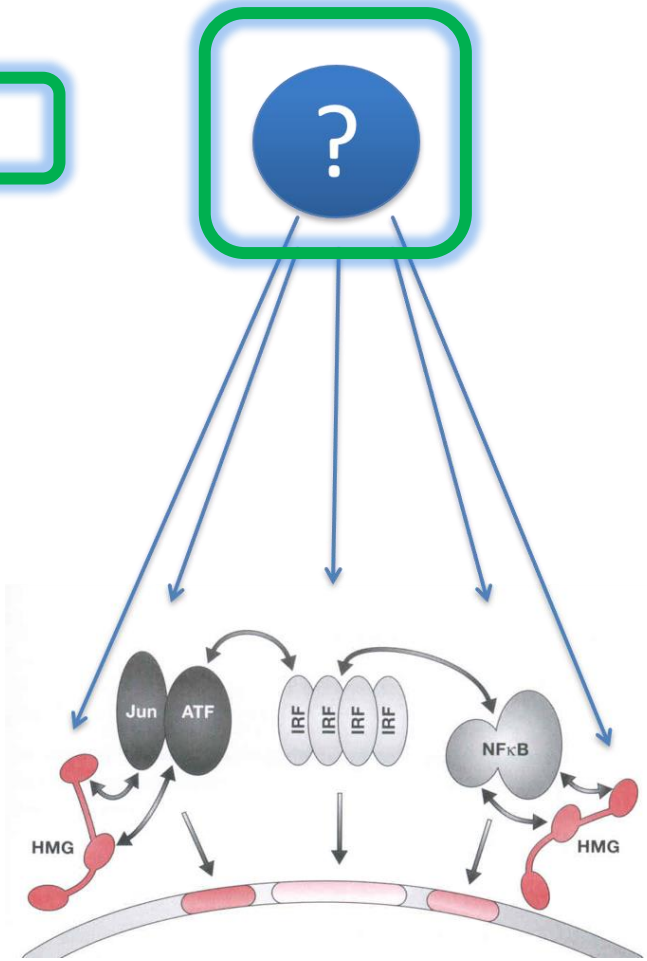
Targets

Day 7

info@genexplain.com

Upstream analysis – From motifs in promoters to drug targets

Drug target ?



Master regulator

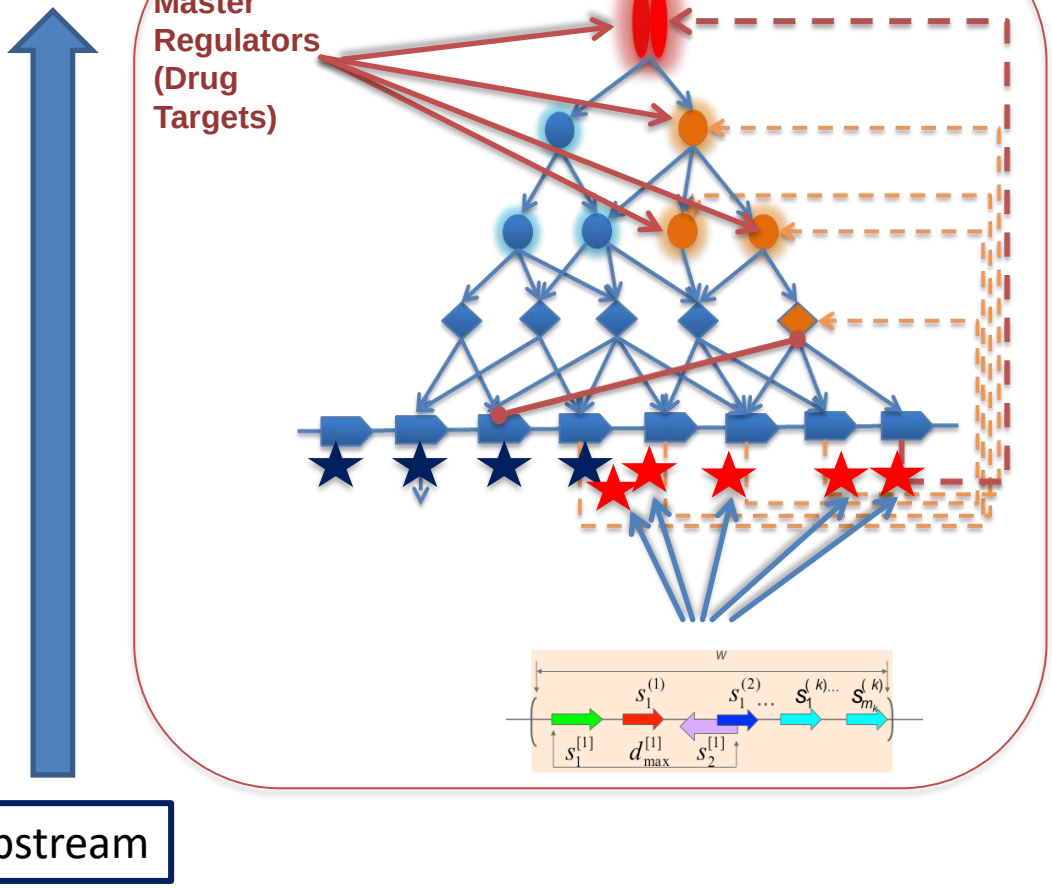
Promoter motifs

Upstream Analysis Concept

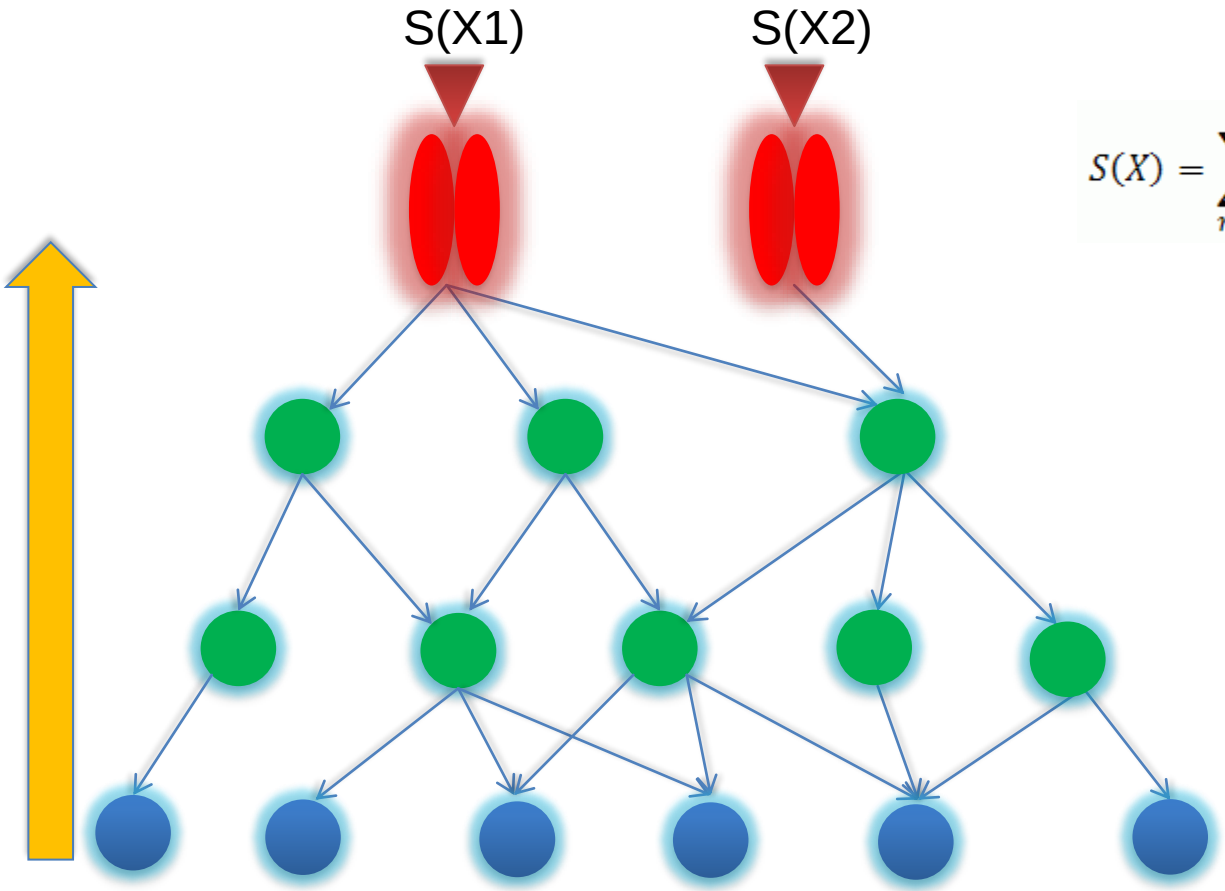
The concept of upstream analysis includes two steps, promoter analysis and network analysis.

Promoter analysis is done as TFBS search using proprietary F-Match and CMA algorithms together with the unique collection of TRANSFAC® matrices.

Network analysis is done with a proprietary graph-analyzing algorithm and the unique collection of TRANSPATH® reactions. As a result, the master regulators in the networks are identified.



Search for master regulators



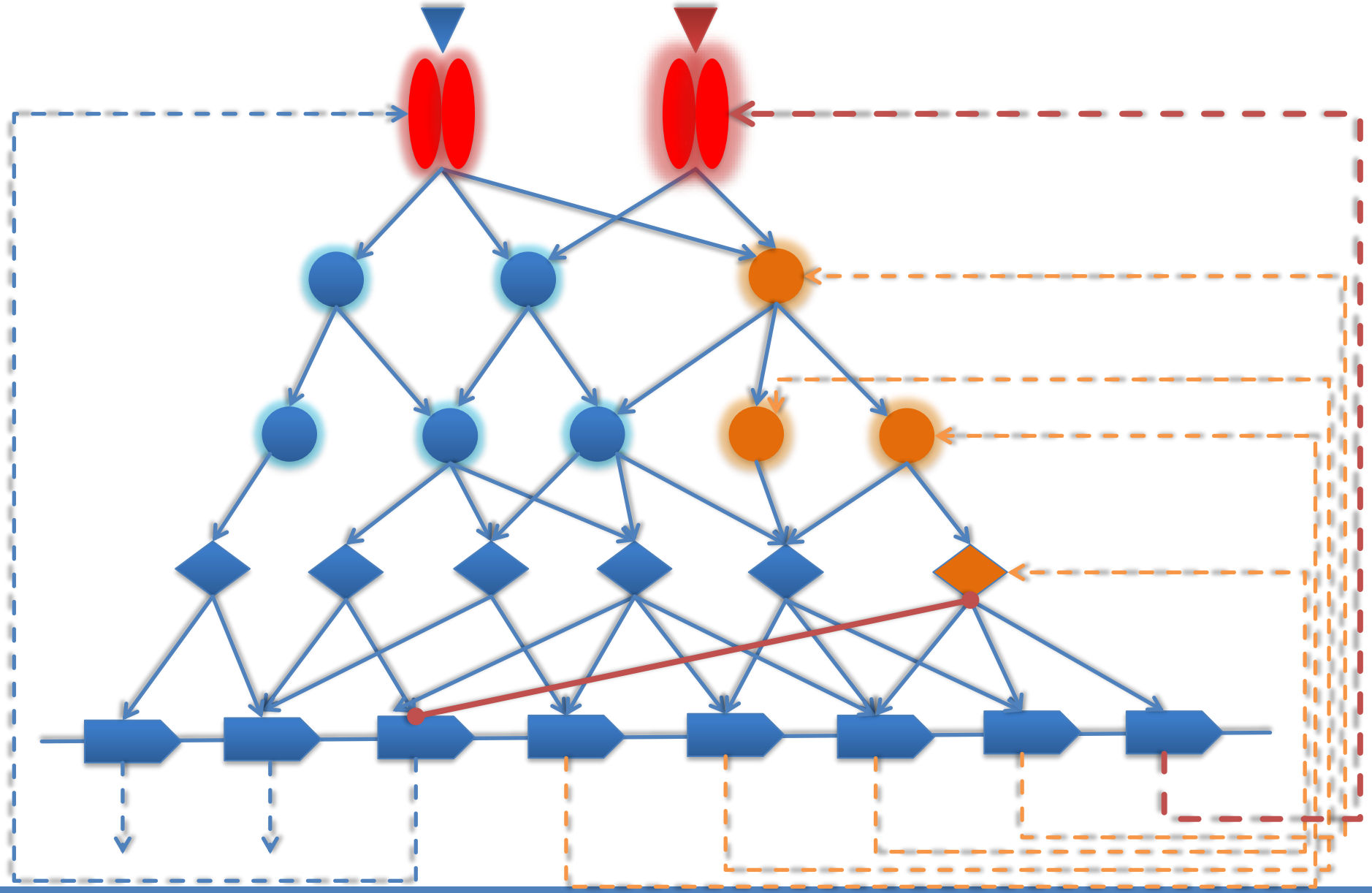
$$S(X) = \sum_{r=1}^{R} \frac{M(X,r)}{M_{max}(r)} \cdot \frac{1}{1 + pN(X,r)/N_{max}(r)}$$

Where:
R - Max radius (input parameter)
p - Penalty (input parameter)
N(X,r) - total number of molecules reachable from key molecule X within the radius r.
N_{max}(r) - maximal value of *N(X,r)* over all key molecules X found for this radius.
M(X,r) - sum of *w(X)* for all hits reachable from key molecule X within the radius r, where *w(X)* - weight of hit X.
M_{max}(r) - maximal value of *M(X,r)* over all key molecules X found for this radius.

Kel, A., Voss, N., Jauregui, R., Kel-Margoulis, O. and Wingender, E.: Beyond microarrays: Find key transcription factors controlling signal transduction pathways BMC Bioinformatics 7(Suppl. 2), S13 (2006). Link

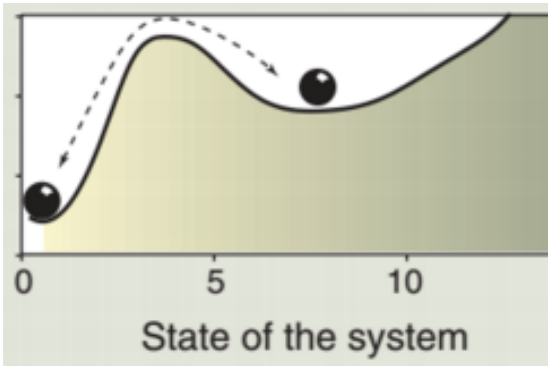
info@genexplain.com

Walking pathways



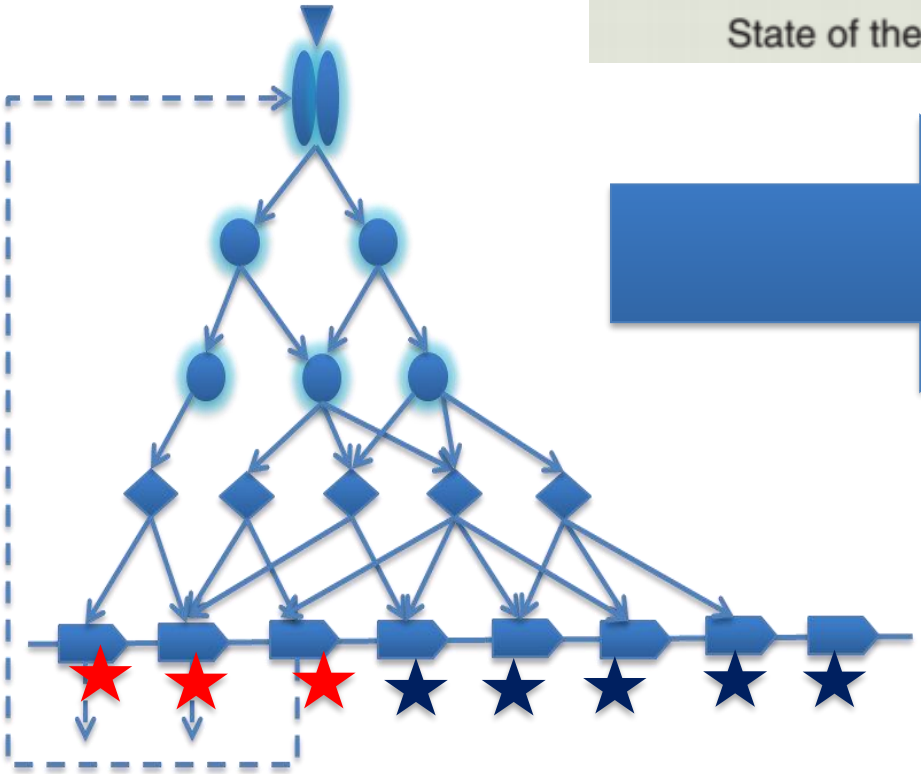
info@genexplain.com

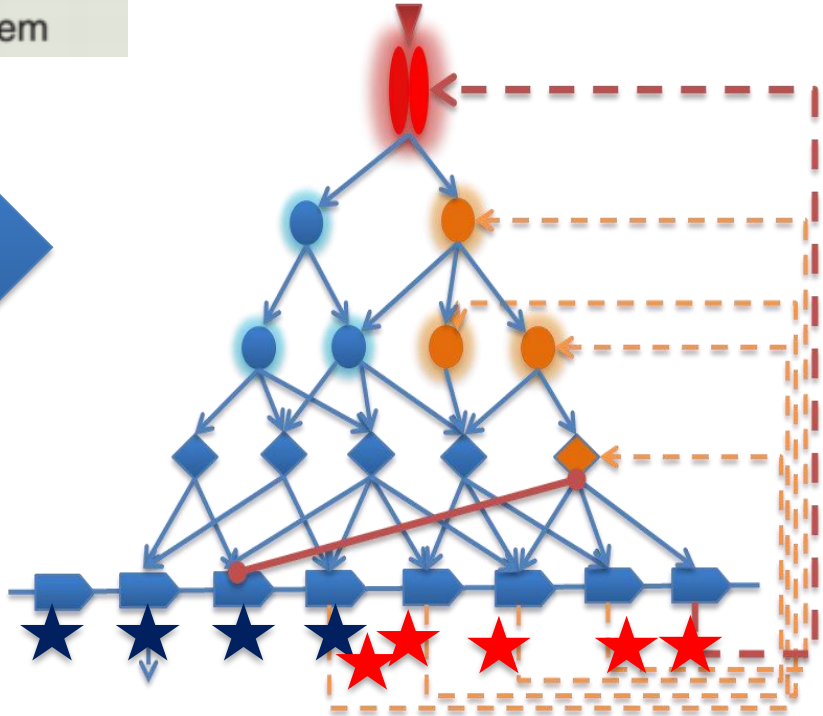
„Walking pathways“

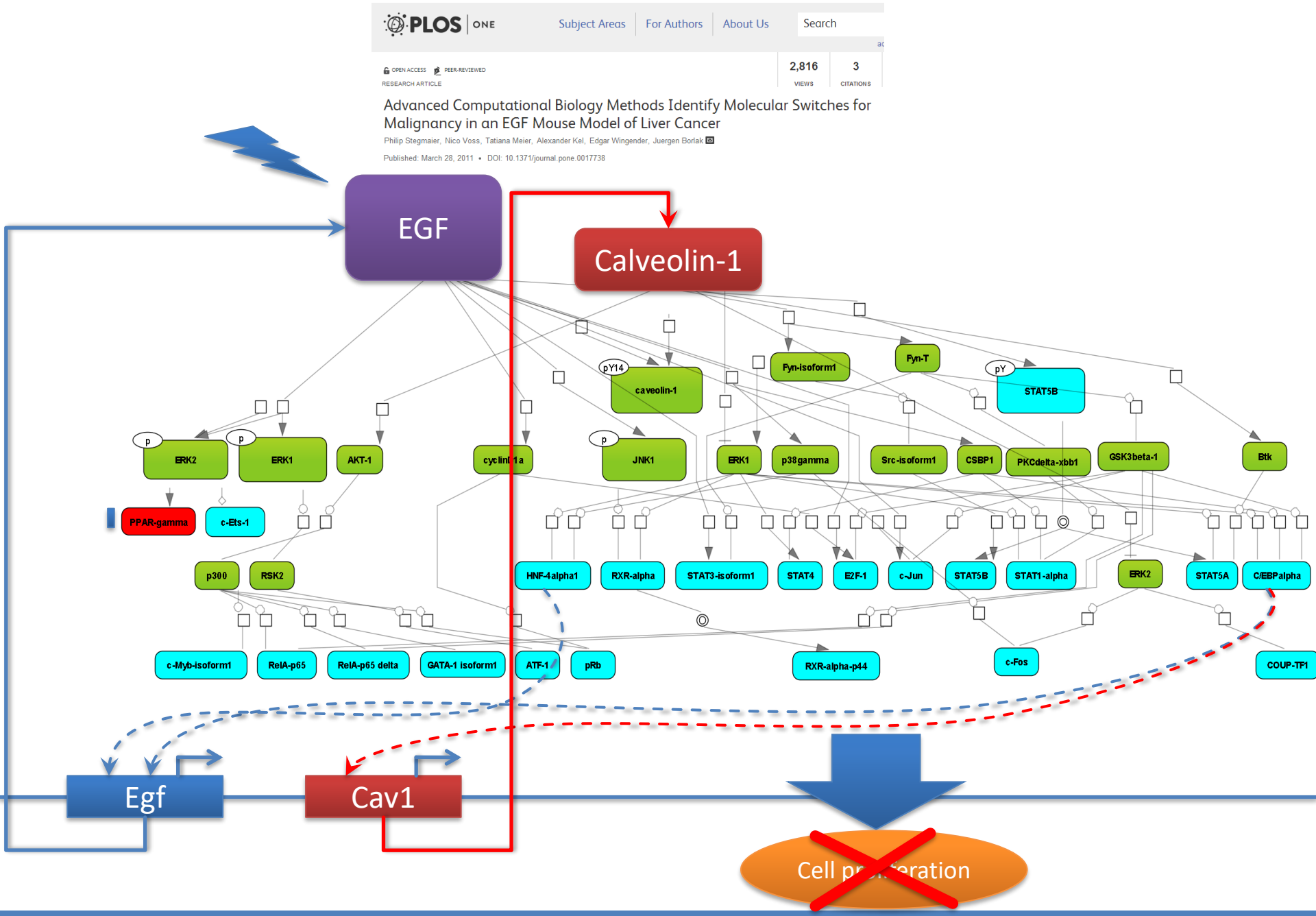


Healthy

Disease






PLOS ONE
Subject Areas
For Authors
About Us
Search
OPEN ACCESS
PEER-REVIEWED
RESEARCH ARTICLE
2,816
VIEWS
3
CITATIONS
Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer
Philip Stegmaier, Nico Voss, Tatiana Meier, Alexander Kel, Edgar Wingender, Juergen Borlak
Published: March 28, 2011 • DOI: 10.1371/journal.pone.0017738
EGF
Calveolin-1
pY14
caveolin-1
Fyn-isoform1
Fyn-T
pY
STAT5B
p
ERK2
p
ERK1
AKT-1
cyclin D1a
p
JNK1
ERK1
p38gamma
Src-isoform1
CSBP1
PKCdelta-xbb1
GSK3beta-1
Btk
PPAR-gamma
c-Ets-1
p300
RSK2
HNF-4alpha1
RXR-alpha
STAT3-isoform1
STAT4
E2F-1
c-Jun
STAT5B
STAT1-alpha
ERK2
STAT5A
C/EBPalpha
c-Myb-isoform1
RelA-p65
RelA-p65 delta
GATA-1 isoform1
ATF-1
pRb
RXR-alpha-p44
c-Fos
COUP-TF1
Egf
Cav1
Cell proliferation

PLOS ONE | Subject Areas | For Authors | About Us | Search

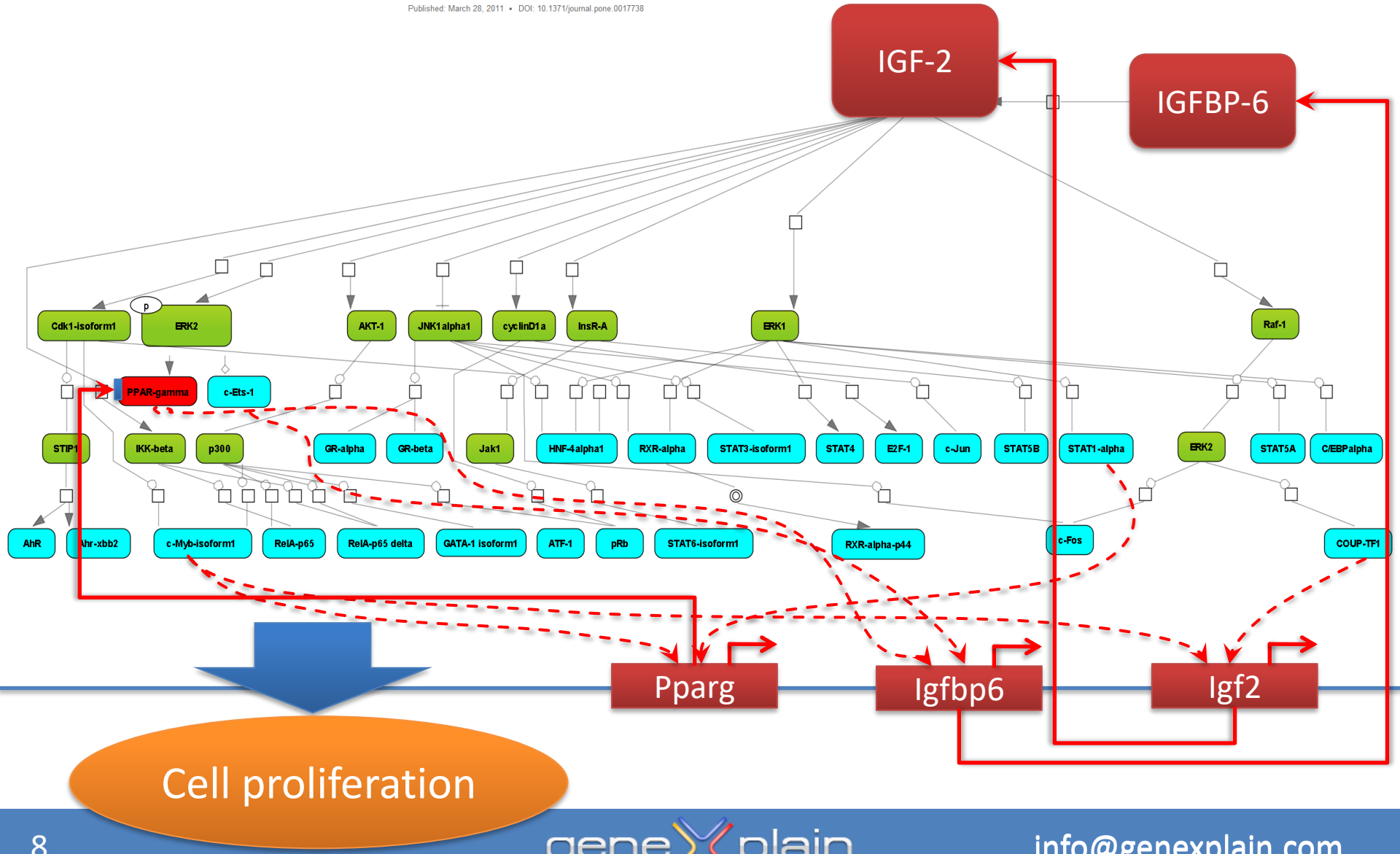
OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

2,816 VIEWS 3 CITATIONS

Advanced Computational Biology Methods Identify Molecular Switches for Malignancy in an EGF Mouse Model of Liver Cancer

Philip Stegmaier, Nico Voss, Tatiana Meier, Alexander Kel, Edgar Wingender, Juergen Borlak

Published: March 28, 2011 • DOI: 10.1371/journal.pone.0017738



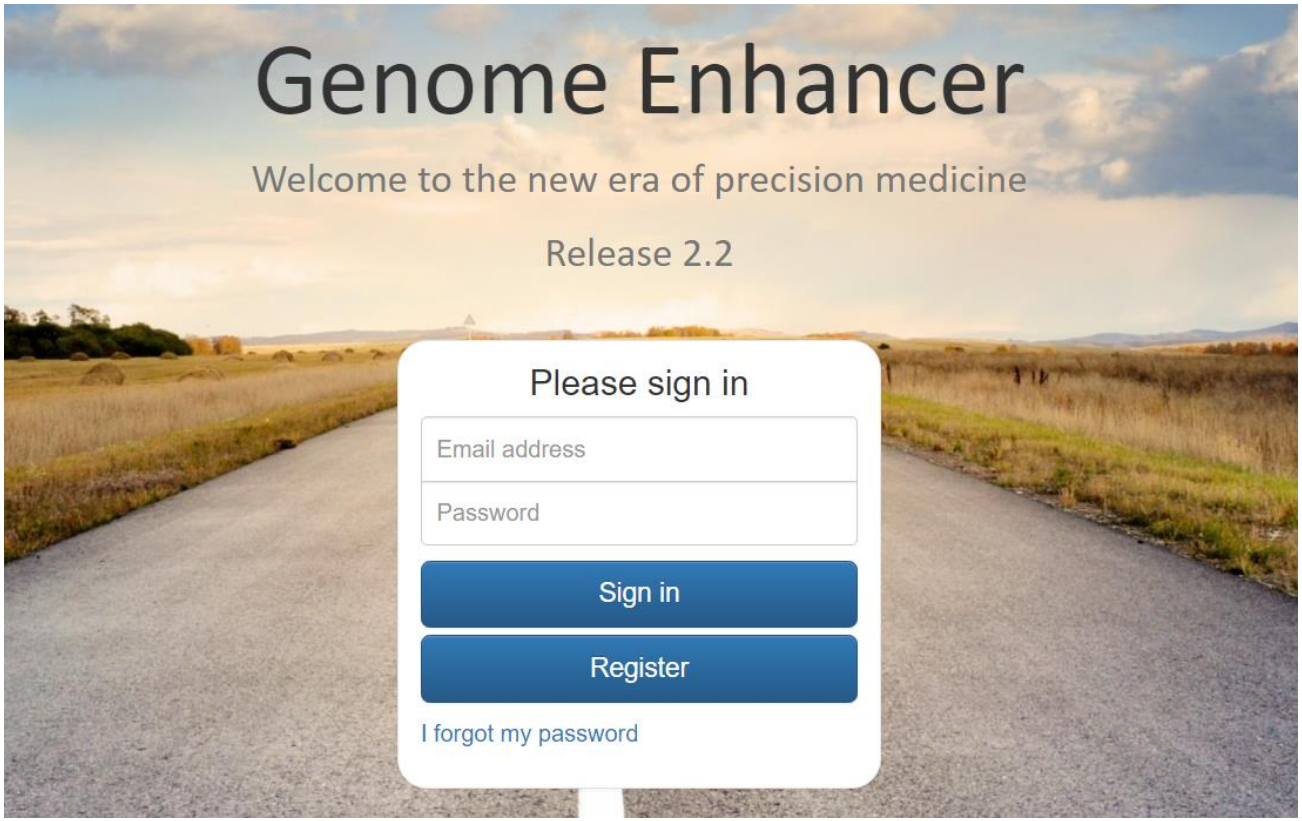
info@genexplain.com

Introduction into Genome Enhancer

RSCI training

Day 7

info@genexplain.com



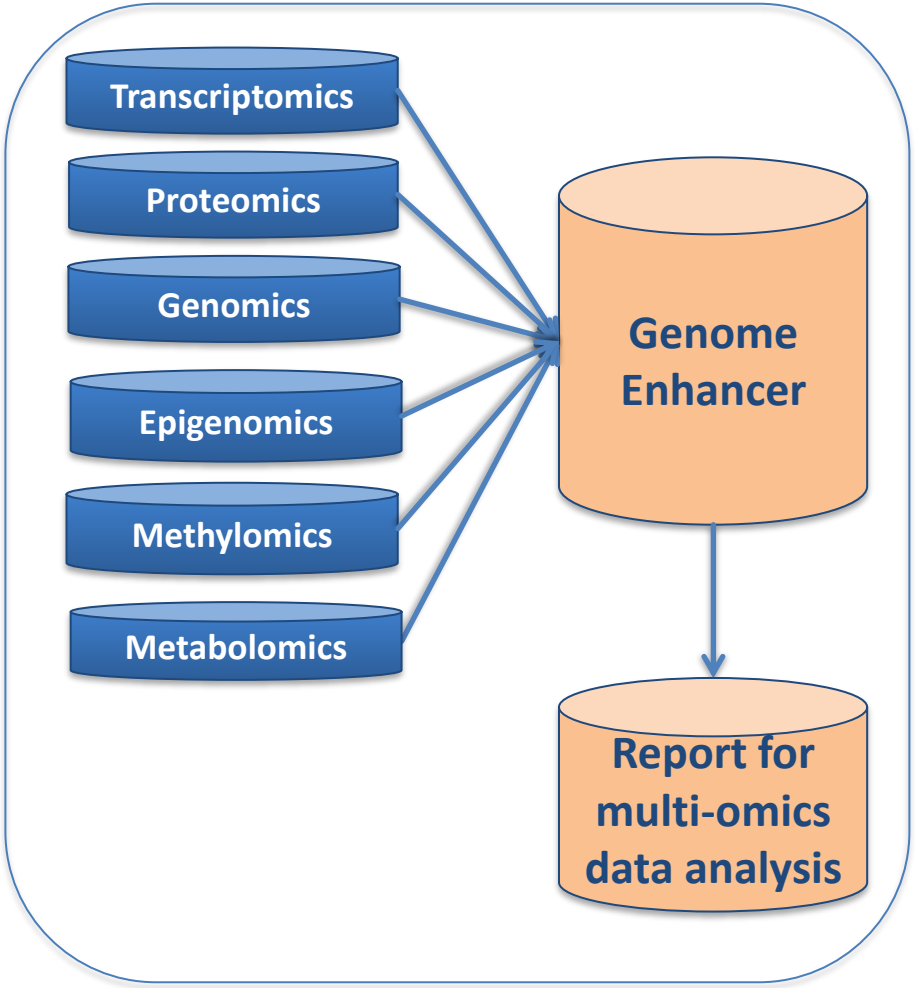
https://ge.genexplain.com/

Multi-omics data analysis

GE provides a unique possibility for multi-omics analysis.

GE accepts input of several different human omics data at the same time to be analyzed together.

The report is generated for multi-omics data analysis.



Input data formats

As input, GE accepts a wide range of data formats, starting with different raw formats https://genexplain.com/genome-enhancer/#section3 and going to pre-calculated data as Excel tables, csv, and other tabulated formats.

Researchers are welcome to import the data formats they already have in their hands.

Transcriptomics (RNA-seq, microarrays)

*.txt, *.csv, *.xls (table with gene identifiers)

*.CEL (affymetrix)

*.txt (special agilent format)

*.txt (special illumina format)

*.fastq

Epigenomics (ChIP-seq)

*.fastq

*.bam (hg38 only)

*.bed (hg38 only)

.txt (table with illumina methylation probe ids, cg)

Genomics

*.vcf

*.txt, *.csv, *.xls (table data with SNP identifiers, rs*)

*.fastq

Proteomics

*.txt, *.csv, *.xls (table with protein identifiers)

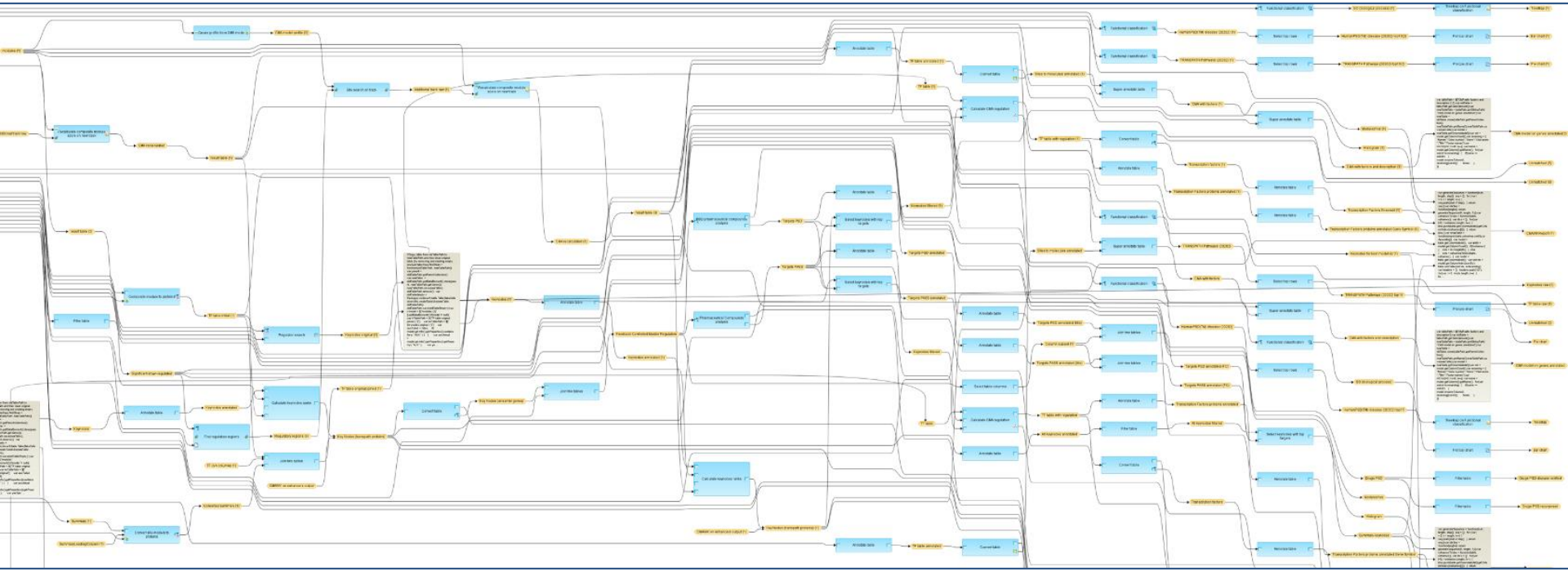
Metabolomics

*.txt, *.csv, *.xls (table with the list of metabolites from chebi database, e.g. CHEBI:57316)

Files of one data format can be uploaded in a .zip archive

Automatically generated workflow (pipeline)

Depending on the submitted omics data and file format, GE automatically generates a pipeline.
An example illustrates a fragment of a pipeline for transcriptomics data analysis.

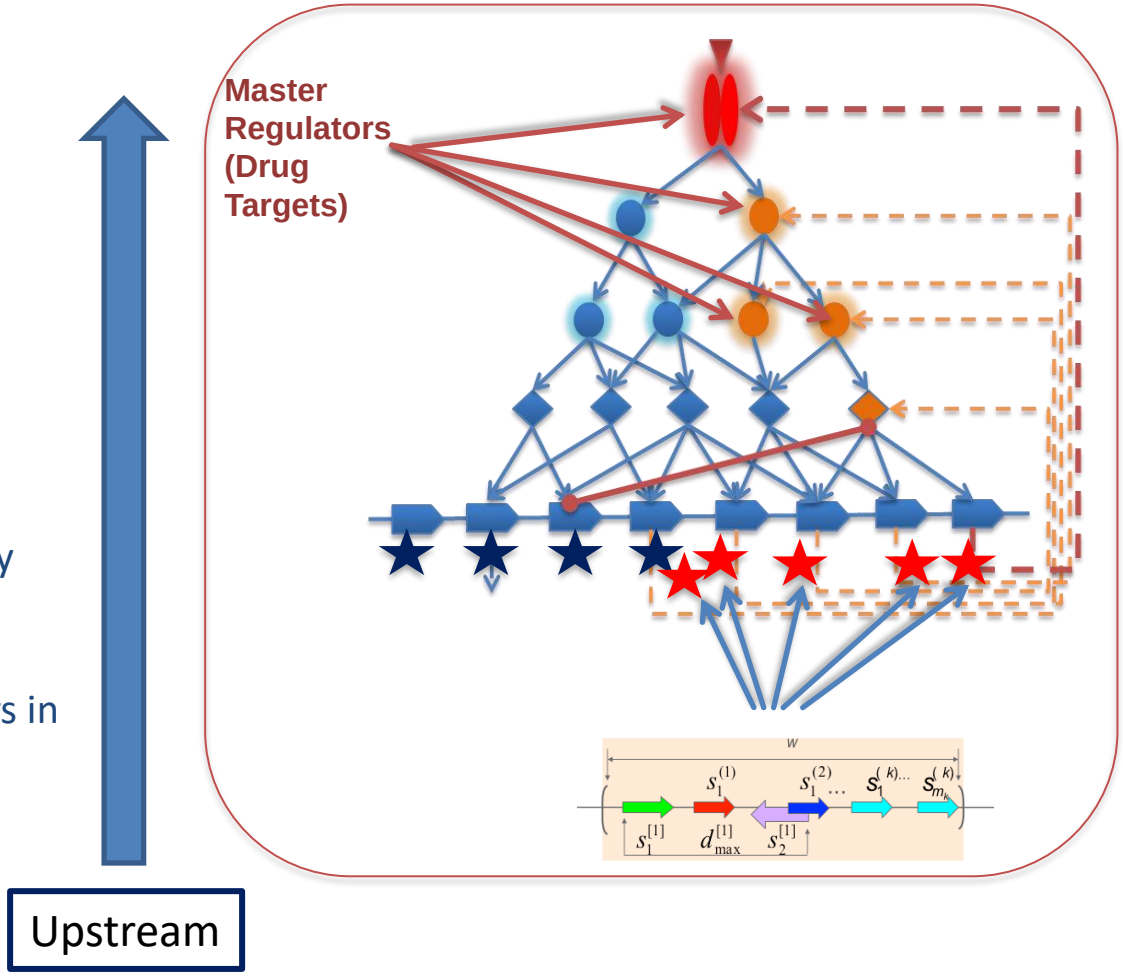


Upstream Analysis Concept

The concept of upstream analysis includes two steps, promoter analysis and network analysis.

Promoter analysis is done as TFBS search using proprietary F-Match and CMA algorithms together with the unique collection of TRANSFAC® matrices.

Network analysis is done with a proprietary graph-analyzing algorithm and the unique collection of TRANSPATH® reactions. As a result, the master regulators in the networks are identified.



Genomic variations in gene regulatory regions (regulatory SNPs)

GE includes a unique algorithm for analysis of genomic variations in gene regulatory regions, also known as regulatory SNPs.

Changes of TFBS scores resulting from particular genomic variations are part of the report generated by GE.

Table 3. Mutations revealed in genes carrying SNP variations

See full table →

ID	Gene symbol	Gene schematic representation	Number of variations
ENSG00000196735	HLA-DQA1		81
ENSG00000130164	LDLR		46
ENSG00000161888	SPC24		30
ENSG00000165029	ABCA1		30
ENSG00000128918	ALDH1A2		26
ENSG00000166035	LIPC		24
ENSG00000169174	PCSK9		20
ENSG00000136573	BLK		19
ENSG00000175445	LPL		18
ENSG00000084674	APOB		16

Table 4. PWMs whose sites were lost or gained due to mutations in genes carrying SNP variations

See full table →

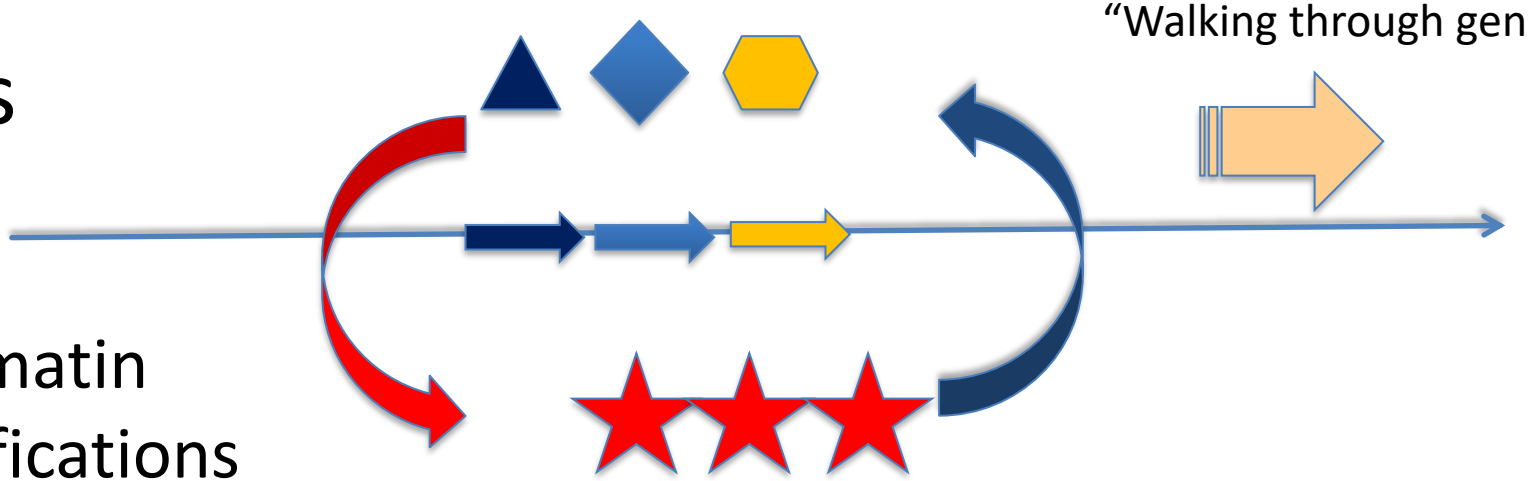
ID	P-value (gains)	P-value (losses)	yesCount (gains)	yesCount (losses)
V$BTEB2_Q3_01	1.79E-11	1.01E-12	14	21
V$SNAI2_01	1.79E-11	2.11E-12	14	21
V$STAT4_01	3.37E-14	6.82E-13	17	24
V$TFE_Q6	3.37E-14	1.26E-13	17	25
V$AML1_Q4	2.62E-16	5.58E-12	20	22
V$AP1_02	2.62E-16	5.58E-12	20	22
V$CETS1_Q6	2.62E-16	5.58E-12	20	22
V$LEF1_Q5	2.62E-16	3.6E-13	20	24
V$LEF1_Q5_01	2.62E-16	3.6E-13	20	24
V$AHR_Q6	1.3E-16	3.32E-11	21	22
V$AML1_01	1.3E-16	3.32E-11	21	22

info@genexplain.com

Epigenetic “walking”

“Walking through genome”

TFs



Chromatin modifications

Druggability of the Master Regulators (drug targets)

GE calculates, in a unique way, and provides a report section for the druggability of the identified master regulators (drug targets).

Additionally, the clinical trials for the suggested drugs are included in the report as well.

These data are included based on HumanPSD database.

ID	Gene symbol	Gene description	Druggability score
ENSG00000125538	IL1B	interleukin 1 beta	11
ENSG00000097007	ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase	8
ENSG00000109320	NFKB1	nuclear factor kappa B subunit 1	8
ENSG00000136244	IL6	interleukin 6	6
ENSG00000197122	SRC	SRC proto-oncogene, non-receptor tyrosine kinase	6
ENSG00000259207	ITGB3	integrin subunit beta 3	4
ENSG00000112062	MAPK14	mitogen-activated protein kinase 14	3
ENSG00000140992	PDPK1	3-phosphoinositide dependent protein kinase 1	3
ENSG00000143322	ABL2	ABL proto-oncogene 2, non-receptor tyrosine kinase	3
ENSG00000103653	CSK	c-src tyrosine kinase	2

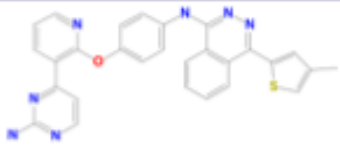
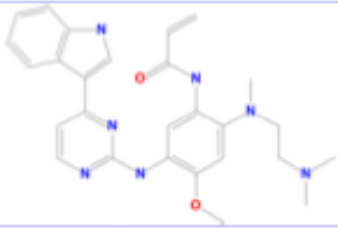
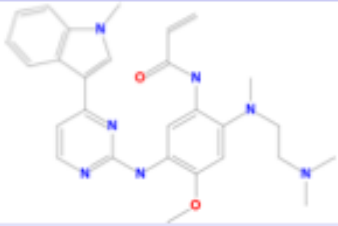
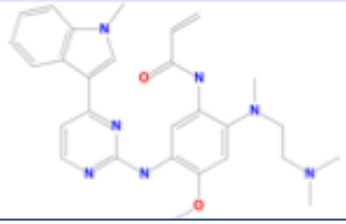
ID	Name	Target names	Target activity score	NA	Phase 1	Phase 2	Phase 3	Phase 4
DB00482	Celecoxib	PDPK1	1	Inflammation, Adenoma, Adenomatous Polyposis Coli, Adenomatous Polyps, Alzheimer Disease, Antiphospholipid Syndrome, Arthritis...	Inflammation, Apnea, Brain Abscess, Breast Neoplasms, Carcinoma, Carcinoma, Non-Small-Cell Lung, Carcinoma, Small Cell...	Adenocarcinoma, Adenoma, Adenomatous Polyposis Coli, Adenomatous Polyps, Amyotrophic Lateral Sclerosis, Angiomyoma, Apnea...	Adenoma, Adenomatous Polyposis Coli, Adenomatous Polyps, Alzheimer Disease, Anterior Cruciate Ligament Injuries, Arthritis, Arthritis, Gouty...	Inflammation, Inflammatory Bowel Diseases, Adenoma, Adenomatous Polyposis Coli, Adenomatous Polyps, Ankle Injuries, Arteriosclerosis...
DB01017	Minocycline	IL1B	1	Acne Vulgaris, Acute Kidney Injury, Alopecia, Angelman Syndrome, Anxiety,	Inflammation, Acne Vulgaris, Acute Kidney Injury, Affect, Alcohol Drinking	Inflammation, Acne Vulgaris, Alcohol Drinking, Alopecia, Alzheimer Disease	Acne Vulgaris, Affect, Alopecia, Amphetamine-Related Disorders, Amyotrophic	Acne Vulgaris, Affect, Alopecia, Autistic Disorder, Bacterial Infections

Calculations done by the cheminformatics software PASS are part of the report

GE report includes suggestions for drug re-purposing (off label drug applications), and even for new chemical substances, to target identified master regulators.

This part is based on the calculations done with the Cheminformatics software PASS.

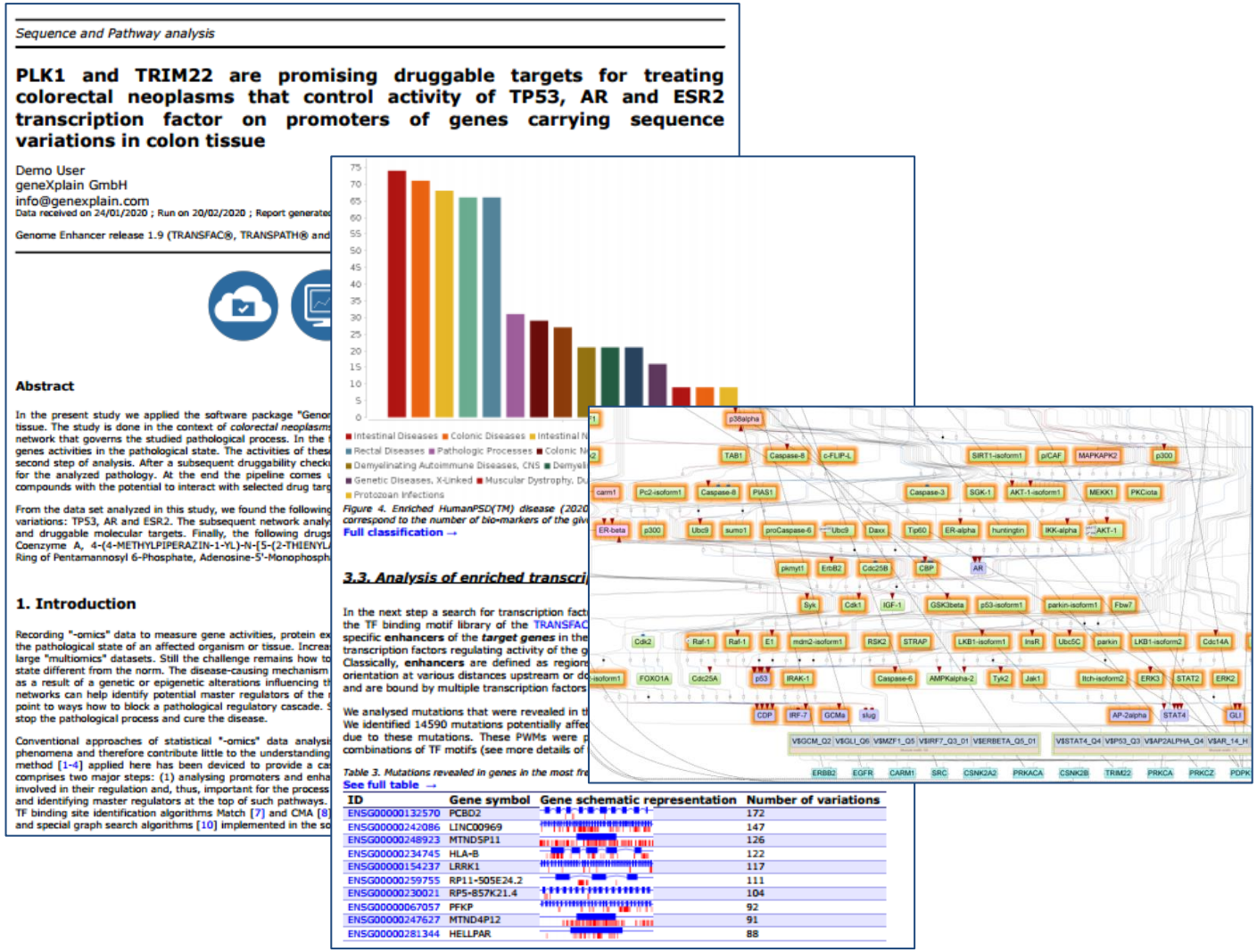
Integration of bioinformatics and Cheminformatics approaches is the next specific and unique GE feature.

Name	Structure	Target names	Target activity score
AMG 900		ADAM17, MAP4K4, SETD7, PINK1, ABL1, PTPN6, MAPK14...	11.67
AZ5104		ADAM17, MAP4K4, PINK1, ABL1, MAPK14, IRAK2, SGK1...	8.52
Osimertinib		ADAM17, MAP4K4, PINK1, ABL1, MAPK14, IRAK2, SGK1...	7.83
AZD9291		ADAM17, MAP4K4, PINK1, ABL1, MAPK14, IRAK2, SGK1...	7.83

The Report is generated automatically and can be downloaded

As a result of GE run, a report is automatically generated, which can be downloaded in a comfortable format, as PDF file or HTML page.

All primary results, including all intermediate results, are available as tables or graphs, depending on data type, and can be downloaded in a number of formats.



Sequence and Pathway analysis

PLK1 and TRIM22 are promising druggable targets for treating colorectal neoplasms that control activity of TP53, AR and ESR2 transcription factor on promoters of genes carrying sequence variations in colon tissue

Demo User
geneXplain GmbH
info@genexplain.com
Data received on 24/01/2020 ; Run on 20/02/2020 ; Report generated

Genome Enhancer release 1.9 (TRANSFAC®, TRANSPATH® and

Abstract

In the present study we applied the software package "Genom
tissue. The study is done in the context of *colorectal neoplasms*
network that governs the studied pathological process. In the
genes activities in the pathological state. The activities of these
second step of analysis. After a subsequent druggability checki
for the analyzed pathology. At the end the pipeline comes u
compounds with the potential to interact with selected drug targ

From the data set analyzed in this study, we found the following
variations: TP53, AR and ESR2. The subsequent network analy
and druggable molecular targets. Finally, the following drugs
Coenzyme A, 4-(4-METHYLPIPERAZIN-1-YL)-N-[5-(2-THIENYL
Ring of Pentamannosyl 6-Phosphate, Adenosine-5'-Monophosph

1. Introduction

Recording "-omics" data to measure gene activities, protein ex
the pathological state of an affected organism or tissue. Increa
large "multiomics" datasets. Still the challenge remains how to
state different from the norm. The disease-causing mechanism
as a result of a genetic or epigenetic alterations influencing th
networks can help identify potential master regulators of the
point to ways how to block a pathological regulatory cascade.
stop the pathological process and cure the disease.

Conventional approaches of statistical "-omics" data analysi
phenomena and therefore contribute little to the understanding
method [1-4] applied here has been deviced to provide a ca
comprises two major steps: (1) analysing promoters and enha
involved in their regulation and, thus, important for the process
and identifying master regulators at the top of such pathways.
TF binding site identification algorithms Match [7] and CMA [8
and special graph search algorithms [10] implemented in the so

*Figure 4. Enriched HumanPSD(TM) disease (2020
correspond to the number of bio-markers of the giv*
Full classification →

3.3. Analysis of enriched transcri

In the next step a search for transcription fact
the TF binding motif library of the TRANSFAC
specific **enhancers** of the ***target genes*** in the
transcription factors regulating activity of the g
Classically, **enhancers** are defined as region
orientation at various distances upstream or do
and are bound by multiple transcription factors

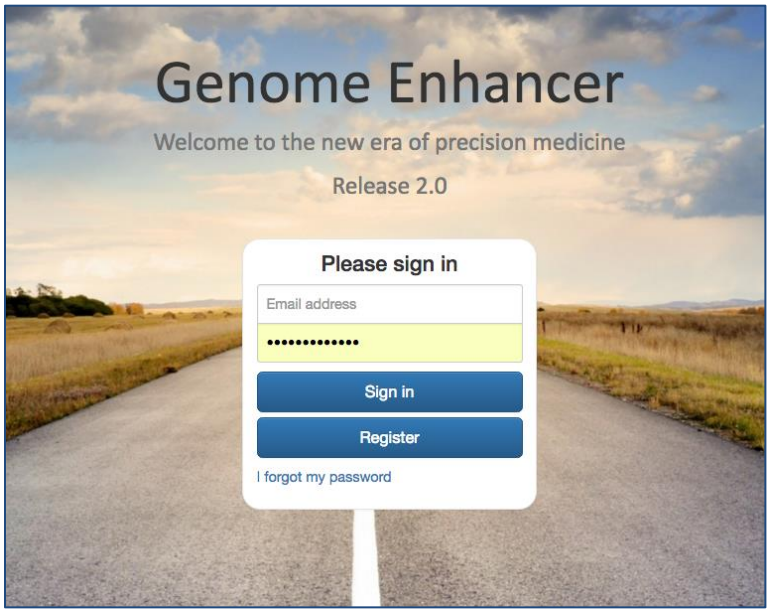
We analysed mutations that were revealed in t
We identified 14590 mutations potentially affec
due to these mutations. These PWMs were p
combinations of TF motifs (see more details of

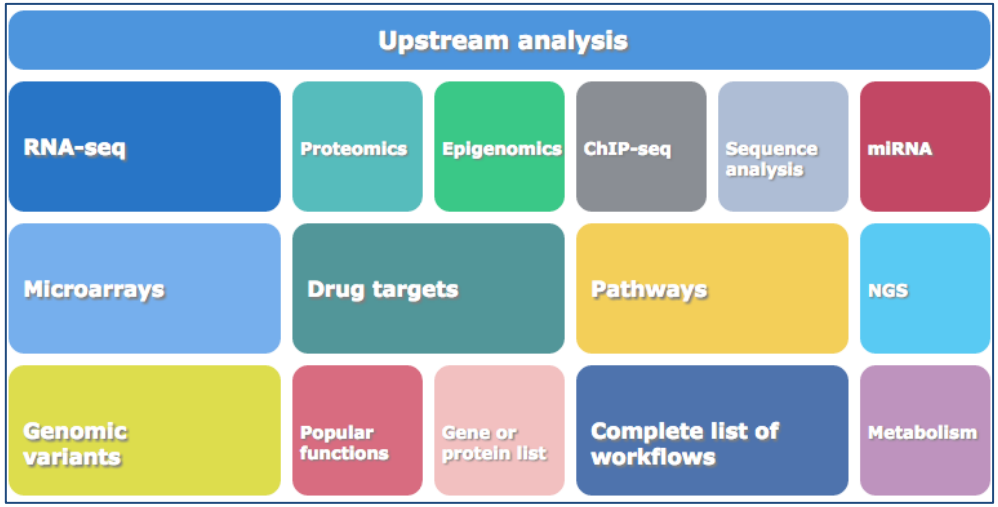
Table 3. Mutations revealed in genes in the most fr
See full table →

ID	Gene symbol	Gene schematic representation	Number of variations
ENSG00000132570	PCBD2		172
ENSG00000242086	LINC00969		147
ENSG00000248923	MTND5P11		126
ENSG00000234745	HLA-B		122
ENSG00000154237	LRRK1		117
ENSG00000259755	RP11-505E24.2		111
ENSG00000230021	RP5-857K21.4		104
ENSG00000067057	PFKP		92
ENSG00000247627	MTND4P12		91
ENSG00000281344	HELLPAR		88

Genome Enhancer Expert

Synergism between the automatic pipeline of Genome Enhancer and a comprehensive bioinformatics toolbox of the geneXplain platform





info@genexplain.com

Genome Enhancer Expert

More functionalities for the Genome Enhancer users:

- ✓ Post-processing the results of the automatic GE pipeline
- ✓ Integrated databases and analysis tools
- ✓ Ready-made workflows for an easy start
- ✓ Genomes for several different organisms
- ✓ Java API access
- ✓ Simulation engine inside