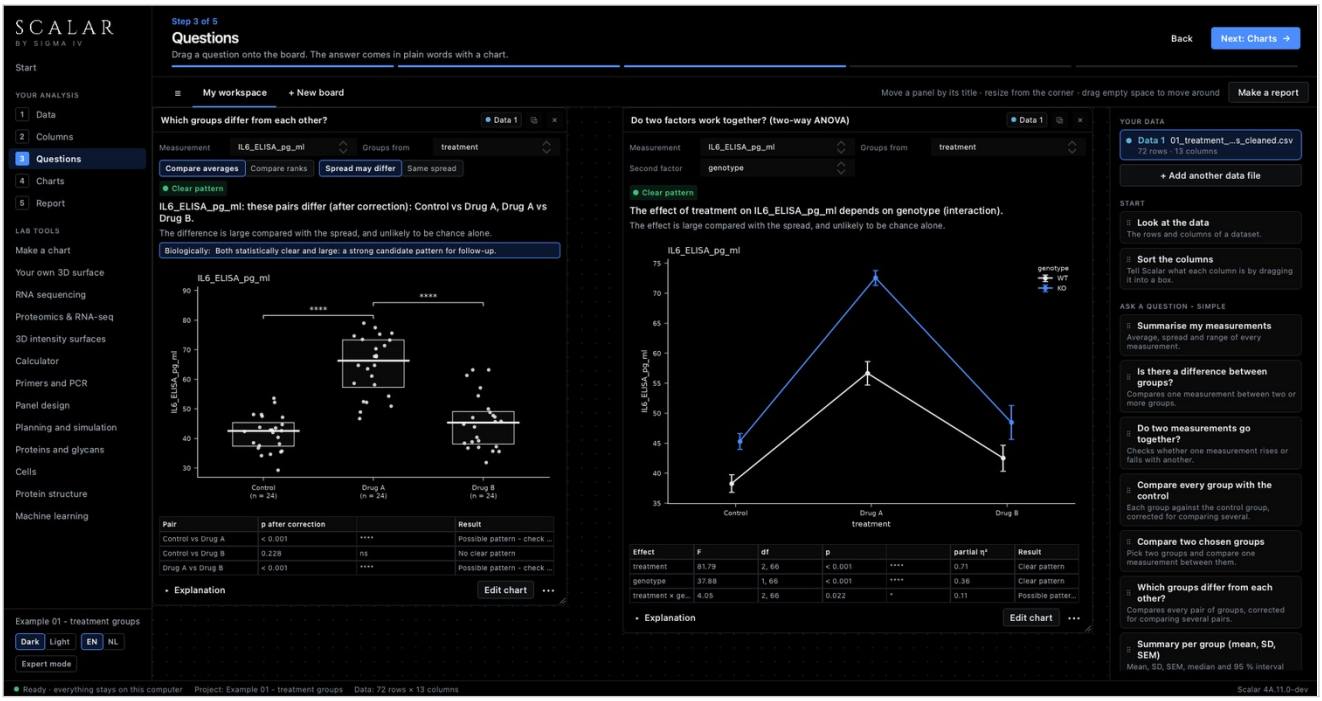
SCALAR

BY SIGMA IV

Guide and worked examples

How Scalar works, what it can do in version 4A.11, and a synthetic data set for every test, question and lab tool - so each of them can be tried and checked.



The board in step 3: two answered questions on example file 01 - a sentence, a label, a chart with stars and the numbers.

21	24	4	25	6	22 + 3
questions in plain words	tests you can choose yourself	pattern analyses	lab tools	new lab pages in 4A.11	synthetic data files and images

> **The data are invented.** Every file that comes with this guide was generated by a script with a fixed seed; nothing comes from an experiment, a patient or a donor. Every sentence, number and chart shown here was produced by Scalar 4A.11 itself on those files, so you should see exactly the same when you repeat it.

November 2026 · Scalar by SIGMA IV · results are exploratory: a clear pattern is a reason to repeat an experiment, not a conclusion.

1 · OVERVIEW

What Scalar can do now

One program for the road from a spreadsheet to a report: check the data, ask a question in plain words or pick a named test, read the answer with its chart, see every step of the calculation, and take it to R or Excel to check.

	What you can do	Where
Your data	CSV or Excel. Scalar checks the file first and repairs what is safe to repair; your original file is never changed. Row 2 can say what each column is, and six standard templates show a good layout.	Step 1 · Data
Columns	Drag each column into a box: measurements, groups, subgroups, sample name, patient / donor, time point, not used. Calculate new columns: log2, ratio, % of control, z-score, qPCR ΔΔCt.	Step 2 · Columns
Questions in plain words	21 questions in three tiers, from 'Is there a difference between groups?' to survival curves, ROC and mixed models. Scalar picks the test from the design of the question, never by looking at the data.	Step 3 · Questions
Choose a test yourself	24 named tests, from the Welch t-test to Kaplan-Meier, each offered when the columns allow it.	Step 3 · Questions
Find patterns	PCA, Ward clustering, cluster stability and cluster profiles over all measurements at once.	Step 3 · Questions
Every answer	A sentence, a label by one fixed rule, a chart with significance stars, the numbers, an explanation in plain words, and 'See the calculation': each step with formula, numbers filled in and result, an Excel check and R code that runs on its own.	Each panel
Charts	Templates, stars, points per sample, n per group. Copy at 300 dpi, or save as PNG (300 / 600 dpi), TIFF (600 dpi), SVG or PDF.	Step 4 · Charts
Report	A PDF of a board: summary, what was measured and done per question, what it means, how sure it is, the charts and an appendix with every test and number.	Step 5 · Report
Proteomics & RNA-seq	Which proteins or genes differ between two groups: limma for intensities (MaxQuant, DIA-NN or any table), DESeq2 for raw counts. Volcano, PCA, heat map, full table and R code.	Lab tools
3D intensity surfaces	Fluorescence images as landscapes: side by side on one scale, overlaid, subtracted, and measured (mean, integrated intensity, area above a threshold, CTCF, pixel correlation).	Lab tools
Make a chart	Type or paste numbers, 31 chart types and 18 looks, every detail of the design, a significance test that fits the table with the stars drawn in.	Lab tools
Your own 3D surface	X, Y and Z values as a surface, wireframe, points, bars, contour or heat map; turn it with the mouse in every direction.	Lab tools
RNA sequencing	From a count table: library quality, normalisation, sample similarity, the top genes per group compared, DESeq2 for every pair (with a donor or batch factor, or an interaction), gene sets by over-representation or GSEA.	Lab tools
Machine learning	A neural network or four other models trained on your own table, judged by cross-validation against guessing and shuffled labels, kept in your own model library, used for new samples.	Lab tools
Protein structure	A PDB or mmCIF file: chains, sequences, size, backbone angles, contacts, disulfides, glycosylation sites, pLDDT; the backbone turns with the mouse.	Lab tools
Calculator, primers, panels, planning, proteins, cells	25 tools: dilutions and solutions, primers, flow cytometry and immunofluorescence panels, sample size, bootstrap and permutation, protein properties, glycosylation sites and glycan masses, doubling time, cell counts and seeding. Each tool explains itself and fills in a worked example.	Lab tools
Privacy and language	Everything stays on this computer. English and Dutch, dark and light.	Everywhere

New in 4A.11

- **Make a chart** from numbers you type or paste: 31 chart types, 18 looks, every design detail, and a fitting significance test with the stars drawn in.

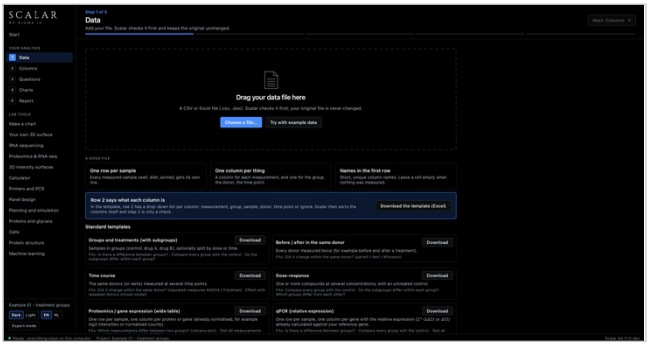
- **Your own 3D surface** from X, Y and Z values; every 3D view (also the intensity surfaces and protein structures) now turns with the mouse.
- **RNA sequencing**: quality, normalisation, top genes per group compared, DESeq2 for every pair with a donor factor or an interaction, GSEA.
- **Machine learning**: a neural network and four other models, honest cross-validation, a personal model library, also on your project's data.
- **Protein structure**, **proteins and glycans** and **cells**: new lab pages and tools.
- **Every lab tool explains itself** and fills in a worked example; the file pages have a guide card with an example and a template.
- **One start example** on which every question and every test answers (file 13).

2 · HOW IT WORKS

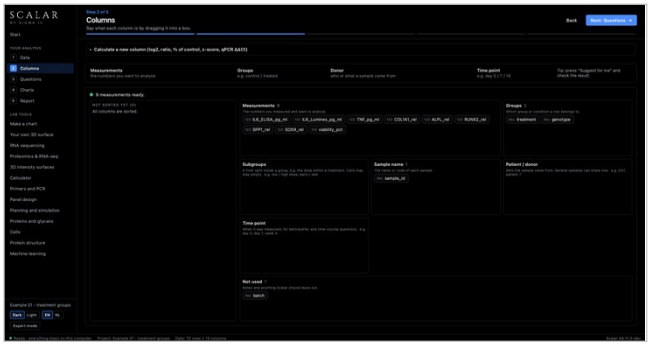
Five steps from file to report

The sidebar shows the route. You can go back at any time; earlier answers are kept.

	Step	What happens
1	**Data**	Drop a CSV or Excel file. Scalar shows what it found and what it would repair (empty rows, decimal commas, text in number columns).
2	**Columns**	Check the boxes. With a role row in the file they are already filled. Calculate new columns here.
3	**Questions**	Drag a question, a test or a pattern analysis onto the board and choose the columns. The answer appears in the panel.
4	**Charts**	Every answered question has a chart. Change its look, save a template, copy or export it.
5	**Report**	Make a PDF report of a board.



Step 1 - Data: drop a file; templates show a good layout.



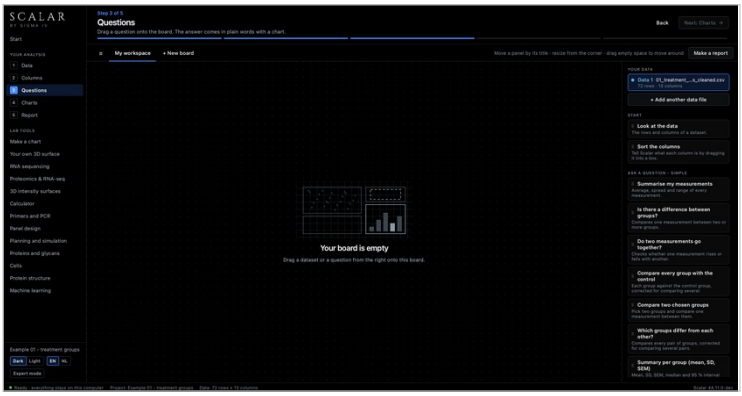
Step 2 - Columns: every column in its box (file 01).

A file Scalar understands at once

One row per sample, one column per thing you know or measured. Row 2 may name the role of each column - then step 2 is already done. All example files are built this way. The first rows of **01_treatment_groups.csv**:

sample_id	treatment	genotype	IL6_ELISA_pg_ml	IL6_Luminex_pg_ml	TNF_pg_ml	viability_pct	batch
sample	group	group	measurement	measurement	measurement	measurement	ignore
S001	Control	WT	29.3	31.5	20.51	93.1	B1
S002	Control	WT	40.4	47.2	16.02	90.5	B2
S003	Control	WT	36.8	42.8	28.86	92.9	B3

Row 2 (blue) is the role row: sample, group, subgroup, measurement, donor, time point or ignore. Six of the thirteen columns are left out here.



Step 3 - Questions: the board, with the questions, tests and pattern analyses to drag onto it.

The board

Sections of the list: Start, Ask a question (simple, a step further, advanced), Choose a test yourself, Find patterns, Compare, Other.

In a panel you choose the measurement and the groups, and where it applies 'Compare averages' or 'Compare ranks' and 'Spread may differ' or 'Same spread'.

A board can hold many panels; 'Make a report' turns the whole board into a PDF.

3 · READING AN ANSWER

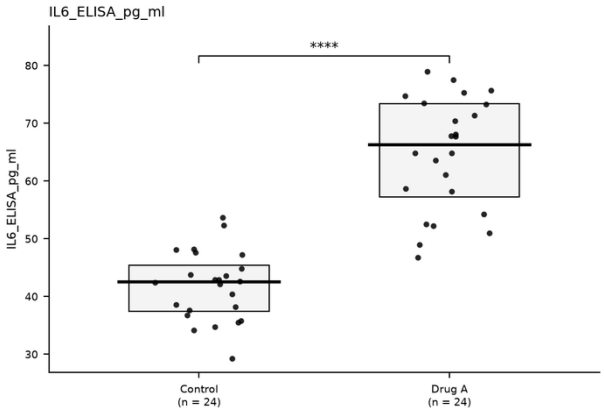
A sentence, a label, a chart - and every step behind it

The example: the Welch t-test on file 01, IL6_ELISA_pg_ml, Control against Drug A.

IL6_ELISA_pg_ml is higher in Drug A than in Control.

● **Clear pattern** The difference is large compared with the spread, and unlikely to be chance alone.

Typical value (median)	**Control 42.50, Drug A 66.25**
Number of samples	**Control 24, Drug A 24**
p	**< 0.001**
Effect size	**d = -2.81 (large)**
Difference in averages (95% CI)	**-22.84 (-27.59 – -18.08)**



Why this test. Two groups and 'Compare averages' with 'Spread may differ': the Welch t-test compares the two averages without assuming that both groups spread equally. That is the safe standard.

How sure. p is below 0.01 and the effect is large (d = -2.81), so by the fixed rule: Clear pattern.

How big. Drug A is on average 55 % higher than Control (64.64 vs 41.80). Fold change 1.55. 100 % of its samples lie above the middle value of Control.

The label rule

One fixed rule for every answer: **clear pattern** = p < 0.01 with at least a medium effect size; **possible pattern - check again** = p < 0.05, or p < 0.10 with at least a medium effect; **no clear pattern** otherwise; **can't tell yet** with fewer than 3 values in a group. Purely descriptive answers get the label **description**. With many measurements at once the corrected p-value counts. Stars in a chart: * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001. A label is a reading aid, not a verdict.

The explanation in plain words

'Explanation' in a panel says what every number means, without jargon:

Number	Value	What it means
Typical value (median)	Control 42.50, Drug A 66.25	The middle value per group: half of the samples lie above it, half below.
Number of samples	Control 24, Drug A 24	How many samples counted. Empty cells are left out.
p	< 0.001	Very small: a difference like this hardly ever arises by chance.
Effect size	d = -2.81 (large)	How big the difference is, regardless of how many samples there are: large.
Difference in averages (95% CI)	-22.84 (-27.59 – -18.08)	The difference between the two averages. The true difference most likely lies between the two numbers in brackets.

What you see. Highest middle value: Drug A (66.25); lowest: Control (42.50). The spread differs: widest in Drug A, narrowest in Control. A wide spread makes a difference harder to see.

What it may mean. Both statistically clear and large: a strong candidate pattern for follow-up.

3 · READING AN ANSWER

See the calculation

The menu (…) of a panel opens the calculation. Nothing is recomputed for display: every number comes from the trace saved with the analysis. The first steps of this Welch t-test:

Step	Formula	Filled in	Result
Size of each group	n_1, n_2	24, 24	**24 / 24**
Average of each group	$\bar{x} = \Sigma x / n$	Control: 41.8042 Drug A: 64.6417	**41.8042 / 64.6417**
Variance of each group	$s^2 = \Sigma(x - \bar{x})^2 / (n - 1)$	822.97 / 23 and 2220.84 / 23	**35.7813 / 96.5582**
Difference in averages	$\bar{x}_1 - \bar{x}_2$	41.8042 − 64.6417	**-22.8375**
Standard error of the difference	$SE = \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	√(35.7813/24 + 96.5582/24)	**2.34822**
Degrees of freedom (Welch-Satterthwaite)	$df = (s_1^2/n_1 + s_2^2/n_2)^2 / [(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)]$	(1.49089 + 4.02326)² / …	**37.9879**

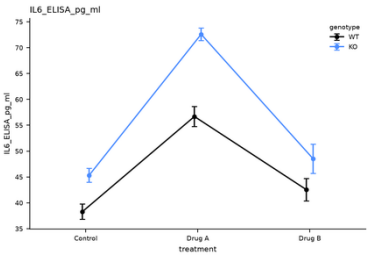
'Copy R code' gives code with the data written in, so it runs on its own (values shortened here):

```
# Welch t-test
y <- c(29.3, 40.4, 36.8, 44.8, 35.5, 47.6, 35.8, 34.7, 38.2, 43.6, 38.6, 34.2, 43.8, 37.6, 4 …)
g <- factor(c(rep("Control", 24), rep("Drug A", 24)), levels = c("Control", "Drug A"))
t.test(y ~ g, var.equal = FALSE)
```

'Excel check…' saves a workbook with the values used and Excel's own formulas next to Scalar's numbers, with ✓ where they agree. 'Save R script…' writes the same code to a file.

Some answers come as a table

Two-way ANOVA on file 01: IL6_ELISA_pg_ml, effect of treatment, second factor genotype.



The effect of treatment on IL6_ELISA_pg_ml depends on genotype (interaction).

Effect	F	df	p		partial η²	Result
treatment	81.79	2, 66	< 0.001	****	0.71	**Clear pattern**
genotype	37.88	1, 66	< 0.001	****	0.36	**Clear pattern**
treatment × genotype	4.05	2, 66	0.022	*	0.11	**Possible pattern - check again**

Each line is one effect with its F, degrees of freedom, p, stars and effect size, and gets its own label by the same rule. The chart shows the group means with their intervals: lines that are not parallel are the interaction.

4 · THE EXAMPLE DATA

Twenty-two files, three images, and inputs for the lab tools

Scalar_4A.11_example_data.zip. Each file is small, has a role row, and is built for a design that certain tests need. README.txt in the zip repeats the tables of this guide as plain text.

File	Design	Made for
01_treatment_groups.csv	72 samples: treatment (Control, Drug A, Drug B) × genotype (WT, KO), 9 measurements	Two measurements of IL-6 (ELISA and Luminex), TNF, five expression markers and viability. The all-round file: group comparisons, two-way ANOVA, correlation, regression, method agreement, testing many measurements, and the four pattern analyses.
02_before_after_paired.csv	18 donors, each measured before and after	CRP, IL-6 and a pain score per donor at two moments. For the paired tests.
03_time_course_repeated.csv	14 donors, each measured on day 0, 3, 7 and 14	ALP activity, calcium and cell count over time. For repeated measures and the mixed model.
04_dose_response.csv	40 wells: viability at 10 doses, four wells per dose	A falling dose-response curve with a true IC50 of 2.0 µM.
05_survival.csv	90 patients: Control and Drug, time in months, event 1 / 0	Follow-up times with censoring. For Kaplan-Meier and the log-rank test.
06_categories_chi_square.csv	80 mice: genotype, phenotype and sex as categories	Counts per category. For chi-square / Fisher's exact test.
07_biomarker_roc.csv	80 samples: healthy and disease, two candidate markers	Marker A separates well, marker B weakly. For ROC curves and AUC.
08_recovery_one_sample.csv	20 quality-control samples, recovery in %	Recovery sits below 100 %. For the one-sample tests against a fixed value.
09_qpcr_ct_values.csv	16 samples: control and osteogenic, raw Ct values of 3 genes + GAPDH	For the qPCR ΔΔCt calculation (new columns) and the test that follows.
10_groups_with_subgroups.csv	50 samples: three treatments with dose and time as subgroups	For subgroup charts and 'Do the subgroups differ within each group?'.
11_proteomics_intensities.csv	600 proteins × 10 samples (Ctrl_1–5, Treat_1–5), intensities	For Lab tools → Proteomics & RNA-seq (limma).
12_rnaseq_counts.csv	1500 genes × 8 samples (Ctrl_1–4, Treat_1–4), raw counts	For Lab tools → Proteomics & RNA-seq (DESeq2).
13_every_question_study.csv	24 donors × 3 treatments × day 0 / 7 / 14 (72 rows), with dose subgroups	The start example of 4A.11: every question and every named test answers on this one file (also 'Try with example data' on the start page).
14_chart_columns.csv	Three groups typed as columns	For Lab tools → Make a chart (one group per column) and its significance tests.
15_chart_xy.csv	Time with three series of replicates	For Make a chart in the XY format (lines, regression).
16_chart_grouped.csv	Conditions × two series	For Make a chart in the grouped format (grouped bars, two-way ANOVA).
17_chart_parts.csv	Five parts of a whole	For Make a chart: pie, donut, stacked share; chi-square against an even split.
18_surface_xyz.csv	Dose × time with a response (36 points)	For Lab tools → Your own 3D surface (a list of X, Y, Z).
19_surface_matrix.csv	The same as a Z matrix	For Your own 3D surface (X along the top, Y down the side).
20_rnaseq_three_groups_counts.csv	1200 genes × 12 samples: three groups, four donors, a Length column; + gene sets (.gmt)	For Lab tools → RNA sequencing: quality, top genes, DESeq2 per pair with the donor factor, over-representation and GSEA.

File	Design	Made for
21_machine_learning_patients.csv	120 made-up patients: five measurements, a site and the response	For Lab tools → Machine learning (neural network and four other models).
22_two_helices_made_up.pdb	A made-up two-helix protein (PDB)	For Lab tools → Protein structure.

Also in the zip

images/ control.tif, treated_low.tif, treated_high.tif	Three 16-bit images of the same field, slightly shifted against each other; brighter spots with treatment. For 3D intensity surfaces.
lab_tools/ primer_template_900bp.fasta	A random 900 bp template for 'Design primers'.
lab_tools/ flow_panel_markers.txt, if_panel_targets.txt	Marker lists to paste into the panel designers.
lab_tools/ protein_sequence_made_up.fasta, cell_counts_over_time.txt	Inputs for the protein and cell tools.
lab_tools/ example_inputs.txt	What to type into each of the 25 lab tools, with the result to expect.
README.txt	Every file, every test and question with what to choose and the answer Scalar gives.

Why a file per design, and one for everything. A paired test needs the same donors measured more than once; a survival curve needs a time and an event column of 1 and 0; a ROC curve needs exactly two groups. Scalar only offers what the data can answer. Files 01-12 each show one design on its own; file 13 (the start example) holds all designs at once, so every question and test answers on it.

5 · CHOOSE A TEST YOURSELF

All 24 tests, each with a file that runs it

Board, section 'Choose a test yourself'. Drag the test onto the board, choose what the table says, and compare the answer.

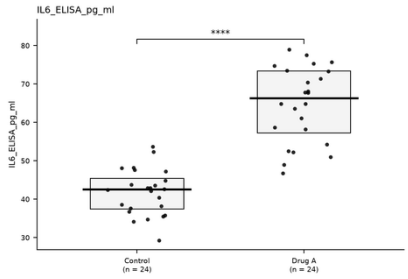
Test	File	What to choose	Scalar's answer	Label
Welch t-test	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; Group A Control, Group B Drug A	IL6_ELISA_pg_ml is higher in Drug A than in Control.	**Clear**
Student t-test	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; Group A Control, Group B Drug A	IL6_ELISA_pg_ml is higher in Drug A than in Control.	**Clear**
Mann-Whitney U test	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; Group A Control, Group B Drug A	IL6_ELISA_pg_ml is higher in Drug A than in Control.	**Clear**
Welch ANOVA + Games-Howell	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment	IL6_ELISA_pg_ml: these pairs differ (after correction): Control vs Drug A, Drug A vs Drug B.	**Clear**
One-way ANOVA + Tukey HSD	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment	IL6_ELISA_pg_ml: these pairs differ (after correction): Control vs Drug A, Drug A vs Drug B.	**Clear**
Kruskal-Wallis + pairwise Mann-Whitney	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment	IL6_ELISA_pg_ml: these pairs differ (after correction): Control vs Drug A, Drug A vs Drug B.	**Clear**
Dunnett (every group vs control)	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; control group Control	IL6_ELISA_pg_ml differs from Control in: Drug A.	**Clear**
Two-way ANOVA	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; second factor genotype	The effect of treatment on IL6_ELISA_pg_ml depends on genotype (interaction).	**Clear**
Paired t-test	**02** before after paired	Measurement CRP_mg_L; donor and moment are filled in	CRP_mg_L went down from before to after.	**Clear**
Wilcoxon signed-rank test	**02** before after paired	Measurement CRP_mg_L; donor and moment are filled in	CRP_mg_L went down from before to after.	**Clear**
Repeated-measures ANOVA	**03** time course repeated	Measurement ALP_activity; donor and day are filled in	ALP_activity changes over day: day 0 1.06 → day 3 1.24 → day 7 1.97 → day 14 2.69.	**Clear**
Friedman test	**03** time course repeated	Measurement ALP_activity; donor and day are filled in	ALP_activity changes over day: day 0 1.11 → day 3 1.20 → day 7 2.05 → day 14 2.77.	**Clear**
Linear mixed model	**03** time course repeated	Measurement ALP_activity; donor; factor day (also works on file 02)	day has an effect on ALP_activity, taking repeated donors into account.	**Clear**
Pearson correlation	**01** treatment groups	IL6_ELISA_pg_ml (x) and IL6_Luminex_pg_ml (y)	Higher IL6_ELISA_pg_ml goes together with higher IL6_Luminex_pg_ml.	**Clear**
Spearman correlation	**01** treatment groups	ALPL_rel and SPP1_rel	Higher ALPL_rel goes together with higher SPP1_rel.	**Clear**
Linear regression	**01** treatment groups	IL6_ELISA_pg_ml (x) and IL6_Luminex_pg_ml (y)	IL6_Luminex_pg_ml changes by 1.07 per unit of IL6_ELISA_pg_ml.	**Clear**

Test	File	What to choose	Scalar's answer	Label
Bland-Altman	**01** treatment groups	IL6_ELISA_pg_ml (x) and IL6_Luminex_pg_ml (y)	On average IL6_Luminex_pg_ml is 8.70 compared with IL6_ELISA_pg_ml; 95 % of the differences lie between 0.83 and 16.58.	**Description**
One-sample t-test	**08** recovery one sample	Measurement recovery_pct; fixed value 100	recovery_pct is lower than 100.00.	**Clear**
One-sample Wilcoxon test	**08** recovery one sample	Measurement recovery_pct; fixed value 100	recovery_pct is lower than 100.00.	**Clear**
Chi-square / Fisher's exact test	**06** categories chi square	Categories genotype and phenotype	genotype and phenotype are related.	**Clear**
Dose-response (4PL, EC50/IC50)	**04** dose response	Dose dose_uM, measurement viability_pct	IC50 = 1.99 (95 % interval 1.79 – 2.20).	**Description**
Kaplan-Meier + log-rank	**05** survival	Time time_months, event event, groups from treatment	Survival differs between the groups of treatment.	**Clear**
ROC curve and AUC	**07** biomarker roc	Marker marker_A_ng_ml, groups from status	marker_A_ng_ml separates disease from healthy with an AUC of 0.89 (95 % interval 0.81 – 0.96).	**Clear**
Many tests + Benjamini-Hochberg	**01** treatment groups	Groups from treatment (all 9 measurements are tested)	8 of 9 measurements differ clearly between the groups, 0 possibly (after correction for testing many).	**Clear**

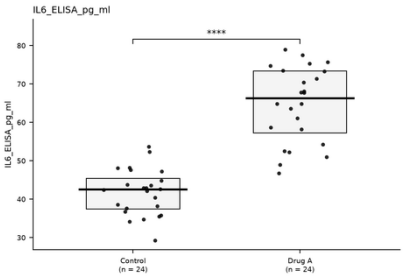
The Student t-test now carries Cohen's d (pooled SD), so it is labelled by the same rule as Welch's. The 4PL fit and Bland-Altman describe and do not test, hence 'Description'.

THE CHARTS OF THESE TESTS

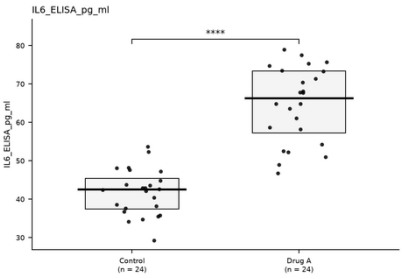
Drawn by Scalar in the light chart theme, on the files of the table.



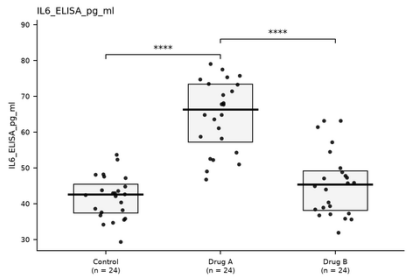
Welch t-test · file 01



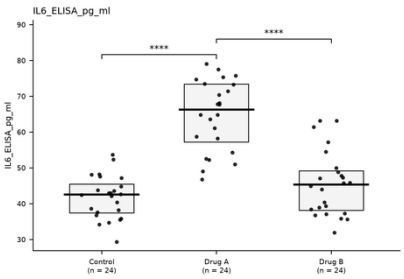
Student t-test · file 01



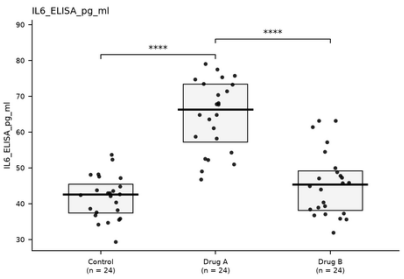
Mann-Whitney U test · file 01



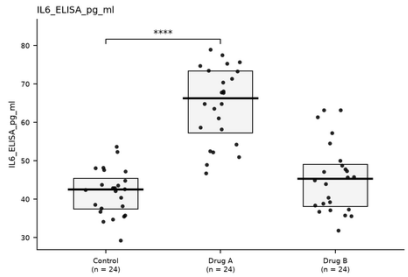
Welch ANOVA + Games-Howell · file 01



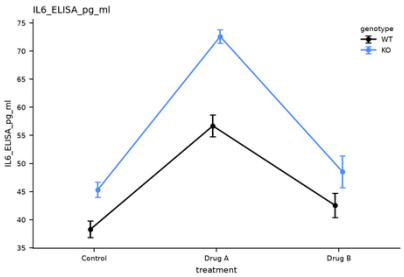
One-way ANOVA + Tukey HSD · file 01



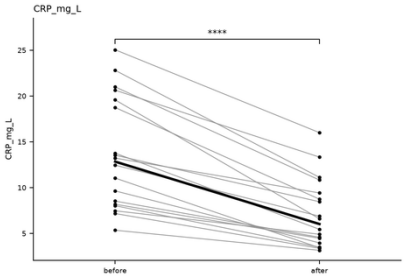
Kruskal-Wallis + pairwise Mann-Whitney · file 01



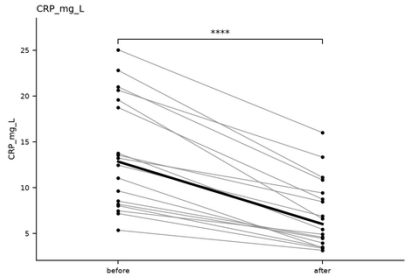
Dunnett (every group vs control) · file 01



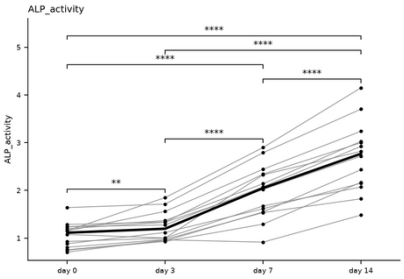
Two-way ANOVA · file 01



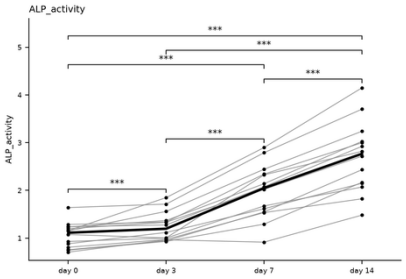
Paired t-test · file 02



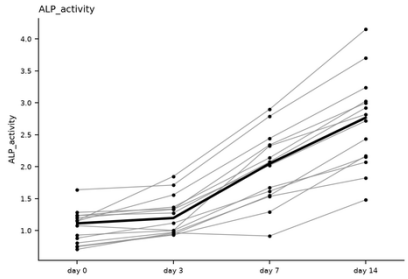
Wilcoxon signed-rank test · file 02



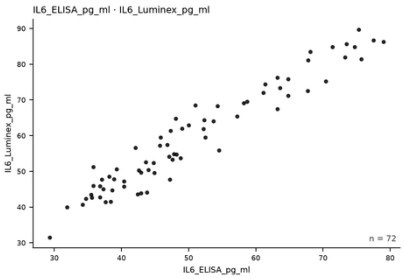
Repeated-measures ANOVA · file 03



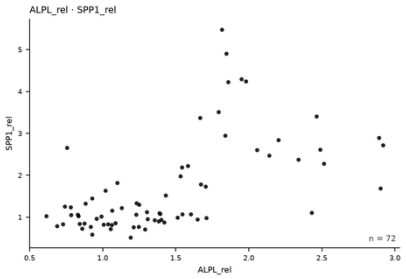
Friedman test · file 03



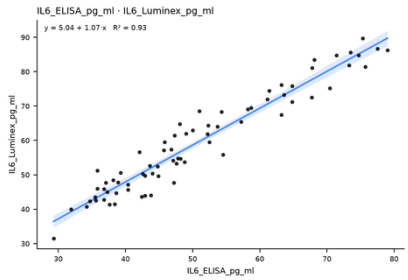
Linear mixed model · file 03



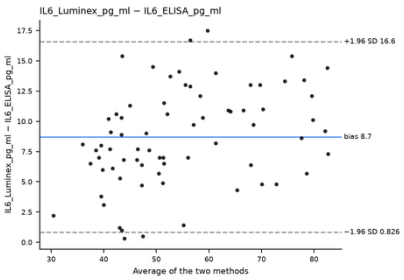
Pearson correlation · file 01



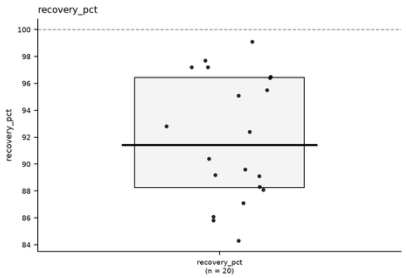
Spearman correlation · file 01



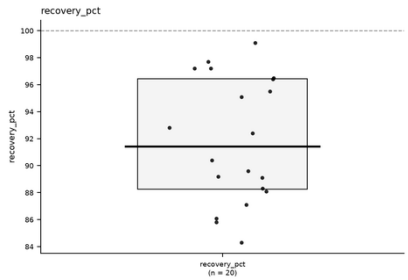
Linear regression · file 01



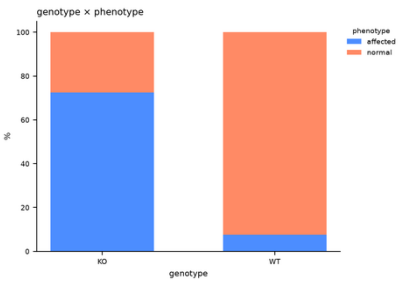
Bland-Altman · file 01



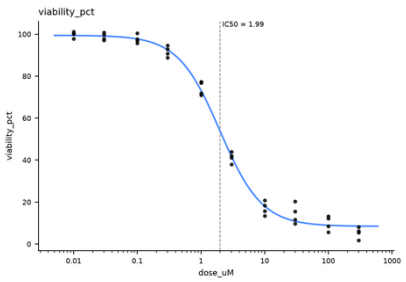
One-sample t-test · file 08



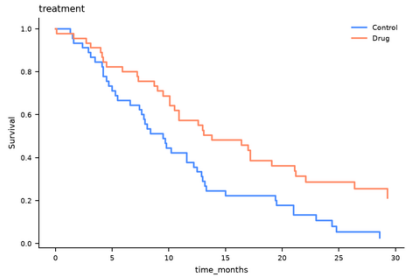
One-sample Wilcoxon test · file 08



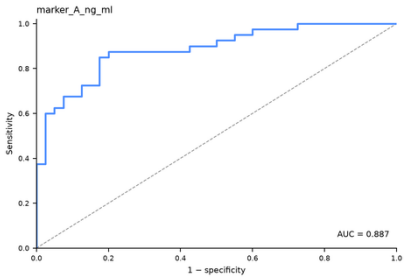
Chi-square / Fisher's exact test · file 06



Dose-response (4PL, EC50/IC50) · file 04



Kaplan-Meier + log-rank · file 05



ROC curve and AUC · file 07

Many tests + Benjamini-Hochberg: the answer is a table

File 01, groups from treatment: every measurement is tested and the p-values are corrected together. **8 of 9 measurements differ clearly between the groups, 0 possibly (after correction for testing many).**

Measurement	Largest change vs Control	p	p after correction	Result
IL6_ELISA_pg_ml	Drug A +55 %	< 0.001	< 0.001	**Clear pattern**
IL6_Luminex_pg_ml	Drug A +53 %	< 0.001	< 0.001	**Clear pattern**
TNF_pg_ml	Drug B -42 %	< 0.001	< 0.001	**Clear pattern**
ALPL_rel	Drug A +118 %	< 0.001	< 0.001	**Clear pattern**
RUNX2_rel	Drug A +58 %	< 0.001	< 0.001	**Clear pattern**
SPP1_rel	Drug A +160 %	< 0.001	< 0.001	**Clear pattern**
SOX9_rel	Drug B +105 %	< 0.001	< 0.001	**Clear pattern**
viability_pct	Drug B -9 %	< 0.001	< 0.001	**Clear pattern**
COL1A1_rel	Drug A +8 %	0.381	0.381	**No clear pattern**

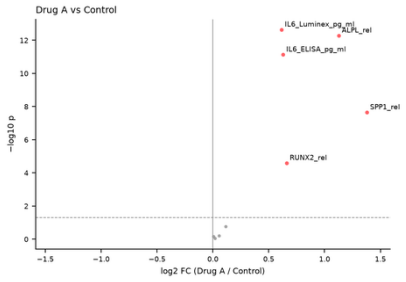
6 · QUESTIONS IN PLAIN WORDS

The 21 questions on the board

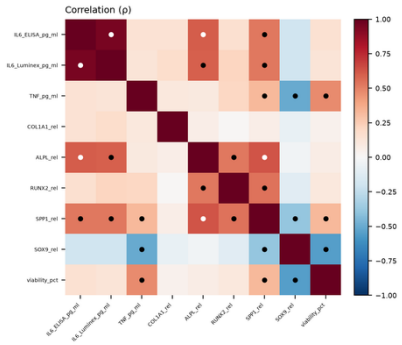
The same statistics, asked the way you would ask a colleague. Scalar chooses the test from the design: independent groups, the same donor measured again, or two measurements - and from your choice of averages or ranks.

Question	File	What to choose	Scalar's answer	Label
ASK A QUESTION - SIMPLE				
Summarise my measurements	**01** treatment groups	nothing to choose	Summary of 9 measurements.	**Description**
Is there a difference between groups?	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment	IL6_ELISA_pg_ml differs between the groups: highest in Drug A, lowest in Control.	**Clear**
Do two measurements go together?	**01** treatment groups	IL6_ELISA_pg_ml (x) and IL6_Luminex_pg_ml (y)	Higher IL6_ELISA_pg_ml goes together with higher IL6_Luminex_pg_ml.	**Clear**
Compare every group with the control	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; control Control	IL6_ELISA_pg_ml differs from Control in: Drug A.	**Clear**
Compare two chosen groups	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; Group A Control, Group B Drug A	IL6_ELISA_pg_ml is higher in Drug A than in Control.	**Clear**
Which groups differ from each other?	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment	IL6_ELISA_pg_ml: these pairs differ (after correction): Control vs Drug A, Drug A vs Drug B.	**Clear**
Summary per group (mean, SD, SEM)	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment	IL6_ELISA_pg_ml in 3 groups: mean, SD, SEM and 95 % interval.	**Description**
Is it different from a fixed value?	**08** recovery one sample	recovery_pct; fixed value 100	recovery_pct is lower than 100.00.	**Clear**
ASK A QUESTION - A STEP FURTHER				
Which measurements differ between two groups?	**01** treatment groups	Groups from treatment; Control against Drug A	5 of 9 measurements differ clearly between Control and Drug A, 0 possibly (after correction for testing many).	**Clear**
Do two factors work together? (two-way ANOVA)	**01** treatment groups	Measurement IL6_ELISA_pg_ml, groups from treatment; second factor genotype	The effect of treatment on IL6_ELISA_pg_ml depends on genotype (interaction).	**Clear**
Does one measurement predict another? (regression)	**01** treatment groups	IL6_ELISA_pg_ml (x) and IL6_Luminex_pg_ml (y)	IL6_Luminex_pg_ml changes by 1.07 per unit of IL6_ELISA_pg_ml.	**Clear**
Dose-response curve (EC50 / IC50)	**04** dose response	Dose dose_uM, measurement viability_pct	IC50 = 1.99 (95 % interval 1.79 – 2.20).	**Description**
Are two categories related? (chi-square / Fisher)	**06** categories chi square	genotype and phenotype	genotype and phenotype are related.	**Clear**
Do two methods agree? (Bland-Altman)	**01** treatment groups	IL6_ELISA_pg_ml (x) and IL6_Luminex_pg_ml (y)	On average IL6_Luminex_pg_ml is 8.70 compared with IL6_ELISA_pg_ml; 95 % of the differences lie between 0.83 and 16.58.	**Description**
Do the subgroups differ within each group?	**10** groups with subgroups	Measurement OPG, groups from treatment	OPG: the subgroups differ within: Drug A, Drug B.	**Clear**

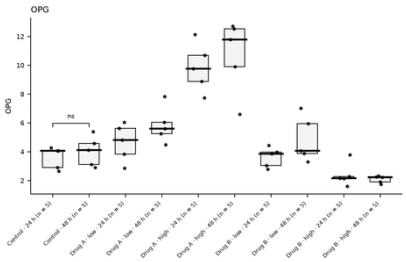
Question	File	What to choose	Scalar's answer	Label
Which measurements go together?	**01** treatment groups	nothing to choose	14 of 36 pairs go together (after correction); the strongest: IL6_ELISA_pg_ml and IL6_Luminex_pg_ml (ρ = 0.95).	**Clear**
Did it change within the same donor?	**02** before after paired	Measurement CRP_mg_L (file 03 gives the four-day version)	CRP_mg_L went down from before to after.	**Clear**
Test all measurements at once	**01** treatment groups	Groups from treatment	8 of 9 measurements differ clearly between the groups, 0 possibly (after correction for testing many).	**Clear**
ASK A QUESTION - ADVANCED				
Survival curves (Kaplan-Meier)	**05** survival	time_months, event, treatment	Survival differs between the groups of treatment.	**Clear**
How well does a marker separate two groups? (ROC)	**07** biomarker roc	marker_A_ng_ml, status	marker_A_ng_ml separates disease from healthy with an AUC of 0.89 (95 % interval 0.81 – 0.96).	**Clear**
Effect with repeated donors (mixed model)	**03** time course repeated	Measurement ALP_activity, donor, factor day	day has an effect on ALP_activity, taking repeated donors into account.	**Clear**



Which measurements differ between two groups? · file 01



Which measurements go together? · file 01



Do the subgroups differ within each group? · file 10

Summary per group: the table for a paper or a lab meeting

File 01, IL6_ELISA_pg_ml, groups from treatment. **IL6_ELISA_pg_ml in 3 groups: mean, SD, SEM and 95 % interval.**

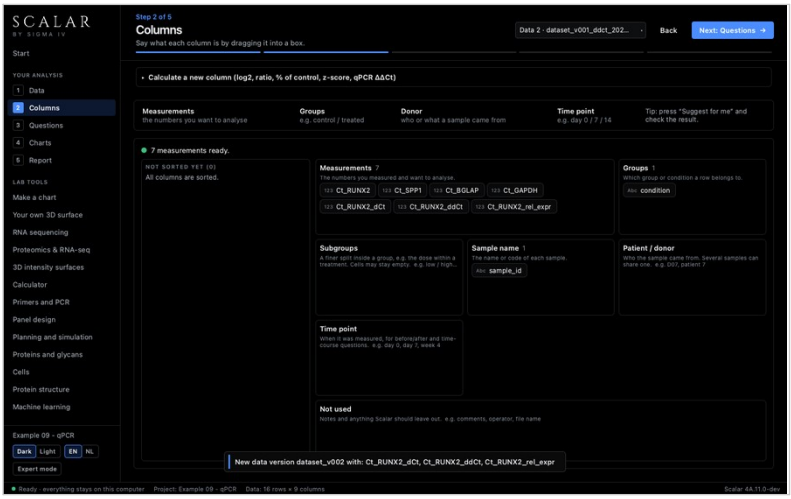
Group	n	Mean	SD	SEM	95 % interval	Median (IQR)
Control	24	41.80	5.98	1.22	39.28 – 44.33	42.50 (37.40 – 45.40)
Drug A	24	64.64	9.83	2.01	60.49 – 68.79	66.25 (57.23 – 73.35)
Drug B	24	45.52	9.07	1.85	41.69 – 49.35	45.30 (38.12 – 49.10)

7 · NEW COLUMNS AND PATTERNS

Calculate a column, then ask the question

Step 2, 'Calculate a new column'. Each calculation makes a new data version; your own columns and earlier answers stay as they are.

Calculation	Example	New column(s)
log2	File 01, column TNF_pg_ml	TNF_pg_ml_log2
% / fold of control	File 01, column ALPL_rel, groups from treatment, control Control	ALPL_rel_fold_vs_Control
qPCR ΔΔCt	File 09, gene Ct_RUNX2, reference gene Ct_GAPDH, groups from condition, control 'control'	Ct_RUNX2_dCt, Ct_RUNX2_ddCt, Ct_RUNX2_rel_expr
ratio, z-score	Any two measurements / any measurement	one new column each



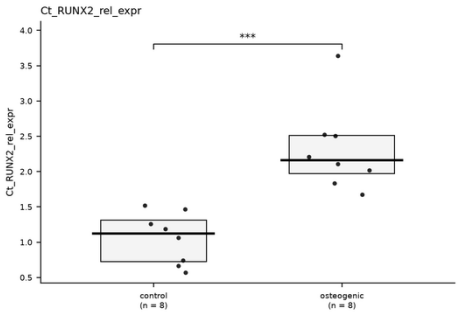
Step 2 after the ΔΔCt calculation on file 09: three new columns in a new data version.

qPCR from Ct values

ΔCt = Ct(Ct_RUNX2) − Ct(Ct_GAPDH); ΔΔCt = ΔCt − mean ΔCt of control (7.395); relative expression = 2^−ΔΔCt (Livak & Schmittgen 2001). Test on ΔCt (normally distributed); show 2^−ΔΔCt.

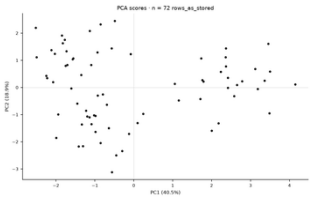
Then 'Compare two chosen groups' on Ct_RUNX2_rel_expr:

Ct_RUNX2_rel_expr is higher in osteogenic than in control.

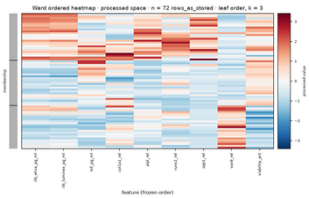


Find patterns

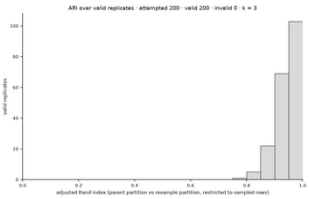
Four analyses that look at all measurements together, here on the 9 measurements of file 01. They describe structure; they do not test a hypothesis.



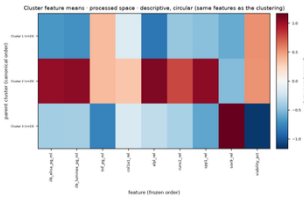
PCA · samples in the first two components



Ward clustering · ordered heat map



Cluster stability · 200 resamples



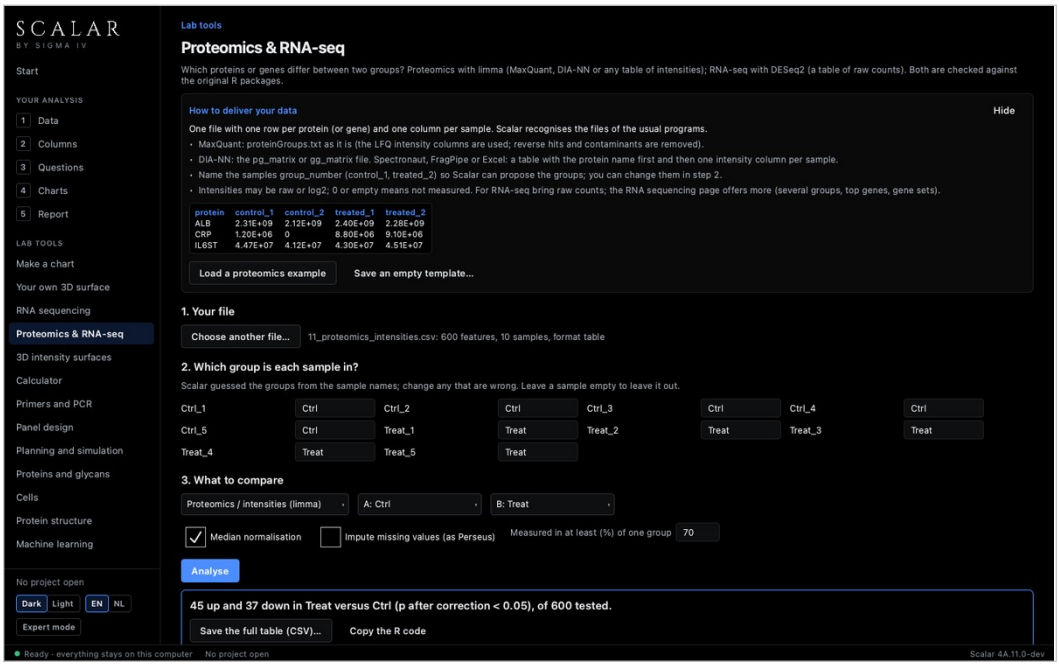
Profile · what marks each cluster

8 · LAB TOOLS

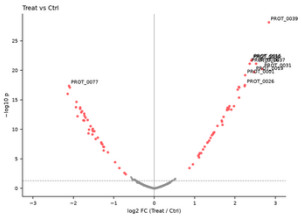
Proteomics & RNA-seq

Which proteins or genes differ between two groups? Choose a file, check the groups Scalar read from the sample names, press Analyse.

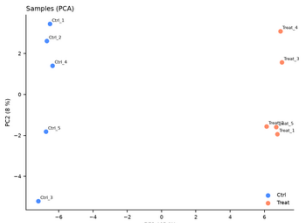
File	Method and steps	Scalar's answer (Treat versus Ctrl)	Strongest difference
11_proteomics_intensities.csv	limma: log2, filter on valid values, median normalisation, moderated t (empirical Bayes), Benjamini-Hochberg	**45 up and 37 down in Treat versus Ctrl (p after correction < 0.05), of 600 tested.**	PROT_0039: log2 fold change +2.83, corrected p 4.4e-26
12_rnaseq_counts.csv	DESeq2: filter low counts, size factors (median of ratios), negative binomial model with shrunken dispersions, Wald test, Benjamini-Hochberg	**55 up and 59 down in Treat versus Ctrl (p after correction < 0.05), of 1500 tested.**	GENE_0032: log2 fold change +2.64, corrected p 7.3e-18



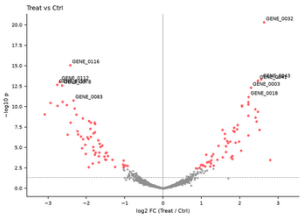
The proteomics example in the program: groups, settings, answer, volcano plot and table.



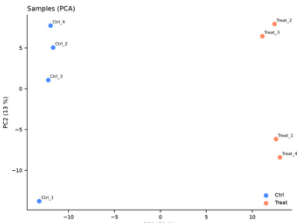
Proteomics · volcano



Proteomics · PCA of the samples



RNA-seq · volcano



RNA-seq · PCA of the samples

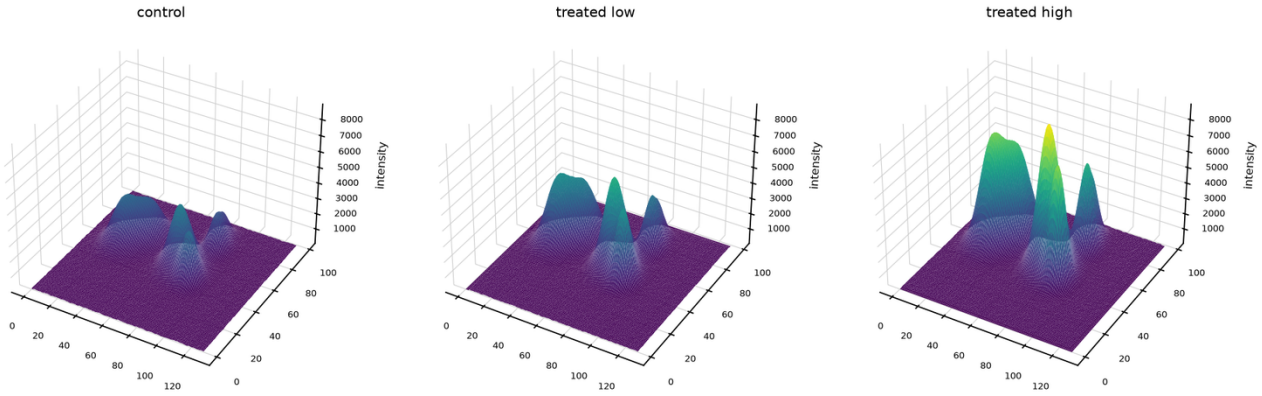
Also read directly: MaxQuant proteinGroups.txt, DIA-NN pg_matrix.tsv and featureCounts output. Missing values can be imputed (as Perseus). The full result table can be saved as CSV, and the R code copied.

> A plain table in which every value is a whole number is taken for RNA-seq counts. If your intensities happen to be whole numbers, set the mode on the page yourself. The example proteomics file has one decimal for that reason.

8 · LAB TOOLS

3D intensity surfaces

Fluorescence images as landscapes: the height is the intensity. Add the three example images in the order control, treated_low, treated_high.



Side by side on one height scale. Settings used: background = low percentile 1, smoothing σ = 1 px, align the images, threshold 500.

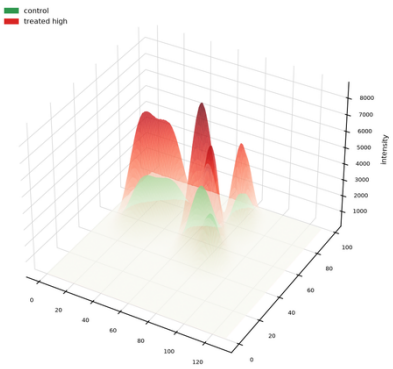
Image	Mean	Median	Max	Integrated	Area above 500
control	413.5	108.8	3944	2.15e+07	20.1 %
treated low	552.1	107.3	5654	2.87e+07	23.0 %
treated high	821.8	108.4	8945	4.27e+07	26.2 %

treated low vs control: mean intensity +33.5 %, pixel correlation r = 1.000.

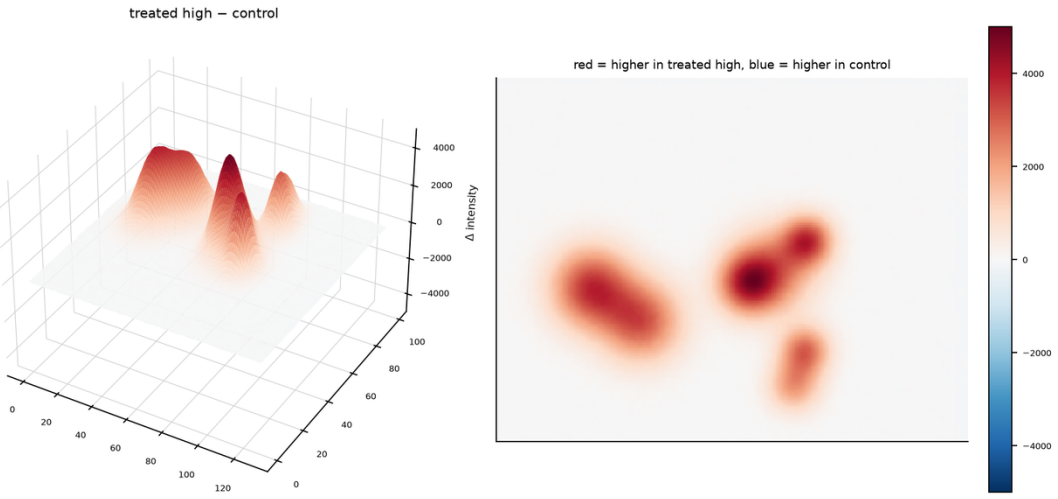
treated high vs control: mean intensity +98.8 %, 66 % of the pixels higher, largest increase 5003.

Processing, recorded with the result: background − 227 (percentile 1); negative values set to 0; Gaussian smoothing σ = 1 px; treated low aligned to control: shifted -4 rows and 6 columns; treated high aligned to control: shifted 3 rows and -7 columns.

One image per condition is a description, not a test. For a test, measure several images per group and compare their means on the board.



Overlay: control and treated high in one view



Difference B − A (treated high − control): as a surface and as a flat map.

8 · LAB TOOLS

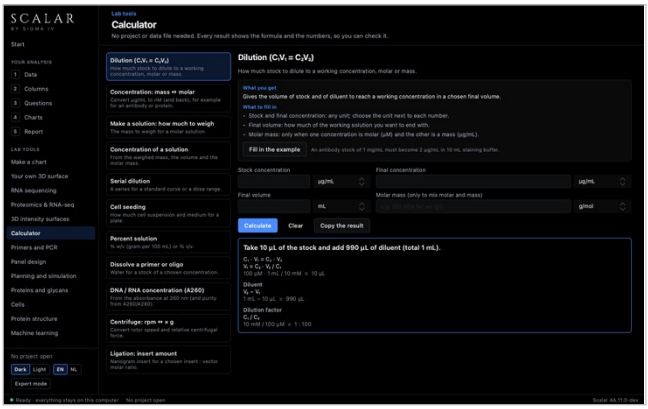
Calculator, primers, panels, planning, proteins and cells

25 tools that need no project. In the program every tool says what you get and what to fill in, and 'Fill in the example' runs it. Type the example input and compare; example_inputs.txt in the zip has the same list with the full tables.

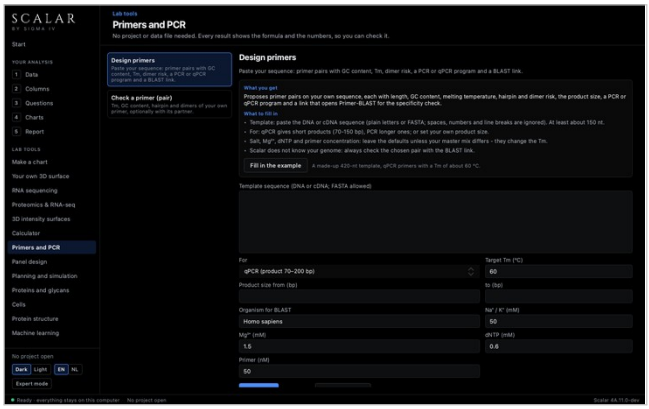
Tool	Example input	Result
CALCULATOR		
Dilution ($C_1V_1 = C_2V_2$)	Stock concentration: 10 mM; Final concentration: 100 µM; Final volume: 1 mL	Take 10 µL of the stock and add 990 µL of diluent (total 1 mL).
Concentration: mass ↔ molar	Concentration: 1 µg/mL; Molar mass: 150 kDa	1 µg/mL = 6.667 nM
Make a solution: how much to weigh	Concentration: 50 mM; Volume: 250 mL; Molar mass: 121.14 g/mol	Weigh 1.514 g and dissolve to 250 mL.
Concentration of a solution	Mass: 5 mg; Volume: 10 mL; Molar mass: 180.16 g/mol	The solution is 2.775 mM (500 µg/mL).
Serial dilution	Start concentration: 100 µM; Dilution factor (2 = 1:2): 3; Number of steps: 8; Volume per tube: 200 µL	Put 133.3 µL of diluent in each tube, then move 66.67 µL from tube to tube (final volume 200 µL per tube, except the last: 200 µL + 66.67 µL to discard).
Cell seeding	Cell count: 1.2 ×10^6 cells/mL; Cells per well: 10000; Number of wells: 96; Medium per well: 100 µL; Extra (%) for pipetting loss: 10	Mix 0.88 mL of cell suspension with 9.68 mL medium (10.56 mL in total) and put 100 µL in each well.
Percent solution	Percentage: 4; Volume: 500 mL; Kind: % w/v (solid)	Dissolve 20 g and make up to 500 mL.
Dissolve a primer or oligo	Amount on the tube: 25 nmol; Stock concentration: 100 µM	Dissolve in 250 µL of water or TE.
DNA / RNA concentration (A260)	A260: 0.42; A280 (optional): 0.21; Dilution factor: 10; Kind: RNA (40)	168 ng/µL RNA
Centrifuge: rpm ↔ × g	Convert: rpm → × g; Speed: 13000; Rotor radius (cm): 8.4	1.3e+04 rpm = 1.587e+04 × g
Ligation: insert amount	Vector (ng): 50; Vector length (bp): 5400; Insert length (bp): 900; Insert : vector (molar): 3	Use 25 ng insert with 50 ng vector (3 : 1 molar).
PRIMERS AND PCR		
Design primers	Template sequence: the sequence in primer_template_900bp.fasta; For: qPCR (product 70–200 bp)	Best pair: TGGGAATACGCGGCTATTCTAC / GCGATGCCATTCCGTAAGTAAC — product 170 bp, Tm 60 / 60 °C, dimer risk low.
Check a primer (pair)	Primer 5′→3′: TGACTGAACTCTCGAATAATT	TGACTGAACTCTCGAATAATT: 21 nt, GC 33.3 %, Tm 52.3 °C.
PANEL DESIGN		
Flow cytometry panel	Markers, one per line: the lines of flow_panel_markers.txt; Instrument: 4 lasers (405, 488, 561, 640)	Suggested 8-colour panel for a 4-laser instrument (405, 488, 561, 640 nm).
Immunofluorescence panel	Markers, one per line: the lines of if_panel_targets.txt; Nuclear stain: DAPI; Far-red: yes	Suggested 3-marker immunofluorescence panel with DAPI.
PLANNING AND SIMULATION		
How many samples do I need? (t-test)	Expected difference between the groups: 14; Standard deviation: 9; α (significance level): 0.05; Power (1 − β): 0.8	8 per group (16 in total) for 80 % power to find d = 1.56 at α = 0.05.
How many samples do I need? (ANOVA)	Number of groups: 3; Effect size f (0.10 / 0.25 / 0.40): 0.25; α: 0.05; Power: 0.8	53 per group (159 in total) for 80 % power in a one-way ANOVA with 3 groups (f = 0.25).

Tool	Example input	Result
Bootstrap confidence interval	Group A and B: the two lists of 12 values in example_inputs.txt; Statistic: mean; Resamples: 10000	the difference in mean (B − A): 18.4, 95 % bootstrap interval 13.93 to 23.
Permutation test (Monte Carlo)	Group A and B: the two lists of 12 values in example_inputs.txt; Permutations: 20000	Difference in means (B − A) = 18.4; permutation p = 5.000e-05 (two-sided).
PROTEINS AND GLYCANS		
Protein from its sequence	Protein sequence (one-letter code; FASTA allowed): the sequence in protein_sequence_made_up.fasta	84 residues, 9.7401 kDa, pI 7.92, A280 of 1 mg/mL = 1.435 (1 cm path).
N-glycosylation sites	Protein sequence (one-letter code; FASTA allowed): the sequence in protein_sequence_made_up.fasta	2 possible N-glycosylation site(s) (N-X-S/T, X not Pro) in 84 residues.
Glycan mass from its composition	Hex (Man, Gal, Glc): 5; HexNAc (GlcNAc, GalNAc): 4; Fuc (deoxyhexose): 1; NeuAc (sialic acid): 0; NeuGc: 0; Pent (Xyl): 0; HexA (GlcA, IdoA): 0; Sulfate groups: 0; Phosphate groups: 0; Label on the reducing end: 2-AB; Show as: [M+2H]2+	Hex5 HexNAc4 Fuc1: monoisotopic mass 1906.7188 Da, [M+2H]2+ at m/z 954.3667.
CELLS		
Doubling time from counts	Time and count, one time point per line (e.g. '0 50000'): the lines of cell_counts_over_time.txt; Time unit: hours	Doubling time 25.5 h (growth rate 0.0272 per h).
Cell count with trypan blue	Live (unstained) cells counted: 182; Dead (blue) cells counted: 14; Large squares counted: 4; Dilution factor (2 for trypan blue 1:1): 2; Volume of the suspension (mL): 5	0.91 ×10⁶ live cells/mL, viability 92.9 %.
How many cells to seed	Cells you need: 2000000; Hours until you need them: 72; Doubling time (hours): 24; Density of your suspension (cells/mL): 900000	Seed 2.5e+05 cells to have about 2.00e+06 after 72 h.

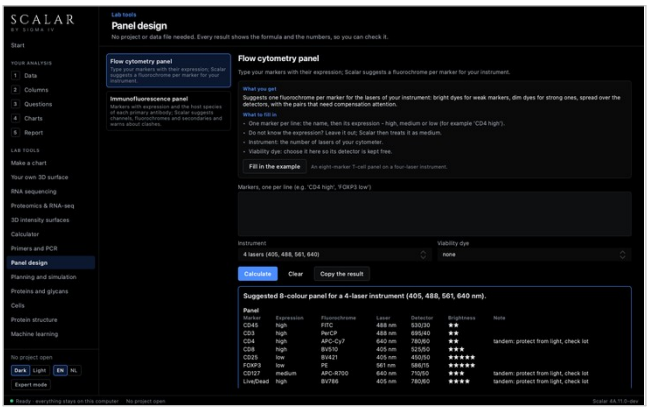
THE LAB TOOLS IN THE PROGRAM



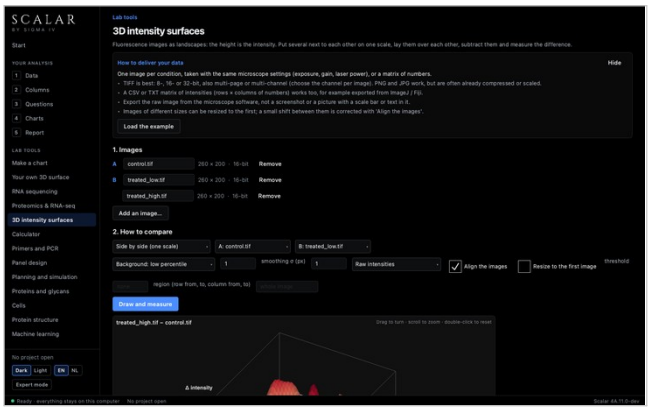
Calculator: a dilution, with the steps.



Primers and PCR: primer pairs for the example template.



Panel design: an 8-colour flow cytometry panel.



3D intensity surfaces: the difference view.

What the tools tell you besides the number

- **Primers**: length, GC content, Tm, product size, a PCR programme and the risk of dimers; melting temperatures match Primer3.
- **Panels**: the brightest fluorochromes go to markers with low expression; estimated spillover is listed. Always run single-stain and FMO controls on your own instrument.
- **Sample size**: the same answer as R's power.t.test, plus a check by simulation with the seed shown.
- **Bootstrap and permutation**: fixed seed, so the result repeats exactly.

9 · THE START EXAMPLE

One file on which every question and every test answers

'Try with example data' on the start page loads file 13: 24 made-up donors in three treatments, each measured on day 0, 7 and 14, with dose subgroups, two IL-6 methods, CRP and responders, follow-up and events, a dose-response pair and genes with and without an effect. Row 2 sorts the columns.

Every one of the 45 questions and named tests on the board answers with its first column choice (45 of 45 completed in the verification run). The patterns (PCA, clustering, stability, profile) run as well.

Question or test	Scalar's answer on file 13	Label
Summarise my measurements	Summary of 15 measurements.	**Description**
Is there a difference between groups?	IL6_ELISA_pg_mL differs between the groups: highest in Drug A, lowest in Drug B.	**Clear**
Do two measurements go together?	Higher IL6_ELISA_pg_mL goes together with higher IL6_Luminex_pg_mL.	**Clear**
Compare every group with the control	IL6_ELISA_pg_mL differs from Control in: Drug A, Drug B.	**Clear**
Compare two chosen groups	IL6_ELISA_pg_mL is higher in Drug A than in Control.	**Clear**
Which groups differ from each other?	IL6_ELISA_pg_mL: these pairs differ (after correction): Control vs Drug A, Control vs Drug B, Drug A vs Drug B.	**Clear**
Which measurements differ between two groups?	6 of 15 measurements differ clearly between Control and Drug A, 0 possibly (after correction for testing many).	**Clear**
Summary per group (mean, SD, SEM)	IL6_ELISA_pg_mL in 3 groups: mean, SD, SEM and 95 % interval.	**Description**
Is it different from a fixed value?	IL6_ELISA_pg_mL is higher than 0.00.	**Clear**
Do two factors work together? (two-way ANOVA)	The effect of treatment on IL6_ELISA_pg_mL depends on time (interaction).	**Clear**
Does one measurement predict another? (regression)	IL6_Luminex_pg_mL changes by 1.08 per unit of IL6_ELISA_pg_mL.	**Clear**
Dose-response curve (EC50 / IC50)	EC50 = 1.48 (95 % interval 1.35 – 1.62).	**Description**
Are two categories related? (chi-square / Fisher)	treatment and responder are related.	**Clear**
Do two methods agree? (Bland-Altman)	On average IL6_Luminex_pg_mL is 1.88 compared with IL6_ELISA_pg_mL; 95 % of the differences lie between -0.35 and 4.11.	**Description**
Survival curves (Kaplan-Meier)	No clear difference in survival between the groups of treatment.	**None**
How well does a marker separate two groups? (ROC)	IL6_ELISA_pg_mL separates yes from no with an AUC of 0.58 (95 % interval 0.35 – 0.81).	**None**
Do the subgroups differ within each group?	IL6_ELISA_pg_mL: the subgroups differ within: Drug A, Drug B.	**Possible**
Which measurements go together?	27 of 105 pairs go together (after correction); the strongest: IL6_ELISA_pg_mL and IL6_Luminex_pg_mL (ρ = 0.99).	**Clear**
Did it change within the same donor?	IL6_ELISA_pg_mL changes over time: day 0 20.62 → day 7 22.91 → day 14 25.04.	**Possible**
Test all measurements at once	9 of 15 measurements differ clearly between the groups, 0 possibly (after correction for testing many).	**Clear**
Effect with repeated donors (mixed model)	treatment has an effect on IL6_ELISA_pg_mL, taking repeated donors into account.	**Clear**
Welch t-test	IL6_ELISA_pg_mL is higher in Drug A than in Control.	**Clear**

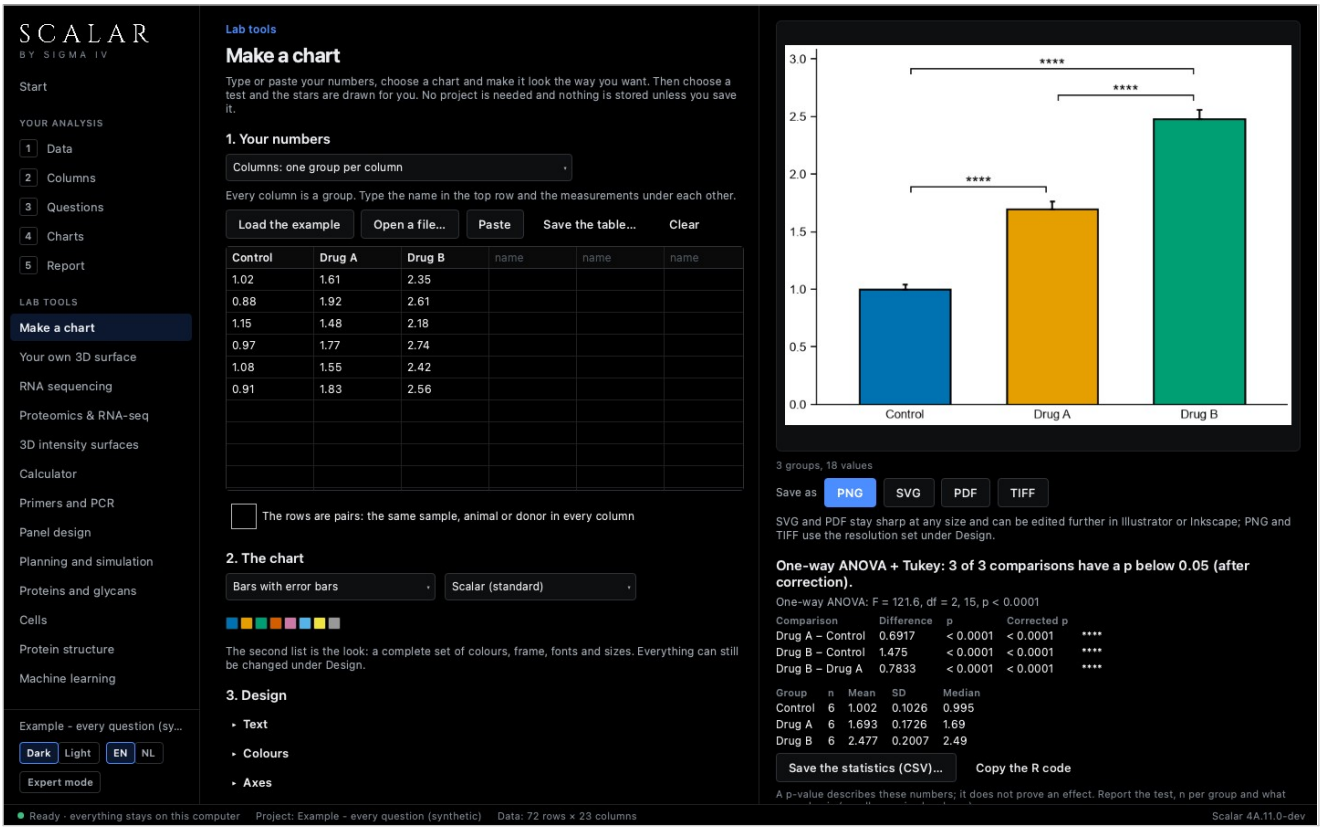
Question or test	Scalar's answer on file 13	Label
Student t-test	IL6_ELISA_pg_mL is higher in Drug A than in Control.	**Clear**
Mann-Whitney U test	IL6_ELISA_pg_mL is higher in Drug A than in Control.	**Clear**
Welch ANOVA + Games-Howell	IL6_ELISA_pg_mL: these pairs differ (after correction): Control vs Drug A, Control vs Drug B, Drug A vs Drug B.	**Clear**
One-way ANOVA + Tukey HSD	IL6_ELISA_pg_mL: these pairs differ (after correction): Control vs Drug A, Drug A vs Drug B.	**Clear**
Kruskal-Wallis + pairwise Mann-Whitney	IL6_ELISA_pg_mL: these pairs differ (after correction): Control vs Drug A, Control vs Drug B, Drug A vs Drug B.	**Clear**
Paired t-test	IL6_ELISA_pg_mL went up from day 0 to day 7.	**Possible**
Wilcoxon signed-rank test	IL6_ELISA_pg_mL went up from day 0 to day 7.	**Possible**
Repeated-measures ANOVA	IL6_ELISA_pg_mL changes over time: day 0 20.62 → day 7 22.91 → day 14 25.04.	**Possible**
Friedman test	No clear change in IL6_ELISA_pg_mL over time.	**None**
Pearson correlation	Higher IL6_ELISA_pg_mL goes together with higher IL6_Luminex_pg_mL.	**Clear**
Spearman correlation	Higher IL6_ELISA_pg_mL goes together with higher IL6_Luminex_pg_mL.	**Clear**
Linear mixed model	treatment has an effect on IL6_ELISA_pg_mL, taking repeated donors into account.	**Clear**
Dunnett (every group vs control)	IL6_ELISA_pg_mL differs from Control in: Drug A, Drug B.	**Clear**
Two-way ANOVA	The effect of treatment on IL6_ELISA_pg_mL depends on time (interaction).	**Clear**
One-sample t-test	IL6_ELISA_pg_mL is higher than 0.00.	**Clear**
One-sample Wilcoxon test	IL6_ELISA_pg_mL is higher than 0.00.	**Clear**
Chi-square / Fisher's exact test	treatment and responder are related.	**Clear**
Linear regression	IL6_Luminex_pg_mL changes by 1.08 per unit of IL6_ELISA_pg_mL.	**Clear**
Dose-response (4PL, EC50/IC50)	EC50 = 1.48 (95 % interval 1.35 – 1.62).	**Description**
Kaplan-Meier + log-rank	No clear difference in survival between the groups of treatment.	**None**
ROC curve and AUC	IL6_ELISA_pg_mL separates yes from no with an AUC of 0.58 (95 % interval 0.35 – 0.81).	**None**
Bland-Altman	On average IL6_Luminex_pg_mL is 1.88 compared with IL6_ELISA_pg_mL; 95 % of the differences lie between -0.35 and 4.11.	**Description**
Many tests + Benjamini-Hochberg	9 of 15 measurements differ clearly between the groups, 0 possibly (after correction for testing many).	**Clear**

Donor-level outcomes (responder, follow-up, event) are filled in on the day-0 row only, so each donor counts once; PCA and clustering therefore use the 24 day-0 rows, where every measurement is filled in.

10 · MAKE A CHART

Your own numbers, any chart, a fitting test

Lab tools → Make a chart. Type or paste the numbers (or open a file), pick a chart and a look, change any detail under Design, then choose a test: only tests that fit the table are offered, and the stars are drawn in the chart.



File 14 with a one-way ANOVA + Tukey: the stars come from the test on the right.

Four ways to type a table

- **Columns** (file 14): one group per column - bars, box, violin, raincloud, before-after, histogram, Q-Q and more.
- **XY** (file 15): an X column and Y series; replicate columns share a name - lines with error bars, scatter with a fit.
- **Grouped** (file 16): rows × series - grouped and stacked bars, grouped box plots.
- **Parts of a whole** (file 17): pie, donut, stacked share.

Tests offered for file 14: Welch ANOVA + Games-Howell, One-way ANOVA + Tukey, Kruskal-Wallis + Dunn (Holm), One-way ANOVA + Dunnett, Pairwise Welch t-tests (Holm), Pairwise t-tests, pooled SD (Bonferroni).

Result for file 14: One-way ANOVA + Tukey: 3 of 3 comparisons have a p below 0.05 (after correction). One-way ANOVA: F = 121.6, df = 2, 15, p < 0.0001

18 looks (journal styles with NPG, AAAS, NEJM and Lancet colours, poster, greyscale, thesis ...) are starting points; every colour, axis, error bar, point, font and size can then be changed. Save as PNG, SVG, PDF or TIFF; the statistics as CSV; the R code is copied with one click.

11 · YOUR OWN 3D SURFACE

X, Y and Z values as a landscape that turns with the mouse

Lab tools → Your own 3D surface. A list of points (file 18) or a Z matrix (file 19). Drag to turn the surface in every direction, scroll to zoom, double-click to reset; the saved picture uses the view you see.



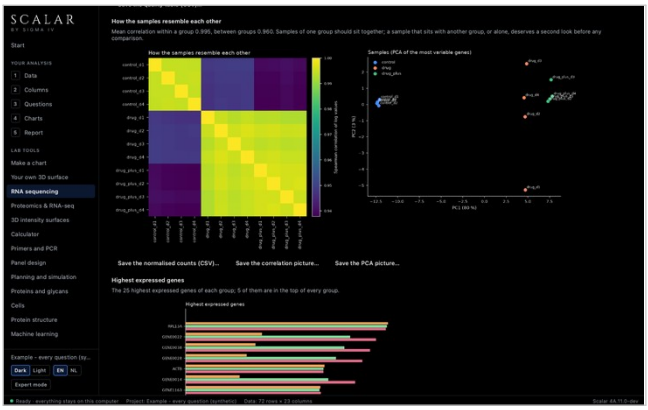
File 18 as a surface, turned with the mouse.

- 36 points; Z from 0 to 81.8.
- The points form a complete grid: the surface goes exactly through your values.
- Kinds: Surface, Surface with grid lines, Wireframe, Surface with a contour map underneath, 3D points, 3D bars, 3D contour lines, Contour map (flat), Heat map (flat).
- The intensity surfaces of fluorescence images and the protein backbone turn the same way; images side by side turn together.

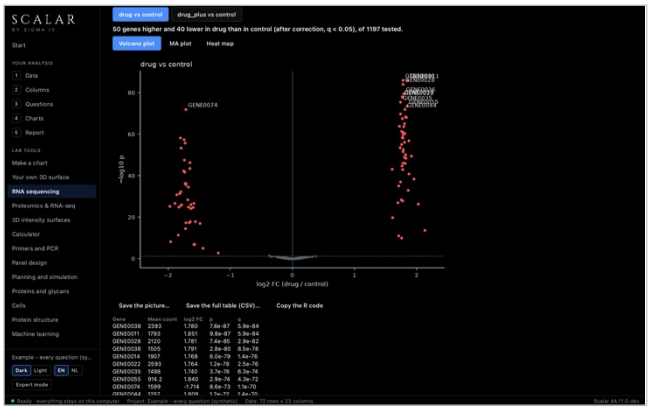
12 · RNA SEQUENCING

From a count table to genes, groups and gene sets

Lab tools → RNA sequencing. Bring the table of raw counts from your pipeline (one row per gene, one column per sample). Scalar does not align reads (FASTQ).



Explore: how the samples resemble each other (correlation and PCA).



Compare: DESeq2 for every group against the first, within donor.

What it does on file 20

- 1200 genes × 12 samples read from 20_rnaseq_three_groups_counts.csv.
- 1186 genes kept of 1198; DESeq2's median of ratios (the usual choice for comparing samples).
- The 25 highest expressed genes of each group; 5 of them are in the top of every group.
- 50 genes higher and 40 lower in drug than in control (after correction, q < 0.05), of 1197 tested. (~batch + condition)
- 50 genes higher and 40 lower in drug_plus than in control (after correction, q < 0.05), of 1197 tested. (~batch + condition)
- Over-representation: 2 of 2 gene sets are over-represented (q < 0.05) among 90 genes from drug vs control.
- GSEA: 2 of 3 gene sets are enriched at the top or bottom of the ranked list (q < 0.05) in drug vs control.

Gene set (GSEA)	Genes	NES	q
MADE_UP_UP_WITH_DRUG	50	1.94	4.4e-04
MADE_UP_DOWN_WITH_DRUG	40	-2.11	4.4e-04
MADE_UP_UNCHANGED	100	-0.75	0.948

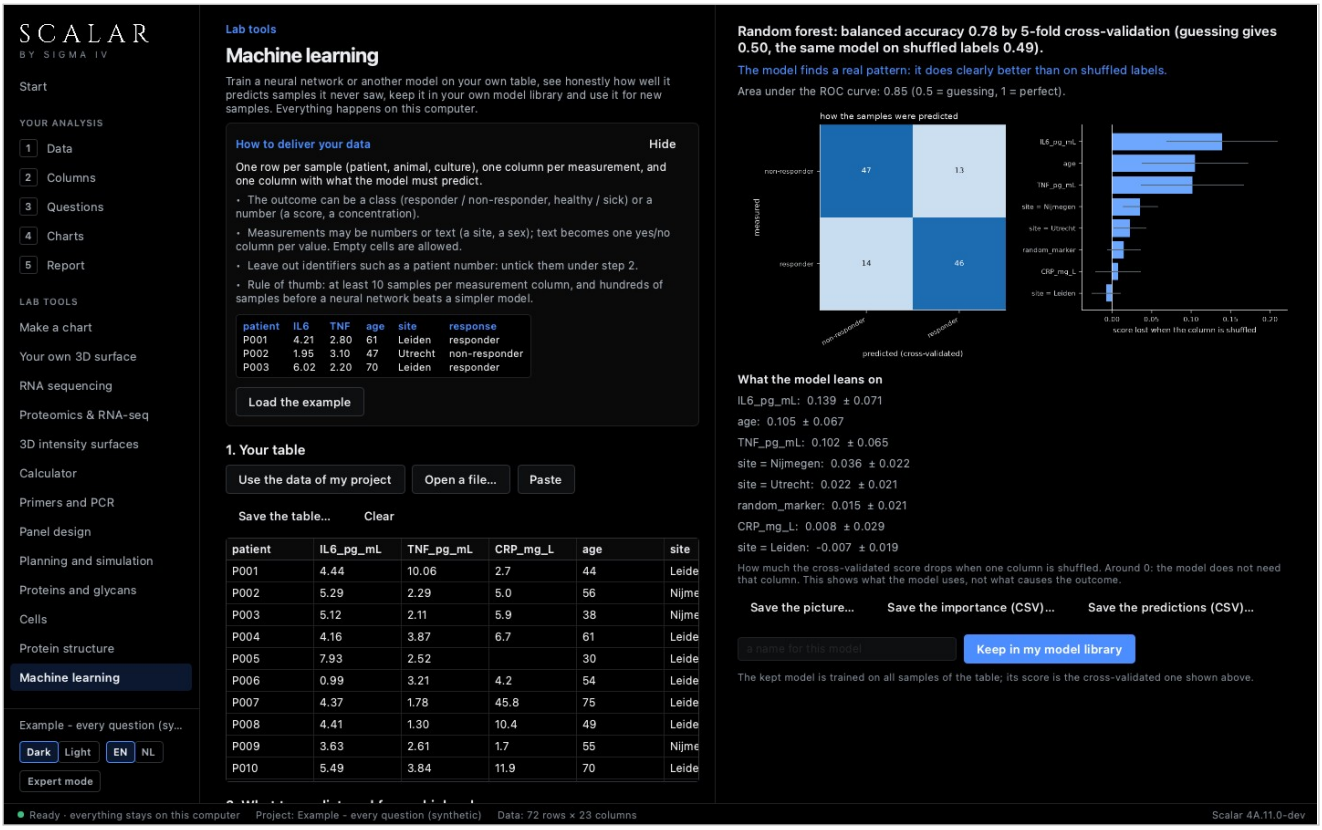
Also on this page

- Normalisation by DESeq2's median of ratios, CPM or TPM (with a Length column); leave out genes such as MT- or RPL.
- The top genes of every group side by side with their rank, and the genes most specific to one group.
- A donor or batch factor in the model; a PCA after limma's removeBatchEffect; an interaction test (does the difference depend on a second factor?).
- Your own genes per group; the overlap between comparisons; gene IDs turned into names with your own table; the normalised table as CSV.
- Checked against R: DESeq2 (also with a donor factor and an interaction), limma's removeBatchEffect, phyper for over-representation.

13 · MACHINE LEARNING

A model trained on your own table, judged honestly

Lab tools → Machine learning. One row per sample, one column per measurement, and the outcome to predict. Everything stays on this computer.



File 21: a random forest predicts the response; the check is cross-validated.

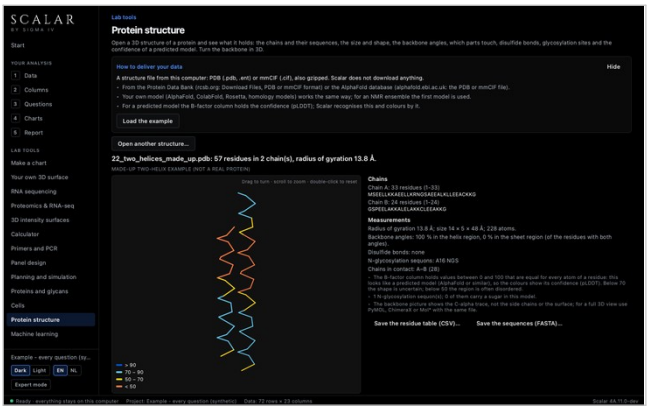
Model on file 21	Balanced accuracy	Shuffled labels	AUC	Verdict
Linear model	0.79	0.47	0.91	The model finds a real pattern: it does clearly better than on shuffled labels.
Random forest	0.78	0.49	0.85	The model finds a real pattern: it does clearly better than on shuffled labels.
Neural network	0.72	0.49	0.80	The model finds a real pattern: it does clearly better than on shuffled labels.

- Every sample is predicted by a model that never saw it (k-fold cross-validation); the same model on shuffled labels shows what chance gives.
- Permutation importance shows which columns the model leans on - what it uses, not what causes the outcome.
- Keep a model in your own library and predict new samples (same column names, outcome left empty). A saved model is only opened when its file is unchanged (SHA-256).
- 'Use the data of my project' loads the sorted measurements and groups of the open project.

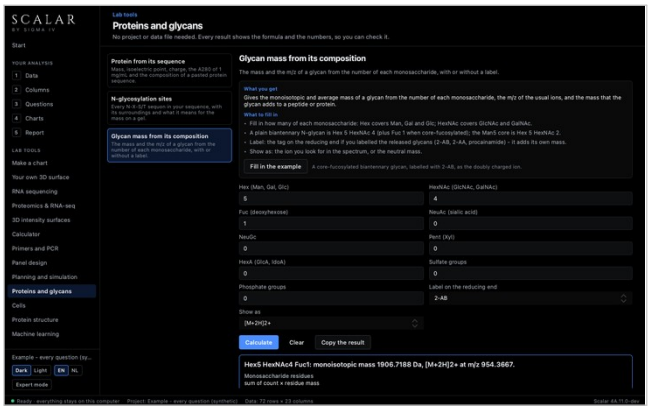
14 · PROTEIN STRUCTURE, PROTEINS AND GLYCANS, CELLS

Three new lab pages

Protein structure opens a PDB or mmCIF file from this computer; Proteins and glycans and Cells are tools that need no file.



File 22: the backbone turns with the mouse, coloured by pLDDT.



Glycan mass: the worked example filled in with one click.

- 22_two_helices_made_up.pdb: 57 residues in 2 chain(s), radius of gyration 13.8 Å.
- The parser was checked against Biopython (atoms, sequences, coordinates, backbone angles).
- Protein properties match Biopython's ProtParam (mass, pI, extinction coefficient, GRAVY); Man5GlcNAc2 gives 1234.4334 Da.
- Cells: doubling time from counts (with a 95 % interval), a haemocytometer count with trypan blue, how many cells to seed.

15 · CHARTS, REPORT AND LIMITS

From answer to figure and report

Charts (step 4)

- Every answered question has a chart; only the look changes - a chart always shows the same rows the test used.
- Show the groups in different ways, with or without significance stars, a point per sample, n per group, slanted group names.
- Save a look as your own template and use it for all new charts.
- Copy at 300 dpi for Word, PowerPoint or e-mail; save as PNG (300 or 600 dpi), TIFF (600 dpi), SVG or PDF.

Report (step 5)

'Make a report' turns a board into a PDF: a short summary of all answers; per question what was measured and done, what we see and what it means, and how sure it is by the fixed label rule; your charts in the look you chose; and an appendix with every test and number. The report is written in the language the program is set to. Expert mode adds detailed reports of a single saved analysis.

What Scalar does not do (yet)

- Alignment of raw sequencing reads (FASTQ): bring the count table from your pipeline (nf-core/rnaseq, STAR + featureCounts, salmon).
- Neural networks on images, and protein design (for example RFdiffusion): not in Scalar.
- Gene sets are not downloaded: bring them as a .gmt file (MSigDB, Enrichr or your own).
- 3D surfaces describe images; they do not test. A test needs several images per group, measured and then compared on the board.

Good to know in version 4A.11

- **Offered means it fits.** A question or test is only offered when the data can answer it: paired and mixed models need a donor that occurs more than once, Kaplan-Meier a 1 / 0 event column, ROC a column with exactly two groups, the dose-response fit a dose-like column.
- **The first column choice** of a question follows the names of the columns (dose, follow-up, event); always check it in the panel.
- **Paired t-test and Wilcoxon** on three or more time points compare the two you choose in the panel.
- **Machine learning** needs many samples; with fewer than about a hundred, a random forest or a linear model is usually as good as a neural network, and Scalar says so.
- **Exploratory.** These results describe these samples and show where it is worth looking further. Confirm a pattern in new material before it is used as a conclusion.

A ten-minute tour

1. Open Scalar and drop **01_treatment_groups.csv** on step 1. Read the check and go on.
2. Step 2: the boxes are already filled from row 2. Go on to step 3.
3. Drag **Is there a difference between groups?** onto the board; choose IL6_ELISA_pg_ml and treatment.
4. Read the sentence and the label, open **Explanation**, then the menu (…) and **See the calculation**. Copy the R code and run it in R.
5. Drag **Two-way ANOVA** from 'Choose a test yourself' next to it: treatment and genotype.
6. Step 4: give a chart another look and save it as PNG or SVG. Step 5: **Make PDF report**.
7. Lab tools: open **Proteomics & RNA-seq** with file 11, and **3D intensity surfaces** with the three images.