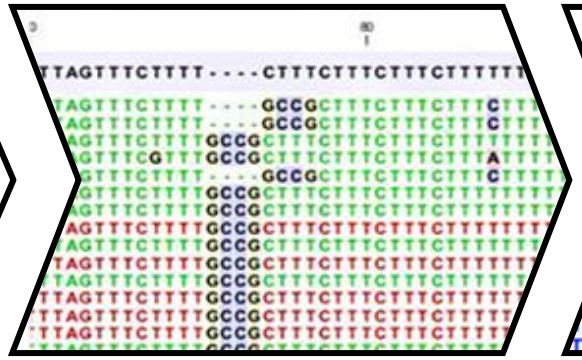
ANNOVAR, functional annotation of genetic variants from high-throughput sequencing data, is an efficient command line Perl program to functionally annotate genetic variants from diverse genomes (including human genome hg18, hg19, as well as mouse, worm, fly, yeast and many others). High-throughput sequencing platforms are generating massive amounts of genetic variation data, and it remains a » read more

NGS Analysis Pipeline

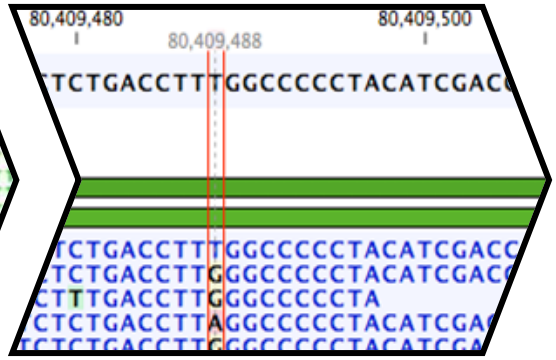
Sequencing Assembly



Alignment

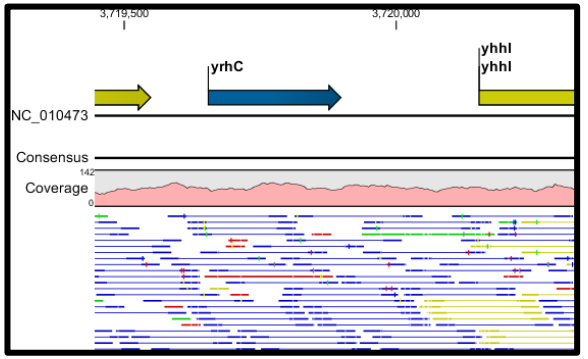


Variant detection



Map variations to genomic features
to analyze significance of variations

GFF, BED format



Upload data tracks

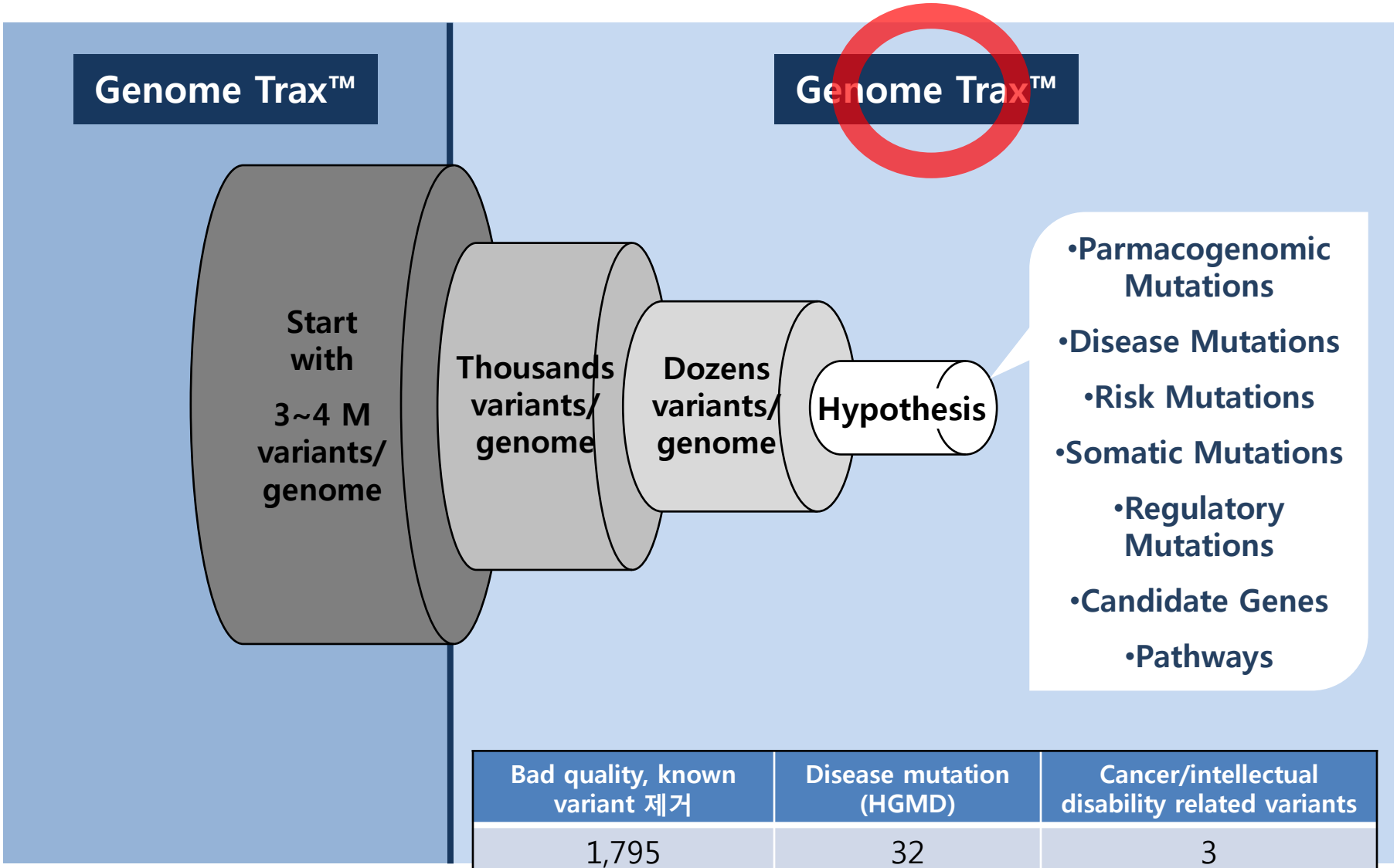
AGCTTGACTCCATGATGATT
AGCTTGACGCCATGATGATT
AGCTTGACTCCCTGATGATT
AGCTTGACGCCCTGATGATT
AGCTTGACTCCATGATGATT
AGCTTGACGCCATGATGATT
AGCTTGACTCCCTGATGATT
AGCTTGACGCCCTGATGATT

Regulatory sites

Novel mutation

Disease mutation

Variants Filtering



Bad quality, known variant 제거	Disease mutation (HGMD)	Cancer/intellectual disability related variants
1,795	32	3

Genome Trax™ for Next Generation Sequencing

Features

•알려진 variant 필터링 뿐만 아니라 관련있는 변이를 직접 알 수 있음.

•1억 6500만개 이상의 annotation

•대립유전자 frequency, 단백질 서열에 대한 효과와 functional annotation 등을 기반으로 새로운 variant의 영향 예측

•ANNOVAR, UCSC Genome Browser, Gaxaxy, CLC Genomics Workench, Avadis NGS, Alamut®, GeneData SimulConsult, Cartagenia BENCH, Microsoft Excel와 완벽한 호환

•Hg18/NCBI36, hg19/GRCh19, hg38/GRCh38 레퍼런스 지놈을 지원

Benefit

- 대용량 시퀀싱에서 mutation을 우선으로 함
- 질병에 관련된 mutation을 확인
- 변이 분석 비용과 시간을 절감
- 아직 정보가 없는 새로운 mutation을 발견, 영향 예측
- Variant에 관련된 질병이나 약물 pathway찾기

Users

1. Researchers
2. Pathologist
3. Diagnostic lab

Database includes...

1. HGMD®에서 방대한 데이터베이스
2. PGMD™ 전문적인 약물유전체 variant
3. TRANSFAC® regulatory site
4. PROTEOME™ 질병관련 유전자, drug target pathway

Contents Comparison

INSILICOGEN
www.insilicogen.com

Contents	GenomeTrax	HGMD
Disease/phenotype	○	○
Reference(Journal)		○
HGVS	○	○
Base Change	○	○
Amino-acid Change	○	○
OMIM	○	○
Refseq	○	○
dbSNP	○	○
Entrez ID	○	○
Phenbase		○
1000G Frequency		○
Variant Class	○	○
Sequence	○	○

Contents	GenomeTrax	HGMD
HGNC	○	○
Ensembl ID	○	○
dbNSFP2 predictions	○	○
PGMD	○	
ClinVar	○	
Pathway	○	
Transfac	○	
TSS	○	
Drug	○	
Microsatellites	○	
CpG island	○	
HGMD hyperlink	○	○

Genome Trax licensing

BIOBASE
BIOLOGICAL DATABASES
A QIAGEN® Company

Download page for Genome Trax data files

Genome Trax Documentation

BIOBASE Genome Trax File Documentation Version 2015.4

Genome Trax data files for hg18 (2015-12-31)

File:	Description:	Size:
Genome_Trax_hg18_gff.tar.gz	Parseable NGS hg18 GFF data files (for use in genome analysis tools, optimized for CLCBio Genomics Workbench)	14,659,724,493 Bytes
Genome_Trax_hg18_bed.tar.gz	NGS hg18 BED files (for use in genome browsers, optimized for UCSC Genome Browser)	4,287,319,026 Bytes
Genome_Trax_hg18_tsv.tar.gz	NGS hg18 PGMD data in tab delimited format	43,869,694 Bytes

Genome Trax data files for hg19 (2015-12-31)

File:	Description:	Size:
Genome_Trax_hg19_gff.tar.gz	Parseable NGS hg19 GFF data files (for use in genome analysis tools, optimized for CLCBio Genomics Workbench)	13,058,659,011 Bytes
Genome_Trax_hg19_bed.tar.gz	NGS hg19 BED files (for use in genome browsers, optimized for UCSC Genome Browser)	4,150,274,829 Bytes
Genome_Trax_hg19_vcf.tar.gz	NGS hg19 VCF files (for use in genome browsers and analysis tools that support VCF)	198,012,896 Bytes
Genome_Trax_hg19_tsv.tar.gz	NGS hg19 PGMD data in tab delimited format	42,291,618 Bytes

Genome Trax data files for hg38 (2015-12-31)

File:	Description:	Size:
Genome_Trax_hg38_gff.tar.gz	Parseable NGS hg38 GFF data files (for use in genome analysis tools, optimized for CLCBio Genomics Workbench)	13,323,879,358 Bytes
Genome_Trax_hg38_bed.tar.gz	NGS hg38 BED files (for use in genome browsers, optimized for UCSC Genome Browser)	4,271,744,589 Bytes
Genome_Trax_hg38_vcf.tar.gz	NGS hg38 VCF files (for use in genome browsers and analysis tools that support VCF)	200,655,605 Bytes
Genome_Trax_hg38_tsv.tar.gz	NGS hg38 PGMD data in tab delimited format	43,832,445 Bytes

Genome Trax MySQL dump (2015-12-31)

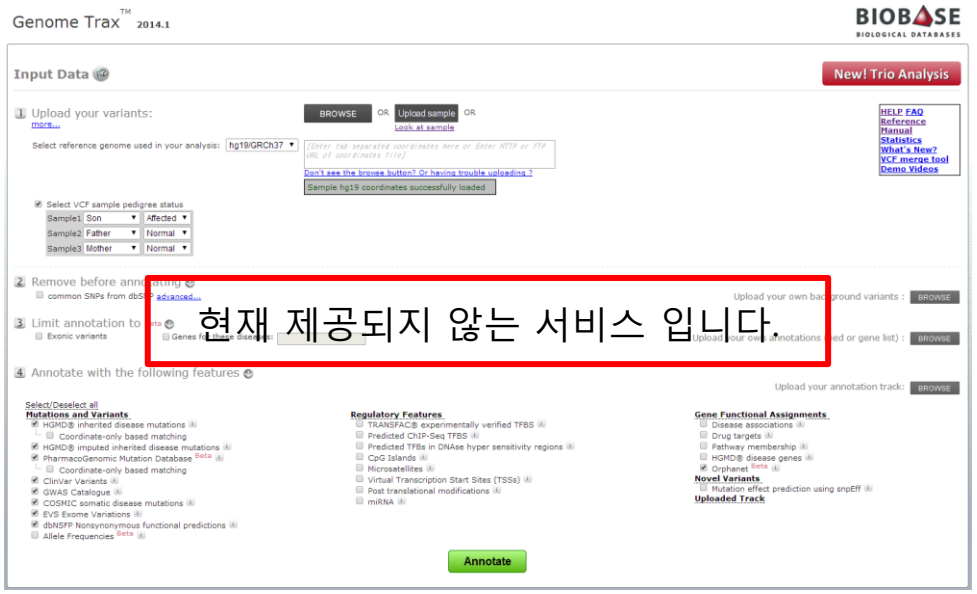
File:	Description:	Size:
Genome_Trax_2015.4.sql.gz	MySQL dump of the Genome Trax database	45,578,126,756 Bytes

Genome Trax data files for HGMD and PGMD (2015-12-31)

File:	Description:	Size:
Genome_Trax_HGMD_2015.4.tar.gz	All files pertaining to HGMD datasets.	50,531,153 Bytes
Genome_Trax_PGMD_2015.4.tar.gz	All files pertaining to PGMD datasets.	129,989,490 Bytes

HGMD coordinate free mutations (2015-12-31)

File:	Description:	Size:
Genome_Trax_HGMD_unmapped_gff.tar.gz	The mutations in HGMD that do not have a valid genomic positions are listed here. This file is formatted as a gff file with all the attributes and annotation, with no chromosome, start and end positions given. Please note that this file is not expected to work with other genome browsers or tools as they do not have a genomic valid position.	823,290 Bytes
Genome_Trax_HGMD_unmapped_bed.tar.gz	The mutations in HGMD that do not have a valid genomic positions are listed here. This file is formatted as a bed file with no data for chromosome, start and end position. Please note that this file is not expected to work with other genome browsers or tools as they do not have a valid genomic position.	152,159 Bytes
File:	Description:	Size:
md5sum.txt	md5sum's for the distribution files	



현재 제공되지 않는 서비스 입니다.

Add Custom Tracks

clade Mammal ▾ genome Human ▾ assembly Dec. 2013 (GRCh38/hg38) ▾

Display your own data as custom annotation tracks in the browser. Data must be formatted in bigBed, bigGenePred, bigWig, BAM, VCF, BED, BED detail, bedGraph, broadPeak, GFF, GTF, MAF, narrowPeak, Personal Genome SNP, PSL, or WIG formats. To configure the display, set track and browser line attributes as described in the User's Guide. Data in the bigBed, bigWig, bigGenePred, BAM and VCF formats can be provided via only a URL or embedded in a track line in the box below. Examples are here.

Error **Unrecognized binary data format in file chip-hg19.gff**

Paste URLs or data: Or upload: 파일 선택 sample_chip-….linking.bed Submit

Clear

Optional track documentation: Or upload: 파일 선택 선택된 파일 없음

Clear

Click here for an HTML document template that may be used for Genome Browser track descriptions.

Loading Custom Tracks

An annotation data file in one of the supported custom track formats may be uploaded by any of the following methods:

- (*Preferred*) Enter one or more URLs for custom tracks (one per line) in the data text box. The Genome Browser supports both the HTTP and FTP (passive-only) protocols.
- Click the "Browse" button directly above the URL/data text box, then choose a custom track file from your local computer, or type the pathname of the file into the "upload" text box adjacent to the "Browse" button. The custom track data may be compressed by any of the following programs: gzip (*.gz*), compress (*.Z*), or bzip2 (*.bz2*). Files containing compressed data must include the appropriate suffix in their names.
- Paste the custom annotation text directly into the URL/data text box.
- Data provided by a URL may need to be proceeded by a separate line defining type=<*track_type*> required for some tracks, for example such as "track type=broadPeak".

If a login and password is required to access data loaded through a URL, this information can be included in the URL using the format *protocol://user:password@server.com/somepath*. Only Basic Authentication is supported for HTTP. Note that passwords included in URLs are **not** protected. If a password contains a non-alphanumeric character, such as $, the character must be replaced by the hexidecimal representation for that character. For example, in the password *mypwd$wk*, the $ character should be replaced by %24, resulting in the modified password *mypwd%24wk*.

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Help | About Us

Manage Custom Tracks

genome: Human **assembly:** Feb. 2009 (GRCh37/hg19) [hg19]

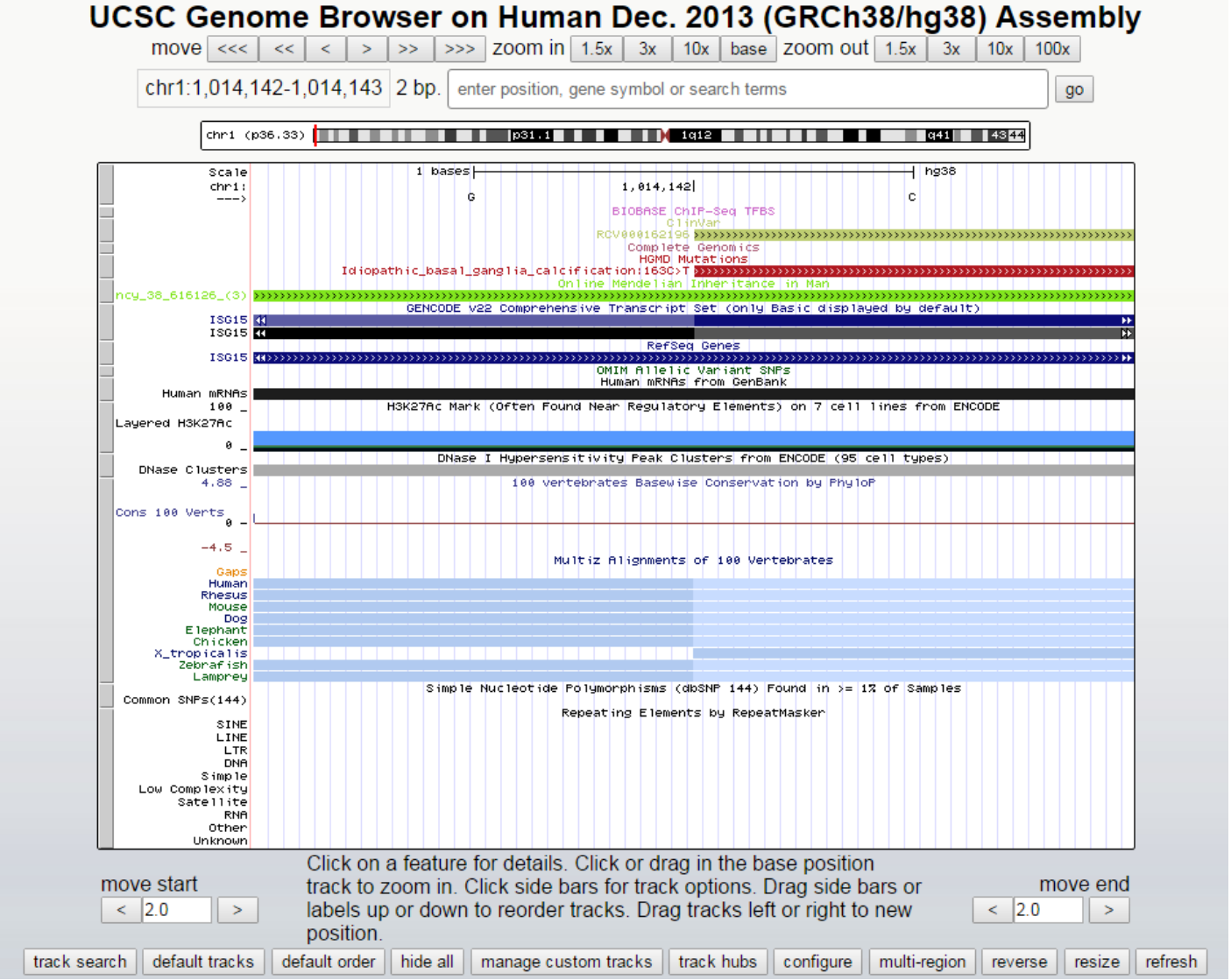
Name	Description	Type	Doc	Items	Pos	delete	update
HGMD Mutations	HGMD_Mutations	bed		13	chr2:		
COSMIC	COSMIC	bed		4	chr2:		
GWAS Catalogue	GWAS_Catalogue	bed		9	chr2:		
EVS	Exome_Variation_Server	bed		164	chr2:		
DbNSFP	Functional_predictions_for_non_synonymous_snps	bed		50	chr2:		
HGMD imputed	HGMD_imputed	bed		3	chr2:		
ClinVar	NCBI_ClinVar	bed		14	chr2:		
PGMD	PharmacoGenomic_Mutation_Database	bed		14	chr2:		
orpha	orpha	bed		153	chr2:		
					check all / clear all	+ -	+ -

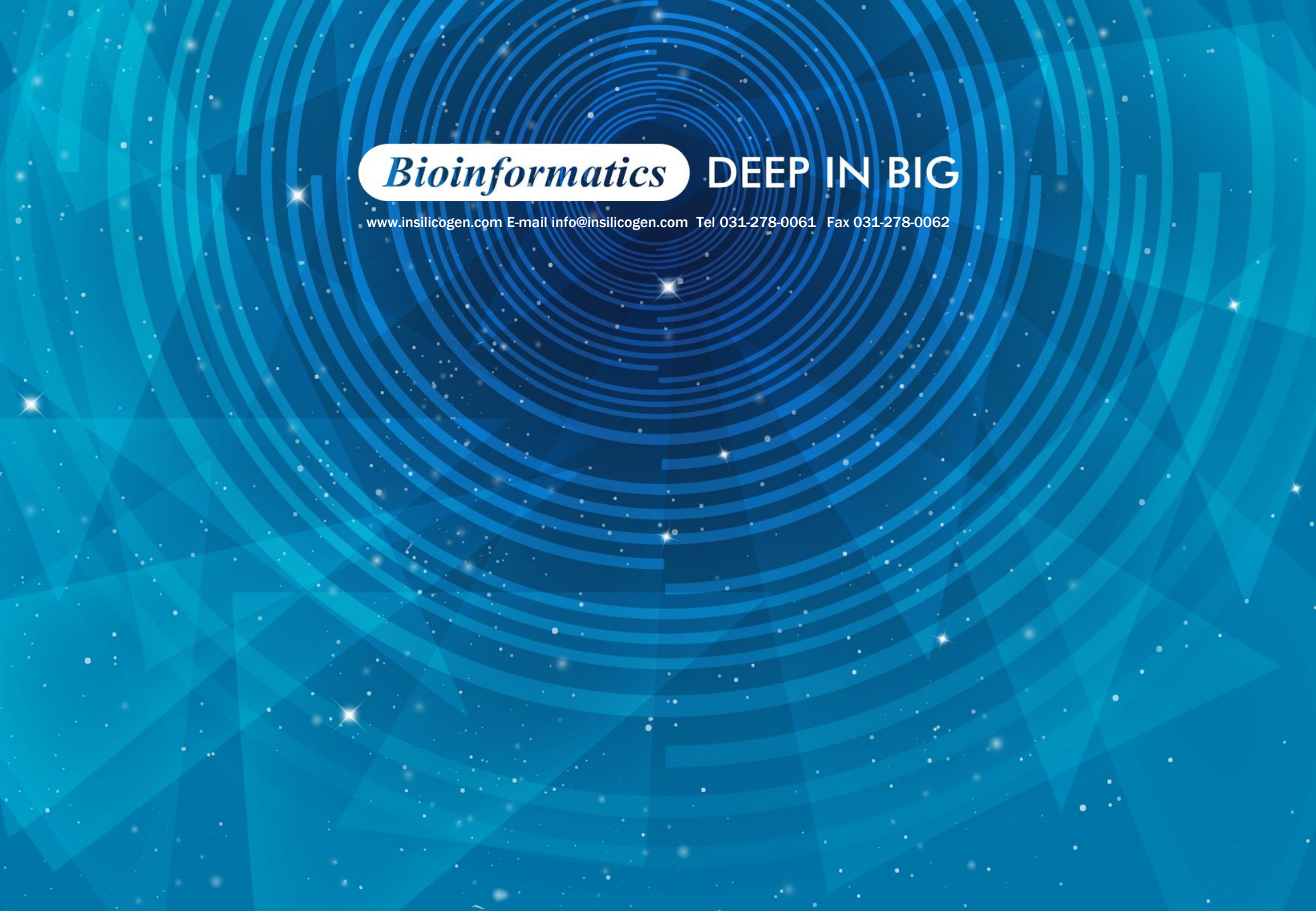
add custom tracks
go to genome browser
go to table browser
go to variant annotation integrator

Managing Custom Tracks

This section provides a brief description of the columns in custom track management table. For more details about managing custom tracks, see the Genome Browser User's Guide.

- **Name** - a hyperlink to the update page where you can edit your track data.
- **Description** - the value of the "description" attribute from the track line, if present. If no description is included in the input file, this field contains the track name.
- **Type** - the track type, determined by the Browser based on the format of the data.
- **Doc** - displays "Y" (Yes) if a description page has been uploaded for the track; otherwise the field is blank.
- **Items** - the number of data items in the custom track file. An item count is not displayed for tracks lacking individual items (e.g. wiggle format data).
- **Pos** - the default chromosomal position defined by the track file in either the browser line "position" attribute or the first data line. Clicking this link opens the Genome Browser or Table Browser at the specified position (note: only the chromosome name is shown in this column). The Pos column remains blank if the track lacks individual items (e.g. wiggle format data) and the browser line "position" attribute hasn't been set.




Bioinformatics
DEEP IN BIG
www.insilicogen.com E-mail info@insilicogen.com Tel 031-278-0061 Fax 031-278-0062