GOLDEN HELIX

Enabling Precision Medicine

Splice Site Algorithms for Clinical Genomics

CIOReview

20 most promising Biotech Technology Providers

pharma TECH OUTLOOK

Top 10 Analytics Solution Providers

Gartner

Hype Cycle for Life sciences

Thanks to NIH & Stakeholders

- NIH Grant Supported
  - Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R43GM128485. PI is Dr. Andreas Scherer, CEO Golden Helix. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

- ACMG Guidelines Author Collaborator:
  - Dr. Elaine Spector (Childrens Colorado, USA)
- Stakeholders:
  - Dr. Abdallah Elias (Shodair Children's Hospital, USA)
  - Dr. Ahmed Alfares, King Abdul Aziz Medical City, Saudi Arabia),
  - Dr. Bailey Glen (Medical University of South Carolina, USA)
  - Dr. Jim Weber (PreventionGenetics, USA)
  - Dr. Qin Hao and Dr. Line Larsen (Amplexa, Denmark)
  - Dr. Val Hyland (Pathology Queensland, Australia).

Q & A

Please enter your questions into your GoToWebinar Panel

Golden Helix – Who We Are

Golden Helix is a global bioinformatics company founded in 1998.

GlaxoSmithKline

VarSEQ®

- Variant Calling
- Filtering and Annotation
- Clinical Reports
- CNV Analysis
- Pipeline: Run Workflows

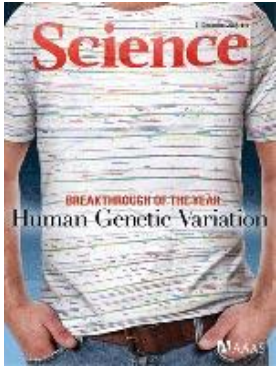
VARIANT WAREHOUSE

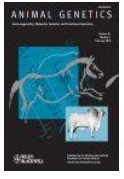
- Variant Warehouse
- Centralized Annotations
- Hosted Reports
- Sharing and Integration

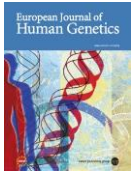
SNP & VARIATION SUITE

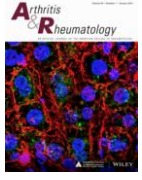
- GWAS
- Genomic Prediction
- Large-N-Population Studies
- RNA-Seq
- Large-N CNV-Analysis

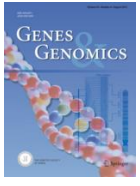
Cited in over 1100 peer-reviewed publications

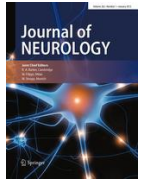












Over 350 customers globally

SickKids
THE HOSPITAL FOR SICK CHILDREN

Cincinnati Children's
Hospital Medical Center

The Children's Hospital of Philadelphia

Stanford University

UBC

Ucla

Yale

The Feinstein Institute for Medical Research
North Shore-Long Island Jewish Health System
North Shore LIJ

PREVENTION GENETICS
DISEASE PREVENTION THROUGH GENETIC TESTING

MAYO CLINIC

LINEAGEN

MOFFITT CANCER CENTER

Dartmouth-Hitchcock
NORRIS COTTON CANCER CENTER

BAYER

Lilly

abbvie

TEVA

GOLDEN HELIX
Enabling Precision Medicine

NATIONAL CANCER INSTITUTE

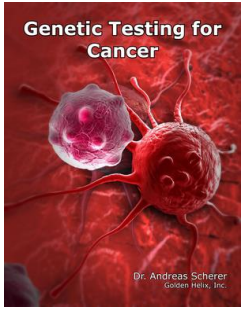
USDA
USDA Agricultural Research Service

FDA

Golden Helix – Who We Are

When you choose a Golden Helix solution, you get more than just software

- REPUTATION
- TRUST
- EXPERIENCE



- INDUSTRY FOCUS
- THOUGHT LEADERSHIP
- COMMUNITY

- TRAINING
- SUPPORT
- RESPONSIVENESS

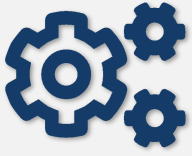
- INNOVATION and SPEED
- CUSTOMIZATIONS

Genetic Testing Process

Golden Helix Clinical Suite

Sample Prep

Sequencing



Align & Call

Sentieon & VS-CNV

Annotate & Filter

VarSeq

Variant Interpretation

VSClinical

Report

VSReports

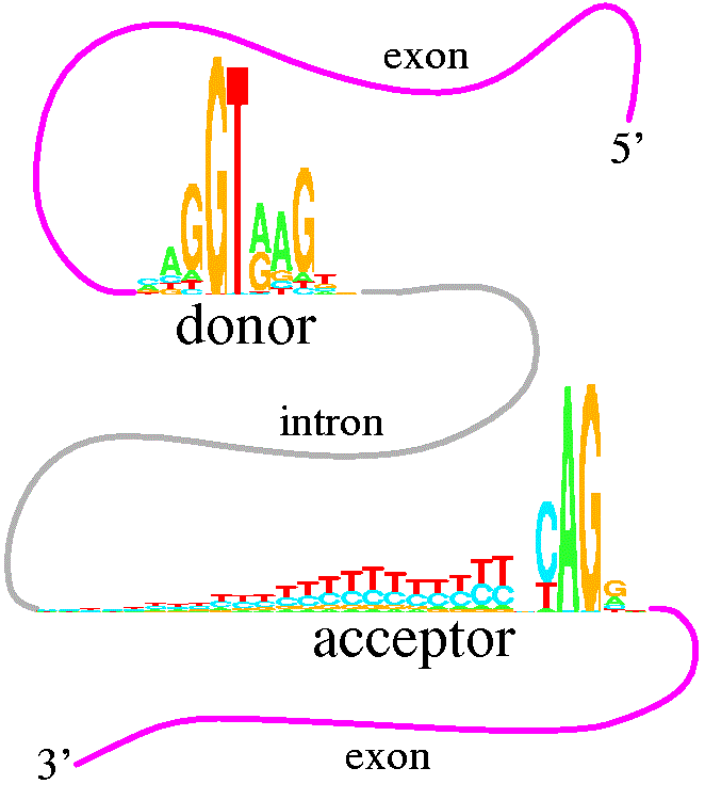
VSClinical

- **Complete Support for ACMG Guideline Workflow:**
  - Implements a guided workflow for following the ACMG guideline scoring and classifying
  - Place criteria into conceptually related groups, paired with their opposites, and formatted as answerable question.
- **Aggregate and Automate:**
  - Questions have supporting evidence presented with rich and interactive visuals
  - Automatically computed recommendations for questions that have explicit bioinformatic evidence, with supporting reasons for each answer.
- **Expert and Beginner Friendly:**
  - Start with descriptive summaries and recommendations for a variant
  - Deep dive into Population Catalogs, Gene Impact, Published Studies and Clinical tabs
  - Integrated documentation, readings on scoring criteria and citations

VSClinical®

Splice Sites

- **Introns have distinct nucleotide pairs at each end**
  - GT at the 5’ end (Donor Site)
  - AG at the 3’ end (Acceptor Site)
- **Sequences around splice sites are highly variable**
- **Machine learning and probabilistic methods are used to identify sites**



Algorithms

- **VSClinical supports four splice site prediction algorithms**
  - PWM: Uses position weight matrix similar to SpliceSiteFinder and Human Splice Finder
  - MaxEntScan: Approximates sequence motifs using Maximum Entropy Distribution
  - NNSplice: Identifies splice sites using neural networks
  - GeneSplicer: Uses Markov models combined with maximal dependence decomposition

Experiments

- **Algorithms were compared in terms of**
  - Accuracy
  - Sensitivity
  - Specificity
  - Precision
- **The test data set was generated as follows:**
  - 20,000 Known splice sites were extracted from the 1000 Genomes GRCh38 reference sequence using exon boundaries specified by NCBI RefSeq Genes
  - This was combined with 20,000 false splice sites from the HS3D splice site dataset

Donor Splice Site Results

	Accuracy	Sensitivity	Specificity	Precision
PWM	79.7	82.4	79.5	28.0
MaxEntScan	87.0	72.9	94.9	88.9
NNSplice	87.1	83.6	94.4	96.9
GeneSplicer	88.2	85.4	93.8	96.6

Acceptor Splice Site Results

	Accuracy	Sensitivity	Specificity	Precision
PWM	82.7	81.8	84.4	91.3
MaxEntScan	89.5	89.0	90.3	95.1
NNSplice	81.5	77.4	90.0	94.2
GeneSplicer	92.0	92.9	90.1	95.1

Discussion

- **GeneSplicer has exceptional performance for all metrics**
- **MaxEntScan has high accuracy and is competitive with GeneSplicer in terms of specificity**
- **NNSplice also performs well and is competitive with MaxEntScan on donor data**
- **PWM has a high false positive rate and performs poorly on donor data**

[Demo in VarSeq]

Q & A

Please enter your questions into your GoToWebinar Panel