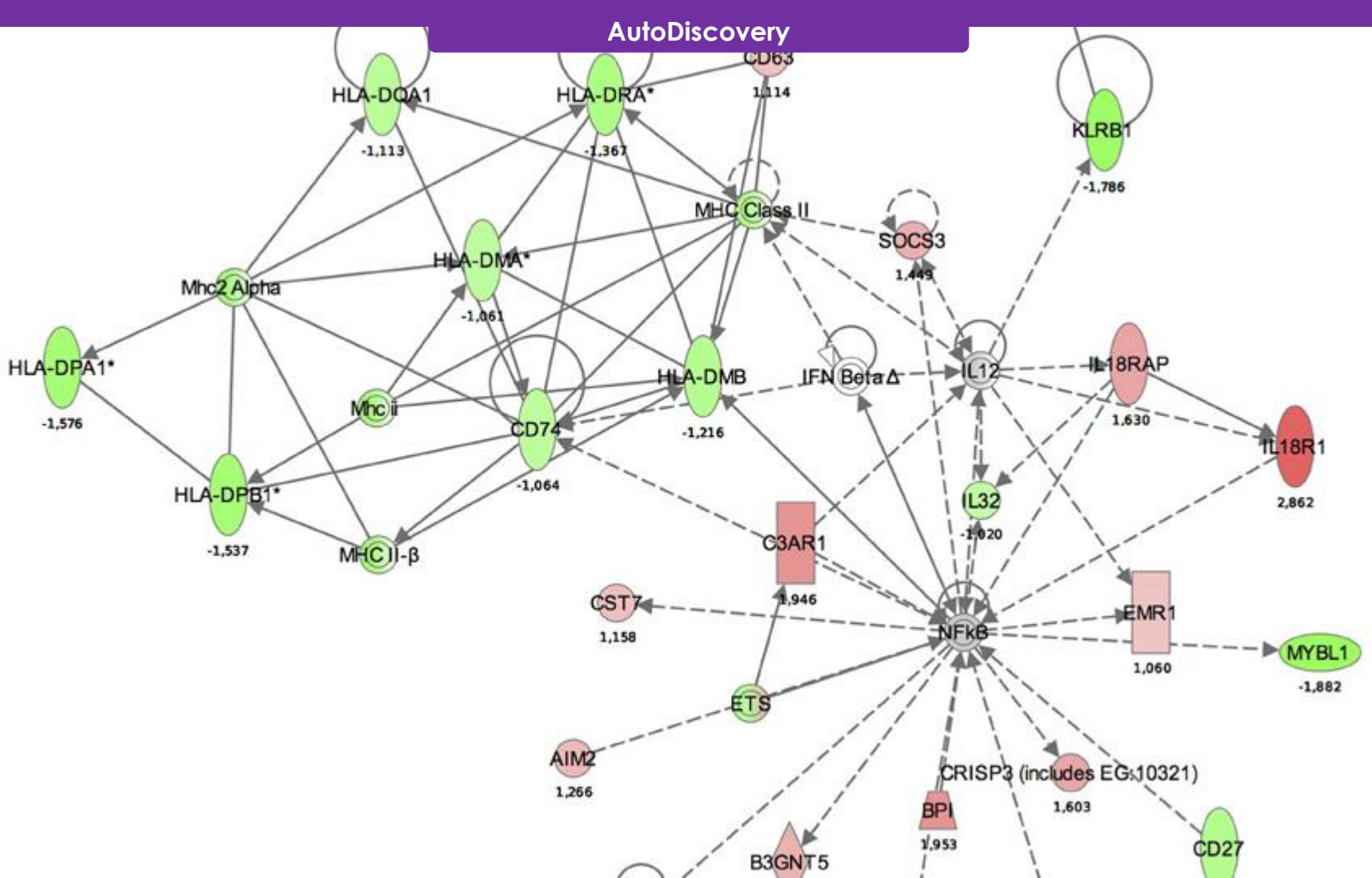
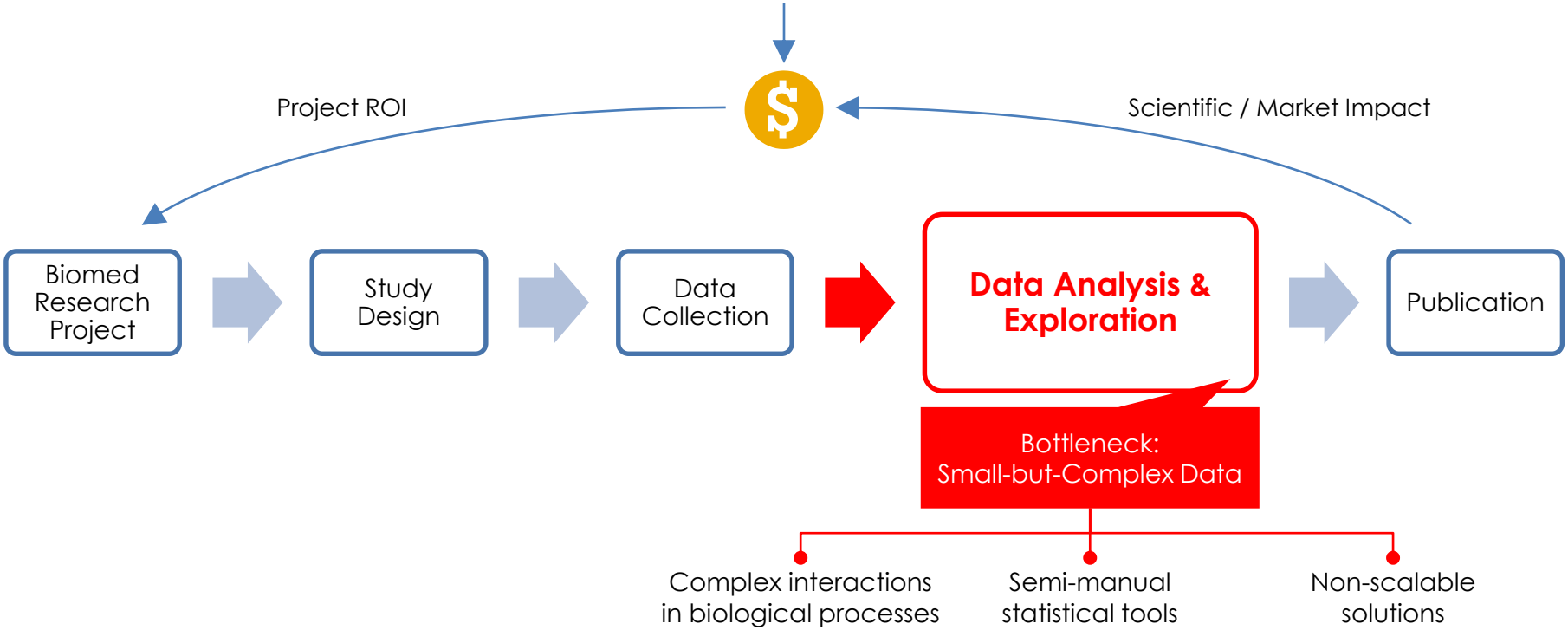

Butler Scientifics
What will you discover today?


HLA-DQA1
-1,113
HLA-DRA*
-1,367
CD63
1,114
KLRB1
-1,786
MHC Class II
SOCS3
1,449
HLA-DMA*
-1,061
Mhc2 Alpha
HLA-DPA1*
-1,576
HLA-DMB
-1,216
IFN Beta Δ
IL12
IL18RAP
1,630
IL18R1
2,862
Mhc ii
CD74
-1,064
IL32
-1,020
HLA-DPB1*
-1,537
MHC II-β
C3AR1
1,946
CST7
1,158
NFkB
EMR1
1,060
MYBL1
-1,882
ETS
AIM2
1,266
CRISP3 (includes EG:10321)
1,603
BPI
1,953
B3GNT5
CD27



Classical statistical tools cannot face the increasing complexity of biomed research projects. Ultimately, that **impacts the project ROI**.

AutoDiscovery

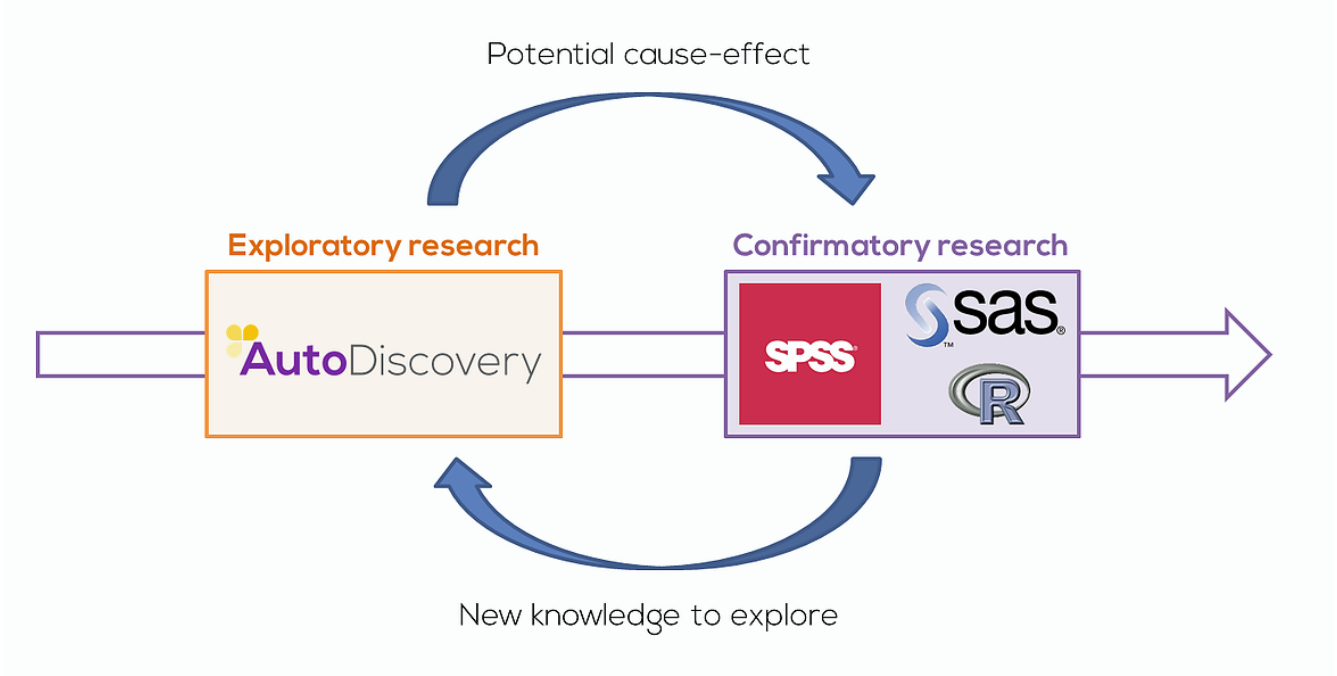
AutoDiscovery is an **intelligent automated exploratory data analysis software** that helps biomedical researchers to unveil clinically relevant associations hidden in the data files of their scientific experiments and studies.

Exploratory Research

Exploratory research is the stage of the research process that aims at connecting ideas as to unveil the why's of potential cause/effect relationships. This occurs when researchers get started at understanding what they are actually "observing" when in the process of building cause/effect models.

Confirmatory research (a.k.a. hypothesis testing) is where researchers have a pretty good idea of what's going on. That is, researcher has a theory (or several theories), and the objective is to find out if the theory is supported by the facts.

Exploratory Research



Exploratory and confirmatory research are two **complementary items** of the same goal: to discover relevant findings in the most efficient, reliable, replicable, applicable manner.

An exploratory study should always be designed and executed in order to answer a number of a-priori **exploratory questions**.

Our experience in dozens of scientific projects has allowed us identify the **5 mostly shared kinds of questions** that researchers ask themselves.

#1 **Role**

Aimed at getting to understand the role that a certain group of known input factors has over the behavior of our system or part of it (responses).

Example:

What is the role that the neuronal structure has over learning and memory performance indicators of our given experimental subjects?

Read More ...

#2 Prediction

Aimed at unveiling what factors will help us modelling certain responses of our system.

Example:

What gen signatures will help us predicting the evolution of the tumor size in our cancer model?

Read More …

#3 Characterization

Aimed at getting to know and understand better that set of factors that better describe our experimental groups.

Example:

What are the key potential factors characterizing the respondents of my study by gender, age, sexual abuses, psychological profile and emotional condition?

Read More …

#4 Differentiation

Aimed at identifying what responses are the most different according to a certain already known factor.

Example:

What proteins express differently in group control patients?

#5 Thresholds

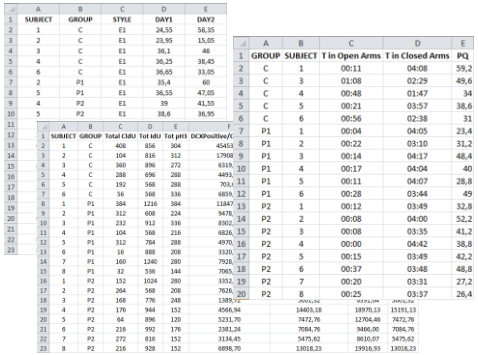
Aimed at getting to know what threshold values are the most clinically relevant in a certain biological process.

Example:

At what levels of magnesium concentration we observe different dynamics in the patients under study?

How It Works?

1. Consolidate

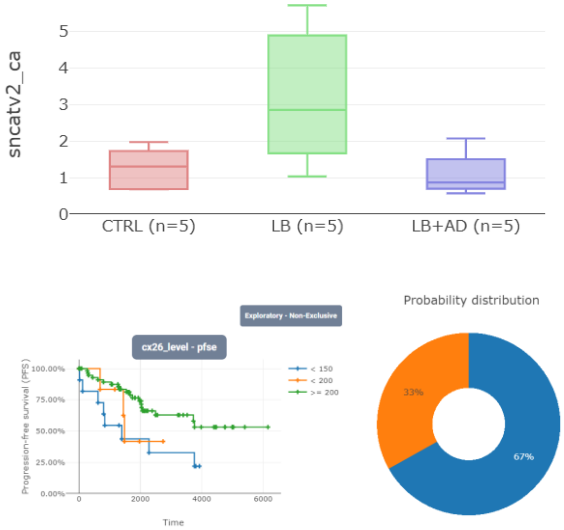


SUBJECT	GROUP	STYLE	DAY1	DAY2
1	C	E1	24,55	58,35
2	C	E1	23,95	15,05
3	C	E1	36,1	46
4	C	E1	36,25	38,45
6	C	E1	36,65	33,05
2	P1	E1	35,4	60
5	P1	E1	36,55	47,65
4	P2	E1	39	41,55
5	P2	E1	38,6	36,95

GROUP	SUBJECT	T in Open Arms	T in Closed Arms	PQ
C	1	00:11	04:08	59,2
C	3	01:08	02:29	49,6
C	4	00:48	01:47	34
C	5	00:21	03:57	38,6
C	6	00:56	02:38	31
P1	1	00:04	04:05	23,4
P1	2	00:22	03:10	31,2
P1	3	00:14	04:17	48,4
P1	4	00:17	04:04	40
P1	5	00:11	04:07	28,8
P1	6	00:28	03:44	49
P2	1	00:12	03:49	32,8
P2	2	00:08	04:00	52,2
P2	3	00:08	03:35	41,2
P2	4	00:00	04:42	38,8
P2	5	00:15	03:49	42,2
P2	6	00:37	03:48	48,8
P2	7	00:20	03:31	27,2
P2	8	00:25	03:37	26,4

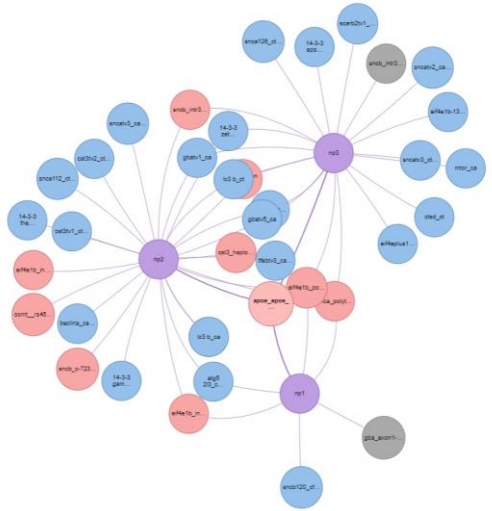
Multiple data files are merged by the software into a **single table**.

2. Discover



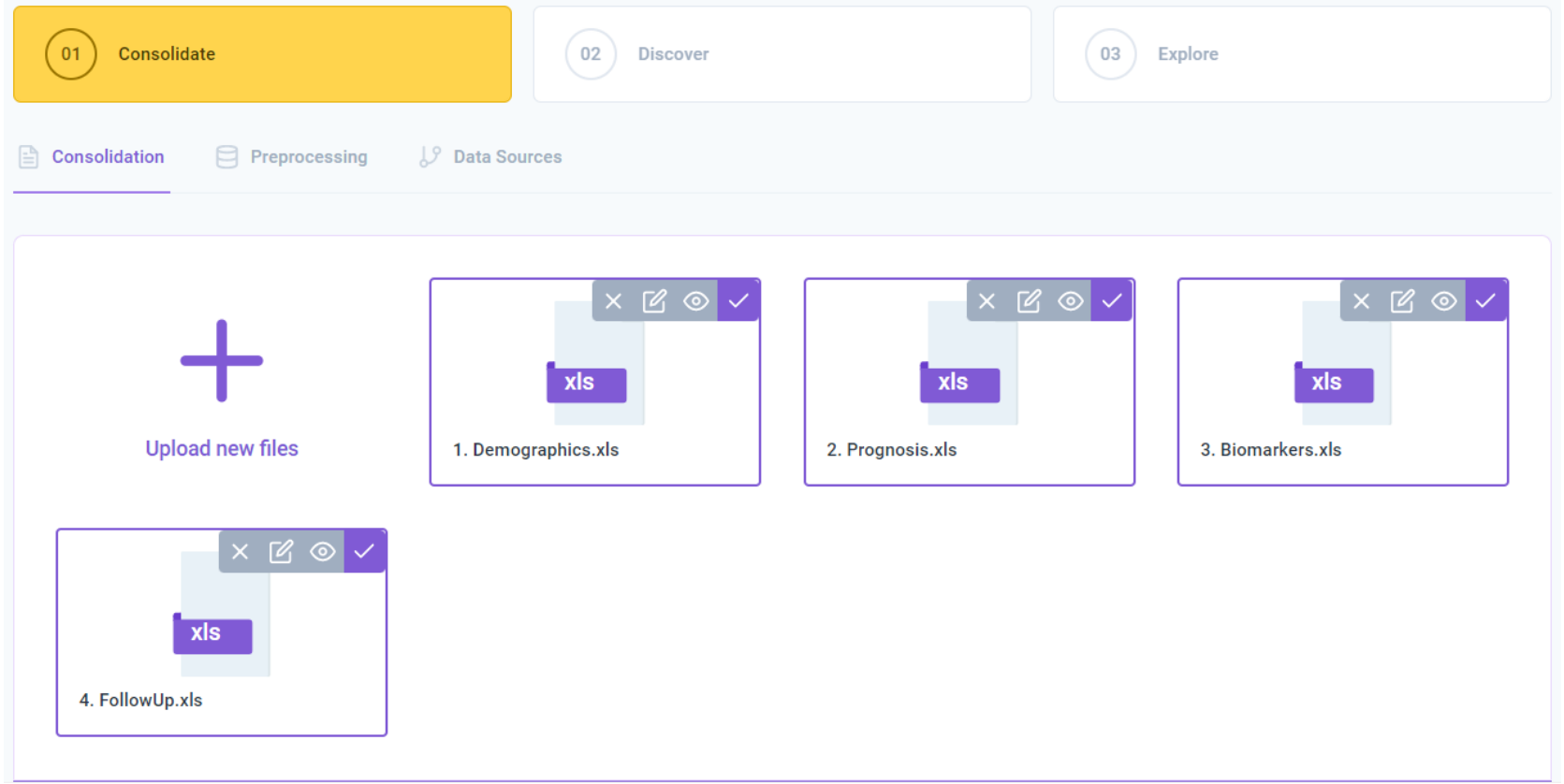
An **exhaustive statistical procedure** is automatically applied to unveil potential cause-effect relationships.

3. Explore



Relationships are browsed following the **original exploratory question**.

1. Consolidate



1. Consolidate

Consolidation_20220221153634
101 rows x 48 columns

Number of rows: 10 rows | Download

ROW #	NUM_BIOPSIA	VOLUMEN	SEXE	EDAT_AL_DX	LOCALITZACIO	ESTADIAJE_DX	ESTADIATGE	HISTOLOGIA	INFILTRADO	INV_NERVI_OPTIC	MITOSIS__10CGA	NEOVASCULARI
1		150	home	77.0	coroides	12B08663	IIIA	IV	mixta	minimo	no	1.0
2		65	home	76.0	cos ciliar	12B06756	IIIA	IV	mixta	no	no	4.0
3		98	home	63.0	coroides	11B20119	IIIA	IIIA	mixta	no		0.0
4		91	dona	80.0	coroides	11B11803	IIIC	IIIC	mixta	minimo	si	0.0
5		104	dona	54.0	coroides	11B09888	IIIA	IV	epiteloide	minimo	no	4.0
6		42	home	52.0	coroides	10B09237	IIA	IIA	fusiforme	minimo		1.0
7		160	dona	84.0	cos ciliar	10B00563	IIIB	IIIB	mixta	minimo	si	
8		180	home	64.0	cos ciliar	09B501288	IIIB	IIIB	fusiforme	minimo	no	5.0
9		140	home	63.0	coroides	09B22693	IIIA	IV	mixta	intenso	no	3.0
10		70	home	77.0	coroides	09B22673	IIIA	IIIA	mixta	no	no	1.0

1. Consolidate

Data consolidation is the process of joining different related data files into a single table.

AutoDiscovery works with the Excel files in which the data of the patients, experiment trials, etc. are stored.

2. Discover

Configurations

Exploratory Analysis

Variables to explain Remove all

ki67 ×

Variables to ignore Remove all

num_biopsia ×

Subject groups Remove all

sexe × inv_nervi_optic × stage_01 ×

Discovery Plan

Stratums

Stratums	Will generate **416** stratums

Analyses

Spearman's Ranks	Will perform **45** calculations
ANOVA 1 Ways	Will perform **163** calculations

Details

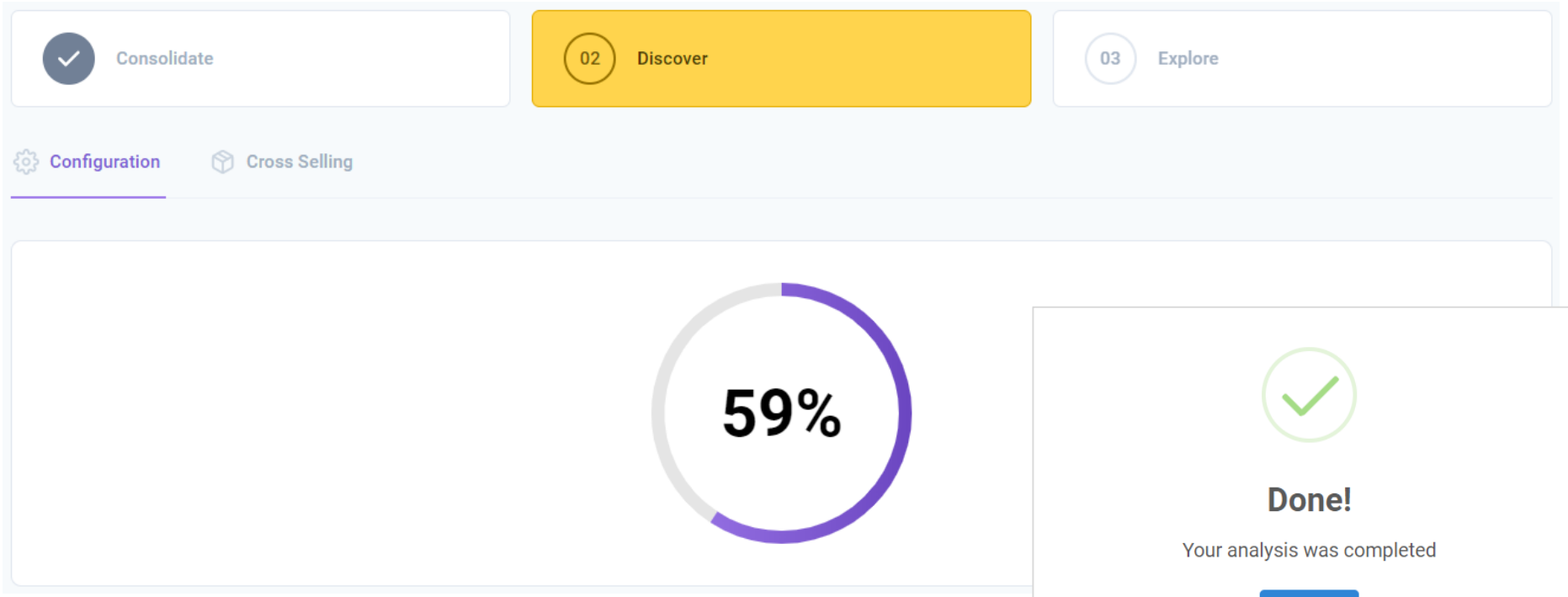
Discover Now

2. Discover

The discovery process consists in evaluating the statistical associations between the variables of interest, in a range of subgroups of the samples.

Advanced configuration tools are available to fine tune the scope and performance of the discovery process.

2. Discover


Consolidate
02
Discover
03
Explore
Configuration
Cross Selling
59%
Done!
Your analysis was completed
OK

2. Discover

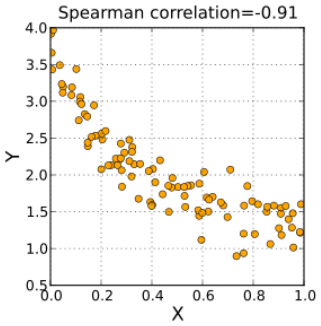
Depending on the nature of the data of each variable, a particular flexible **statistical test is automatically selected** and computed to assess how these variables are associated.

This process is also performed in specific subsets of your data (e.g. groups of patients or animals) and "subintervals" (e.g. patients older than ...).

The exclusiveness post-analysis (including **False Discovery Rates**) assesses the clinical relevance and also the statistical significance of the associations evaluated, that is, its likelihood to become a confirmed novel finding or an exploratory result to be tested in a further confirmatory phase of the experiment.

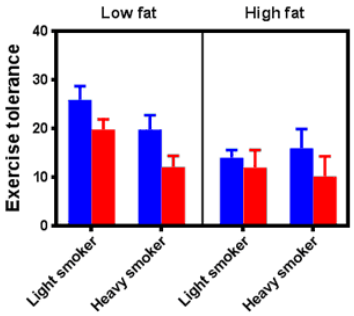
How It Works

Spearman's Rank Correlation



Monotonic positive/negative association.

ANOVA 1-way | U Mann Whitney | Kruskal Wallis



Significant **differences** in average between categories.

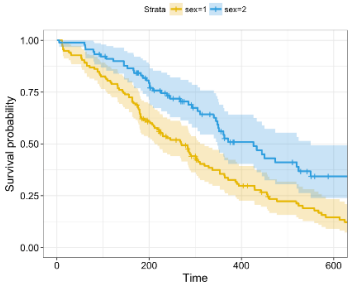
Xi^2 + Cramér's V

	Pass	Fail	Total
Males	46	56	102
Females	68	30	98
Total	114	86	200

Contingency table

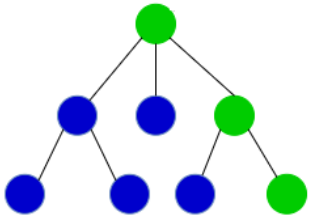
Associations between **categorical variables**.

Survival Analysis



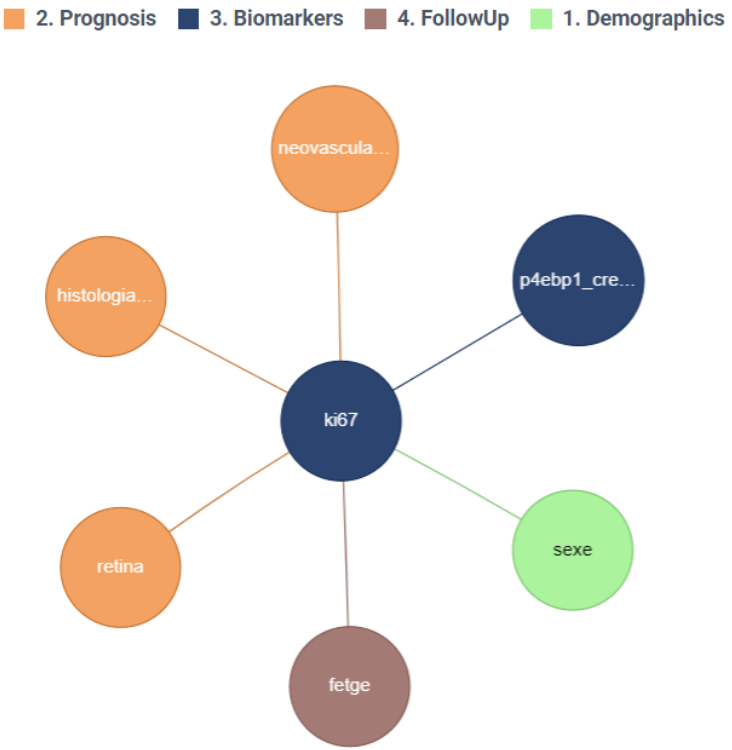
Discover subgroups with **maximized survival rates**.

ML Classification Models

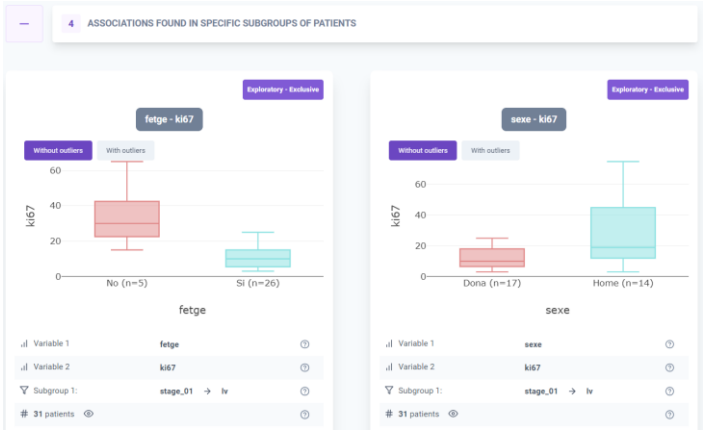


Builds user-friendly apps to learn and predict the **classification risk of patients**.

3. Explore



	KI67		
VARIABLE NAME	WHOLE POPULATION (1)	SPECIFIC SUBGROUPS (4)	OTHER SUBGROUPS (2)
fetge		☆☆	
sexe		☆☆	
p4ebp1_creus		☆☆	
neovascularitzacio	☆		☆☆
retina		☆	
histologia			☆



3. Explore

The **Discovery Map** and **Hypo Booster** tools facilitate browsing the list of associations detected between the variables.

A detailed table of associations arranged by their relevance and significance is provided.

Plots show the subset of data samples used by AutoDiscovery to evaluate every individual association, which enables the traceability of the results.

Graphs, plots and tables can be exported to share.

Summary

Summary of Features

Smart Exhaustivity

- Automatic selection of statistical tests
- Configurable range of combinatorial analysis
- Exclusivity and prioritization features

Biomed-Specific

- Automatic data integration
- Stratified analysis (groups, treatments, …)
- False Discovery Rate
- Traceability of results assured

PI-Convenient

- Complement to hypothesis testing to improve impact
- High-level insights in a few minutes
- Personalised services or individual licenses
- Agile technical support

AutoDiscovery

If you want to know more on how it can help in you research projects, contact our **Data Science Director** at ray@butlerscientifics.com


Butler Scientifics
What will you discover today?