

iFuse

Structural Variants from *Next Generation Sequencing* data

Manual & Specifications

Version 1.0 approved

Prepared by Jos van Nijnatten

Department of Bioinformatics

Erasmus Medical Center, Rotterdam

1/1/2012

Table of Contents

Revision History	5
1. Introduction	7
2. Features and dependencies	7
2.1 Features	7
2.2 Dependencies.....	8
3. Installing & Configuration	8
3.1. Installing Apache, PHP, MySQL and SED on Windows.....	8
3.1.1 Apache	8
3.1.2MySQL	8
3.3.3 PHP	8
3.3.4 SED	9
3.2. Installing Apache, PHP and MySQL on linux	9
3.3 Installing iFuse.....	9
3.4 iFuse Configuration	10
3.4.1 Annotation tables.....	10
3.4.2 Sequence retrieval	11
3.4.3 Database setup	11
3.4.4 Time before deleting uploaded projects.....	12
3.4.5 User login time length.....	12
4. iFuse Program Structure	13
5. Using iFuse	13
5.1 Registration and login	13
5.3 Upload or Open session	14
5.3.1 Upload and open files	14
5.3.2 Filetypes.....	15
5.4 Analysis page.....	19
5.4.1 Menu	20
5.4.2 Error bar	22
5.4.3 Event Overview	23
5.4.4 Legend.....	24
5.4.5 Event menu	24

5.4.6 Details	25
6. Quick Tutorial, Finding fusion genes.....	26

Revision History

Date	Name	Reason For Changes	Version
08 / 02 2012	First release	N/A	v1.0
... / ... 20 ..			

1. Introduction

Multiple groups at Erasmus Medical Center are using Next Generation Sequencing techniques to find unknown events in the human genome. The software packages delivered with these techniques and robots are not designed for specific tasks such as finding fusion genes, returning summaries of genesets and giving sequences of events. However, it is possible to do this using the raw data. But to manually find valid fusion genes takes forever and assembling sequences is difficult and time consuming.

iFuse, the Integrated FUSiongene Explorer, is a software package developed at Erasmus Medical Center, Department of Bioinformatics, in Rotterdam. It is written in PHP, R and therefore very mobile. Its purpose is to explore next generation sequencing data and view events as possible fusion genes and other types of events such as deletions, insertions and inversions, etc.

2. Features and dependencies

2.1 Features

iFuse uses University of California, Santa Cruz (UCSC) genome browser and table data to annotate Complete Genomics event data. Because it is annotated by UCSC, iFuse can therefore calculate and retrieve several new attributes, such as;

-  Gene name and accession number
-  Shared, related genes and related junctions
-  iFuse can give event DNA, RNA and protein sequences
-  iFuse generates a picture of the event, containing the promoter, introns, exons, junction site and the length of the event sequence.
-  Several options to sort and filter on, such as;
 - Chromosomes on either side of the event
 - Different genes on either side of the event
 - Event Type, e.g. deletion or insertion
 - Gene Orientation
-  And more...

2.2 Dependencies

iFuse is scripted in PHP and used Apache to display itself over the web and MySQL for user management. iFuse uses UCSC Tables that are stored into the ./R directory and either the UCSC DAS server or the downloaded genomes. See *iFuse configuration* for details.

3. Installing & Configuration

Before you start, one must first download and install two packages, namely Apache and PHP.

Apache is an open source web server for Windows, Mac, Linux and other Unix-like operating systems. PHP: Hypertext Preprocessor is a scripting language, originally inspired by other scripting languages like Perl and Python. The syntax of PHP looks mostly like that of C but object oriented programming is possible since its most recent version (PHP5). Using the *mod_php* extension, Apache can use PHP to dynamically generate web pages.

iFuse minimal requirements are Apache2, PHP5 and MySQL 4. The paragraphs below describe how to set up a clean new web server running Apache, PHP and MySQL. A standard installation like this is not fully secure. For advanced configuration to make the server secure, please read the corresponding manuals of Apache, PHP and MySQL.

3.1. Installing Apache, PHP, MySQL and SED on Windows

3.1.1 Apache

1. Download the MSI installer for the latest version of Apache2 from the Apache website (<http://httpd.apache.org/>).
2. Double click on the file to execute the installer. Follow the installer.
For *Network Domain* and *Server Name* the i.p. address of the computer is sufficient.
3. When the installer is done, browse with Internet Explorer to `localhost` or `127.0.0.1` to check if the installer worked (it will show "It works!")

3.1.2 MySQL

1. Download the MySQL MSI Installer from the website of MySQL. (<http://dev.mysql.com/downloads/mysql/>)
2. Double click on the file to execute the installer. Follow the installer.

3.3.3 PHP

1. Download the PHP compressed zip file from the website of PHP. (<http://windows.php.net/>)
2. Extract the file on your system, e.g. in `c:/php`
3. When the installer is done, open `./conf/httpd.conf` in the apache-directory
4. Add the following lines to the file, right below the *LoadModule* section

```
LoadModule php5_module "c:/php /php5apache2.dll"
AddType application/x-httpd-php .php .phtml .inc .php3
AddType application/x-httpd-php-source .phps
```

5. Restart the apache service
6. The web server document root is `./htdocs` in the apache-directory

3.3.4 PHP ImageMagick extension

1. Download ImageMagick from www.imagemagick.org/script/binary-release.php (binary, static, 16 bits per pixel), and install it.
2. Make sure the path to the ImageMagick program is in the environment variables of Windows under the key 'MAGICK_HOME', and in the PATH environment.
3. Download (<http://valokuva.org/builds/>) the correct extension for PHP and place it in the extension directory of PHP
4. PHP.ini file needs to be updated. Identify the extension directory for PHP correctly and place the DLL file inside the directory. Update the PHP.ini file with this extension.
5. Restart Apache

3.3.4 SED & AWK

1. Download SED as a zip from GNUWin32 (<http://gnuwin32.sourceforge.net/packages/sed.htm>)
2. Download AWK as a zip from GNUWin32 (<http://gnuwin32.sourceforge.net/packages/gawk.htm>)
3. Place it in C:\Windows\System32

3.2. Installing Apache, PHP and MySQL on Linux

1. For Debian Linux distributions (Debian and Ubuntu), open the terminal
Execute the following snippet:
`sudo apt-get install apache2 php5 libapache2-mod-php5 mysql-server php5-mysql`
2. This will require your password, type it in
3. The web server document root is `/var/www`

By installing PHP on Linux systems, you also install PECL. PECL can be used to install packages extending PHP. One example iFuse uses is ImageMagick

1. First install ImageMagick and its developers package:
`sudo apt-get install ImageMagick ImageMagick-devel`
2. Install the php extension (follow the on screen guide)
`pecl install imagick`

iFuse should now be able to display the images in PNG format

3.3 Installing iFuse

1. Download the latest version of iFuse from www-bioinf.ethz.ch/~erik/iFuse/.
2. The downloaded file is a RAR-file and needs to be unpacked. This can be done on a windows machine using WINRAR (www.winrar.nl) and on Linux systems using 'unrar'.
 - a. Linux: `unrar e iFuse.rar`
3. Move all the files into the web server document root directory.
4. On a Linux system, CHMOD the TMP directory to 777, so apache can create, write and delete files in this folder.

5. Create the MySQL schema and insert the following two SQL tables:

<pre> DROP TABLE IF EXISTS `users`; SET @saved_cs_client = @@character_set_client; SET character_set_client = utf8; CREATE TABLE `users` (`id_user` mediumint(8) unsigned zerofill NOT NULL auto_increment, `user_name` varchar(45) NOT NULL, `user_password` varchar(45) NOT NULL, `user_email` varchar(100) NOT NULL, `user_last_pageview` int(10) default NULL, `user_authkey` varchar(45) default NULL, PRIMARY KEY (`id_user`), UNIQUE KEY `user_name_UNIQUE` (`user_name`), UNIQUE KEY `user_email_UNIQUE` (`user_email`)) ENGINE=InnoDB AUTO_INCREMENT=5 DEFAULT CHARSET=latin1; SET character_set_client = @saved_cs_client; </pre>	<pre> DROP TABLE IF EXISTS `sessions`; SET @saved_cs_client = @@character_set_client; SET character_set_client = utf8; CREATE TABLE `sessions` (`session_id` varchar(40) NOT NULL default '0', `ip_address` varchar(16) NOT NULL default '0', `user_agent` varchar(120) NOT NULL, `last_activity` int(10) unsigned NOT NULL default '0', `user_data` text NOT NULL, PRIMARY KEY (`session_id`), KEY `last_activity_idx` (`last_activity`)) ENGINE=InnoDB DEFAULT CHARSET=latin1; SET character_set_client = @saved_cs_client; </pre>
--	---

6. Configure iFuse to connect to the MySQL database, see *section 3.4.3, Database Setup*.
7. You should now be able to see the login page of iFuse when browsing to this machine's name or ip address

iFuse most important (program) files are listed below..:

<pre> -- R [...] -- TMP [...] -- application -- cache -- config -- constants.php '-- [...] -- controllers -- analyse.php '-- continues.php -- delete.php '-- download.php '-- fastaction.php '-- form.php '-- index.html </pre>	<pre> [...] -- login.php -- logout.php -- main.php -- open.php -- register.php '-- upload.php -- core -- errors -- helpers [...] -- hooks -- language [...] -- libraries -- cli.php '-- ifusefilevalidator.php -- ifuseloader.php </pre>	<pre> [...] -- index.html -- libraries [continued] -- r_handler.php -- sequenceloader.php -- svg_gene.php '-- template.php '-- userfiles.php -- models [...] -- third_party '-- views -- analyse.php -- continues.php -- default [...] -- delete.php '-- download.php </pre>	<pre> [...] -- form '-- files.php '-- sort.php '-- index.html '-- login.php '-- logout.php '-- main.php '-- register.php '-- upload.php '-- css [...] '-- img [...] '-- js [...] '-- manual [...] `-- system [...] </pre>
--	---	---	--

3.4 iFuse Configuration

3.4.1 Annotation tables

Every so many years a new version of the human genome is released. To use these new versions, you should download the table from UCSC tables and save it as `ucscgenes[hg-version].txt` in the `./R` directory. (e.g. `./R/ucscgeneshg19.txt`) The settings for a proper table are;

-  Clade: Mammal
-  Genome: Human
-  Assembly: [hg-version]
-  Group: Genes and Gene Prediction Tracks
-  Track: RefSeq Genes

Table: refGene
Region: Genome

clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)
 group: Genes and Gene Prediction Tracks track: RefSeq Genes add
 table: refGene describe table schema
 region: genome position chr21:33031597-33041570 lookup define regions

3.4.2 Sequence retrieval

iFuse gives the sequence of events on DNA, RNA and protein level. For this it requires access to the internet or access to files that contain the reference genome. iFuse can use three methods:

Description	
File retrieval	If you want to reduce the amount of downloads done by iFuse, download the reference genome (HG18, HG19, etc) and put it into iFuse/R/hg##/ http://hgdownload.cse.ucsc.edu/goldenPath/hg18/chromosomes/ http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/ a. Unpack the genomes b. Remove the header from each file c. Convert the genome into one big string per chromosome iFuse will automatically detect the presence of the genomic files and use it. <i>Recommended option!</i>
UCSC DAS retrieval (all events)	If in the PHP configuration (PHP.ini) allow_url_fopen is set on or 1, iFuse will download all the sequences it should have to the server. The sequences are retrieved from the UCSC DAS server. <i>Not recommended since it uses a lot of bandwidth, space and is limited to small upload files only.</i>
UCSC DAS retrieval (only active events)	If neither the PHP configuration (PHP.ini) allow_url_fopen is set on nor the reference genome is located on the server, iFuse will try to access the files via directly via the internet. The sequences are also retrieved from the UCSC DAS server. <i>Not recommended at all.</i>

3.4.3 Database setup

iFuse uses MySQL to manage its users and sessions. To connect to the database, you need to edit the database configuration file. This file is found in ./application/config/database.php.

```
// Find these lines and replace the values with the right values
$db['default']['hostname'] = 'localhost';
$db['default']['username'] = 'ifuse-user';
$db['default']['password'] = 'ifuse-password';
$db['default']['database'] = 'ifuse-database';
```

3.4.4 Time before deleting uploaded projects

Standard uploaded files are saved in a folder. The name of this folder specifies how long the project should be kept. This is standard five years from uploading the first file. This can be changed by modifying a constant located in `./application/config/constants.php`. Look for

```
define('USER_FILES_EXTRA_TIME_ON_SERVER', (60*60*24*365*5));
```

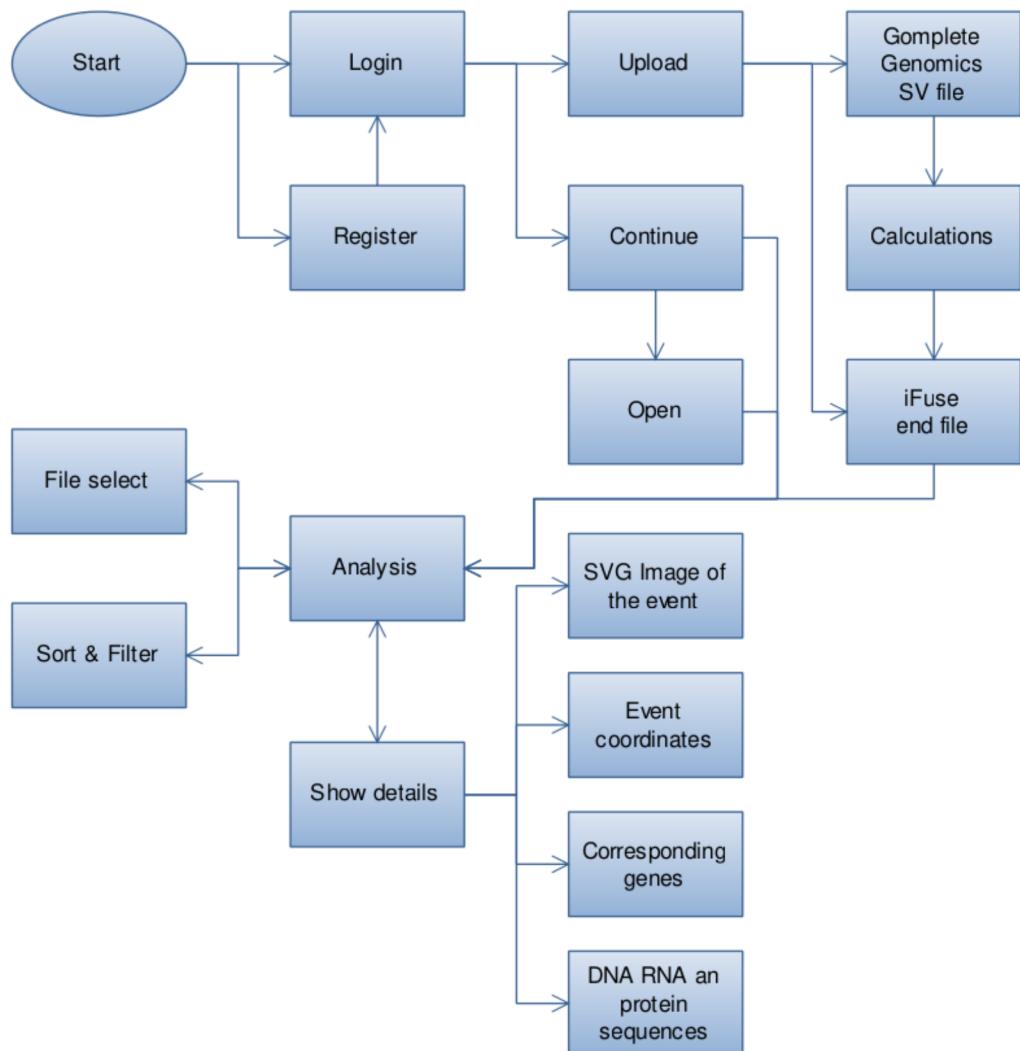
and change `(60*60*24*365*5)` to the amount of seconds you wish to keep the project.

3.4.5 User login time length

iFuse has a user management system so that new project and thus new uploads will be stored under one username. If a user logs in, he or she will be online until fifteen minutes of inactivity. The time of inactivity can be changed by editing `./application/config/config.php` and change the value of the following line. The value is in seconds.

```
$config['user_online_time'] = (60*60);
```

4. iFuse Program Structure



5. Using iFuse

5.1 Registration and login

To use iFuse, registration is required.

We require your name (or alias), a valid email address and a password. It

is as simple as that!

Register	
Username	<input type="text"/>
E-Mail address	<input type="text"/>
Password	<input type="password"/>
Retype password	<input type="password"/>
<input type="button" value="Register"/> <input type="button" value="Reset"/>	

Login	
Username	<input type="text"/>
Password	<input type="password"/>
Remember? <input type="radio"/> yes <input checked="" type="radio"/> no	
<input type="button" value="Login"/> <input type="button" value="Reset"/> REGISTER	

This enables the program to remember the files you have uploaded under the given username. So when logging in next time, it shows the uploaded files on the upload page

Data provided in the register form must be valid, e.g. no empty fields, email address must have a valid structure and both passwords must match with each other. As with all the other user input in iFuse, everything is being escaped.

After creating a new account, you are redirected to the login page. A newly created account can be used right away. A login requires you to sign in with the username and password you provided. A logged in session lasts 1 hour without a page request, unless at the login page `yes` was checked for `remember`. Then the session will last a week.

After logging in, you are taken to the upload page.

5.3 Upload or Open session

5.3.1 Upload and open files

iFuse requires you to upload your structural variant-file for it to work.

This can be done using the upload form at the main page. You need to specify a file to upload, its format, whether or not there is a *short* header in the file and what reference genome should be used to gain sequences from.

– NGS input file: [Browse...](#) ?

Select input format

True False

Input has a first-line header

Reference genome

SUBMITTING

The file is being submitted and analysed. Please wait and don't close your browser. This process usually takes 10 minutes per ~4000 events.'

File formats currently supported by iFuse are Complete Genomics Structural Variant files and the raw output of iFuse, see next paragraph.

The reference genome specified for the input is used to annotate the uploaded file. Also the sequences provided by iFuse are reference genome specific. Extra information about the uploaded files can be found under the question mark next to the file select button.

After pressing submit, the user should not break the connection with the server by refreshing or stopping the load since the analysis will be stopped then.

5.3.2 Filetypes

Complete genomics

iFuse can read and load Structural Variants files from Complete Genomics. The files are located in the ASM/SV directory of a Complete Genomics analysis.

The files located in the ASM directory describe and annotate the genome assembly with respect to the reference genome. The ASM directory contains the primary results of the assembly within several files. Each file includes a description of all loci where the assembled genome differs from the reference genome, but the files differ in format.

Small Variations and Annotations Files

The files in the ASM directory describe and annotate the sample's genome assembly with respect to the Reference genome, including:

-  Variations: The primary results of the assembly describing variant and non-variant alleles found.
-  Master Variations: Results of the assembly describing variant and non-variant alleles found, with annotation information in a one-line-per-locus format.
-  Genes: Annotated variants within known protein coding genes.
-  ncRNAs: Annotated variants within non-coding RNAs
-  Gene Variation Summary: Count of variants in known genes.
-  DB SNP: Variations in known dbSNP loci.

 Variations and Annotations Summary: Statistics of sequence data to assess genome quality.

Datafile format

The data files iFuse can read are located in the ASM/SV folder and are tab delimited. The first few rows contain file specific header information. These contain information about the run such as assembly id, time of generation and software version. There first character on the line is a dash ('#').

```
#ASSEMBLY_ID GS19240-ASM
#BUILD 1.7
#DBSNP_BUILD dbSNP build 129
#GENERATED_AT 2010-Jan-21 13:42:57.076648
#GENERATED_BY callannotate
#GENE_ANNOTATIONS NCBI build 36.3
#GENOME_REFERENCE NCBI build 36
#TYPE GENE-VAR-SUMMARY-REPORT
#VERSION 0.6

>column-headers
Data
```

The next line contains the headers of the data. Its first character is a bigger then sign ('>') followed by the columns, delimited by a tab. The column descriptions are given below.

After this there is the data, also tab delimited.

	Column Header	Description
1	JunctionId	Identifier for junction that this DNA nano Ball (DNB) alignment supports. Junction Ids are consistent across all junction files for a given assembly.
2	Slide	Identifier for the slide from which data for this DNB was obtained.
3	Lane	Identifier for the lane within the slide from which data for this DNB was obtained.
4	FileNumInLane	The file number of the reads file describing this DNB.
5	DnbOffsetInLaneFile	Record within data for the slide lane in reads_[SLIDE-LANE]_00X.tsv.bz2 that corresponds to this DNB
6	LeftDnbSide	Identifies the side of the DNB that was associated with the "left" (that is, earlier in the reference; on lower-numbered chromosome or with smaller offset within the same chromosome) side of the cluster. L if the left side of the DNB belongs to the left side of the cluster R if the right side of the DNB belongs to the left side of the cluster For the simple case of junctions that connect "+" strand sequence to "+" strand sequence, the left side of DNB belongs to the left side of the cluster if the DNB was produced from the "+" strand of the genomic DNA.
7	LeftStrand	The strand of the half-DNB, "+" or "-", expressed relative to the reference genome.
8	LeftChromosome	Left chromosome name in text: chr1, chr2,..., chr22, chrX, chrY. The mitochondrion is represented as chrM, though this may be absent from SV analyses. The pseudoautosomal regions within the sex chromosomes X and Y are reported at their coordinates on chromosome X.
9	LeftOffsetInReference	The chromosomal position on the reference genome at which the half-DNB starts (as seen on the "+" strand).

10	LeftAlignment	The alignment of the half-DNB to the left section of junction, provided in an extended CIGAR format (see “Alignment CIGAR Format”).
11	LeftMappingQuality	A Phred-like encoding of the probability that this half-DNB mapping is incorrect, encoded as a single character with ASCII-33. The Phred score is obtained by subtracting 33 from the ASCII code of the character.
12	RightDnbSide	Identifies the side of the DNB that was associated with the right side of the cluster.
13	RightStrand	The strand of the half-DNB, “+” or “-”, expressed relative to the reference genome.
14	RightChromosome	Left chromosome name in text: chr1, chr2,..., chr22, chrX, chrY. The mitochondrion is represented as chrM, though this may be absent from SV analyses. The pseudoautosomal regions within the sex chromosomes X and Y are reported at their coordinates on chromosome X.
15	RightOffsetInReference	The chromosomal position on the reference genome at which the half-DNB starts (as seen on the “+” strand).
16	RightAlignment	The alignment of the half-DNB to the right section of junction, provided in an extended CIGAR format (see “Alignment CIGAR Format”).
17	RightMappingQuality	A Phred-like encoding of the probability that this half-DNB mapping is incorrect, encoded as a single character with ASCII-33. The mapping quality is related to the existence of alternate mappings; the Phred score is obtained by subtracting 33 from the ASCII code of the character.
18	EstimatedMateDistance	Estimate of the distance between the left and right arm of the DNB in the assayed genome, taking the junction into account.
19	Sequence	Sequence of the DNB arm bases in the DNB order (same as in the reads_[SLIDE-LANE]_00X.tsv.bz2 file).
20	Scores	Phred-like error scores for DNB bases in the DNB order, not separated (same as in the reads_[SLIDE-LANE]_00X.tsv.bz2 file).

Further specifications for the Complete Genomics data files can be downloaded from;
<http://www.completemicrogenomes.com/customer-support/documentation/100357139.html>

iFuse Raw file

A file processed by iFuse can be uploaded contains some old columns from the input file. But most are new and calculated. The file contains a first line with the header of all the columns (tab-separated) after which the events are described (one per line).

	Column Header	Description
1	Junction CG.ID	A unique ID given by Complete Genomics to this junction/event.
2	Related Junctions	Id for junctions that are within 100bp of other junctions
3	Associated Junctions	Id for junctions that land within the same gene
4	Shared Genes	Id for junctions that have the same genes on either the left or right side
5	Gene Mismatch	Junctions with different genes on either the left or right side
6	Single Event	Event type. E.g. Deletion, inversion, interchromosomal, translocation.
7	Fusion Gene	Whether the junction has genes on the same strand ('same direction').
8	Left Position in CDS	Position of the junction of the left strand is in a coding region
9	Right Position in CDS	Position of the junction of the right strand is in a coding region
10	Left Position in Exon	Position of the junction of the left strand is in an exon
11	Right Position in Exon	Position of the junction of the right strand is in an exon
12	Gene Left.name2	Alias for the left gene name
13	Gene Right.name2	Alias for the right gene name
14	Gene Left.name	Accession number for the left gene
15	Gene Left.chrom	Chromosome id of the gene left of the junction
16	Gene Left.strand	Strand of the gene left of the junction
17	Gene Left.txStart	Transcription start of the gene left of the junction
18	Gene Left.txEnd	Transcription end of the gene left of the junction
19	Gene Left.cdsStart	Coding region start of the gene left of the junction
20	Gene Left.cdsEnd	Coding region end of the gene left of the junction
21	Gene Left.exonStarts	Start positions of the exons in the gene left of the junction
22	Gene Left.exonEnds	End positions of the exons in the gene left of the junction
23	Gene Right.name	Accession number for the right gene
24	Gene Right.chrom	Chromosome id of the gene right of the junction
25	Gene Right.strand	Strand of the gene right of the junction
26	Gene Right.txStart	Transcription start of the gene right of the junction
27	Gene Right.txEnd	Transcription end of the gene right of the junction
28	Gene Right.cdsStart	Coding region start of the gene right of the junction
29	Gene Right.cdsEnd	Coding region end of the gene right of the junction
30	Gene Right.exonStarts	Start positions of the exons in the gene right of the junction
31	Gene Right.exonEnds	End positions of the exons in the gene right of the junction
32	Junction LeftChr	Chromosome id of the DNA sequence left of the junction
33	Junction LeftStrand	Strand of the DNA sequence left of the junction
34	Junction LeftPosition	Position breakpoint of the DNA sequence left of the junction
35	Junction LeftStart	Junction site start of the DNA sequence left of the junction
36	Junction LeftEnd	Junction site end of the DNA sequence left of the junction
37	Junction RightChr	Chromosome id of the DNA sequence right of the junction
38	Junction RightStrand	Strand of the DNA sequence right of the junction
39	Junction RightPosition	Position breakpoint of the DNA sequence right of the junction
40	Junction RightStart	Junction site start of the DNA sequence right of the junction
41	Junction RightEnd	Junction site end of the DNA sequence right of the junction
42	Junction LeftLength	Length of the junction site, left of the actual junction
43	Junction RightLength	Length of the junction site, right of the actual junction

44	Junction TransitionLength	Length of the transition sequence between the left and right part of the junction
45	Junction TransitionSequence	Sequence between the left and right part of the junction.
46	Junction AssembledSequence	Assembled sequence of the junction

FusionMap

Output generated by FusionMap can also be visualized in iFuse.

	Column Header
1	Fusionid
2	UnmappedDatasetP2SimulatedReads_from_tophat.fastq.UniqueCuttingPositionCount
3	UnmappedDatasetP2SimulatedReads_from_tophat.fastq.SeedCount
4	UnmappedDatasetP2SimulatedReads_from_tophat.fastq.RescuedCount
5	Strand
6	Chromosome1
7	Position1
8	Chromosome2
9	Position2
10	KnownGene1
11	KnownTranscript1
12	KnownExonNumber1
13	KnownTranscriptStrand1
14	KnownGene2
15	KnownTranscript2
16	KnownExonNumber2
17	KnownTranscriptStrand2
18	FusionJunctionSequence
19	SplicePattern

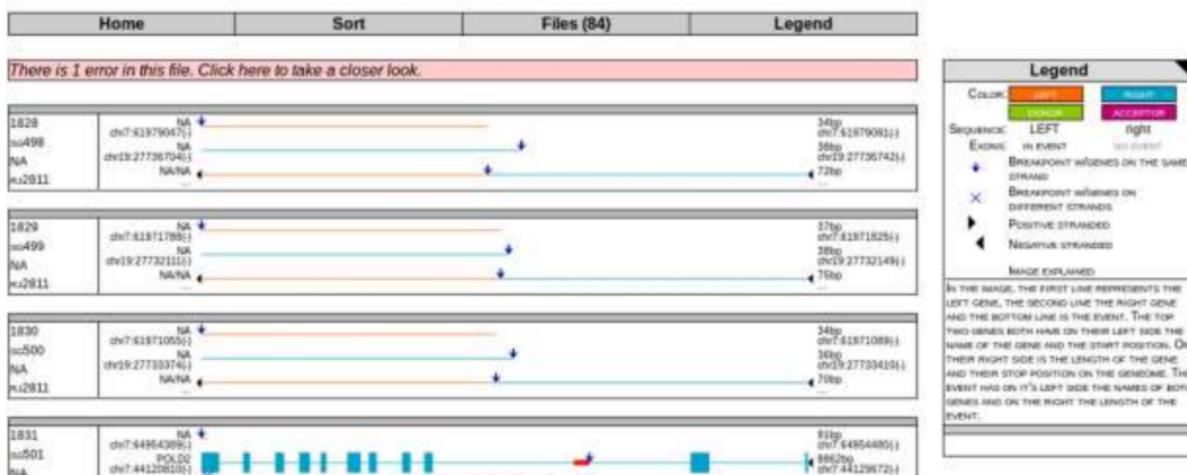
For more information, see: <http://www.omicsoft.com/fusionmap>.

5.4 Analysis page

The analysis page shows all the events per file. Events can be sorted, filtered and details can be shown.

Per page, only one file can be shown and only 10 events per page by default. Without manually sorting, the events are sorted on gene fusions or none.

iFuse structural variants from NGS data



5.4.1 Menu

The menu bar at the top has four options;

Home

Resets the filter and sort pagination options. If that is done, clicking one more time will redirect the user to the start/upload page.

Sort

When clicking the sort-menu option, a box will appear on top of your page containing a form to sort the page.

The figure shows a screenshot of the 'Sort' dialog box. It has a list of columns on the left and two sorting columns on the right. The top column is 'By Event ID' and the bottom column is 'By Start'. The 'Start' column has a dropdown menu with options like 'Asc' and 'Desc'.

The left box contains columns that are not being used to sort on, the right does have columns that are being used to sort on. The columns in the right box are in order of sorting, the top column is used to sort on first, the lowest column is used as last. This means when you primary want to sort on column A, this column must be at the bottom of the list. If between results with the same value you would like to sort on column B, this column must be above column A.

Ordering, adding and deleting columns can be done with the arrow files between the fields. Select a column in the left box and click the arrow to the right to add this column to the list. To re move it, click it in the right box and click on the arrow to the right. To reorder the columns in the right box, click the column to move and the arrows for up or down.

After making the sort order, submit the form to reorganize the analysis results. Without any form of sorting, fusion genes are shown first.

Files

When clicking the files-menu option, a box will appear on top of your page containing a list with all the files in your current session, sorted by upload time and part number.

#	Name	Options	Upload Time
1	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_Trap (part 2)	Format: CSV Header: None Reference: None Part: None	33 hours ago
2	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_2.mer (part 3)	Format: CSV Header: None Reference: None Part: None	33 hours ago
3	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_3.mer (part 3)	Format: CSV Header: None Reference: None Part: None	33 hours ago
4	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_4.mer (part 4)	Format: CSV Header: None Reference: None Part: None	33 hours ago
5	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_5.mer (part 5)	Format: CSV Header: None Reference: None Part: None	33 hours ago
6	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_6.mer (part 6)	Format: CSV Header: None Reference: None Part: None	33 hours ago
7	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_7.mer (part 7)	Format: CSV Header: None Reference: None Part: None	33 hours ago
8	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_8.mer (part 8)	Format: CSV Header: None Reference: None Part: None	33 hours ago
9	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_9.mer (part 9)	Format: CSV Header: None Reference: None Part: None	33 hours ago
10	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_10.mer (part 10)	Format: CSV Header: None Reference: None Part: None	33 hours ago
11	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_11.mer (part 11)	Format: CSV Header: None Reference: None Part: None	33 hours ago
12	HCC2318_A02_allAnnotations-HCC2318- H-250-97-A02-N1_12.mer (part 12)	Format: CSV Header: None Reference: None Part: None	33 hours ago

The first column contains the order id, followed by a column with the original name of the file and part number. The files are not saved with these names so that the user can upload files with the same name, even though this is not recommended. The third column contains the options given during the upload process; e.g. format of the file, whether or not the files contains a header and the reference genome being used. The final column contains shows the user when the files was uploaded.

The current active file the current analysis is based upon has a soft green background. When clicking right mouse button upon a row, a menu will appear. Using that menu, the file the row is for, can be deleted or activated to use at the analyze page. By refreshing the page, the results of a newly activated file will be visible.

Files can be downloaded in its right mouse button menu.

Legend

Clicking the legend-menu option allows the user to hide or show the Legend temporary on the right side of the analysis page. The black triangle on the legend header can permanently show or hide the Legend.

The legend is designed to help the user to understand the graphs on the analysis page more easily and to show help messages when hovering with the mouse over certain components of the page.

The legend consists of three parts, the top, middle and the bottom. See section 5.4.4, *Legend* for detailed information about the legend.

5.4.2 Error bar

There is 1 error in this file. Click here to take a closer look.

The error bar shows the amount of formatting errors specific for the current file. Examples of errors can be too few columns on a line (e.g. *Line 250 is not an array or does not have the same column count as the header (1!=46)*) or a specific column can't be validated (e.g. *Line 101 column 2 (Associated.Junctions) cannot be validated using REGEX('/^(aj([0-9]+)|NA)\$/')*).

5.4.3 Event Overview

2159	EXOC2 chr6:430137(-)		207972bp chr6:638109(-)
sg671	MANIA1 chr6:119540965(-)		171660bp chr6:119712625(-)
NA	EXOC2/MANIA1 ...		186220bp ...
pu508			

In minimalistic view, a event is described as above. The left cell of the table has from top to bottom the event ID, the shared gene id, the associated junction id and the related junction id. (described in section [5.3.2, filetypes](#).



The second cell contains an image of the sequences on the left and right side of the junction and the fusion. The first row visualizes the original left sequence of the junction. Top-left of the image contains the name of the gene on the left side of the junction, if applicable. The top right contains the length of the sequence visualized. The bottom left and bottom right show the coordinates of the beginning and end of the visualization, including the strand.

The image itself shows the introns and exons on the sequence visualized. The promoter is visualized on the outer left or outer right of the image as an arrow or triangle. The breakpoint has an arrow hovering above it, showing the precise position of the base that connects to the right part of the event. The junction site has a static length and is shown as a red block over the sequence.

The color of the visualization is explained in section [5.4.4, Legend](#).

The following row is visualization for the right side of the sequence, but has the same construction.



The bottom visualization shows how the event is constructed out of the previous two sequences. Top left of the image is the name of the two genes on the left and right side of the event, respectively if applicable. The top right shows the total length of the event. The left part of the visualization is the left part of the event and left of it shows the direction the promoter is being read. The same goes for the right part of the visualization. The junction can be visualized either with an arrow hovering above it identifying the genes on the same strand or a cross if it is not.

5.4.4 Legend

As previously said, the legend consists of a top, middle and bottom part. The top is made specific for the visualization of the event.

The colors are equivalent to the colors of an event. When there is no specific donor/acceptor gene, the left part of the event will be orange and the right part will be blue. When there is a promoter-donor gene, that gene will be green and the non-promoter-region will be purple.

The sequences in the details of an event, see section 5.4.5, *Details*, can be either uppercase or lower case, depending on the side it is on (e.g. left or right).

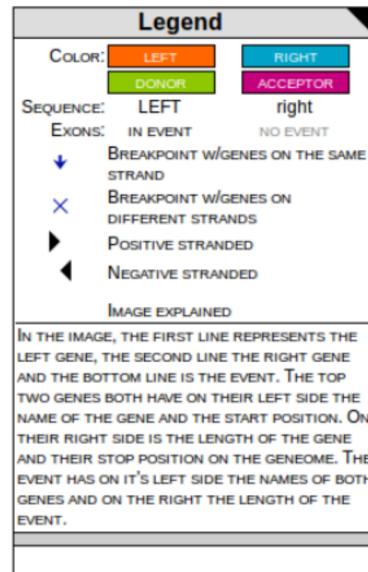
The exons in the details of an event, see section 5.4.5, *Details*, can have a black or gray font. Black coordinates represent exons that are in the event while gray coordinates are not.

The arrow and cross are shown above the base pair that represents the breakpoint. The arrow is shown when the genes on both sides of the junction are on the same strand after joining. The cross is shown when it is not.

The arrows to the left and right are a way to show where the promoter is and on what strand the gene is.

The middle part of the legend is a short description of how the picture is constructed. It is a short description equal to section 5.4.3, *Event Overview*.

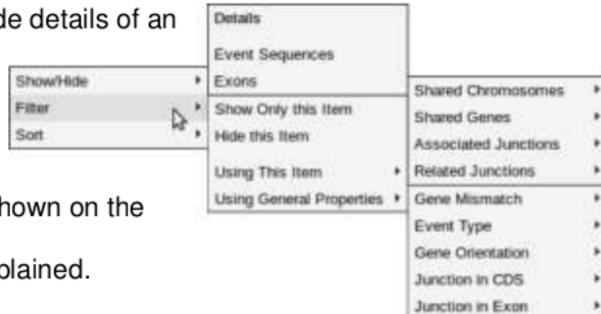
The bottom part of the legend shows help messages when hovering some parts of the page and thus can be seen as a status-bar.



5.4.5 Event menu

The right mouse button can be used to show or hide details of an event. It can also be used to filter or sort events with.

Right clicking on an event shows a submenu, as shown on the right. Below most options are written down and explained.



Show/Hide

- **Details**
Show or hide all details from the right mouse clicked event. See also section 5.4.6, *Details*.
- **Event Sequences**
After showing the event details, hide or show the Sequences section
- **Exons**
After showing the event details, hide or show the Exon section. *Note:* Exons are only present if there is a gene on one or both sides of the event.

Filter

Filter out uninteresting events

- **Show Only this Item**
Show all the details from this event on a new page.
- **Hide This Item**
Hide this event from the current browser window. After clicking the Home-button or reentering the URL, this will be undone.
- **(Filter...) Using this Item**
Filter all results; use properties related to this event. E.g. if this event has a associated junctions id 'aj001', and you filter on associated junctions > only associated junctions. Only associated junctions are shown.
- **Using General Properties**
Filter all results; use general properties. General properties are columns with limited values, e.g. yes/no.

Sort

Sort events by specific columns. See section 5.4.1 for details on the sort function.

For detailed descriptions of the columns, see section 5.3.2, *filetypes*.

5.4.6 Details

Underneath the image of the event, a lot more details can be shown via the right mouse button menu or by viewing only one event.

Left	
Gene Name	EXOC2 (NM_018303)
Coding Sequence	chr6:431670-582818(-)
Transcript	chr6:430137-638109(-)
Junction Site	chr6:469569-469614(-)
Junction Position	chr6:469569

The **first section**, right underneath the visualization of the event, contains details of the left and right part of the event. The color on the title bar connects this information to the image.

The information given in this section is the gene name, including the accession number, and the coordinates of the coding sequence, transcript, junction site and junction within the genome.

The **next section** contains the ensemble sequences.

These are derived from the reference genome.

Event Se	
DNA	>EXOC2_- chr6:469569,6381094MAN1A1_- chr6:119540965,119558645 DNA(+) GAAGTGCCTGGCCTGGCGGCCGCGGAGUCUUCGGCAGAGACCUAECGGGCCUGUCUGUC [...]
RNA	>EXOC2_- chr6:469569,6381094MAN1A1_- chr6:119540965,119558645 RNA(+) GAAGUGCCGGCGCUUGCGGCCGCGGAGUCUUCGGCAGAGACCUAECGGGCCUGUCUGUC [...]
Predicted Protein	>EXOC2_- chr6:469569,6381094MAN1A1_-chr6:119540965,119558645_mR MSRSRQPPPLVLTGIPNEGLPWTKVITRGENLGTGPTDLIGLTICGHNCLLTAEwMSASKIVCRV PLSLRPANPGLTEKEKSFSQKOLEMFLHGMISADFTSENFSAAwYLIENHSNTSFEQLKMAVTN [...]

The DNA sequence contains either the sequence that resembles the event. This is constructed of the sequence of the left gene and the right gene. The left sequence is uppercase and the right sequence is lowercase. The right column contains a shortened version of this sequence containing 500bp upstream and 500bp downstream of the sequence.

The RNA sequence is derived from the DNA sequence and the coordinates of the promoter-donor-gene. If there is no promoter donating gene, no RNA can be given. Also when the coding sequence start codon is not in the sequence of the donating gene, no sequence can be given. The left sequence is uppercase and the right sequence is lowercase. The right column contains a shortened version of this sequence containing 500bp upstream and 500bp downstream of the sequence.

The protein sequence is derived from the RNA and the transcript coordinates of the donor region. Only the sequence between the start and stop codon is shown.

The third and last section contains the coordinates of the exons within the gene. Exons in the event are black and those that are not are gray, as described in section 5.4.4, *legend*. If the left and right part of the event both don't have a gene, this section is not shown.

28 Exons	
1	chr6:638109-638018(-)
2	chr6:582061-582700(-)
3	chr6:578117-577940(-)
27	chr6:434038-433878(-)
28	chr6:431764-430137(-)

6. Quick Tutorial, Finding fusion genes.

A test file can be downloaded from <http://bioinf-ifuse.eurasmusmc.nl/TestFile.tsv>¹. This is a structural variants file from cell line HCC1187, using HG18 as reference genome. Details can be found at the website of Complete Genomics.

1. Go to <http://bioinf-ifuse.eurasmusmc.nl/> and Browse for the file
2. Leave the Input format, first-line header and Reference genome fields as it is. And click submit
You'll have to wait for a minute or so
3. Right click on an event, Filter > Using General Properties > Gene Mismatch > Yes to view all the events with different genes on either side of the junction.

¹ Originally from Complete genomics at

ftp://ftp2.completegenomics.com/Cancer_pairs/ASM_Build36_2.0.2/HCC1187/GS00258-DNA_E01/

4. Right click on an event, Filter > Using General Properties > Gene Orientation > Both on same strand to view all the events that have gene sequences on the same strand and thus of which genes are actually fusing.
These are fusion genes
5. Right click on a fusion gene and click Show/Hide to view all details or Filter > Show Only this Item to have that event only on this page. (To go back to the list, click back in the browsers menu.)
Blast some sequences if you would like to test whether or not the outcome is correct.
6. To see all the events on one gene, right click on it and select Filter > Using This Item > Shared Genes > Only These Shared Genes. (To go back to the list, click back in the browsers menu.)
This might give a good idea of all the breakpoints in a gene.

7. Frequently Asked Questions

1. Where can I find iFuse?

You can find iFuse at; <http://bioinf-ifuse.erasmusmc.nl/>. This is the place where you can login and thereafter see you previously uploaded files.

2. I do not have an account, where can I create one?

You can create an account by following the 'Register' link on the front page or by going to <http://bioinf-ifuse.erasmsmc.nl/index.php/register>. You need to provide a valid and unique email address and username. You also need to provide a password. After creating an account, you can use it right away to login to iFuse.

3. How do I logout from iFuse?

This is not necessary but if you want to, go to <http://bioinf-ifuse.erasmusmc.nl/index.php/logout>.

4. What are you filetypes I can upload and where can I find them?

A *Complete Genomics Event File* is a file provided by Complete Genomics for each sequenced sample. The files are located in the ASM/SV directory of a Complete Genomics analysis and have the word "event" in the filename. The specifics for this file are shown in paragraph 5.3.2: *filetypes*.

The *Tab Delimited File* is derived from this file and generated by the R-script used by iFuse. When you don't want to wait between multiple uploads, it is helpful to download the R-script and run it for multiple files at the same time. This reduces time to wait significantly, since you can run the script overnight. The specifics for this filetype are also shown in paragraph 5.3.2: *filetypes*.

5. That is meant with the different options at the upload page?

See paragraph 5.3.1

6. When I upload a file, I'm redirected to the upload page without seeing the uploaded file

This means that you uploaded a file that is not understood by iFuse. Please check the format of your file.

7. When I upload a file, I get an error saying that the file can not be read

This means that you uploaded a file that is not understood by iFuse. Please check the format of your file.

8. After uploading, there seems to be an error in the file according to the error bar.

Once clicking on the error bar, you'll see the error. It might occur that the column count is not the same as the header. If the column count is one, the line might be empty. If this is not the case, you might be able to find the line in your uploaded file using the line number.

It might also be the case a column cannot be validated by iFuse. iFuse does not show the contents of the cell that cannot be validated but this cell can be found using the provided line number and column number/name. Also the regular expression that is used to validate the column is provided as extra technical information.

9. According to the analysis page, there are more files than I count in the file page?

The analysis page shows the amount of files iFuse has created after upload (files are split per 500 events). Also deleted files are not subtracted from this number.

10. I would like to not see the legend, permanently. How do I do this?

Click the black corner of the Legend. As long as your browser session continues, the Legend will not be shown, unless you click the 'Legend' option of the Top Menu.

11. What is shown in the analysis page?

Please see paragraph *5.4.3 Event Overview* and paragraph *5.4.6 Event Details*.

12. How can I navigate through my results

Use a combination of filtering (*5.4.5 Event Menu*), sorting (*5.4.1 Sort* and [right mouse button] *5.4.5 Event Menu*), selecting files (*5.4.1 Files*)and pagiation.

13. I would like to show details of this event, how do I do this?

Choose [Right mouse button] > *details* > *show/hide*

14. I would like to hide an event, how do I do this?

Choose [Right mouse button] > *filter* > *hide this item*. The item will not be removed from the dataset, though.

15.