Shannon Human Splicing Pipeline Documentation

Version 2.0

Last Revised: Oct 22, 2013

www.cytognomix.com

700 Collip Circle #150

London Ontario N6G 4X8

Canada

info@cytognomix.com

Table of Contents

Installation Guide:

Shannon Human mRNA Splicing Pipeline

Pre-installation requirements: Either CLC Genomics Workbench or CLC Genomics Workbench and CLC Genomics server. Perl and gcc were previously required, however the plugin now contains a compiled binary to alleviate the need for those two dependencies.

Summary of changes in version 2.0

- Runs on Windows Vista, 7, or 8 (64 bit)
- Is approximately 12 times faster (analyze complete genomes in 10-15min)
- Handles indels (beta)
- Does gene set overrepresentation analysis (pathway analysis)
- Preset filters to simplify post-run analysis
- Updated to include RefSeq gene annotations

Standalone CLC Bio Genomics Workbench Version (Server license or support either undesired or unavailable)

1. Open CLC-Bio Workbench and select the "Plug-ins" button from the toolbar
2. Uninstall previous version(s) of plugin, if present
3. We will now install the Shannon pipeline plugin. This can be done in two different ways: "Workbench automatic install" - Download and install automatically through the CLC bio Genomics Workbench or "Manual download and install" - Download plugin from the CLC bio website and install manually.

   Installation methods - choose one of the following:
   Workbench automatic install
   - Within the Genomics Workbench, click the *Plug-ins* button in the top toolbar.
   - Select the *Download Plug-ins* tab.
   - Select the plugin "**Cytognomix Shannon Pipeline**" and click *Download and Install*.

   Manual download and install
   - Download the plugin "**Shannon Human Splicing Pipeline**" from http://www.clcbio.com. The downloaded plugin will be a cpa file.
   - Within the Genomics Workbench, click the *Plug-ins* button in the top toolbar.
   - Click the *Install from File* button, browse computer to find the cpa file you previously downloaded and click *Install*.

4. Necessary files containing genomic annotations (Ensembl gene, RefSeq, dbSNP, and hg reference sequence) must be installed for use by the main pipeline. Like the

software, these plug-ins are available as part of your purchase. There are separate 'dependencies' plug-ins for each genome build available. Currently, dependencies plug-ins are available for hg18/NCBI36 and hg19/GRCh37.

- If you intend to examine variants that are mapped to hg19 coordinates, install the plugin "**Cytognomix Dependencies HG19**" using the same methods found in step 3.
- If you intend to examine variants that are mapped to hg18 coordinates as well as hg19, the plugin "**Cytognomix Dependencies HG18**" (~2 Gb) must be installed in the same manner.
- If variant data sources are mixed, ie. from both hg18 and hg19, both "**Cytognomix Dependencies HG18**" and "**Cytognomix Dependencies HG19**" must be installed.
- Keep in mind that at least one dependencies file must be installed for the pipeline to function. Otherwise an error is generated at run time.

5. Restart Genomics Workbench to complete installation

Installation of Genomics Workbench Client-Genomics Server Version (both CLC Workbench and Server licenses are active)

1. First, install the CLC Bio Workbench version - as outlined above - under the heading "Standalone Client Version". It is not necessary to install "**Cytognomix Dependencies HG18**" or "**Cytognomix Dependencies HG19**" on the client Workbench if you never intend to use the client computer to run the pipeline. (ie: every run will take place on the server). If you want to run the Shannon pipeline computations on both the Workbench or the server, then dependency files must be installed on the respective computers. The "Dependencies" plug-ins must be installed on the Workbench when calculations are performed locally (without running the software on the server).
2. Ensure the CLC-Bio Genomic Server is running. For more instructions on how to set up and access the server, see CLC-Bio's Genomics server documentation
3. Access the server through your web-browser and log in to the Server
4. Select the Plug-ins option in the Admin tab
5. Uninstall any previous version(s) of the Shannon pipeline for human splicing mutations plug-in
6. Install the main pipeline plug-in "**Shannon Human Splicing Pipeline Server"**.
   - Download the plugin "**Shannon Human Splicing Pipeline Server**" from http://www.clcbio.com. The downloaded plugin will be a cpa file.
   - Select the Plugins option in the Admin tab.
   - Under the *Install new plugin* heading click *Choose File*, browse computer to find the cpa file you previously downloaded and Click *Install*.

7. Necessary files containing genomic annotations (Ensembl gene and dbSNP, hg reference sequence) must be installed for use by the main pipeline on the Genomics Server. Like the software, these plug-ins are available at the CLC Bio website as part of your purchase. There are separate 'dependencies' plug-ins for each genome build. Currently, dependencies plug-ins are available for hg18/NCBI36 and hg19/GRCh37.
   - The dependencies plug-ins are large (~2Gb each). By default, the CLC-Bio server does not allow plug-ins of this size to be installed. To change the settings, under the Main configuration tab->HTTP settings: modify the 'Max upload size (MB)' value to ≥ 3000 Mb.
   - If you intend to examine variants that are mapped to hg19 coordinates, install the plugin "**Cytognomix Dependencies HG19**" (~2 Gb).
   - If you intend to examine variants that are mapped to hg18 coordinates as well as hg19, the file "**Cytognomix Dependencies HG18**" (~2 Gb) must be installed.

- If variant data sources are mixed, ie. from both hg18 and hg19, both "**Cytognomix Dependencies HG19**" and "**Cytognomix Dependencies HG18**" must be installed.
- Keep in mind that at least one dependencies file must be installed for the pipeline to function. Otherwise an error is generated at run time.

8. If you have not previously defined *File system locations* for your server as outlined in CLC bio's server documentation, this must be done before the Shannon pipeline can be run. The Server must have access to a CLC bio folder on your hard drive where CLC objects and files are defined in advance of running the plug-in. Please refer to the CLC bio Genomics Server installation guide for details on Server set up.

* Requirements

The Cytognomix Shannon human mRNA splicing plug-in runs in standalone mode on the CLC Genomics Workbench V6.5 or with both the Workbench and CLC Genomics Server V6.5 (as a standalone server using grid). Released for the following 64 bit operating systems: **Windows (Vista, 7, 8)**, **Linux,** and **MacOSX**. This plugin requires at least 4Gb of RAM.

Support

CLC Bio Customer Support (primary)

Cytognomix Inc. (secondary)
700 Collip Circle #150, London ON N6G 4X8 Canada
info@cytognomix.com ♦ www.cytognomix.com

Trial/Web Server Login

The trial version of the server does not report all of the results that the fully licensed version does:
-Any number of variants may be submitted.
-The set of variants shown is randomly selected from a larger set of results.
-Up to 20% of all possible results are shown.
-Only a few inactivating or leaky variants are returned.
-No more than 500 variants with changes in information content are displayed.
-User access is anonymous, but IP addresses are tracked.
-The only Server function that is enabled is the Shannon pipeline plugin.
-The full version of the plugin contains none these limitations.

***Please Note**: The server may be accessed without completing steps 1 and 3 below. The server contains results generated from the full version of the plugin which can be viewed without installation of the trial. However, if you would like to examine your own variants (or analyze the sample variants in the _SAMPLE-VARIANTS folder) on the trial server, all of the following steps are required:

1. Uninstall any previously installed Shannon Human Splicing Pipeline plugins.

2. Installation of a CLC bio plugin is required to allow the workbench to connect to a server. This plugin is not created by Cytognomix, it is a CLC bio plugin and simply allows the workbench to connect to any server. If you can already connect to other servers, this step is unnecessary.
- From within the Genomics Workbench, click the plugins button in the top toolbar
- Select “Download Plugins”
- Install the plugin named “CLC Workbench Client Plugin”

3. We will now install the client plugin used to submit and retrieve Shannon pipeline data from the trial server.
- From within the Genomics Workbench, click the plugins button in the top toolbar
- Select “Download Plugins”
- Install the plugin named “Cytognomix Shannon Pipeline Client”

4. Connect to the server
- From within the Genomics Workbench, select file->CLC Server Login
- If necessary, expand the advanced option to uncover 'Server host' and 'Server port'
- Login using the following credentials.
    - User name: trial
    - Password: Cytognomix
    - Server host: 208.75.74.35
    - Server port: 7777

- Click login

5. If desired, you can view some results generated by the full version of the plugin in the folder _SAMPLE-RESULTS-FROM-FULL-VERSION.

6. Your own data in VCF format may be imported at this time. Alternatively, this step may be skipped and sample variants will be examined instead. To examine your own VCF files the data **must** be imported using the following method.

VCF files must be imported. The file must be a standard VCF file with at least the first five columns present. The necessary fields are CHROM, POS, ID, REF, ALT in that order. File headers are not necessary and will be ignored if present. When specifying indels, the reference nucleotide field must include the base preceding the event, which must also be reflected in the position field.
For example, the following lines are acceptable:
5 148835675 . C T
5 148989410 ID1 A G,T
5 148989435 CAGT C (deletion)
5 148989435 C CAAA (insertion)

To import the data, click the import button on the taskbar in the CLC-Bio workbench and select 'Standard Import'. Select the file to be imported and select **force import as type: Shannon Pipeline VCF Format**. The imported object can now be used as input for the Shannon pipeline.

For more instructions regarding the import process, please consult the Shannon Human Splicing Pipeline documentation (Quick start section) on how to import either VCF or Shannon pipeline basic variant. For simplicity, the steps given below will use pre-imported data that are already resident on the Trial Server. However if you import your own data, simply use your imported data as pipeline input instead of sample variants in step 9.

7. In the Genomics Workbench toolbox, expand Shannon Human Splicing Pipeline and double click 'Launch Pipeline'.

8. A wizard will pop up. Select CLC Server and click next.

9. Expand the folder _SAMPLE-VARIANTS to view its contents. Select 'Pre-Imported-hg19-Variants-Ready-to-be-examined-by-the-Shannon-Pipeline' and move it to the 'Selected Elements' region of the wizard. Click next.

10. The pre-imported variants are hg19, so ensure hg19 is selected in 'Genome Build' and click next.

11. Pipeline results should be saved and not opened. This will be selected by default so click next on the Result handling wizard screen.

12. Create a folder to store your results. To do this, highlight the Trial_Server_Data folder and press the +folder button. Name the folder whatever you would like. Highlight your newly created folder and click finish to begin your run.

13. With this data (approximately 5000 variants on 3 different chromosomes), the run will take approximately 30 seconds to complete (if using the pre-imported variants), so be sure to check back to review your results.

Quick Start

This page contains information on how to run the plugin. For an overview of information theory please view the "Review" section.
Note: This guide assumes that the Shannon Human Splicing plugin has been installed. For installation help, please consult the installation guide (above).

Importing Data

Before analysis can take place, input data containing variants is needed. The data to be examined must be in one of the two formats described below. Version 2.0 introduces a beta version of indel analysis* (please view “Note on indel analysis” section below).

Import option 1 - VCF files (recommended method)

VCF files must be imported. The file must be a standard VCF file with at least the first five columns present. The necessary fields are CHROM, POS, ID, REF, ALT in that order. File headers are not necessary and will be ignored if present. When specifying indels, the reference nucleotide field must include the base preceding the event, which must also be reflected in the position field.
For example, the following lines are acceptable:
5 148835675 . C T
5 148989410 ID1 A G,T
5 148989435 CAGT C (deletion)
5 148989435 C CAAA (insertion)

To import the data, click the import button on the taskbar in the CLC-Bio workbench and select 'Standard Import'. Select the file to be imported and **select force import as type: Shannon Pipeline VCF Format**. The imported object can now be used as input for the Shannon pipeline.

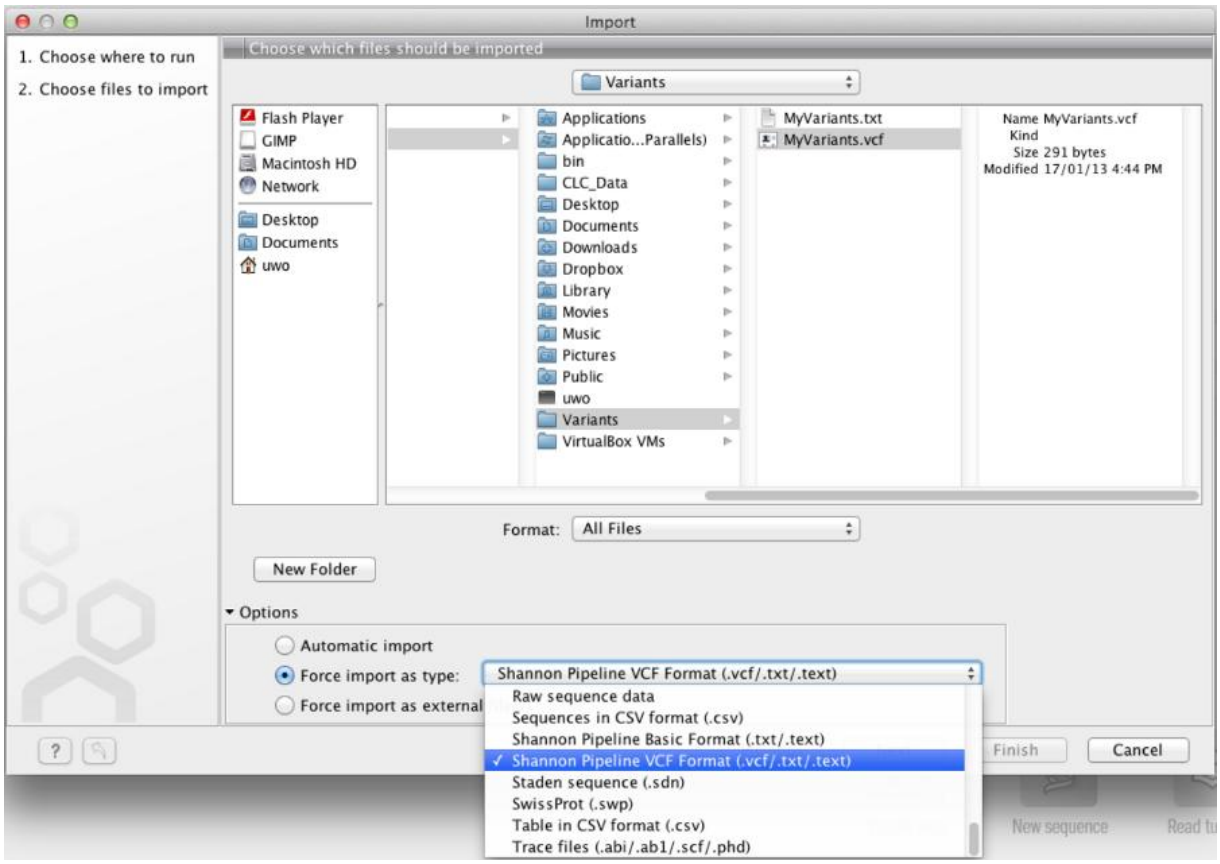


Figure 1. A demonstration of the Option 1 (VCF) import process. Select force import as: Shannon Pipeline VCF Format

Import option 2 - Variant tracks

CLC-Bio variant tracks can be used as input for the pipeline. If the variants to be examined are already located within a variant track object in the CLC-Bio environment, no further import is necessary. Otherwise, a VCF file can be imported as a variant track.
To import a VCF file as a variant track, click the import button on the taskbar in the CLC-Bio workbench and select 'Tracks'. Under type of files to import, select VCF. You will be asked to choose the file to be imported as a variant track as well as a reference track. For additional help regarding generating a reference track, please refer to the CLC-Bio documentation covering track import. When executing the Shannon pipeline with track input, the run will take slightly longer because the track is exported and imported behind the scenes to reorder data.

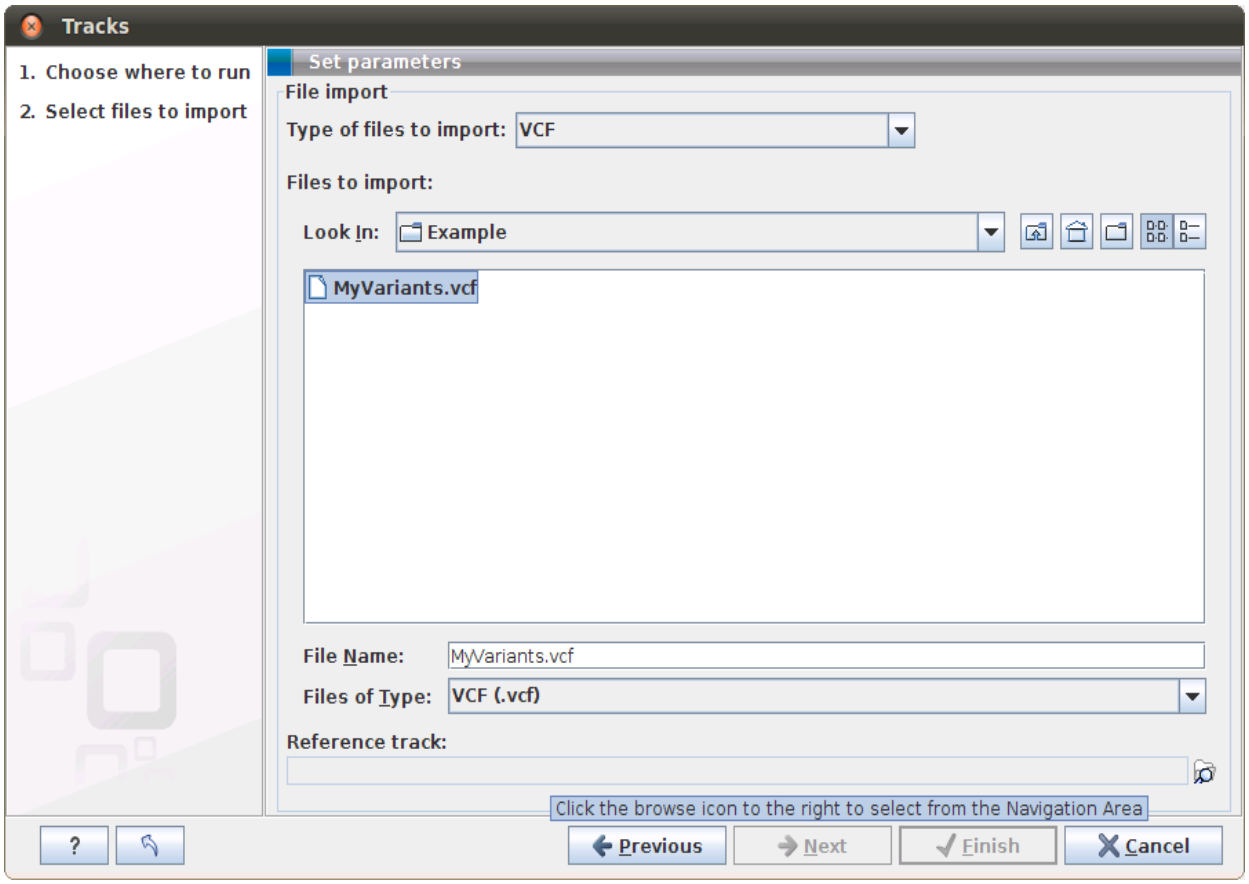


Figure 2. A demonstration of the Option 2 (Track) import process. The imported variant track may be used as input for the Shannon pipeline.

Running multiple samples together in a single run

The ID field in a VCF file can be used as a sample label to facilitate separating the samples after the run. To label all variants from the same sample, place the name of the sample in the ID field in a VCF file. Do this for all samples and place them in the same VCF file. Import the single VCF file containing multiple samples using Shannon Pipeline VCF Import. This enables multiple samples to be analyzed in a single run. Table filtering can then be used when the run is complete to examine results of each sample separately.

Example data

Example data was automatically placed in the directory "ExampleData_ShannonPipeline" upon plugin installation. Three descriptively named objects can be found in the directory:

- 1. Pre-Imported-hg19-Variants_Ready-to-be-examined-by-the-Shannon-Pipeline: The result of either importing file 2 using Option 1 in the import section above or importing file 3 using Option 2. This object may be used to test the Shannon Pipeline.
- 2. SampleBasicFormat_Would-be-imported-by-forcing-import-as-Shannon-Pipeline-Basic-Format.txt: This is an example of the Shannon Pipeline Basic format before importing.

- 3. SampleVCF_Would-be-imported-by-forcing-import-as-Shannon-Pipeline-VCF.vcf: An example of the Shannon Pipeline VCF format before importing

This example data was included to act as a starting point for first time users of the plugin. If it is no longer needed, the ExampleData_ShannonPipeline may be deleted. If the ExampleData_ShannonPipeline folder was accidentally deleted, please reinstall the plugin as the example files are placed in the directory upon installation.

Running the pipeline

After importing (data must be imported) your data and clicking on "Launch Shannon Pipeline" (located in the toolbox, under Shannon Human Splicing Pipeline), a wizard will appear asking whether you would like the analysis to take place on the server or the workbench. Each step of this wizard is described below:

- Step 1: Select the desired analysis location (workbench, CLC server, or grid). If using the trial server (only for trial version of the plugin), select CLC Server.
- Step 2: Select your imported variant data (Data must be a track or imported using one of the methods above. See the importing and example data sections above for help. The file Pre-Imported-hg19-Variants_Ready-to-be-examined-by-the-Shannon-Pipeline in the ExampleData_ShannonPipeline can be used in this step if you have not imported your own data yet.)
- Step 3: Select the desired reference genome (hg19 or hg18) and file for exome annotation (RefSeq or Ensembl 66).
- Step 4: Select whether you would like to save or open your results. We suggest that you select save and check "make log". Delta R_i and Final R_i plot checkboxes are available here to enable or disable their construction.
- Step 5: Select your desired location for the results data
- Upon clicking finish, the pipeline will begin analysis
- Only one analysis can be run at a time on the CLC Genomics Workbench. On the CLC Server, multiple analysis are queued and then run consecutively. Using the licensed grid version, multiple analyses can be run simultaneously.

A pop-up window indicates the run is "Done" when complete.
Your results are located in the directory specified in the final step of the Launch Shannon Pipeline wizard. They include: tabular output split into 4 files (complete, inactivating, leaky, and cryptic), plots for every chromosome and a genome-wide Manhattan style plot which includes all the variants.

Displaying results

In the navigation area of the workbench, double click a tabular or plot results object. The objects will be displayed in an appropriate editor.

* *Note on indel analysis:*

This is a beta version with a known issue which can generate incorrect R_i values (when examining indels). We estimate correct R_i values will be generated for ~95% of indels. The

issue is related to incorrect shifting of coordinate systems generally related to insertions/deletions directly overlapping a splice site. Therefore, please keep in mind that some R_i values related to indels may be incorrect.

Tables

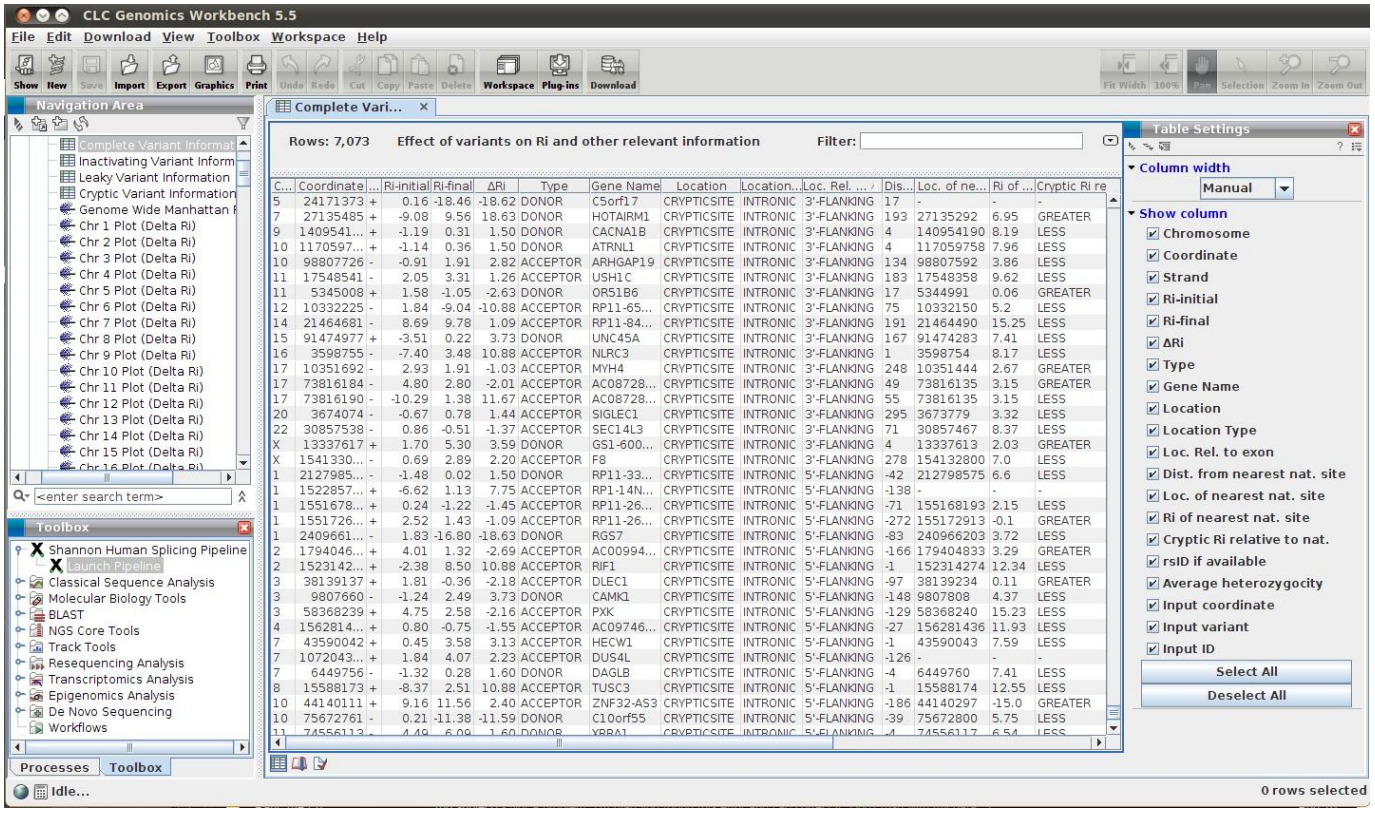


Figure 1. An example of a table generated by the pipeline. Columns which are not of immediate interest can be temporarily removed by unchecking the checkboxes within the sidebar on the right.

The tables contain all information gained through the information analysis on variants. Four tables are generated each time the pipeline is executed, these are:

1. Complete Variant Information
All sites exhibiting a delta R_i of at least 1.0 bits (and less than -1.0 bits) are included here

2. Inactivating Variant Information
Includes natural site variants with an original R_i greater than 1.6 bits and which drop below that value after the variant is introduced

3. Leaky Variant Information
Natural site variants which experience a drop in R_i after the variant is introduced

4. Cryptic Variant Information
Includes only cryptic site variants

Each row of the table represents a single variant. The meaning of each column is described below:

1. Chromosome
Chromosome containing the splice site experiencing a change in R_i

2. Splice site coordinate
Location of the splice site experiencing a change in R_i

3. Strand
Displayed as "+" for positive and "-" for negative strand

4. R_i-initial
R_i of splice site before introducing the variant

5. R_i-final
R_i of splice site after introducing the variant

6. Delta R_i
The change in R_i before and after introducing the variant

7. Type
The site is either an acceptor or a donor. Displayed as "ACCEPTOR" or "DONOR"

8. Gene Name
Name of the gene closest to the location of the variant. If multiple genes overlap the coordinate of the variant, they will all be appear in a comma delimited list.

9. Location
The site is either natural or cryptic. Displayed as "NATURALSITE" or "CRYPTICSITE"

Columns displayed only for cryptic site variants
10. Location Type
If the location of the variant is within an exon it is "EXONIC". Otherwise, it is "INTRONIC"

11. Location relative to exon
If the location of the variant is "INTRONIC" and within 300 base pairs of an exon, depending on its location relative to the exon it is "3'-FLANKING" or "5'-FLANKING"

12. Distance from nearest natural site
If the location of the variant is within 1000 base pairs of a natural site, the number of base pairs separating the two sites is shown here

13. Location of nearest natural site
If the location of the variant is within 1000 base pairs of a natural site, the coordinates of the nearest natural site are shown here

14. R_i of nearest natural site

15. Cryptic R_i relative to natural site R_i
If a cryptic site has a higher R_i than the nearest natural site after the variant is introduced it is

"GREATER", otherwise it is "LESS"

Additional columns displayed for all variants

16. rsID if available
dbSNP135 is examined to determine if the variant in question is a known variant. If it is found within dbSNP135, its rsID is displayed

17. Average heterozygosity
If the variant is a known SNP in dbSNP135, its average heterozygosity is displayed

18. Variant coordinate
Location of SNP which was examined by the pipeline

19. Input variant
Reference and variant nucleotides of SNP

20. Variant type
Denotes the variant as an 'SNV' of 'Indel'.

21. Input ID
ID as specified in the 'ID' column of input track or VCF file followed by a unique number

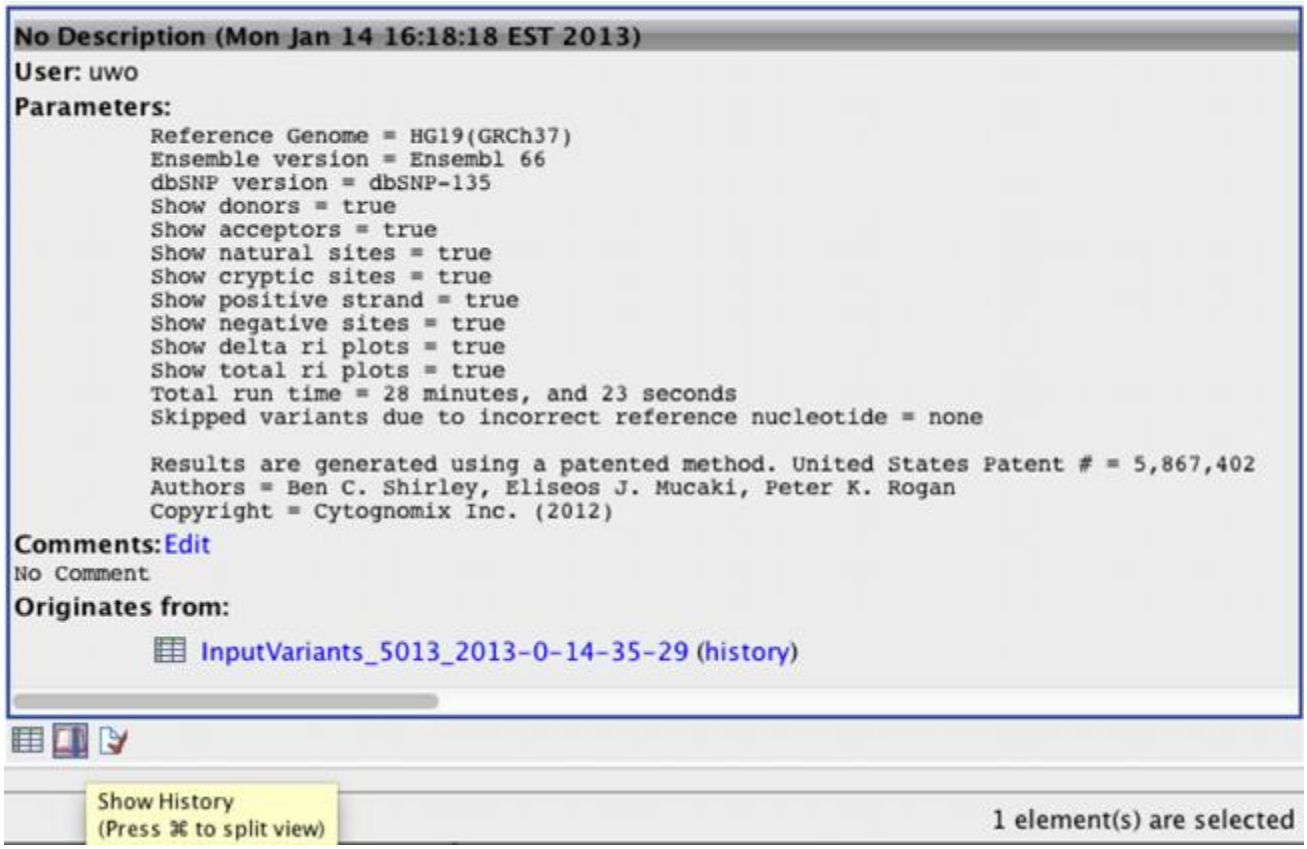


Figure 2. An example of table history generated by the pipeline. While the table editor is open (such as in figure 1), select history at the bottom of the screen. Genome version, Ensembl version, filer options, run time, skipped variants, and the input file for the appropriate pipeline execution can be found here.

Plots

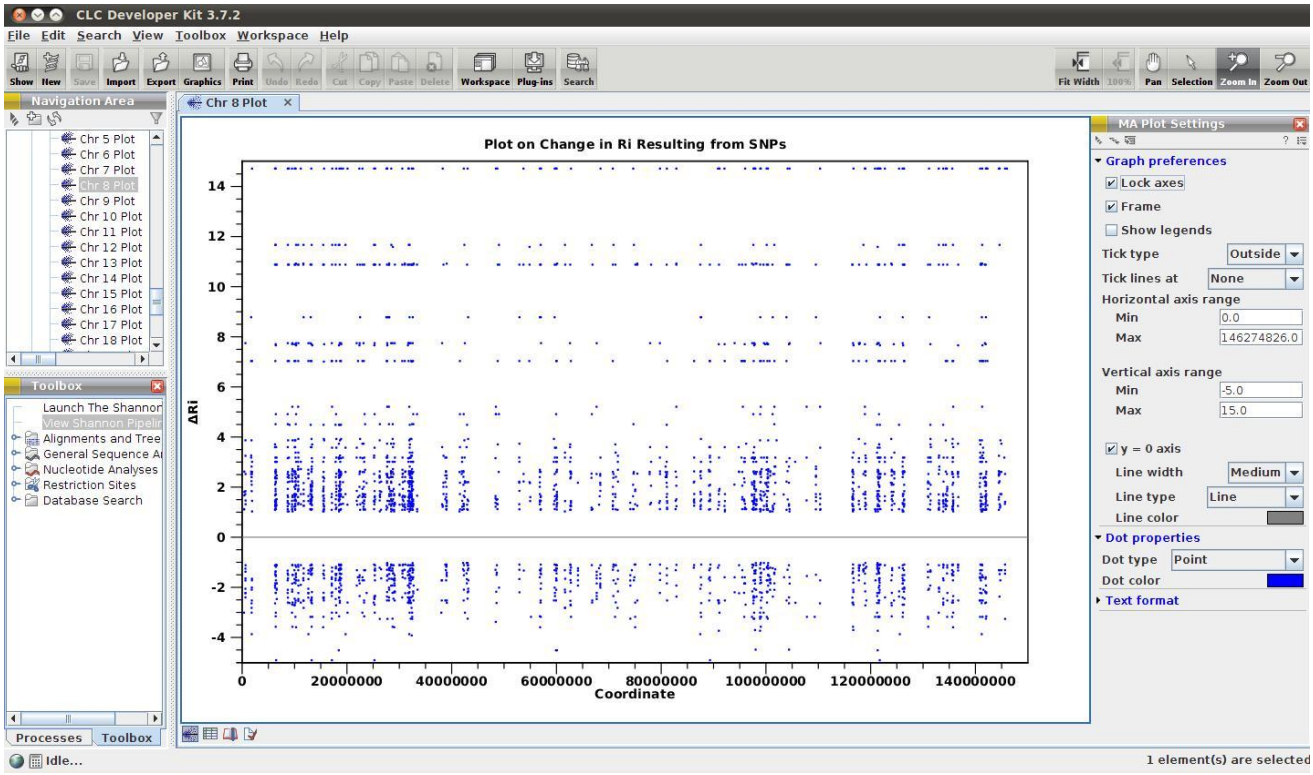


Figure 1. An example of a plot generated by the pipeline.

The plots provide a visual representation of the delta R_i for each variant. The genome wide, Manhattan style plot shows variants across the whole genome. Plots for individual chromosomes provide a closer look at the effect of the variants on a single chromosome.

Hovering the mouse over a plot point will produce a tool-tip containing the following information about the variant:

Chromosome
Coordinate
Delta R_i (change in R_i before and after variant is introduced)
Final R_i (R_i after variant is introduced)
rsID from dbSNP130/135 if available

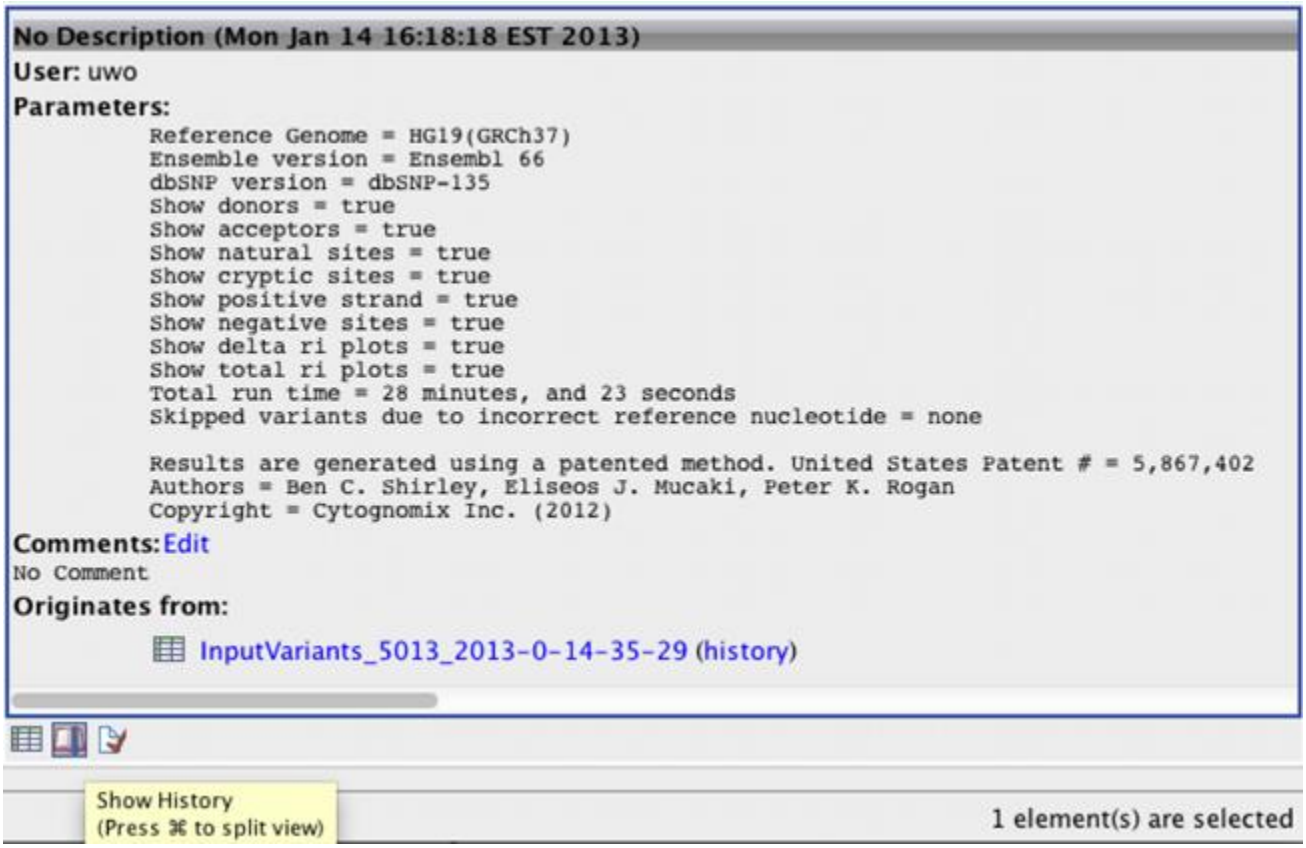


Figure 2. An example of plot history generated by the pipeline. While the plot editor is open (such as in figure 1), select history at the bottom of the screen. Genome version, Ensembl version, filer options, run time, skipped variants, and the input file for the appropriate pipeline execution can be found here.

Tracks

Four BEDGRAPH tracks are generated each time the pipeline is executed, these are:

1. customtrack-positive-acceptor-deltaRi (acceptor sites on the positive strand)
2. customtrack-negative-acceptor-deltaRi (acceptor sites on the negative strand)
3. customtrack-positive-donor-deltaRi (donor sites on the positive strand)
4. customtrack-negative-donor-deltaRi (donor sites on the negative strand)

Each row of a track represents a single variant. Each track has a header automatically included and are ready to be viewed using a genome browser. The header hides other tracks and displays ensGene (Ensembl Gene Predictions).

Example row:
chr1 8863452 8863452 14.7147693634033
On chromosome 1, coordinate 8863452 the predicted result of the input variant is an R_i increase of 14.71 bits. When viewed in a genome browser, a vertical line depicts the change in R_i.

How to view tracks

Tracks can be viewed in either the CLC bio workbench or other online browsers. To view in an online browser follow step 1 below to create a BedGraph file(s). To view in the workbench, follow all steps below.

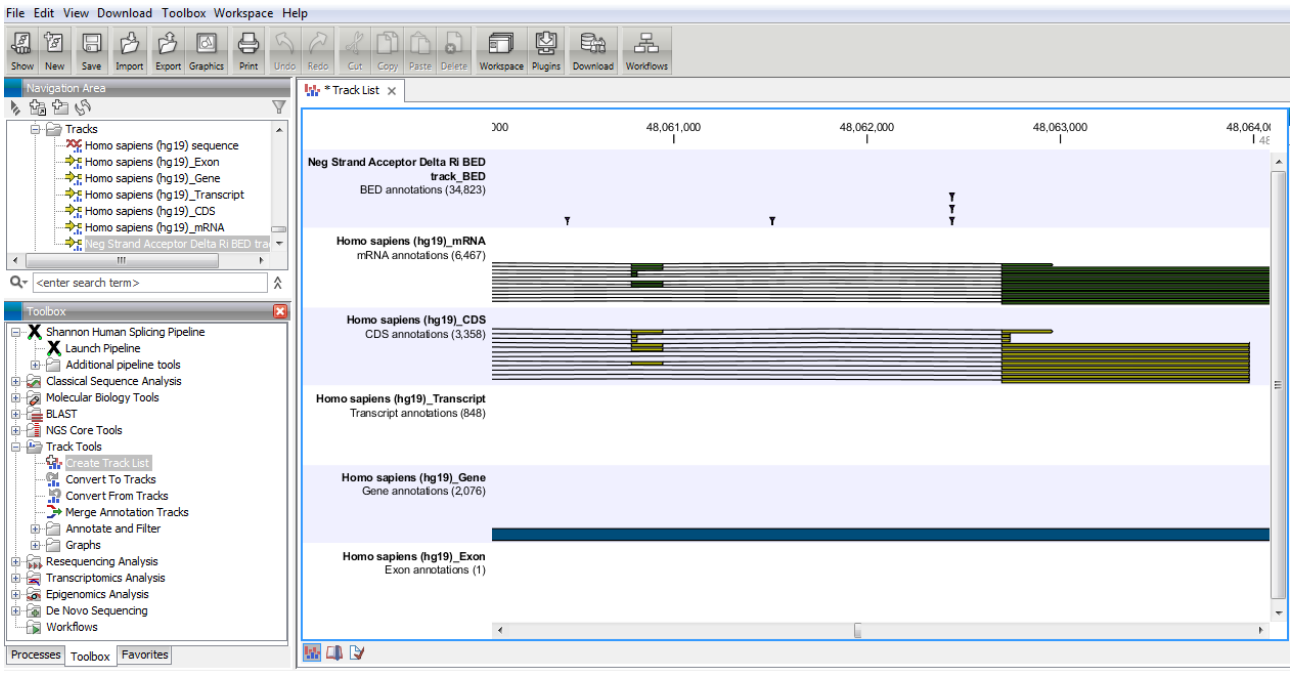
1. Export track
   - Right click on a track (from Shannon pipeline results) you would like to export and click *Export*.
   - A list of export formats will appear, select *Text* and click *Open*.
   - Select the location of your track files (Workbench or CLC Server) and click *Next*.
   - Place the tracks you would like to export in the box on the right side of the window. Multiple tracks can be exported at once. Click *Next*.
   - Select a desired file name and output location. It is helpful at this time to change the extension of the exported tracks to '.bed' to alleviate the need to manually change them later when reimporting (the workbench requires file extensions to match the type of file being imported) and click *Finish*.
   - The BedGraph files are now ready to be uploaded to an online genome browser. If that is your intent, the following steps can be ignored.

2. Establishing a CLC reference genome track
   - The workbench needs a reference track of the human genome in order to import track data. If you already have a human genome reference track, go straight to step 3.
   - Within the workbench, click *Download* from the top toolbar and select *Download genome*.
   - Choose where you would like the reference track to be located (Workbench or CLC Server) and click *Next*.
   - Select Homo sapiens (hg19) or Homo sapiens (hg18) depending on which genome build your data corresponds to and click *Next*.

- You are given the choice of using your own human reference or downloading one. If you have the appropriate genome on your computer you can select *Use existing genome sequence track*. Otherwise, select *Download genome sequence* and click *Next*.
- You may also want to download annotation tracks to view alongside your data. We would recommend downloading at least the Gene annotation track at this step. To do so, check the box beside the desired tracks and click *Next*.
- Downloaded tracks should be saved and not opened, so select *Save* and click *Next*.
- Select a location to save the tracks and click *Finish*.

3. Import tracks in CLC format
   - Within the CLC workbench, click *Import* in the top toolbar and select *Tracks*. Select where you would like to import the tracks (Workbench or CLC Server). Click *next*.
   - Under *Type of files to import* select BED and browse your computer to find the file(s) exported in step 1 (Note that if the files have the file extension '.txt' they must be changed to '.bed' before they can be imported using this method).
   - A reference track needs to be defined. Click the browse icon next to *Reference track* and browse to find the human genome sequence downloaded in step 2. Click *Next*.
   - Select *Save* and click *Next*. Choose the location to save your imported tracks and click *Finish*.

4. View tracks
   - In the Genomics workbench toolbox (bottom left of screen), expand *Track Tools* and double click *Create Track List*. Place all tracks you would like to view together in the box on the right side of the screen and click *Finish*. This will display your results in the same manner as the screenshot below.



Filtering Tips

A method to further reduce the number of variants found in Shannon pipeline results is described here. For additional specifics and explanations of the filters please refer to the following paper: Shirley BC, Mucaki EJ, Whitehead T, Costea PI, Akan P, Rogan PK. Interpretation, stratification and evidence for sequence variants affecting mRNA splicing in complete human genome sequences. Genomics Proteomics Bioinformatics. 2013 Apr;11(2):77-85.

In general, further filtering is not required for inactivating or leaky variants. All cryptic site variants are reported, however some of these variants are less likely to alter splicing. The following steps describe a method to filter out those variants and keep only variants most likely to functionally relevant. To perform the following filtering steps, it is expected that you have the 'Cryptic Variant Information' table open in the CLC-Bio Workbench. Filtering options are located near the top-right of the table. In particular, click the arrow labelled 'Advanced filter' on mouseover to access the filtering options described below. In the 'Advanced filter' display, you will see two dropdown boxes. The first box represents the columns in the table, the second allows several filtering options to be applied. Very similar filters can also be applied outside of the workbench in spreadsheet software.

1. Eliminate cryptic sites which experience a decrease in R_i
delta Ri > 0 (note: in the workbench, 'delta' will appear as a triangle)
Select delta R_i from the column dropdown box in 'Advanced filter' options. In the second box, we are interested in thos variants which results in an increased in R_i, so we select '>'. Finally, in the text field we enter 0 and click the apply button to execute the filter.
It is generally assumed that natural sites are used unless the natural site is weakened or a nearby cryptic site is strengthened. This filter removes those variants which contribute to a decrease in R_i of a cryptic site as it is unlikely this will contribute to deleterious splicing.

2. Eliminate cryptic sites weaker than a nearby natural site
Cryptic Ri relative to nat. = GREATER
To create a new filter while preserving previous filters, click the green plus sign button adjacent to the current filter.
As a result of similar reasoning to filter 1, we are generally only interested in those cryptic sites predicted to be stronger than a nearby natural site.

3. Eliminate cryptic sites too far away from an exon
Filters 1 and 2 will greatly reduce the number of rows in the table. If additional filtering is required, cryptic sites may be eliminated which are not sufficiently close to an exon. If a cryptic site is more than 300bp away from an exon, it is quite unlikely the cryptic site will be used.

Two filters to be applied separately are required for this step. First, the following filter can be applied to eliminate variants not within 300bp of an exon:
Loc. Rel. to exon contains FLANKING
This filters works because a variant will only be annotated as 3'-FLANKING or 5'-FLANKING if it is intronic and within 300bp of an exon. After all the filters up to this point have been applied, intronic cryptic sites most likely to be functionally relevant will be displayed. Exonic cryptic sites must also be taken into account however. Exonic cryptic sites are not annotated as 3'-

FLANKING or 5'-FLANKING because they are within the exon. To display exonic sites, the 'FLANKING' filter should be removed and the following filter can be used:
Location Type = EXONIC

Others filters which may be of interest
rsID if available doesn't contain rs
This filter will display variants not present in dbSNP135.

Acceptable cryptic site distance from the nearest natural site can be narrowed if desired.
Dist. from nearest nat. site abs value < 100
Note that 100 can be changed to any value under 300.

FAQ

This FAQ will be updated with answers to common questions.

Q: How fast is the Shannon Human Splicing Pipeline?
A: In our testing, 5,000 variants took ~30sec to analyze. Increasing the number of variants leads to an approximately linear increase in computation time (5,000,000 variants in ~20min on an I7 CPU with 16Gb RAM).

Q: When I attempt to run the pipeline, I can't see any data which can be used as input?
A: Your data must be imported before it can be analyzed by the Shannon pipeline. Please see the Quick Start section of this documentation.

Q: How can I export Shannon pipeline plots?
A: Plots cannot be exported using the standard export function. Instead, open the plot you wish to export and click on *Graphics* in the top workbench toolbar. Select *Export visible area* and follow the rest of the wizard to name the exported file and choose a file extension.

Q: How does the Shannon pipeline annotate the gene name field for a variant location overlapped by more than one gene?
A: The gene name field will contain multiple genes separated by commas.

Q: Does the Shannon pipeline handle indels?
A: Yes, but please keep in mind that the indel analysis portion of the plugin is currently in the beta stages of development.

Pathway Analysis

Overview

This is an add-on module for the Shannon Pipeline. It analyzes results from the Shannon Pipeline's output, specifically the gene annotations, and performs a pathway over-representation analysis.

The over-representation analysis performs a one-tailed Fisher's Exact Test on unique gene hits from a Shannon Pipeline output table. It will only include those genes that are within the Ensembl gene database, so as to exclude mRNA transcripts from the analysis.

The user is required to input two parameters in order for the analysis to be run. The user must first provide the number of unique genes present within the global genome the test-data was derived from. The default value is 20,750, the current number of coding genes estimated to be within the human genome by the Ensembl Genome Annotation System(1). Cancer genome gene numbers may differ significantly from this value, as such users who are able to estimate their genomes global gene count differs from the standard human genome should change this value. Changing this value will result in a change of the significance of results.

The second parameter the user is required to provide is the significance level they wish to filter results at. It is recommended to be kept at 0.05, and not to be placed higher, as this could cause a great number of results to be produced which will be of less value to the user.

Module Input, and File Import

This module takes two different formats of input, either a Shannon Pipeline output table, or a text-file containing a list of genes. To use the Shannon Pipeline output table, upon selecting the Pathway Analysis option, just bring the table in question over from the left hand side of available inputs, to the right hand side indicating that you would like to run Pathway Analysis on this table.

In order to run the Pathway Analysis upon a text-file, you will need to import the file in question. First it is essential that the text-file is the correct format. It should contain a list of the genes you want to analyze, each on a separate line. Incorrectly named genes, or genes not separated by a new-line will not be examined by the pipeline.

To import the properly formatted file, click on *Import* in the top toolbar, or from the *File* menu, select *Import*. Select *Workbench* from the wizard that pops up, click on *Next*, then select the file in question from the file browser. You want the import type to be Automatic. Click through now until you can select *Finish*. Upon clicking *Finish* the wizard should close, and your file should be listed in your CLC Data folder.

In-Depth Guide to Running

Select *Pathway Analysis* from the available Shannon Pipeline modules, this should open a CLC bio Wizard. Ensure that Workbench is selected, then click *Next*. The next screen is where you will input the data you wish to analyze. The only inputs that the module will take are Shannon Pipeline output tables, and a new-line delimited list of genes, imported from a text-file.

NOTE: If you choose to analyze anything with an imported text-file, the Pipeline will ignore any filter options you input. Although the module will currently allow you to set filter options, if a text-file was selected the program will ignore all filter options set. If the user chooses one or more tables, all filter options will be preserved.

After bringing all inputs you want from the left column to the right column, select *Next*. The following page allows you to set certain parameters for the run. The three necessary parameters are the significance of the results, the number of genes you believe to be within your genome, and whether you would like to look at Natural sites, Cryptic sites, or both. If you select Natural sites or Cryptic sites, the program will ignore all inputs that don't match these criteria. This can concentrate the pool of results you are looking for. E.g., if you have a results table with 500 unique genes, but only 100 of those belong to Natural sites, if you select Natural sites, the program will treat your input as only those 100 unique genes, and not the complete set of 500.

The next set of input options ask you whether you would like to filter results by Ri values. This will allow the user to only look for those genes that are over-represented that match a certain initial and final Ri criteria. If the results match the criteria, they will be eligible for over-representation. Note: Unlike the filtering options above, even if a certain gene's variant's do not match the Ri threshold information, this will not exclude them from being included in the analysis, as they still contribute to genes that are present, and affect the over-representation analysis.

After setting all the necessary options, click *Next*. If you want the table to just be opened in the Workbench, and not save the results for later examination, select *Open*, followed by *Finish*. If however you wish to save the results, select *Save*, followed by *Next*. The final page will ask you where you wish the results to be saved, through a File explorer window. Once you have correctly entered the location, click *Finish*, and the module will begin analysis.

Exporting Results

Upon completing the Pathway Analysis, you will be presented with a table in the Workbench, containing all your results. The results are organized in the table by the Pathway in question, followed by the significance of its over-expression, followed next by the number of genes it found within this pathway, the number of genes in the pathway total, and a comma delimited list of the genes found.

Users may find that the table does not adequately allow them to see all results, or may wish to export the data in order to manipulate it further/extract specific results. To accomplish this simply click on the *Export* button in the top toolbar, or select *Export* from the *File* menu. In the window that pops-up select the name you wish to call the file, the location you would like the file to be exported to, and the format of the outputted file (it should be .txt by default). After inputting this information, select *Save*, this will begin the Export process.

Pathway Analysis FAQ

- **Q: What is the run time for this analysis?**

  A: We have measured the analysis for 1,600,000 variants at 53 seconds. This number should increase in a roughly linear manner with the number of variants produced in a table.

- **Q: What are the different input types for the module?**

  A: The module will run on either a Shannon Pipeline output table, or an imported text-file. The text-file that is to be imported must be formatted as a list of genes, with each gene name on a separate line. The gene names do not have to be unique, as the program will handle this.

- **Q: How do I import a text-file?**

  A: Select *Import* from the top toolbar, or from the File menu. Ensure *Workbench* is selected, then click *Next*. On the next window, navigate to the file you wish to import, and ensure that the import type is selected as *Automatic*. Click through now until you hit *Finish*.

- **Q: How do I export my results?**

  A: Select *Export* from the top toolbar, or from the File menu. In the window that appears, enter the name you wish the exported file to be call, the location you would like it to go, and the file type you would like to export as. It will be .txt by default. When this is completed, click *Save* and the file will begin exporting.

Preset Filtering

Overview

This is an add-on module for the Shannon Pipeline. It examines Shannon Pipeline output and filters it based on a preset standard filter, or by selecting individual options.

Selecting objects to be filtered

This module filters Shannon Pipeline output. Appropriate objects will be named "Complete Variant Information" if not renamed after pipeline execution. Double click on "Filter Pipeline Output" in the "Additional Pipeline Tools" subfolder within the "Shannon Human Splicing Pipeline" folder. After selecting the location in which the filtering will be executed, you will be asked to select a Shannon Pipeline results object. Multiple objects can be filtered simultaneously by placing multiple Shannon Pipeline results objects into the "Selected elements" window on the right side of the wizard.

Explanation of Filtering Options

Note: These filters define those variants to be preserved (not eliminated) in the resulting table. For example, filtering for all variants with initial R_i above 1.6 will result in a filtered table containing variants with initial R_i greater than 1.6.

A user can either filter output with their own criteria or using preset filters that have been created by Cytognomix. Combining the following filtering criteria in various ways will reduce the number of variants reported.

- Preset filters: A preset filter can be selected using the dropdown menu. When a filter is selected, all options below are automatically modified according to the selected preset. If desired, filters set automatically by selecting a preset filter can be manually modified. To reset all filters to default, select the preset "None".
- Filters based on Ri: Filters related to initial, final, and change in Ri values can be set here. For each of these numeric filters, you may filter for those values higher/lower than a specified value by selecting the appropriate option from the dropdown menu. If the option "Any" is selected, any value placed in the adjacent text box will be ignored. Otherwise, enter the desired value in the adjacent text box.
- Natural site change in R_i: You may filter for natural sites increasing in R_i or decreasing in R_i
- Novel and known variants: A known variant contains an rsID in the rsID column of Shannon Pipeline output. Filtering by novel variants will eliminate any variant with an rsID. Filtering by known variants will eliminate all variants without one.
- Average heterozygosity: Similar to the filters based on R_i, filters can be defined for average heterozygosity above or below a specified value.
- Strand: Filter by positive or negative strand
- Donors/Acceptors: Filter by splice site donors or acceptors
- Intronic cryptic site distance from nearest natural site: Generally, variants most likely to be potentially deleterious are found nearby a natural site. Use this filter to define how far away an intronic cryptic site can be from a natural site.

- Cryptic site strength relative to nearest natural site: Generally, only those cryptic sites with R_i greater than a nearby natural site have the potential to be deleterious. Selecting "Cryptic site R_i greater than nearest natural site R_i" will preserve these variants.
- Cryptic sites within introns and exons: This option will filter for cryptic sites within exons or introns only.

Discussion of the Standard Preset Filter

This preset emulates the filtering methods used in the following paper: Shirley BC, Mucaki EJ, Whitehead T, Costea PI, Akan P, Rogan PK. Interpretation, stratification and evidence for sequence variants affecting mRNA splicing in complete human genome sequences. Genomics Proteomics Bioinformatics. 2013 Apr;11(2):77-85. Further explanations for each filter can be found there.

These filters are designed to keep those variants most likely to be potentially deleterious. When considering natural sites, we are only interested in those decreasing in R_i since (in general) a natural site increasing in R_i will only widen the gap in R_i between itself and nearby potential cryptic sites. Similarly, we are only interested in cryptic sites with an R_i greater than a natural site within 300bp. We allow variants with an rsID, but only those with an average heterozygosity less than 5%. This eliminates common rsIDs which are therefore unlikely to be deleterious.

Exporting Results

You may wish to export your filtered tables to examine using external software. To do so, select *Export* from the top toolbar in the workbench or from the File menu. In the window that appears enter a desired file name, save location, and file type. Click *Save* and file export will begin.

References

These are references for information theory based splice site analysis.

1. Abstracts of the the American Society for Neurochemistry 37th annual meeting (Portland, Oregon USA), 11-15 march 2006. *J Neurochem*. 2006;96 Suppl 1:1-150.

2. - *Annals of Neurology*. (- 5):- 625. doi: - 10.1002/ana.21027.

3. Aggarwal S, Jinda W, Limwongse C, Atchaneeyasakul LO, Phadke SR. Run-on mutation in the PAX6 gene and chorioretinal degeneration in autosomal dominant aniridia. *Mol Vis*. 2011;17:1305-1309.

4. Aissat A, Becdelièvre A, Golmard L, et al. Combined Computational–Experimental analyses of CFTR exon strength uncover predictability of Exon‐Skipping level. *Hum Mutat*. 2013.

5. Akiyama M, Titeux M, Sakai K, et al. DNA-based prenatal diagnosis of harlequin ichthyosis and characterization of ABCA12 mutation consequences. *J Invest Dermatol*. 2007;127(3):568-573. doi: 10.1038/sj.jid.5700617.

6. Alcantara-Ortigoza MA, Belmont-Martinez L, Vela-Amieva M, Gonzalez-Del Angel A. Analysis of the CTNS gene in nephropathic cystinosis mexican patients: Report of four novel mutations and identification of a false positive 57-kb deletion genotype with LDM-2/exon 4 multiplex PCR assay. *Genet Test*. 2008;12(3):409-414. doi: 10.1089/gte.2008.0014.

7. Allikmets R, Wasserman WW, Hutchinson A, et al. Organization of the ABCR gene: Analysis of promoter and splice junction sequences. *Gene*. 1998;215(1):111-122.

8. Anczukow O, Buisson M, Salles MJ, et al. Unclassified variants identified in BRCA1 exon 11: Consequences on splicing. *Genes Chromosomes Cancer*. 2008;47(5):418-426. doi: 10.1002/gcc.20546.

9. Anglani F, Fabris A, Torregrossa R, Cristofaro R, Gambaro C, D'Angelo A. Varianti rare del genoma e suscettibilità alle malattie multifatto-riali. L'Esempio del rene con midollare a spugna. *G Ital Nefrol*. 2013;28(3):246.

10. Anglani F, Fabris A, Torregrossa R, Cristofaro R, Gambaro G, D'Angelo A. Rare genomic variants and susceptibility to multifactorial diseases. the example of medullary sponge kidney. *G Ital Nefrol*. 2011;28(3):246-248.

11. Aoyama Y, Ozer I, Demirkol M, et al. Molecular features of 23 patients with glycogen storage disease type III in turkey: A novel mutation p.R1147G associated with isolated glucosidase deficiency, along with 9 AGL mutations. *J Hum Genet*. 2009;54(11):681-686. doi: 10.1038/jhg.2009.100; 10.1038/jhg.2009.100.

12. Arita K, Wessagowit V, Inamadar AC, et al. Unusual molecular findings in kindler syndrome. *Br J Dermatol*. 2007;157(6):1252-1256. doi: 10.1111/j.1365-2133.2007.08159.x.

13. Arnould I, Schriml LM, Prades C, et al. Identifying and characterizing a five‐gene cluster of ATP‐binding cassette transporters mapping to human chromosome 17q24: A new subgroup within the ABCA subfamily. *GeneScreen*. 2001;1(3):157-164.

14. Atwood CS, inventorMethods of Assessing Risk of Alzheimer's Disease in a Patient. 2009, .

15. Bacci C, Sestini R, Provenzano A, et al. Schwannomatosis associated with multiple meningiomas due to a familial SMARCB1 mutation. *Neurogenetics*. 2010;11(1):73-80. doi: 10.1007/s10048-009-0204-2.

16. Baralle M, Baralle D. Splicing mechanisms and mutations in the NF1 gene. In: *Neurofibromatosis type 1.* Springer; 2012:135-150.

17. Baralle D, Lucassen A, Buratti E. Missed threads. the impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep*. 2009;10(8):810-816. doi: 10.1038/embor.2009.170.

18. Baralle MA, Baralle D. From bedside to bench: How to analyze a splicing mutation. *Alternative pre-mRNA Splicing: Theory and Protocols*. 2012;DOI: 10.1002/9783527636778.ch11.

19. Beetz C, Schule R, Deconinck T, et al. REEP1 mutation spectrum and genotype/phenotype correlation in hereditary spastic paraplegia type 31. *Brain*. 2008;131(Pt 4):1078-1086. doi: 10.1093/brain/awn026.

20. Ben Selma Z, Yilmaz S, Schischmanoff PO, et al. A novel S115G mutation of CGI-58 in a turkish patient with dorfman-chanarin syndrome. *J Invest Dermatol*. 2007;127(9):2273-2276. doi: 10.1038/sj.jid.5700860.

21. Ben-Salem S, Begum MA, Ali BR, Al-Gazali L. A novel aberrant splice site mutation in RAB23 leads to an eight nucleotide deletion in the mRNA and is responsible for carpenter syndrome in a consanguineous emirati family. *Mol Syndromol*. 2013;3(6):255-261. doi: 10.1159/000345653; 10.1159/000345653.

22. Bertola F, Filocamo M, Casati G, et al. IDUA mutational profiling of a cohort of 102 european patients with mucopolysaccharidosis type I: Identification and characterization of 35 novel alpha-L-iduronidase (IDUA) alleles. *Hum Mutat*. 2011;32(6):E2189-210. doi: 10.1002/humu.21479; 10.1002/humu.21479.

23. Bertolini S, Pisciotta L, Rabacchi C, et al. Spectrum of mutations and phenotypic expression in patients with autosomal dominant hypercholesterolemia identified in italy. *Atherosclerosis*. 2013;227(2):342-348. doi: 10.1016/j.atherosclerosis.2013.01.007; 10.1016/j.atherosclerosis.2013.01.007.

24. Bi C, Rogan PK. Information theory as a model of genomic sequences. *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. 2005.

25. Bi C, Rogan P. *Determining thresholds for binding site sequence models using information theory.* ; 2005:1290.

26. Bloethner S, Mould A, Stark M, Hayward NK. Identification of ARHGEF17, DENND2D, FGFR3, and RB1 mutations in melanoma by inhibition of nonsense-mediated mRNA decay. *Genes Chromosomes Cancer*. 2008;47(12):1076-1085. doi: 10.1002/gcc.20598.

27. Bocchi L, Pisciotta L, Fasano T, et al. Multiple abnormally spliced ABCA1 mRNAs caused by a novel splice site mutation of ABCA1 gene in a patient with tangier disease. *Clin Chim Acta*. 2010;411(7-8):524-530. doi: 10.1016/j.cca.2010.01.008.

28. Bogaerts V, Nuytemans K, Reumers J, et al. Genetic variability in the mitochondrial serine protease HTRA2 contributes to risk for parkinson disease. *Hum Mutat*. 2008;29(6):832-840. doi: 10.1002/humu.20713.

29. Bonafe L, Giunta C, Gassner M, Steinmann B, Superti‐Furga A. A cluster of autosomal recessive spondylocostal dysostosis caused by three newly identified DLL3 mutations segregating in a small village. *Clin Genet*. 2003;64(1):28-35.

30. Bonnart C. *Etude fonctionnelle de LEKTI et de sa nouvelle cible, l'élastase 2 pancréatique*. 2007.

31. Bonnet-Dupeyron MN, Combes P, Santander P, Cailloux F, Boespflug-Tanguy O, Vaurs-Barriere C. PLP1 splicing abnormalities identified in pelizaeus-merzbacher disease and SPG2 fibroblasts are associated with different types of mutations. *Hum Mutat*. 2008;29(8):1028-1036. doi: 10.1002/humu.20758.

32. Borroni B, Archetti S, Alberici A, et al. Progranulin genetic variations in frontotemporal lobar degeneration: Evidence for low mutation frequency in an italian clinical series. *Neurogenetics*. 2008;9(3):197-205. doi: 10.1007/s10048-008-0127-3.

33. Botta E, Nardo T, Orioli D, et al. Genotype-phenotype relationships in trichothiodystrophy patients with novel splicing mutations in the XPD gene. *Hum Mutat*. 2009;30(3):438-445. doi: 10.1002/humu.20912.

34. Brockmoller J, Tzvetkov MV. Pharmacogenetics: Data, concepts and tools to improve drug discovery and drug treatment. *Eur J Clin Pharmacol*. 2008;64(2):133-157. doi: 10.1007/s00228-007-0424-z.

35. Broer S, Bailey CG, Kowalczuk S, et al. Iminoglycinuria and hyperglycinuria are discrete human phenotypes resulting from complex mutations in proline and glycine transporters. *J Clin Invest*. 2008;118(12):3881-3892. doi: 10.1172/JCI36625; 10.1172/JCI36625.

36. Buratti E, Baralle M, Baralle FE. Defective splicing, disease and therapy: Searching for master checkpoints in exon definition. *Nucleic Acids Res*. 2006;34(12):3494-3510. doi: 10.1093/nar/gkl498.

37. Cabral RM, Liu L, Hogan C, et al. Homozygous mutations in the 5' region of the JUP gene result in cutaneous disease but normal heart development in children. *J Invest Dermatol*. 2010;130(6):1543-1550. doi: 10.1038/jid.2010.7.

38. Calandra S, Tarugi P, Bertolini S. Altered mRNA splicing in lipoprotein disorders. *Curr Opin Lipidol*. 2011;22(2):93-99. doi: 10.1097/MOL.0b013e3283426ebc.

39. Caridi G, Dagnino M, Dalgic B, et al. Analbuminemia zonguldak: Case report and mutational analysis. *Clin Biochem*. 2008;41(4-5):288-291. doi: 10.1016/j.clinbiochem.2007.11.016.

40. Cartault F, Nava C, Malbrunot A, et al. A new< i> XPC</i> gene splicing mutation has lead to the highest worldwide prevalence of xeroderma pigmentosum in black mahori patients. *DNA repair*. 2011;10(6):577-585.

41. Castaman G, Giacomelli SH, Mancuso ME, et al. Deep intronic variations may cause mild hemophilia A. *J Thromb Haemost*. 2011;9(8):1541-1548. doi: 10.1111/j.1538-7836.2011.04408.x; 10.1111/j.1538-7836.2011.04408.x.

42. Castiglia D, Zambruno G. Mutation mechanisms. *Dermatol Clin*. 2010;28(1):17-22. doi: 10.1016/j.det.2009.10.002.

43. Caudevilla C, Da Silva-Azevedo L, Berg B, Guhl E, Graessmann M, Graessmann A. Heterologous HIV-nef mRNA< i> trans</i>-splicing: A new principle how mammalian cells generate hybrid mRNA and protein molecules. *FEBS Lett*. 2001;507(3):269-279.

44. Caux-Moncoutier V, Pages-Berhouet S, Michaux D, et al. Impact of BRCA1 and BRCA2 variants on splicing: Clues from an allelic imbalance study. *Eur J Hum Genet*. 2009;17(11):1471-1480. doi: 10.1038/ejhg.2009.89.

45. Cefalu AB, Noto D, Magnolo L, et al. Novel mutations of CETP gene in italian subjects with hyperalphalipoproteinemia. *Atherosclerosis*. 2009;204(1):202-207. doi: 10.1016/j.atherosclerosis.2008.08.031.

46. Chen L, Qin S, Xie J, et al. Genetic polymorphism analysis of CYP2C19 in chinese han populations from different geographic areas of mainland china. *Pharmacogenomics*. 2008;9(6):691-702. doi: 10.2217/14622416.9.6.691.

47. Chen LJ, Tam PO, Tham CC, et al. Evaluation of SPARC as a candidate gene of juvenile-onset primary open-angle glaucoma by mutation and copy number analyses. *Mol Vis*. 2010;16:2016-2025.

48. Cho SY, Ki CS, Park HD, et al. Genetic investigation of patients with undetectable peaks of growth hormone after two provocation tests. *Clin Endocrinol (Oxf)*. 2013;78(2):317-320. doi: 10.1111/j.1365-2265.2012.04514.x; 10.1111/j.1365-2265.2012.04514.x.

49. Clark GR, Crowe P, Muszynska D, et al. Development of a diagnostic genetic test for simplex and autosomal recessive retinitis pigmentosa. *Ophthalmology*. 2010;117(11):2169-77.e3. doi: 10.1016/j.ophtha.2010.02.029.

50. Cleaver J, Collins C, Ellis J, Volik S. Genome sequence and splice site analysis of low-fidelity DNA polymerases H and I involved in replication of damaged DNA. *Genomics*. 2003;82(5):561-570.

51. Cohen B, Chervinsky E, Jabaly-Habib H, Shalev SA, Briscoe D, Ben-Yosef T. A novel splice site mutation of CDHR1 in a consanguineous israeli christian arab family segregating autosomal recessive cone-rod dystrophy. *Mol Vis*. 2012;18:2915-2921.

52. Colombo M, De Vecchi G, Caleca L, et al. Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. *PLoS One*. 2013;8(2):e57173. doi: 10.1371/journal.pone.0057173; 10.1371/journal.pone.0057173.

53. Comin M, Antonello M. Fast computation of entropic profiles for the detection of conservation in genomes. In: *Pattern recognition in bioinformatics.* Springer; 2013:277-288.

54. Concolino P, Vendittelli F, Mello E, et al. Functional analysis of two rare CYP21A2 mutations detected in italian patients with a mildest form of congenital adrenal hyperplasia. *Clin Endocrinol (Oxf)*. 2009;71(4):470-476. doi: 10.1111/j.1365-2265.2008.03517.x.

55. Covaciu C, Grosso F, Pisaneschi E, et al. A founder synonymous COL7A1 mutation in three danish families with dominant dystrophic epidermolysis bullosa pruriginosa identifies exonic regulatory sequences required for exon 87 splicing. *Br J Dermatol*. 2011;165(3):678-682. doi: 10.1111/j.1365-2133.2011.10414.x; 10.1111/j.1365-2133.2011.10414.x.

56. Cox DG, Crusius JB, Peeters PH, Bueno-de-Mesquita HB, Pena AS, Canzian F. Haplotype of prostaglandin synthase 2/cyclooxygenase 2 is involved in the susceptibility to inflammatory bowel disease. *World J Gastroenterol*. 2005;11(38):6003-6008.

57. Cruchaga C, Fernandez-Seara MA, Seijo-Martinez M, et al. Cortical atrophy and language network reorganization associated with a novel progranulin mutation. *Cereb Cortex*. 2009;19(8):1751-1760. doi: 10.1093/cercor/bhn202.

58. Dash DP, George S, O'Prey D, et al. Mutational screening of VSX1 in keratoconus patients from the european population. *Eye (Lond)*. 2010;24(6):1085-1092. doi: 10.1038/eye.2009.217.

59. Day INM. IDDM2 locus: 5' noncoding intron I splicing and translational efficiency effects of INS -23HphI - more than a tag for the INS promoter VNTR. abstracts/hgvs.org/Helsenki/Presentations/Day.ppt. Updated 2006.

60. Deen PM, Dahl N, Caplan MJ. The aquaporin-2 water channel in autosomal dominant primary nocturnal enuresis. *J Urol*. 2002;167(3):1447-1450.

61. Denecke J, Kranz C, Kemming D, Koch H, Marquardt T. An activated 5′ cryptic splice site in the human ALG3 gene generates a premature termination codon insensitive to nonsense‐mediated mRNA decay in a new case of congenital disorder of glycosylation type id (CDG‐Id). *Hum Mutat*. 2004;23(5):477-486.

62. Denson J, Xi Z, Wu Y, Yang W, Neale G, Zhang J. Screening for inter-individual splicing differences in human GSTM4 and the discovery of a single nucleotide substitution related to the tandem skipping of two exons. *Gene*. 2006;379:148-155.

63. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human splicing finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67. doi: 10.1093/nar/gkp215.

64. Di Leo E, Magnolo L, Lancellotti S, et al. Abnormal apolipoprotein B pre-mRNA splicing in patients with familial hypobetalipoproteinaemia. *J Med Genet*. 2007;44(3):219-224. doi: 10.1136/jmg.2006.046359.

65. Di Leo E, Magnolo L, Pinotti E, et al. Functional analysis of two novel splice site mutations of APOB gene in familial hypobetalipoproteinemia. *Mol Genet Metab*. 2009;96(2):66-72. doi: 10.1016/j.ymgme.2008.10.016; 10.1016/j.ymgme.2008.10.016.

66. Di Leo E, Panico F, Tarugi P, Battisti C, Federico A, Calandra S. A point mutation in the lariat branch point of intron 6 of NPC1 as the cause of abnormal pre-mRNA splicing in niemann-pick type C disease. *Hum Mutat*. 2004;24(5):440. doi: 10.1002/humu.9287.

67. Dinakarpandian D, Raheja V, Mehta S, Schuetz E, Rogan P. Tandem machine learning for the identification of genes regulated by transcription factors. *BMC Bioinformatics*. 2005;6:204-204. doi: 10.1186/1471-2105-6-204.

68. Douglas DA, Zhong H, Ro JY, et al. Novel mutations of epidermal growth factor receptor in localized prostate cancer. *Front Biosci*. 2006;11:2518-2525.

69. Doza JP, Topf A, Bentham J, et al. Low-frequency intermediate penetrance variants in the ROCK1 gene predispose to tetralogy of fallot. *BMC genetics*. 2013;14(1):57.

70. Drera B, Floriddia G, Forzano F, et al. Branch point and donor splice-site COL7A1 mutations in mild recessive dystrophic epidermolysis bullosa. *Br J Dermatol*. 2009;161(2):464-467. doi: 10.1111/j.1365-2133.2009.09114.x; 10.1111/j.1365-2133.2009.09114.x.

71. Drogemuller C, Philipp U, Haase B, Gunzel-Apel AR, Leeb T. A noncoding melanophilin gene (MLPH) SNP at the splice donor of exon 1 represents a candidate causal mutation for coat color dilution in dogs. *J Hered*. 2007;98(5):468-473. doi: 10.1093/jhered/esm021.

72. Dua-Awereh MB, Shimomura Y, Kraemer L, Wajid M, Christiano AM. Mutations in the desmoglein 1 gene in five pakistani families with striate palmoplantar keratoderma. *J Dermatol Sci*. 2009;53(3):192-197. doi: 10.1016/j.jdermsci.2008.11.005.

73. Dutrannoy V, Demuth I, Baumann U, et al. Clinical variability and novel mutations in the NHEJ1 gene in patients with a nijmegen breakage syndrome-like phenotype. *Hum Mutat*. 2010;31(9):1059-1068. doi: 10.1002/humu.21315; 10.1002/humu.21315.

74. Eckl KM, de Juanes S, Kurtenbach J, et al. Molecular analysis of 250 patients with autosomal recessive congenital ichthyosis: Evidence for mutation hotspots in ALOXE3 and allelic heterogeneity in ALOX12B. *J Invest Dermatol*. 2009;129(6):1421-1428. doi: 10.1038/jid.2008.409.

75. Eichers ER, Green JS, Stockton DW, et al. Newfoundland rod-cone dystrophy, an early-onset retinal dystrophy, is caused by splice-junction mutations in< i> RLBP1</i>. *The American Journal of Human Genetics*. 2002;70(4):955-964.

76. Ellard S, Patrinos GP, Oetting WS. Clinical applications of Next‐Generation sequencing: The 2013 human genome variation society scientific meeting. *Hum Mutat*. 2013.

77. Ellis JR,Jr, Heinrich B, Mautner VF, Kluwe L. Effects of splicing mutations on NF2-transcripts: Transcript analysis and information theoretic predictions. *Genes Chromosomes Cancer*. 2011;50(8):571-584. doi: 10.1002/gcc.20876; 10.1002/gcc.20876.

78. ElSharawy A, Hundrieser B, Brosch M, et al. Systematic evaluation of the effect of common SNPs on pre-mRNA splicing. *Hum Mutat*. 2009;30(4):625-632. doi: 10.1002/humu.20906.

79. Emmert S, Schneider TD, Khan SG, Kraemer KH. The human XPG gene: Gene architecture, alternative splicing and single nucleotide polymorphisms. *Nucleic Acids Res*. 2001;29(7):1443-1452.

80. Fahey ME, Higgins DG. Gene expression, intron density, and splice site strength in drosophila and caenorhabditis. *J Mol Evol*. 2007;65(3):349-357.

81. Fang S, Guo X, Jia X, Xiao X, Li S, Zhang Q. Novel GPR143 mutations and clinical characteristics in six chinese families with X-linked ocular albinism. *Mol Vis*. 2008;14:1974-1982.

82. Fasano T, Bocchi L, Pisciotta L, Bertolini S, Calandra S. Denaturing high-performance liquid chromatography in the detection of ABCA1 gene mutations in familial HDL deficiency. *J Lipid Res*. 2005;46(4):817-822. doi: 10.1194/jlr.D400038-JLR200.

83. Fasano T, Pisciotta L, Bocchi L, et al. Lysosomal lipase deficiency: Molecular characterization of eleven patients with wolman or cholesteryl ester storage disease. *Mol Genet Metab*. 2011. doi: 10.1016/j.ymgme.2011.12.008.

84. Fasano T, Zanoni P, Rabacchi C, et al. Novel mutations of ABCA1 transporter in patients with tangier disease and familial HDL deficiency. *Mol Genet Metab*. 2012;107(3):534-541. doi: 10.1016/j.ymgme.2012.08.005; 10.1016/j.ymgme.2012.08.005.

85. Fattal-Valevski A, DiMaio MS, Hisama FM, et al. Variable expression of a novel PLP1 mutation in members of a family with pelizaeus-merzbacher disease. *J Child Neurol*. 2009;24(5):618-624. doi: 10.1177/0883073808327833.

86. Faz DB. *Bases geneticas de la conducta / genetic bases of behavior.* Barcelona, Catalonia, Spain: Edidions de la Universitat Oberta de Catalunya; 2009.

87. Fei J, Chen S. Splice site mutation-induced alteration of selective regional activity correlates with the role of a gene in cardiomyopathy. *J Clin Exp Cardiolog S*. 2013;12:2.

88. Ferlini A, Neri M, Gualandi F. The medical genetics of dystrophinopathies: Molecular genetic diagnosis and its impact on clinical practice. *Neuromuscul Disord*. 2013;23(1):4-14. doi: 10.1016/j.nmd.2012.09.002; 10.1016/j.nmd.2012.09.002.

89. Fong K, Rama Devi AR, Lai-Cheong JE, et al. Infantile systemic hyalinosis associated with a putative splice-site mutation in the ANTXR2 gene. *Clin Exp Dermatol*. 2012;37(6):635-638. doi: 10.1111/j.1365-2230.2011.04287.x; 10.1111/j.1365-2230.2011.04287.x.

90. Fornage M, Lee CR, Doris PA, et al. The soluble epoxide hydrolase gene harbors sequence variation associated with susceptibility to and protection from incident ischemic stroke. *Hum Mol Genet*. 2005;14(19):2829-2837. doi: 10.1093/hmg/ddi315.

91. Funghini S, Thusberg J, Spada M, et al. Carbamoyl phosphate synthetase 1 deficiency in italy: Clinical and genetic findings in a heterogeneous cohort. *Gene*. 2012;493(2):228-234. doi: 10.1016/j.gene.2011.11.052; 10.1016/j.gene.2011.11.052.

92. Gaedigk A, Gaedigk R, Leeder JS. < i> CYP2D7</i> splice variants in human liver and brain: Does< i> CYP2D7</i> encode functional protein? *Biochem Biophys Res Commun*. 2005;336(4):1241-1250.

93. Gaedigk A, Baker DW, Totah RA, et al. Variability of CYP2J2 expression in human fetal tissues. *J Pharmacol Exp Ther*. 2006;319(2):523-532. doi: 10.1124/jpet.106.109215.

94. Gaedigk A, Bhathena A, Ndjountche L, et al. Identification and characterization of novel sequence variations in the cytochrome P4502D6 (CYP2D6) gene in african americans. *Pharmacogenomics J*. 2005;5(3):173-182. doi: 10.1038/sj.tpj.6500305.

95. Gaedigk A, Gaedigk R, Leeder JS. CYP2D7 splice variants in human liver and brain: Does CYP2D7 encode functional protein? *Biochem Biophys Res Commun*. 2005;336(4):1241-1250. doi: 10.1016/j.bbrc.2005.08.255.

96. Gaedigk A, Leeder JS. Letter to the editor. *Clin Pharmacol Ther*. 2006;80:558-560.

97. Gaedigk A, Ndjountche L, Leeder JS, Bradford LD. Limited association of the 2988g > a single nucleotide polymorphism with CYP2D641 in black subjects. *Clin Pharmacol Ther*. 2005;77(3):228-30; author reply 230-1. doi: 10.1016/j.clpt.2004.10.014.

98. Gallagher C. Development of an automated identification system for nanocrystal encoded microspheres in flow cytometry. . 2009.

99. Gao S, Zhang N, Zhang L, Duan GY, Zhang T. The human variome project and its progress. *Yi Chuan*. 2010;32(11):1105-1113.

100. Garcia-Blanco M. Alternative splicing: Therapeutic target and tool. In: Jeanteur P, ed. *Alternative splicing and disease.* Berlin, Germany: Springer; 2006:47-64.

101. Garcia-Gonzalez MA, Jones JG, Allen SK, et al. Evaluating the clinical utility of a molecular genetic test for polycystic kidney disease. *Mol Genet Metab*. 2007;92(1-2):160-167. doi: 10.1016/j.ymgme.2007.05.004.

102. Gaweda-Walerych K, Safranow K, Maruszak A, et al. Mitochondrial transcription factor A variants and the risk of parkinson's disease. *Neurosci Lett*. 2010;469(1):24-29. doi: 10.1016/j.neulet.2009.11.037.

103. Gerykov-Bujalkova M, Krivulcik T, Bartosova Z. Novel approaches in evaluation of pathogenicity of single-base exonic germline changes involving the mismatch repair genes MLH1 and MSH2 in diagnostics of lynch syndrome minireview. *Neoplasma*. 2008;55(6):463-471.

104. Gibbons WJ, Yan Q, Li R, Li X, Guan M. Genomic organization, expression, and subcellular localization of mouse mitochondrial seryl-tRNA synthetase. *Biochem Biophys Res Commun*. 2004;317(3):774-778.

105. Godefroid N, Riveira-Munoz E, Saint-Martin C, Nassogne MC, Dahan K, Devuyst O. A novel splicing mutation in SLC12A3 associated with gitelman syndrome and idiopathic intracranial hypertension. *Am J Kidney Dis*. 2006;48(5):e73-9. doi: 10.1053/j.ajkd.2006.08.005.

106. Gozukara EM, Khan SG, Metin A, et al. A stop codon in xeroderma pigmentosum group C families in turkey and italy: Molecular genetic evidence for a common ancestor. *J Invest Dermatol*. 2001;117(2):197-204.

107. Gruber FX, Hjorth-Hansen H, Mikkola I, Stenke L, Johansen T. A novel bcr-abl splice isoform is associated with the L248V mutation in CML patients with acquired resistance to imatinib. *Leukemia*. 2006;20(11):2057-2060. doi: 10.1038/sj.leu.2404400.

108. Hageman GS, Anderson DH, Johnson LV, et al. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular

degeneration. *Proc Natl Acad Sci U S A*. 2005;102(20):7227-7232. doi: 10.1073/pnas.0501536102.

109. Hamada T, Fukuda S, Sakaguchi S, Yasumoto S, Kim SC, Hashimoto T. Molecular and clinical characterization in japanese and korean patients with hailey-hailey disease: Six new mutations in the ATP2C1 gene. *J Dermatol Sci*. 2008;51(1):31-36. doi: 10.1016/j.jdermsci.2008.02.003.

110. Hampson G, Konrad MA, Scoble J. Familial hypomagnesaemia with hypercalciuria and nephrocalcinosis (FHHNC): Compound heterozygous mutation in the claudin 16 (CLDN16) gene. *BMC Nephrol*. 2008;9:12. doi: 10.1186/1471-2369-9-12.

111. Hartmann L, Theiss S, Niederacher D, Schaal H. Diagnostics of pathogenic splicing mutations: Does bioinformatics cover all bases? *Front Biosci*. 2008;13:3252-3272.

112. Hefferon TW, Broackes-Carter FC, Harris A, Cutting GR. Atypical 5′ splice sites cause CFTR exon 9 to be vulnerable to skipping. *The American Journal of Human Genetics*. 2002;71(2):294-303.

113. Hellerud C, Burlina A, Gabelli C, Ellis JR, Nyholm P, Lindsted S. Glycerol metabolism and the determination of triglycerides–clinical, biochemical and molecular findings in six subjects. *Clinical chemistry and laboratory medicine*. 2003;41(1):46-55.

114. Hellerud C, Adamowicz M, Jurkiewicz D, et al. Clinical heterogeneity and molecular findings in five polish patients with glycerol kinase deficiency: Investigation of two splice site mutations with computerized splice junction analysis and Xp21 gene-specific mRNA analysis. *Mol Genet Metab*. 2003;79(3):149-159.

115. Hengen PN, Lyakhov IG, Stewart LE, Schneider TD. Molecular flip-flops formed by overlapping fis sites. *Nucleic Acids Res*. 2003;31(22):6663-6673.

116. Henneman P, Schaap FG, Rensen PC, van Dijk KW, Smelt AH. Estrogen induced hypertriglyceridemia in an apolipoprotein AV deficient patient. *J Intern Med*. 2008;263(1):107-108. doi: 10.1111/j.1365-2796.2007.01889.x.

117. Henriksen AM, Tumer Z, Tommerup N, Tranebjaerg L, Larsen LA. Identification of a novel EYA1 splice-site mutation in a danish branchio-oto-renal syndrome family. *Genet Test*. 2004;8(4):404-406. doi: 10.1089/gte.2004.8.404.

118. Hertecant JL, Ben-Rebeh I, Marah MA, et al. Clinical and molecular analysis of isovaleric acidemia patients in the united arab emirates reveals remarkable phenotypes and four novel mutations in the IVD gene. *Eur J Med Genet*. 2012;55(12):671-676. doi: 10.1016/j.ejmg.2012.08.001; 10.1016/j.ejmg.2012.08.001.

119. Hiller M, Huse K, Szafranski K, et al. Phylogenetically widespread alternative splicing at unusual GYNGYN donors. *Genome Biol*. 2006;7(7):R65. doi: 10.1186/gb-2006-7-7-R65.

120. Hines RN, Koukouritaki SB, Poch MT, Stephens MC. Regulatory polymorphisms and their contribution to interindividual differences in the expression of enzymes influencing drug and toxicant disposition. *Drug Metab Rev*. 2008;40(2):263-301. doi: 10.1080/03602530801952682.

121. Hobson GM, Huang Z, Sperle K, et al. Splice-site contribution in alternative splicing of PLP1 and DM20: Molecular studies in oligodendrocytes. *Hum Mutat*. 2006;27(1):69-77. doi: 10.1002/humu.20276.

122. Houdayer C, Caux‐Moncoutier V, Krieger S, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat*. 2012;33(8):1228-1238.

123. Houdayer C. In silico prediction of splice-affecting nucleotide variants. *Methods Mol Biol*. 2011;760:269-281. doi: 10.1007/978-1-61779-176-5_17; 10.1007/978-1-61779-176-5_17.

124. Houdayer C, Dehainault C, Mattler C, et al. Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat*. 2008;29(7):975-982. doi: 10.1002/humu.20765.

125. Hube F, Guo J, Chooniedass-Kothari S, et al. Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol*. 2006;25(7):418-428. doi: 10.1089/dna.2006.25.418.

126. Inui H, Oh KS, Nadem C, et al. Xeroderma pigmentosum-variant patients from america, europe, and asia. *J Invest Dermatol*. 2008;128(8):2055-2068. doi: 10.1038/jid.2008.48.

127. Ishitsuka Y, Furuta J, Miyashita T, Otsuka F. Splicing aberration in naevoid basal cell carcinoma syndrome. *Acta Derm Venereol*. 2012;92(6):619-620. doi: 10.2340/00015555-1332; 10.2340/00015555-1332.

128. Jeon GW, Kwon MJ, Lee SJ, Sin JB, Ki CS. Clinical and genetic analysis of a korean patient with X-linked chondrodysplasia punctata: Identification of a novel splicing mutation in the ARSE gene. *Ann Clin Lab Sci*. 2013;43(1):70-75.

129. Jimenez NL, Flannick J, Yahyavi M, et al. Targeted 'next-generation' sequencing in anophthalmia and microphthalmia patients confirms SOX2, OTX2 and FOXE3 mutations. *BMC*

Med Genet. 2011;12:172-2350-12-172. doi: 10.1186/1471-2350-12-172; 10.1186/1471-2350-12-172.

130. Johnson AD. Single-nucleotide polymorphism bioinformatics: A comprehensive review of resources. *Circ Cardiovasc Genet*. 2009;2(5):530-536. doi: 10.1161/CIRCGENETICS.109.872010.

131. Johnson DS. *Study of a possible genetic cause of CHARGE association.* [MD]. Glasgow, UK: University of Glasgow; 2010.

132. Jung SC, Park JW, Cho DY, et al. PHEX gene mutations and genotype-phenotype analysis of korean patients with hypophosphatemic ricketts. .

133. Kahn AB, Zeeberg BR, Ryan MC, et al. Ontogenomic study of the relationship between number of gene splice variants and GO categorization. *Bioinformatics*. 2010;26(16):1945-1949. doi: 10.1093/bioinformatics/btq335; 10.1093/bioinformatics/btq335.

134. Kang D, Lee DH, Hong Y, et al. Identification of a novel splicing mutation in the ARSA gene in a patient with late-infantile form of metachromatic leukodystrophy. *The Korean Journal of Laboratory Medicine*. 2010;30(5):516-520.

135. Kannabiran C, Rogan PK, Olmos L, et al. Autosomal dominant zonular cataract with sutural opacities is associated with a splice mutation in the betaA3/A1-crystallin gene. *Mol Vis*. 1998;4:21.

136. Kaput J, Cotton RG, Hardman L, et al. Planning the human variome project: The spain report. *Hum Mutat*. 2009;30(4):496-510. doi: 10.1002/humu.20972.

137. Keren B, Suzuki OT, Gerard-Blanluet M, et al. CNS malformations in knobloch syndrome with splice mutation in COL18A1 gene. *Am J Med Genet A*. 2007;143A(13):1514-1518. doi: 10.1002/ajmg.a.31784.

138. Kern JS. *The molecular basis of dystrophic epidermolysis bullosa: Mutation detection and study of clinical, biochemical, and molecular findings in 29 patients.* Freiburg im Breisgau, Germany: Albert Ludwigs Universitat Freiburg; 2005.

139. Khan SG, Levy HL, Legerski R, et al. Xeroderma pigmentosum group C splice mutation associated with autism and hypoglycinemia. *J Invest Dermatol*. 1998;111(5):791-796. doi: 10.1046/j.1523-1747.1998.00391.x.

140. Khan SG, Muniz-Medina V, Shahlavi T, et al. The human XPC DNA repair gene: Arrangement, splice site information content and influence of a single nucleotide polymorphism in a splice acceptor site on alternative splicing and function. *Nucleic Acids Res*. 2002;30(16):3624-3631.

141. Khurana E, Fu Y, Colonna V, et al. Integrative annotation of variants from 1092 humans: Application to cancer genomics. *Science*. 2013;342(6154):1235587.

142. Kim GH, Ko JM, Lee JJ, Yoo HW. A novel intronic point mutation of CPS1 gene in a korean family with CPS1 deficiency. .

143. Kolsch H, Jessen F, Wiltfang J, et al. Influence of SORL1 gene variants: Association with CSF amyloid-beta products in probable alzheimer's disease. *Neurosci Lett*. 2008;440(1):68-71. doi: 10.1016/j.neulet.2008.05.049.

144. Kolsch H, Jessen F, Wiltfang J, et al. Association of SORL1 gene variants with alzheimer's disease. *Brain Res*. 2009;1264:1-6. doi: 10.1016/j.brainres.2009.01.044; 10.1016/j.brainres.2009.01.044.

145. Kolsch H, Lutjohann D, Jessen F, et al. RXRA gene variations influence alzheimer's disease risk and cholesterol metabolism. *J Cell Mol Med*. 2009;13(3):589-598. doi: 10.1111/j.1582-4934.2009.00383.x.

146. Kolsch H, Lutjohann D, Jessen F, et al. CYP46A1 variants influence alzheimer's disease risk and brain cholesterol metabolism. *Eur Psychiatry*. 2009;24(3):183-190. doi: 10.1016/j.eurpsy.2008.12.005.

147. Koukouritaki SB, Poch MT, Cabacungan ET, McCarver DG, Hines RN. Discovery of novel flavin-containing monooxygenase 3 (FMO3) single nucleotide polymorphisms and functional analysis of upstream haplotype variants. *Mol Pharmacol*. 2005;68(2):383-392. doi: 10.1124/mol.105.012062.

148. Kralovicova J, Lei H, Vorechovsky I. Phenotypic consequences of branch point substitutions. *Hum Mutat*. 2006;27(8):803-813. doi: 10.1002/humu.20362.

149. Královicová J, Gaunt TR, Rodriguez S, Wood PJ. Variants in the human insulin gene that affect pre-mRNA splicing: Is -23HphI a functional single nucleotide polymorphism at IDDM2? *Diabetes*. 2006;55(1):260-4. http://search.proquest.com/docview/216476271?accountid=15115.

150. Krawczak M, Thomas NS, Hundrieser B, et al. Single base‐pair substitutions in exon–intron junctions of human genes: Nature, distribution, and consequences for mRNA splicing. *Hum Mutat*. 2007;28(2):150-158.

151. Kwon MJ, Baek W, Ki CS, et al. Screening of the SOD1, FUS, TARDBP, ANG, and OPTN mutations in korean patients with familial and sporadic ALS. *Neurobiol Aging*. 2012;33(5):1017.e17-1017.e23. doi: 10.1016/j.neurobiolaging.2011.12.003.

152. Kwong AK, Fung CW, Chan SY, Wong VC. Identification of SCN1A and PCDH19 mutations in chinese children with dravet syndrome. *PLoS One*. 2012;7(7):e41802. doi: 10.1371/journal.pone.0041802; 10.1371/journal.pone.0041802.

153. Lacroix M, Lacaze-Buzy L, Furio L, et al. Clinical expression and new SPINK5 splicing defects in netherton syndrome: Unmasking a frequent founder synonymous mutation and unconventional intronic mutations. *J Invest Dermatol*. 2011. doi: 10.1038/jid.2011.366; 10.1038/jid.2011.366.

154. Lamba V, Lamba J, Yasuda K, et al. Hepatic CYP2B6 expression: Gender and ethnic differences and relationship to CYP2B6 genotype and CAR (constitutive androstane receptor) expression. *J Pharmacol Exp Ther*. 2003;307(3):906-922. doi: 10.1124/jpet.103.054866.

155. Lancellotti S, Di Leo E, Calandra S, Tarugi P. Difetto di splicing del pre-mrna dell'apolipoproteina b nel fegato di pazienti con ipobetalipoproteinemia familiare. *Patologia genetica*. 2005.

156. Laššuthová P, Žaliová M, Inoue K, et al. Three new PLP1 splicing mutations demonstrate pathogenic and phenotypic diversity of pelizaeus-merzbacher disease. *J Child Neurol*. 2013.

157. Le Guedard-Mereuze S, Vache C, Molinari N, et al. Sequence contexts that determine the pathogenicity of base substitutions at position +3 of donor splice-sites. *Hum Mutat*. 2009;30(9):1329-1339. doi: 10.1002/humu.21070; 10.1002/humu.21070.

158. Lebel K. *Génétique moléculaire du glaucome: WDR36, un gène modificateur potentiel pour la sévérité du glaucome*. 2008.

159. Leclerc D, Boutros M, Suh D, et al. SLC7A9 mutations in all three cystinuria subtypes. *Kidney Int*. 2002;62(5):1550-1559.

160. Leclerc D, Wu Q, Ellis JR, Goodyer P, Rozen R. Is the SLC7A10 gene on chromosome 19 a candidate locus for cystinuria? *Mol Genet Metab*. 2001;73(4):333-339.

161. Lee SH, Kim S, Noh EB, Oh S, Kim S. Novel deletion mutation (c. 3717del5) in COL7A1 in a patient with recessive dystrophic epidermolysis bullosa. *J Dermatol*. 2013;40(1):59-61.

162. Lee Y, Lee DH, Vockley J, Kim N, Lee YK, Ki C. Different spectrum of mutations of isovaleryl-CoA dehydrogenase (< i> IVD</i>) gene in korean patients with isovaleric acidemia. *Mol Genet Metab*. 2007;92(1):71-77.

163. Lee PP, Chen TX, Jiang LP, et al. Clinical and molecular characteristics of 35 chinese children with wiskott-aldrich syndrome. *J Clin Immunol*. 2009;29(4):490-500. doi: 10.1007/s10875-009-9285-9.

164. Lee PY. *Prioritizing SNPs for disease-gene association studies: Algorithms and systems.* [PhD]. Kingston, Ontario, Canada: Queens University; 2009.

165. Lee ST, Lee J, Lee M, Kim JW, Ki CS. Clinical and genetic analysis of korean patients with congenital insensitivity to pain with anhidrosis. *Muscle Nerve*. 2009;40(5):855-859. doi: 10.1002/mus.21340.

166. Lehtokari V, Pelin K, Herczegfalvi A, et al. Nemaline myopathy caused by mutations in the nebulin gene may present as a distal myopathy. *Neuromuscular Disorders*. 2011;21(8):556-562.

167. Leman AR, Pearce DA, Rothberg PG. Gene symbol: CLN3. disease: Juvenile neuronal ceroid lipofuscinosis (batten disease). *Hum Genet*. 2005;116(3):236.

168. Leverenz JB, Yu CE, Montine TJ, et al. A novel progranulin mutation associated with variable clinical presentation and tau, TDP43 and alpha-synuclein pathology. *Brain*. 2007;130(Pt 5):1360-1374. doi: 10.1093/brain/awm069.

169. Li A, Jiao X, Munier FL, et al. Bietti crystalline corneoretinal dystrophy is caused by mutations in the novel gene< i> CYP4V2</i>. *The American Journal of Human Genetics*. 2004;74(5):817-826.

170. Li X, Guan M. Identification and characterization of mouse< i> GTPBP3</i> gene encoding a mitochondrial GTP-binding protein involved in tRNA modification. *Biochem Biophys Res Commun*. 2003;312(3):747-754.

171. Li X, Zhang LS, Guan M. Cloning and characterization of mouse mTERF encoding a mitochondrial transcriptional termination factor. *Biochem Biophys Res Commun*. 2005;326(2):505-510.

172. Li L, Xiao X, Li S, et al. Detection of variants in 15 genes in 87 unrelated chinese patients with leber congenital amaurosis. *PLoS One*. 2011;6(5):e19458. doi: 10.1371/journal.pone.0019458.

173. Li L, Xiao X, Yi C, et al. Confirmation and refinement of an autosomal dominant congenital motor nystagmus locus in chromosome 1q31.3-q32.1. *J Hum Genet*. 2012;57(12):756-759. doi: 10.1038/jhg.2012.103; 10.1038/jhg.2012.103.

174. Lietman SA. Preimplantation genetic diagnosis for hereditary endocrine disease. *Endocr Pract*. 2011;17 Suppl 3:28-32. doi: 10.4158/EP11056.RA.

175. Lietman SA, Goldfarb J, Desai N, Levine MA. Preimplantation genetic diagnosis for severe albright hereditary osteodystrophy. *J Clin Endocrinol Metab*. 2008;93(3):901-904. doi: 10.1210/jc.2007-2040.

176. Lim BC, Hwang H, Chae JH, et al. SCN1A mutational analysis in korean patients with dravet syndrome. *Seizure*. 2011;20(10):789-794. doi: 10.1016/j.seizure.2011.08.002.

177. Lim BC, Ki CS, Kim JW, et al. Fukutin mutations in congenital muscular dystrophies with defective glycosylation of dystroglycan in korea. *Neuromuscul Disord*. 2010;20(8):524-530. doi: 10.1016/j.nmd.2010.06.005; 10.1016/j.nmd.2010.06.005.

178. Liu J, Zhou X, Shan Z, et al. The association of LRP5 gene polymorphisms with ankylosing spondylitis in a chinese han population. *J Rheumatol*. 2011;38(12):2616-2618. doi: 10.3899/jrheum.111117.

179. Liu Z, Venkatesh SS, Maley CC. Sequence space coverage, entropy of genomes and the potential to detect non-human DNA in human samples. *BMC Genomics*. 2008;9:509. doi: 10.1186/1471-2164-9-509.

180. Locke G, Haberman D, Johnson SM, Morozov AV. Global remodeling of nucleosome positions in C. elegans. *BMC Genomics*. 2013;14(1):284.

181. Lopez-Jimenez E, de Campos JM, Kusak EM, et al. SDHC mutation in an elderly patient without familial antecedents. *Clin Endocrinol (Oxf)*. 2008;69(6):906-910. doi: 10.1111/j.1365-2265.2008.03368.x.

182. Lopezjimenez N, Flannick J, Yahyavi M, et al. Targeted next-generation sequencing in anophthalmia and microphthalmia patients confirms SOX2, OTX2 and FOXE3 mutations. *BMC Med Genet*. 2011;12(1):172. doi: 10.1186/1471-2350-12-172.

183. Lou H, Li H, Yeager M, et al. Promoter variants in the MSMB gene associated with prostate cancer regulate MSMB/NCOA4 fusion transcripts. *Hum Genet*. 2012;131(9):1453-1466. doi: 10.1007/s00439-012-1182-2; 10.1007/s00439-012-1182-2.

184. Luquin N, Yu B, Trent RJ, Morahan JM, Pamphlett R. An analysis of the entire SOD1 gene in sporadic ALS. *Neuromuscul Disord*. 2008;18(7):545-552. doi: 10.1016/j.nmd.2008.04.013.

185. Luquin N, Yu B, Saunderson RB, Trent RJ, Pamphlett R. Genetic variants in the promoter of TARDBP in sporadic amyotrophic lateral sclerosis. *Neuromuscular Disorders*. 2009;19(10):696-700. doi: 10.1016/j.nmd.2009.07.005.

186. Mackay DS, Henderson RH, Sergouniotis PI, et al. Novel mutations in MERTK associated with childhood onset rod-cone dystrophy. *Mol Vis*. 2010;16:369-377.

187. Maddalena A, Bale S, Das S, Grody W, Richards S, ACMG Laboratory Quality Assurance Committee. Technical standards and guidelines: Molecular genetic testing for ultra-rare disorders. *Genet Med*. 2005;7(8):571-583.

188. Magnolo L, Najah M, Fancello T, et al. Novel mutations in SAR1B and MTTP genes in tunisian children with chylomicron retention disease and abetalipoproteinemia. *Gene*. 2013;512(1):28-34. doi: 10.1016/j.gene.2012.09.117; 10.1016/j.gene.2012.09.117.

189. Malueka RG, Takaoka Y, Yagi M, et al. Categorization of 77 dystrophin exons into 5 groups by a decision tree using indexes of splicing regulatory factors as decision markers. *BMC Genet*. 2012;13:23-2156-13-23. doi: 10.1186/1471-2156-13-23; 10.1186/1471-2156-13-23.

190. Mao Söderberg M. *Clinical Pharmacogenetics of Olanzapine: with Focus on FMO Gene Polymorphisms*. 2012.

191. Marchal A, Goffinet L, Charlesworth A, et al. Un cas particulier d'épidermolyse bulleuse dystrophique. . 2011;138(12):A168-A169.

192. Marco EJ, Abidi FE, Bristow J, et al. ARHGEF9 disruption in a female patient is associated with X linked mental retardation and sensory hyperarousal. *BMJ Case Rep*. 2009;2009:bcr06.2009.1999. Epub 2009 Jul 2. doi: 10.1136/bcr.06.2009.1999.

193. Marco EJ, Bristow J, Cotter P, et al. ARHGEF9: Identification of a novel X-linked mental retardation and behavior disorder gene. *Am Neurological Association*. 2006:Poster #S-144.

194. Martoni E, Urciuolo A, Sabatelli P, et al. Identification and characterization of novel collagen VI non-canonical splicing mutations causing ullrich congenital muscular dystrophy. *Hum Mutat*. 2009;30(5):E662-72. doi: 10.1002/humu.21022.

195. Maruszak A, Safranow K, Gustaw K, et al. PIN1 gene variants in alzheimer's disease. *BMC Med Genet*. 2009;10:115. doi: 10.1186/1471-2350-10-115.

196. McGrory J, Cole WG. Alternative splicing of exon 37 of FBN1 deletes part of an 'eight - cysteine'domain resulting in the marfan syndrome. *Clin Genet*. 1999;55(2):118-121.

197. Megremis S, Mitsioni A, Mitsioni AG, et al. Nucleotide variations in the NPHS2 gene in greek children with steroid-resistant nephrotic syndrome. *Genet Test Mol Biomarkers*. 2009;13(2):249-256. doi: 10.1089/gtmb.2008.0083; 10.1089/gtmb.2008.0083.

198. Mintchev N, Zamba-Papanicolaou E, Kleopa KA, Christodoulou K. A novel ALS2 splice-site mutation in a cypriot juvenile-onset primary lateral sclerosis family. *Neurology*. 2009;72(1):28-32. doi: 10.1212/01.wnl.0000338530.77394.60.

199. Mondal A, Das S, Chu W, Sharma N, Elbein S. Genotype and tissue effects on alternative splicing of the TCF7L2 gene in tissues important to type 2 diabetes (T2DM) pathogenesis. . :Abstract #1352.

200. Moriwaki K, Noda K, Furukawa Y, et al. Deficiency of GMDS leads to escape from NK cell-mediated tumor surveillance through modulation of TRAIL signaling. *Gastroenterology*. 2009;137(1):188-98, 198.e1-2. doi: 10.1053/j.gastro.2009.04.002.

201. Mukhopadhyay A, Nikopoulos K, Maugeri A, et al. Erosive vitreoretinopathy and wagner disease are caused by intronic mutations in CSPG2/Versican that result in an imbalance of splice variants. *Invest Ophthalmol Vis Sci*. 2006;47(8):3565-3572. doi: 10.1167/iovs.06-0141.

202. Murphy LC, Leygue E. The role of estrogen receptor-beta in breast cancer. *Semin Reprod Med*. 2012;30(1):5-13. doi: 10.1055/s-0031-1299592; 10.1055/s-0031-1299592.

203. Naiya T, Misra AK, Biswas A, Das SK, Ray K, Ray J. Occurrence of GCH1 gene mutations in a group of indian dystonia patients. *J Neural Transm*. 2012. doi: 10.1007/s00702-012-0777-z.

204. Najah M, Di Leo E, Awatef J, et al. Identification of patients with abetalipoproteinemia and homozygous familial hypobetalipoproteinemia in tunisia. *Clin Chim Acta*. 2009;401(1-2):51-56. doi: 10.1016/j.cca.2008.11.012.

205. Nalla VK, Rogan PK. Automated splicing mutation analysis by information theory. *Hum Mutat*. 2005;25(4):334-342. doi: 10.1002/humu.20151.

206. Naruse H, Ikawa N, Yamaguchi K, et al. Determination of splice-site mutations in lynch syndrome (hereditary non-polyposis colorectal cancer) patients using functional splicing assay. *Fam Cancer*. 2009;8(4):509-517. doi: 10.1007/s10689-009-9280-6; 10.1007/s10689-009-9280-6.

207. Nasim MT, Ogo T, Ahmed M, et al. Molecular genetic characterization of SMAD signaling molecules in pulmonary arterial hypertension. *Hum Mutat*. 2011;32(12):1385-1389. doi: 10.1002/humu.21605.

208. Nyiraneza C, Jouret-Mourin A, Kartheuser A, et al. Distinctive patterns of p53 protein expression and microsatellite instability in human colorectal cancer. *Hum Pathol*. 2011;42(12):1897-1910. doi: 10.1016/j.humpath.2010.06.021; 10.1016/j.humpath.2010.06.021.

209. Oetting WS, Tabone T. The 2004 human genome variation society scientific meeting. *Hum Mutat*. 2005;26(2):160-163. doi: 10.1002/humu.20194.

210. Oh KS, Khan SG, Jaspers NG, et al. Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): Xeroderma pigmentosum without and with cockayne syndrome. *Hum Mutat*. 2006;27(11):1092-1103. doi: 10.1002/humu.20392.

211. Oh SW, Lee JS, Kim MY, Kim SC. COL7A1 mutational analysis in korean patients with dystrophic epidermolysis bullosa. *Br J Dermatol*. 2007;157(6):1260-1264. doi: 10.1111/j.1365-2133.2007.08191.x.

212. Oh SW, Lee JS, Kim MY, Kim SC. Novel keratin 5 mutations in epidermolysis bullosa simplex: Cases with unusual genotype-phenotype correlation. *J Dermatol Sci*. 2007;48(3):229-232. doi: 10.1016/j.jdermsci.2007.07.014.

213. Okubo M, Ishihara M, Iwasaki T, et al. A novel APOA5 splicing mutation IVS2+1g>a in a japanese chylomicronemia patient. *Atherosclerosis*. 2009;207(1):24-25. doi: 10.1016/j.atherosclerosis.2009.03.046; 10.1016/j.atherosclerosis.2009.03.046.

214. Olsen RK, Brøner S, Sabaratnam R, et al. The ETFDH c. 158A> G variation disrupts the balanced binding of ESE and ESS proteins causing missplicing and multiple acyl‐CoA dehydrogenation deficiency. *Hum Mutat*. 2013.

215. Ozaltin F, Ibsirlioglu T, Taskiran EZ, et al. Disruption of PTPRO causes childhood-onset nephrotic syndrome. *Am J Hum Genet*. 2011;89(1):139-147. doi: 10.1016/j.ajhg.2011.05.026.

216. Palomino-Doza J, Rahman TJ, Avery PJ, et al. Ambulatory blood pressure is associated with polymorphic variation in P2X receptor genes. *Hypertension*. 2008;52(5):980-985. doi: 10.1161/HYPERTENSIONAHA.108.113282.

217. Papi L, Putignano AL, Congregati C, et al. A PALB2 germline mutation associated with hereditary breast cancer in italy. *Fam Cancer*. 2010;9(2):181-185. doi: 10.1007/s10689-009-9295-z.

218. Papp J, Kovacs ME, Olah E. Germline MLH1 and MSH2 mutational spectrum including frequent large genomic aberrations in hungarian hereditary non-polyposis colorectal cancer families: Implications for genetic testing. *World J Gastroenterol*. 2007;13(19):2727-2732.

219. Parslow GR. Websites of note. *Biochem Mol Biol Educ*. 2011;39(3):230-232. doi: 10.1002/bmb.20515; 10.1002/bmb.20515.

220. Pasmooij AM. *Revertant mosaicism in epidermolysis bullosa due to different second site mutations in LAMB3.* Groningen, Netherlands: University Medical Center Groningen; 2006.

221. Pasmooij AM, Pas HH, Bolling MC, Jonkman MF. Revertant mosaicism in junctional epidermolysis bullosa due to multiple correcting second-site mutations in LAMB3. *J Clin Invest*. 2007;117(5):1240-1248. doi: 10.1172/JCI30465.

222. Pasvolsky R, Feigelson SW, Kilic SS, et al. A LAD-III syndrome is associated with defective expression of the rap-1 activator CalDAG-GEFI in lymphocytes, neutrophils, and platelets. *J Exp Med*. 2007;204(7):1571-1582. doi: 10.1084/jem.20070058.

223. Pelucchi S, Mariani R, Trombini P, et al. Expression of hepcidin and other iron-related genes in type 3 hemochromatosis due to a novel mutation in transferrin receptor-2. *Haematologica*. 2009;94(2):276-279. doi: 10.3324/haematol.13576.

224. Perez B, Desviat L, Rodrıguez-Pombo P, et al. Propionic acidemia: Identification of twenty-four novel mutations in europe and north america. *Mol Genet Metab*. 2003;78(1):59-67.

225. Pernet C, Bessis D, Savignac M, Tron E, Guillot B, Hovnanian A. Genitoperineal papular acantholytic dyskeratosis is allelic to Hailey–Hailey disease. *Br J Dermatol*. 2012;167(1):210-212.

226. Pink AE, Simpson MA, Desai N, et al. Mutations in the gamma-secretase genes NCSTN, PSENEN, and PSEN1 underlie rare forms of hidradenitis suppurativa (acne inversa). *J Invest Dermatol*. 2012;132(10):2459-2461. doi: 10.1038/jid.2012.162; 10.1038/jid.2012.162.

227. Piva F, Giulietti M, Nardi B, Bellantuono C, Principato G. An improved in silico selection of phenotype affecting polymorphisms in SLC6A4, HTR1A and HTR2A genes. *Human Psychopharmacology-Clinical and Experimental*. 2010;25(2):153-161. doi: 10.1002/hup.1100.

228. Priore Oliva C, Tarugi P, Calandra S, et al. A novel sequence variant in APOA5 gene found in patients with severe hypertriglyceridemia. *Atherosclerosis*. 2006;188(1):215-217. doi: 10.1016/j.atherosclerosis.2006.04.010.

229. Qadah T, Finlayson J, Ghassemifar R. In vitro characterization of the α-thalassemia point mutation HBA2: C. 95 1G> A [IVS-I-1 (G> A)(α2)]. *Hemoglobin*. 2012;36(1):38-46.

230. Qin S, Shen L, Zhang A, et al. Systematic polymorphism analysis of the CYP2D6 gene in four different geographical han populations in mainland china. *Genomics*. 2008;92(3):152-158. doi: 10.1016/j.ygeno.2008.05.004.

231. Remaley AT, Rust S, Rosier M, et al. Human ATP-binding cassette transporter 1 (ABC1): Genomic organization and identification of the genetic defect in the original tangier disease kindred. *Proceedings of the National Academy of Sciences*. 1999;96(22):12685-12690.

232. Rhyne J, Mantaring MM, Gardner DF, Miller M. Multiple splice defects in ABCA1 cause low HDL-C in a family with hypoalphalipoproteinemia and premature coronary disease. *BMC Med Genet*. 2009;10:1. doi: 10.1186/1471-2350-10-1.

233. Riveira-Munoz E, Chang Q, Godefroid N, et al. Transcriptional and functional analyses of SLC12A3 mutations: New clues for the pathogenesis of gitelman syndrome. *J Am Soc Nephrol*. 2007;18(4):1271-1283. doi: 10.1681/ASN.2006101095.

234. Riveira-Munoz E, Devuyst O, Belge H, et al. Evaluating PVALB as a candidate gene for SLC12A3-negative cases of gitelman's syndrome. *Nephrol Dial Transplant*. 2008;23(10):3120-3125. doi: 10.1093/ndt/gfn229.

235. Roca X, Krainer AR, Eperon IC. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev*. 2013;27(2):129-144. doi: 10.1101/gad.209759.112; 10.1101/gad.209759.112.

236. Rogan PK, Faux BM, Schneider TD. Information analysis of human splice site mutations. *Hum Mutat*. 1998;12(3):153-171. doi: 2-I.

237. Rogan PK, Schneider TD. Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Hum Mutat*. 1995;6(1):74-76. doi: 10.1002/humu.1380060114.

238. Rogan PK, Svojanovsky S, Leeder JS. Information theory-based analysis of CYP2C19, CYP2D6 and CYP3A5 splicing mutations. *Pharmacogenetics*. 2003;13(4):207-218. doi: 10.1097/01.fpc.0000054078.64000.de.

239. Rogan PK. *Ab initio exon definition using an information theory-based approach.* ; 2009:852. 10.1109/CISS.2009.5054835.

240. Rossi PI, Vaccari CM, Terracciano A, et al. The metabotropic glutamate receptor 1, GRM1: Evaluation as a candidate gene for inherited forms of cerebellar ataxia. *J Neurol*. 2010;257(4):598-602. doi: 10.1007/s00415-009-5380-3.

241. Roux-Buisson N, Rendu J, Denjoy I, et al. Functional analysis reveals splicing mutations of the CASQ2 gene in patients with CPVT: Implication for genetic counselling and clinical management. *Hum Mutat*. 2011. doi: 10.1002/humu.21537; 10.1002/humu.21537.

242. Russcher H. *Glucocorticoid receptor variants modulate the sensitivity to cortisol.* Rotterdam, Netherlands: Erasmus University; 2006.

243. Russcher H, Smit P, van Rossum EF, et al. Strategies for the characterization of disorders in cortisol sensitivity. *J Clin Endocrinol Metab*. 2006;91(2):694-701. doi: 10.1210/jc.2005-2212.

244. Sabet A, Li J, Ghandour K, et al. Skin biopsies demonstrate MPZ splicing abnormalities in charcot-marie-tooth neuropathy 1B. *Neurology*. 2006;67(7):1141-1146. doi: 10.1212/01.wnl.0000238499.37764.b1.

245. Sanggaard KM, Rendtorff ND, Kjaer KW, et al. Branchio-oto-renal syndrome: Detection of EYA1 and SIX1 mutations in five out of six danish families by combining linkage, MLPA and sequencing analyses. *Eur J Hum Genet*. 2007;15(11):1121-1131. doi: 10.1038/sj.ejhg.5201900.

246. Sankaranarayanan R, Vasavada AR, Praveen MR, et al. Exploration of molecular factors impairing superoxide dismutase (SOD) isoforms activity in human senile cataractous lenses. *Invest Ophthalmol Vis Sci*. 2013. doi: 10.1167/iovs.13-11935; 10.1167/iovs.13-11935.

247. Schneider TD. Twenty years of delila and molecular information theory: The altenberg-austin workshop in theoretical biology biological information, beyond metaphor: Causality, explanation, and unification altenberg, austria, 11-14 july 2002. *Biol Theory*. 2006;1(3):250-260.

248. Schneider TD. Information content of individual genetic sequences. *J Theor Biol*. 1997;189(4):427-441. doi: 10.1006/jtbi.1997.0540.

249. Schneider TD, Rogan PK, inventors; The United States of America as represented by the Department of Health and Human Services, assignee. Computational Analysis of Nucleic Acid Information Defines Binding Sites. patent 5867402. Feb 2, 1999, .

250. Schonfelder EM, Knuppel T, Tasic V, et al. Mutations in uroplakin IIIA are a rare cause of renal hypodysplasia in humans. *Am J Kidney Dis*. 2006;47(6):1004-1012. doi: 10.1053/j.ajkd.2006.02.177.

251. Schwaderer P, Knuppel T, Konrad M, et al. Clinical course and NPHS2 analysis in patients with late steroid-resistant nephrotic syndrome. *Pediatr Nephrol*. 2008;23(2):251-256. doi: 10.1007/s00467-007-0653-5.

252. Shirley BC, Mucaki EJ, Whitehead T, Costea PI, Akan P, Rogan PK. Interpretation, stratification and evidence for sequence variants affecting mRNA splicing in complete human genome sequences. *Genomics Proteomics Bioinformatics*. 2013;11(2):77-85. doi: 10.1016/j.gpb.2013.01.008; 10.1016/j.gpb.2013.01.008.

253. Shultzaberger RK, Bucheimer RE, Rudd KE, Schneider TD. Anatomy of< i> escherichia coli</i> ribosome binding sites. *J Mol Biol*. 2001;313(1):215-228.

254. Simpson MA, Hsu R, Keir LS, et al. Mutations in FAM20C are associated with lethal osteosclerotic bone dysplasia (raine syndrome), highlighting a crucial molecule in bone development. *Am J Hum Genet*. 2007;81(5):906-912. doi: 10.1086/522240.

255. Skipper L, Shen H, Chua E, et al. Analysis of LRRK2 functional domains in nondominant parkinson disease. *Neurology*. 2005;65(8):1319-1321. doi: 10.1212/01.wnl.0000180517.70572.37.

256. Slavotinek AM, Baranzini SE, Schanze D, et al. Manitoba-oculo-tricho-anal (MOTA) syndrome is caused by mutations in FREM1. *J Med Genet*. 2011;48(6):375-382. doi: 10.1136/jmg.2011.089631.

257. Slavotinek AM, Chao R, Vacik T, et al. VAX1 mutation associated with microphthalmia, corpus callosum agenesis, and orofacial clefting: The first description of a VAX1 phenotype in humans. *Hum Mutat*. 2012;33(2):364-368. doi: 10.1002/humu.21658; 10.1002/humu.21658.

258. Smit P. *Factors determining glucocorticoid sensitivity in man.* Rotterdam, Netherlands: Erasmus University; 2006.

259. Song HR, Park JW, Cho DY, Yang JH, Yoon HR, Jung SC. PHEX gene mutations and genotype-phenotype analysis of korean patients with hypophosphatemic rickets. *J Korean Med Sci*. 2007;22(6):981-986.

260. Soran H, Charlton-Menys V, Hegele R, et al. Proteinuria and severe mixed dyslipidemia associated with a novel APOAV gene mutation. *J Clin Lipidol*. 2010;4(4):310-313. doi: 10.1016/j.jacl.2010.06.004.

261. Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM, IARC Unclassified Genetic Variants Working Group. Prediction and assessment of splicing alterations: Implications for clinical testing. *Hum Mutat*. 2008;29(11):1304-1313. doi: 10.1002/humu.20901.

262. Stockley TL, Mendoza-Londono R, Propst EJ, Sodhi S, Dupuis L, Papsin BC. A recurrent EYA1 mutation causing alternative RNA splicing in branchio-oto-renal syndrome: Implications for molecular diagnostics and disease mechanism. *Am J Med Genet A*. 2009;149A(3):322-327. doi: 10.1002/ajmg.a.32679; 10.1002/ajmg.a.32679.

263. Svojanovsky SR, Schneider TD, Rogan PK. Redundant designations of BRCA1 intron 11 splicing mutation; c. 4216-2A>G; IVS11-2A>G; L78833, 37698, A>G. *Hum Mutat*. 2000;16(3):264. doi: 2-1.

264. Sznajer Y, Coldea C, Meire F, Delpierre I, Sekhara T, Touraine RL. A de novo SOX10 mutation causing severe type 4 waardenburg syndrome without hirschsprung disease. *Am J Med Genet A*. 2008;146A(8):1038-1041. doi: 10.1002/ajmg.a.32247.

265. Tartaglia-Polcini A, Bonnart C, Micheloni A, et al. SPINK5, the defective gene in netherton syndrome, encodes multiple LEKTI isoforms derived from alternative pre-mRNA processing. *J Invest Dermatol*. 2006;126(2):315-324. doi: 10.1038/sj.jid.5700015.

266. Tazi J, Durand S, Jeanteur P. The spliceosome: A novel multi-faceted target for therapy. *Trends Biochem Sci*. 2005;30(8):469-478. doi: 10.1016/j.tibs.2005.06.002.

267. The American College of Medical Genetics Laboratory Quality Assurance Committee. Technical standards and guidelines: Molecular genetic testing for rare disorders. . 2005:Paragraph 7.4.1.3.

268. Thompson TE, Rogan PK, Risinger JI, Taylor JA. Splice variants but not mutations of DNA polymerase beta are common in bladder cancer. *Cancer Res*. 2002;62(11):3251-3256.

269. Titeux M, Mejia JE, Mejlumian L, et al. Recessive dystrophic epidermolysis bullosa caused by COL7A1 hemizygosity and a missense mutation with complex effects on splicing. *Hum Mutat*. 2006;27(3):291-292. doi: 10.1002/humu.9406.

270. Torregrossa R, Anglani F, Fabris A, et al. Identification of GDNF gene sequence variations in patients with medullary sponge kidney disease. *Clin J Am Soc Nephrol*. 2010;5(7):1205-1210. doi: 10.2215/CJN.07551009; 10.2215/CJN.07551009.

271. Torregrossa R, Gambaro G, Fabris M, et al. Glial cell-line derived neurotrophic factor: Un gene candidate per la patogenesi del rene con midollare a spugna. http://www.siqu.net/e107_files/downloads/Comunicazioni/Genetica%20delle%20Malattie%20Complesse.pdf. Updated 2006.

272. Tosetto E, Ceol M, Mezzabotta F, et al. Novel mutations of the CLCN5 gene including a complex allele and A 5' UTR mutation in dent disease 1. *Clin Genet*. 2009;76(4):413-416. doi: 10.1111/j.1399-0004.2009.01212.x.

273. Tosetto E, Ghiggeri GM, Emma F, et al. Phenotypic and genetic heterogeneity in dent's disease--the results of an italian collaborative study. *Nephrol Dial Transplant*. 2006;21(9):2452-2463. doi: 10.1093/ndt/gfl274.

274. Tournier I, Vezain M, Martins A, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat*. 2008;29(12):1412-1424. doi: 10.1002/humu.20796.

275. Tram E, Ibrahim-Zada I, Briollais L, Knight JA, Andrulis IL, Ozcelik H. Identification of germline alterations of the mad homology 2 domain of SMAD3 and SMAD4 from the ontario site of the breast cancer family registry (CFR). *Breast Cancer Res*. 2011;13(4):R77. doi: 10.1186/bcr2926.

276. Tsai KN, Chen GW, Chen CY. A novel algorithm for identification of activated cryptic 5' splice sites. *J Biomol Struct Dyn*. 2012;29(5):1089-1099. doi: 10.1080/073911012010525033.

277. Tsai KN, Wang D. Identification of activated cryptic 5' splice sites using structure profiles and odds measure. *Nucleic Acids Res*. 2012;40(10):e73. doi: 10.1093/nar/gks061; 10.1093/nar/gks061.

278. Tunca B, Pedroni M, Cecener G, et al. Analysis of mismatch repair gene mutations in turkish HNPCC patients. *Fam Cancer*. 2010;9(3):365-376. doi: 10.1007/s10689-010-9336-7.

279. Tuohy T, Burt RW. Attenuated familial adenomatous polyposis: Diagnosis, management and future prognosis. In: Rodriguez-Bigas R, Cutait R, Lunch P, Tomlinson I, Vasen H, eds. *Hereditary colorectal cancer.* Springer; 2010.

280. Vinga S. Information theory applications for biological sequence analysis. *Briefings in bioinformatics*. 2013:bbt068.

281. Vockley J, Rogan PK, Anderson BD, et al. Exon skipping in IVD RNA processing in isovaleric acidemia caused by point mutations in the coding region of the IVD gene. *Am J Hum Genet*. 2000;66(2):356-367. doi: 10.1086/302751.

282. von Kodolitsch Y, Berger J, Rogan PK. Predicting severity of haemophilia A and B splicing mutations by information analysis. *Haemophilia*. 2006;12(3):258-262. doi: 10.1111/j.1365-2516.2006.01216.x.

283. von Kodolitsch Y, Nienaber CA, Fliegner M, Rogan PK. Splice site mutations and atherosclerosis: Mechanisms and prediction models. *Z Kardiol*. 2001;90(2):87-95.

284. von Kodolitsch Y, Pyeritz RE, Rogan PK. Splice-site mutations in atherosclerosis candidate genes: Relating individual information to phenotype. *Circulation*. 1999;100(7):693-699.

285. Vorechovsky I. Aberrant 3' splice sites in human disease genes: Mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res*. 2006;34(16):4630-4641. doi: 10.1093/nar/gkl535.

286. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Hum Mutat*. 2009;30(1):107-114. doi: 10.1002/humu.20811.

287. Wadt K, Choi J, Chung JY, et al. A cryptic BAP1 splice mutation in a family with uveal and cutaneous melanoma, and paraganglioma. *Pigment Cell Melanoma Res*. 2012;25(6):815-818. doi: 10.1111/pcmr.12006; 10.1111/pcmr.12006.

288. Wan L, Lee CC, Hsu CM, et al. Identification of eight novel mutations of the acid alpha-glucosidase gene causing the infantile or juvenile form of glycogen storage disease type II. *J Neurol*. 2008;255(6):831-838. doi: 10.1007/s00415-008-0714-0.

289. Wang E, Dimova N, Cambi F. PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. *Nucleic Acids Res*. 2007;35(12):4164-4178. doi: 10.1093/nar/gkm387.

290. Wang J, Sonnerborg A, Rane A, et al. Identification of a novel specific CYP2B6 allele in africans causing impaired metabolism of the HIV drug efavirenz. *Pharmacogenet Genomics*. 2006;16(3):191-198. doi: 10.1097/01.fpc.0000189797.03845.90.

291. Wang P, Guo X, Jia X, Li S, Xiao X, Zhang Q. Novel mutations of the PAX6 gene identified in chinese patients with aniridia. *Mol Vis*. 2006;12:644-648.

292. Watnick TJ, Garcia-Gonzalez MA, Germino GG, Jones JG, inventorsPKD Mutations and evaluation of the same. 2008.

293. Wessagowit V, Kim SC, Woong Oh S, McGrath JA. Genotype-phenotype correlation in recessive dystrophic epidermolysis bullosa: When missense doesn't make sense. *J Invest Dermatol*. 2005;124(4):863-866. doi: 10.1111/j.0022-202X.2005.23650.x.

294. Wessagowit V, McGrath JA. Clinical and molecular significance of splice site mutations in the plakophilin 1 gene in patients with ectodermal dysplasia-skin fragility syndrome. *Acta Derm Venereol*. 2005;85(5):386-388. doi: 10.1080/00015550510011763.

295. Wessagowit V, Nalla VK, Rogan PK, McGrath JA. Normal and abnormal mechanisms of gene splicing and relevance to inherited skin diseases. *J Dermatol Sci*. 2005;40(2):73-84. doi: 10.1016/j.jdermsci.2005.05.006.

296. Willoughby CE, O'Prey D, Simpson DA. Modeling aberrant splicing in mutant genes associated with inherited retinal degeneration and bardet-biedl syndrome. . :Abstract #1356.

297. Wong T, Gammon L, Liu L, et al. Potential of fibroblast cell therapy for recessive dystrophic epidermolysis bullosa. *J Invest Dermatol*. 2008;128(9):2179-2189. doi: 10.1038/jid.2008.78.

298. Xiong Y, Wang M, Fang K, et al. A systematic genetic polymorphism analysis of the CYP2C9 gene in four different geographical han populations in mainland china. *Genomics*. 2011;97(5):277-281. doi: 10.1016/j.ygeno.2010.11.004.

299. Xu X, Li S, Xiao X, Wang P, Guo X, Zhang Q. Sequence variations of GRM6 in patients with high myopia. *Mol Vis*. 2009;15:2094-2100.

300. Yan Q, Bykhovskaya Y, Li R, et al. Human< i> TRMU</i> encoding the mitochondrial 5-methylaminomethyl-2-thiouridylate-methyltransferase is a putative nuclear modifier gene for the phenotypic expression of the deafness-associated 12S rRNA mutations. *Biochem Biophys Res Commun*. 2006;342(4):1130-1136.

301. Yan Q, Guan M. Identification and characterization of mouse< i> TRMU</i> gene encoding the mitochondrial 5-methylaminomethyl-2-thiouridylate-methyltransferase. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*. 2004;1676(2):119-126.

302. Yang M, Solidar A, Wyckoff GJ. Novel method for discerning the action of selection during evolution. *J Biomed Sci Eng*. 2010;3:109-113.

303. Yu B. In silico interpretation of the splicing code and estimating the abundance of expressed mRNA isoforms. *Hum Mutat*. 2013;34(4):v-v.

304. Yu B. Role of in silico tools in gene discovery. *Mol Biotechnol*. 2009;41(3):296-306. doi: 10.1007/s12033-008-9134-8.

305. Yu H, Patel SB. Recent insights into the smith-lemli-opitz syndrome. *Clin Genet*. 2005;68(5):383-391. doi: 10.1111/j.1399-0004.2005.00515.x.

306. Zaffanello M, Taranta A, Palma A, Bettinelli A, Marseglia GL, Emma F. Type IV bartter syndrome: Report of two new cases. *Pediatr Nephrol*. 2006;21(6):766-770. doi: 10.1007/s00467-006-0090-x.

307. Zampieri S, Buratti E, Dominissini S, et al. Splicing mutations in glycogen-storage disease type II: Evaluation of the full spectrum of mutations and their relation to patients' phenotypes. *European Journal of Human Genetics*. 2010;19(4):422-431.

308. Zhang K, Nowak I, Rushlow D, Gallie BL, Lohmann DR. Patterns of missplicing caused by RB1 gene mutations in patients with retinoblastoma and association with phenotypic expression. *Hum Mutat*. 2008;29(4):475-484. doi: 10.1002/humu.20664.

309. Zhang Q, Zulfiqar F, Riazuddin SA, et al. A variant form of oguchi disease mapped to 13q34 associated with partial deletion of GRK1 gene. *Mol Vis*. 2005;11:977-985.

310. Zhao L, Liang T, Xu J, Lin H, Li D, Qi Y. Two novel FBN1 mutations associated with ectopia lentis and marfanoid habitus in two chinese families. *Mol Vis*. 2009;15:826-832.