Humanizing Phenomics 120
psomagen

Raw data
Report

2025-01-14

Dr. John Doe

Olink Explore 3072/384 Panel

Olink

Research use only

1. Project Information

Customer	**Dr. John Doe**
Institute	Example University
Project ID	AB00000123
Olink Panel	Explore 3072/384
Target Proteins	2,943
Number of Samples	88
Number of Plates	1-8
Normalization Method	Intensity

Table of Contents

2. Assay Description

The Olink Explore 3072/384 assay is designed to measure circulating human protein levels using the Olink Proteomics Proximity Extension Assay technology. The Explore 3072/384 is a high-throughput assay which measures expression of 2,943 different proteins split across eight 384-well plates, covering four disease areas: Cardiometabolic I & II, Inflammation I & II, Neurology I & II, and Oncology I & II. The Explore 3072/384 assay is both flexible and scalable and can be run using one plate of interest up to all eight. Data are quantified relatively and called Normalized Protein Expression (NPX) values. Results are available for Research Use Only (RUO).

3. List of Acronyms:

NPX – normalized protein expression

PEA – proximity extension assay

IC – incubation (or immuno) control

EC – extension control

AC – amplification (or detection) control

NC – negative control

PC – plate control

SC – sample control

4. Workflow Summary

4.1 Sample Reception and Exclusion:

A total of 88 samples were received for this project. Upon inspection, 0 samples were excluded from testing. Therefore, a total of 88 samples were analyzed using the Olink Explore 3072 kit or individual panel.

4.2 Randomization Process:

Total randomization may be applied by either the customer or internally at Psomagen. For Psomagen randomized samples, randomization was conducted using an internally developed randomization tool, placing control and treatment samples on each plate. The randomized samples were then divided into incubation batches of 88 samples each and registered in Psomagen's Laboratory Information Management System (LIMS) under 1 distinct order number(s) for further processing.

4.3 Plate Naming:

Each sequencing run could process 88 samples, leading to the generation of 1 unique plate name(s). The plates following a naming convention of [ProjectID]-#, where "#" indicates the plate number in sequential order (1 to n).

4.4 Experimental Procedure:

1. Sample Incubation: The samples were incubated with paired antibodies linked to DNA oligonucleotides for a fixed duration of 18 hours. These antibodies specifically bind to target proteins in the sample.
2. Proximity Extension Assay (PEA): After binding, the oligonucleotides come into proximity, allowing them to hybridize and form new DNA sequences through enzymatic extension.

4.5 Quality Control and Data Analysis:

After all experiments were completed, each plate underwent individual quality control (QC). Upon successful completion of QC, the data were analyzed and consolidated into a single NPX.csv file, which is provided as the final assay results.

5. Data Access and Downloading

5.1 Download Links:

File name	File size	md5sum
AB00000XYZ	xxx	xxxx

Report.zip - This is a zip file of analysis results.

md5sum - In order to verify the integrity of files, md5sum is used. If the values of md5sum are the same, there is no forgery, modification or omission.

Your data will be retained in our server for 3 months. Should you wish to extend the retention period, please contact us.

5.2 Download Instructions:

Data has been transferred via sFTP/Globus/hard drive. Log-in instructions are provided below:

5.3 Folder and Data Structure:

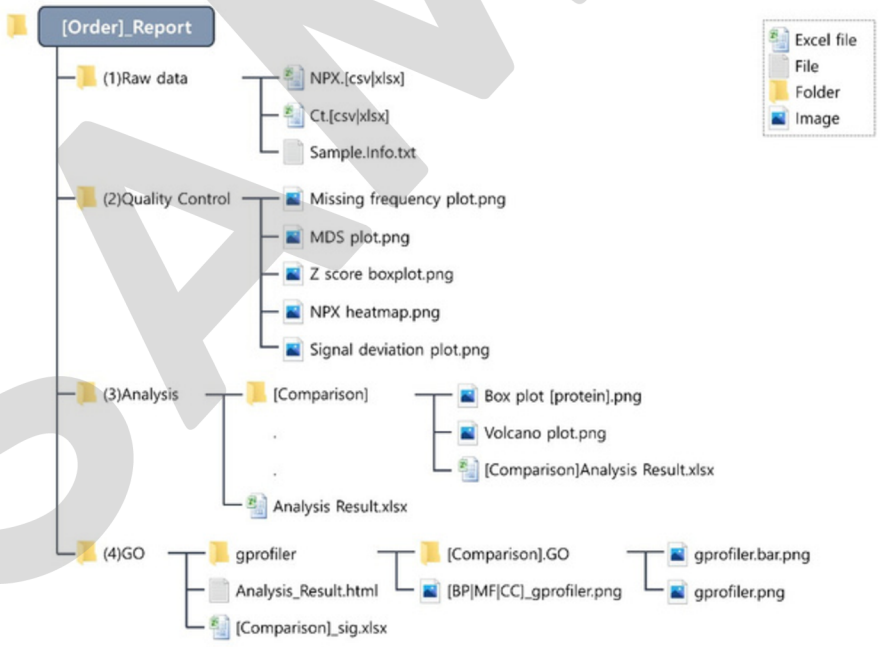

Figure 1: Example data organization.

5.4 Provided Files:

1. Raw data:
    a. group.info.txt
    b. Sample.Info.csv
    c. Sample.Info.txt

2. Quality Control:
    a. MDS plot(multidimensional scaling plot) PlateID.png
    b. MDS plot(multidimensional scaling plot) PlateID.txt
    c. MDS plot(multidimensional scaling plot) total.png
    d. MDS plot(multidimensional scaling plot) total.txt
    e. Missing_frequency_count.txt
    f. Missing_frequency_prop
    g. Signal deviation plot (Plate 1).png
    h. Signal deviation plot (Plate 1).txt
    i. Total NPX heatmap clustering.png
    j. zscore_boxplot.png

3. Differentially Expressed Protein
    a. DEP_ttest_results.txt
    b. DEP_volcano.png
    c. DEP_samples_heatmap.png
    d. GSEA_results.txt
    e. GSEA_heatmap.png
    f. Survival_KM_plot.png

6. Explanation of Quality Controls and Assay Normalization

6.1 Internal Controls:

- **Incubation/Immuno Control (IC)** – non-human antigens (GFP) which monitor potential technical variation in all three steps of the reaction.
- **Extension Control (EC)** – an antibody coupled to a unique pair of DNA oligo tags used for data normalization and NPX value generation. The EC monitors variation in the extension, amplification, and sequencing step for each sample.
- **Amplification/Detection Control (AC)** – a synthetic double stranded DNA which monitors the amplification and sequencing steps and does not require any proximity binding or extension steps.

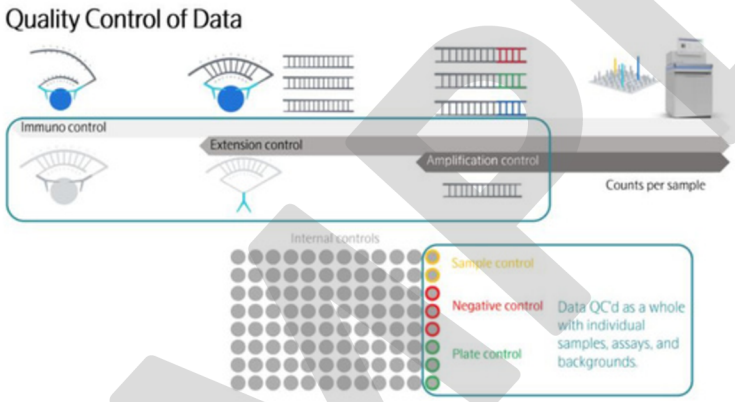

Figure 2: Assay and control design.

6.2 External Controls:

- **Negative Control (NC)** – empty buffer used for monitoring (1) contamination, (2) background noise generated by close-proximity DNA tags without binding the target protein first, and (3) calculating the Limit of Detection (LOD) for each plate.
- **Plate Control (PC)** – pooled human plasma samples from healthy donors used to normalize potential variation between runs and plates to compensate potential variation between runs and plates. The Plate Control is run in three replicates on Explore 3072/384.
- **Sample Control (SC)** – pooled human plasma samples from healthy donors run in duplicate and used to calculate inter and intra Coefficient of Variance.

6.3 Calculation of Normalized Protein Expression (NPX):

The raw output for Olink Explore is counts, where the combination of assay and sample is given an integer value based on the number of copies detected. These raw data counts are converted into NPX values for use in further analysis.

NPX values are calculated in two main steps. (1) The assay counts of a sample are divided by the Extension Control for that sample and block then log2 transformed. The resulting scale has increasing values with increasing concentration for each assay. Values are then normalized by (2a) Plate Control, which subtracts the median of the PCs or (2b) Intensity, which subtracts the median of all samples excluding controls.

For each equation below, i refers to a specific protein assay, j refers to a sample, and ExtNPX defines an extension normalized NPX value:

1. $ExtNPX_{i,j} = \log_2(counts(sample,Assay)/counts(EC))_j$

   -Relate counts to known standard (EC)
   -For all assays and all samples, including NCs, PCs, and SCs
   -Log2 transformation gives more normally distributed data

2a. $NPX_{i,j} = ExtNPX_{i,j} - median(ExtNPX(PC_i))$

   -Normalize by median of PCs
   -For each protein, per plate

2b. $NPX_{i,j} = ExtNPX_{i,j} - median(ExtNPX(Samples))$

   -Normalize by median of samples (excluding controls)
   -For each protein, per plate

6.4 QC Passing:

1. Each plate is evaluated on the number of counts, both total and internal control.
2. Sample quality is determined by the deviation of the ICs and ACs from the plate median for each of those two controls. Samples within ± 0.3 NPX from the plate median pass QC.

6.5 Control QC Warning or Failure:

1. Controls will fail if total assay counts are too low, while samples will be given a warning flag if internal controls have low counts within a block.
2. Plate Controls fail if their internal controls deviate more than ± 30% of the known concentration.
3. Negative Controls fail if assay counts are higher than expected relative to internal controls.
4. A minimum of 2 Sample Control data points should pass for each assay.
5. The precision of the calculated concentration for the interplate controls is evaluated and should have an Intra-CV <30%. If any of the four criteria above is not fulfilled, the assay will be reported as Assay warning: “Warning. Data from assays that do not pass QC should be treated with caution.”
6. A minimum of 2 sample control data points must be valid for each assay – i.e., only one SC replicate may fall outside the LOD.

6.6 Sample QC Warning or Failure:

1. Samples with NPX value deviating more than ± 0.3 from the plate median receive a warning.
2. If total assay counts are too low, while samples will be given a warning flag if internal controls have low counts within a block.

In sample data file (ProjectID_NPX.csv), two columns, SampleQC and AssayQC, indicate the QC status of samples and assays per plate and block (PASS, WARN, FAIL, or NA). “NA” refers to excluded assays, internal controls, or assays where QC cannot be performed.

7. Project QC Results

7.1 QC Summary:

	Total samples (n)	Samples passed QC (n)	Passed samples (%)
Negative Controls	3	3	100
Plate Controls	3	3	100
Sample Controls	2	2	100
Samples	88	86	97.7

7.2 Assay QC Summary:

Plate	Total assays (n):	Assays passed QC (n):	Passed assays (%):
ProjectID_plate1	2,942	2,906	98.8

7.3 Project Percent Coefficient of Variance (%CV) Changes

CV measures sample dispersion, or variation, around the mean within a dataset, with intra-%CV measuring variation within a plate and inter-%CV measuring variation between plates. Note that inter-%CV is omitted in projects with only one plate.

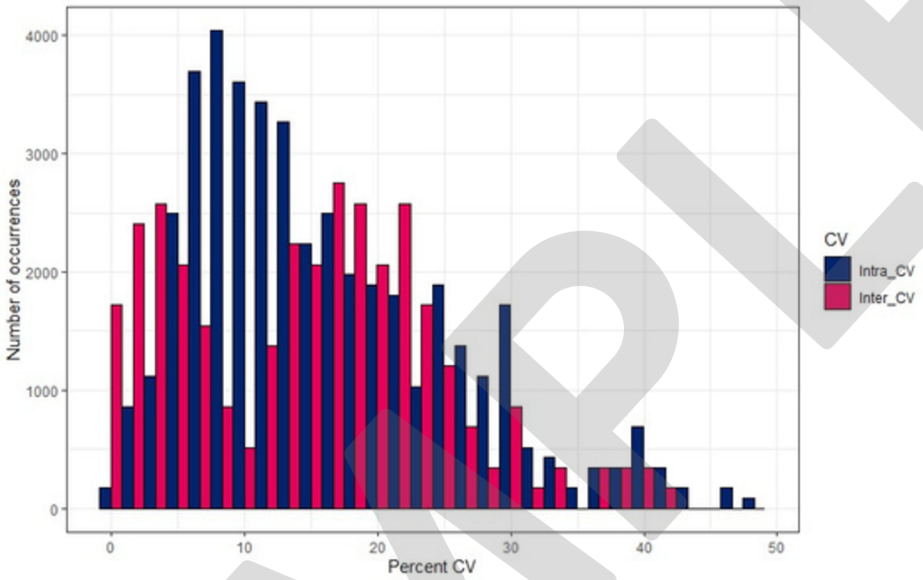

Figure 3: Project Intra- and Inter-%CV.

7.4 Signal deviation from median

Incubation and Amplification Controls are calculated per sample by deviation from the median value of each control. It is possible for individual samples to fall outside the proscribed range of ± 0.3 NPX from the median and still pass QC. This is determined by the fraction of samples within and outside the range. Red colored sample indicates flag (QC warning) sample and grey colored sample indicates the sample that does not have any NPX value.

Calculation of standard deviations for Incubation and Amplification Controls should be within the pre-determined quality threshold (< 0.3). The Signal deviation plots are provided as individual files for each plate run to assess the quality evaluation of both the run and the sample. In this report, only one plot is shown as an example:

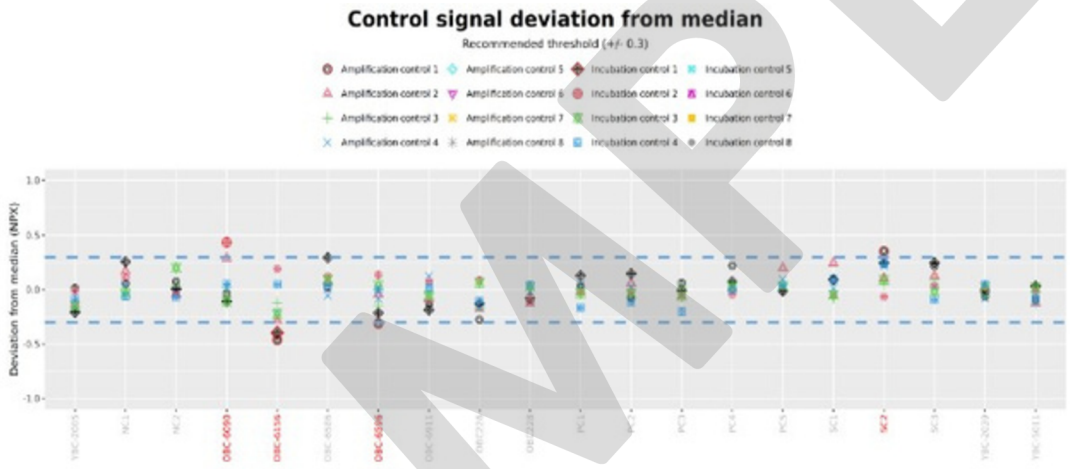

Figure 4: Example signal deviation plot

7.5 Z-score box plot

A visualization of the NPX value from every sample and plate. Each boxplot represents a separate plate where median and quartile values indicated by box edges and error bars are standard deviation.

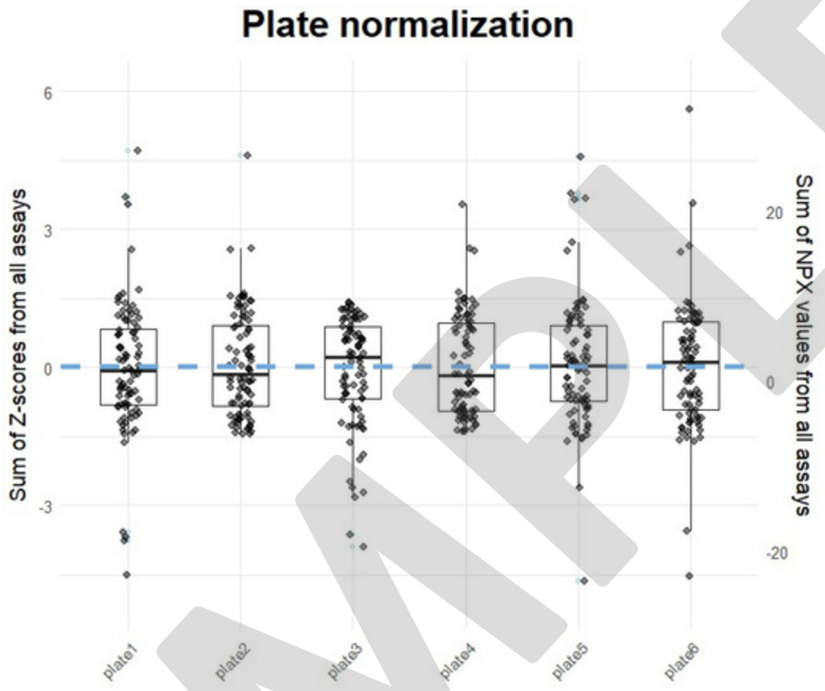

Figure 5: Project Z-score variation of NPX values.

7.6 Principal Component Analysis (PCA) of Controls and Samples

Spatial groupings of controls and samples showing variation within the dataset.

Figure 6: PCA plot of samples and controls.

The previous section included example QC output files of Olink panel. These primary analyses and output files are included free-of-charge for customers.

The below section is an example output of secondary data analysis available from Psomagen, at an additional cost.
Customers who may be interested in such analysis are encouraged to contact our bioinformatics department at "inquire@psomagen.com" to discuss specific requirements or outputs, referring to Quote or Project ID number. Metadata or further information regarding experimental design may be required depending on analysis type(s) requested.

8. Sample results overview

8.1 Differentially Expressed Proteins (DEPs)

Samples are tested for DEPs based on experimental design [treatment versus control]. The full expression table is provided in /Results/DEP_[ttest]_results.txt/

Protein	UniProt.ID	Treated	Untreated	p.value	adj.p.value	method	alternative
TRAIL	P50591	7.4	10	1.85E-06	0.000171	Welch Two Sample t-test	two.sided
SERPINA7	P05543	9.06	12.3	1.86E-06	0.000171	Welch Two Sample t-test	two.sided
CXCL11	O14625	5.62	3.9	2.96E-05	0.00181	Welch Two Sample t-test	two.sided
MMP-10	P09238	9.2	11.3	8.49E-05	0.0039	Welch Two Sample t-test	two.sided
CD6	Q8WWJ7	2.73	1.84	1.06E-04	0.0039	Welch Two Sample t-test	two.sided
Flt3L	P49771	6.13	4.17	1.57E-04	0.00481	Welch Two Sample t-test	two.sided
DPP4	P27487	4.23	6.22	2.37E-04	0.00622	Welch Two Sample t-test	two.sided
TWEAK	O43508	8.13	10.1	4.36E-04	0.01	Welch Two Sample t-test	two.sided
EFEMP1	Q12805	2.09	2.78	1.66E-03	0.0329	Welch Two Sample t-test	two.sided
REG3A	Q06141	7.23	9.49	2.00E-03	0.0329	Welch Two Sample t-test	two.sided
DEFA1	P59665	4.31	3.47	2.01E-03	0.0329	Welch Two Sample t-test	two.sided
ICAM1	P05362	2.5	3.09	2.15E-03	0.0329	Welch Two Sample t-test	two.sided
IL-22 RA1	Q8N6P7	8.99	11.5	2.72E-03	0.0385	Welch Two Sample t-test	two.sided
TCN2	P20062	5.54	8.46	3.49E-03	0.0459	Welch Two Sample t-test	two.sided

DEPs were visualized by volcano plot with significant results are labeled in red and by heatmap (NPX values).

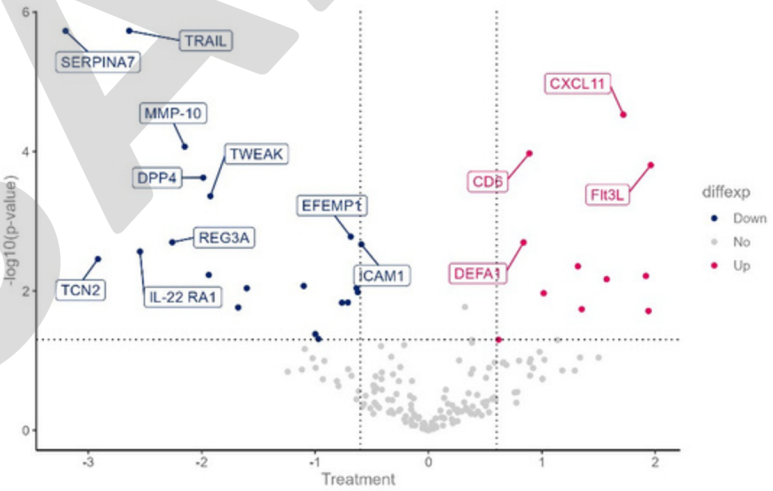

Figure 7: Volcano plot of DEPs

DEPs were visualized by heatmap and classified by treatment group.

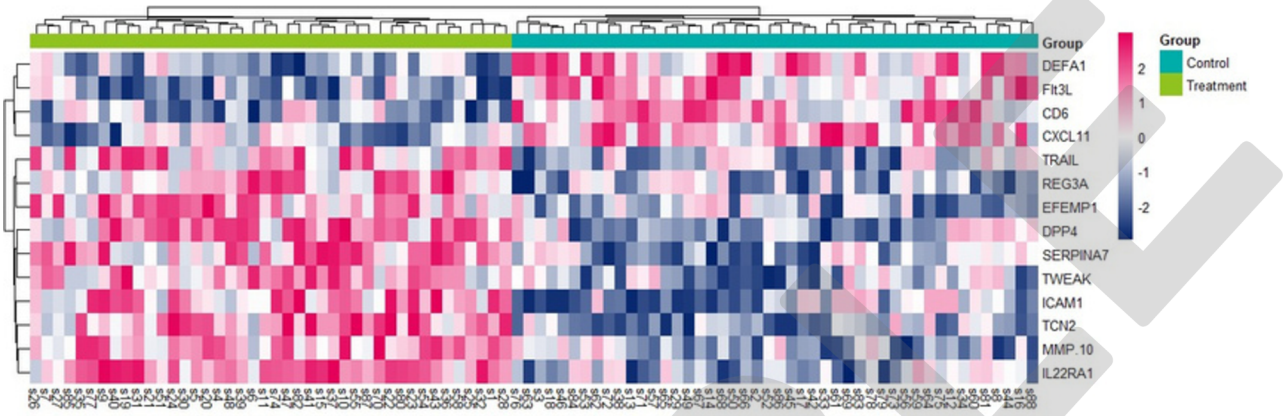

Figure 8: Significant differentially expressed proteins across samples.

8.2 GSEA Pathway Enrichment

DEPs were passed through GSEA for pathway enrichment scores. The top 10 differentially enriched pathways are provided below in table and heatmap format. The full table is provided in /Results/gsea_results.txt/

ID	Set Size	Enrichment Score	NES	P value	P adjust	Q value	rank
LEE_BMP2_TARGETS_UP	18	-0.73251	-1.93225	0.000401	0.224638	0.216615	28
WP_ALLOGRAFT_REJECTION	14	0.689654	1.914638	0.002998	0.29257	0.282121	8
GOBP_RESPONSE_TO_LIPID	26	0.592278	1.891845	0.002019	0.29257	0.282121	40
GOBP_RESPONSE_TO_MOLECULE_OF_BACTERIAL_ORIGIN	22	0.610721	1.87099	0.003738	0.29257	0.282121	30
REACTOME_G_ALPHA_I_SIGNALLING_EVENTS	14	0.668808	1.856767	0.005498	0.29257	0.282121	40
GOBP_NEGATIVE_REGULATION_OF_NUCLEOBASE_CONTAINING_COMPOUND_METABOLIC_PROCESS	10	0.721473	1.79663	0.00733	0.29257	0.282121	13
GOMF_G_PROTEIN_COUPLED_RECEPTOR_BINDING	19	0.602718	1.795635	0.006249	0.29257	0.282121	58
GOMF_CHEMOKINE_ACTIVITY	18	0.608482	1.773397	0.008276	0.29257	0.282121	58
GOMF_CHEMOKINE_RECEPTOR_BINDING	18	0.608482	1.773397	0.008276	0.29257	0.282121	58
KEGG_CHEMOKINE_SIGNALING_PATHWAY	18	0.608482	1.773397	0.008276	0.29257	0.282121	58

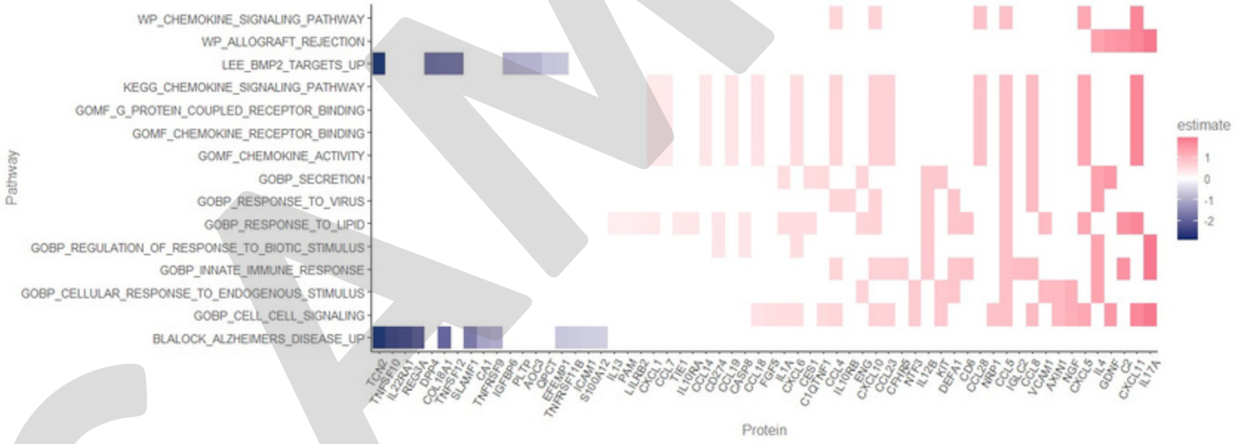

Figure 9: Differentially pathways after GSEA by assay.

8.3 Survival Analysis:

All differentially expressed proteins were checked for significance in survival status based on median NPX value for each protein. Example plots are provided below.

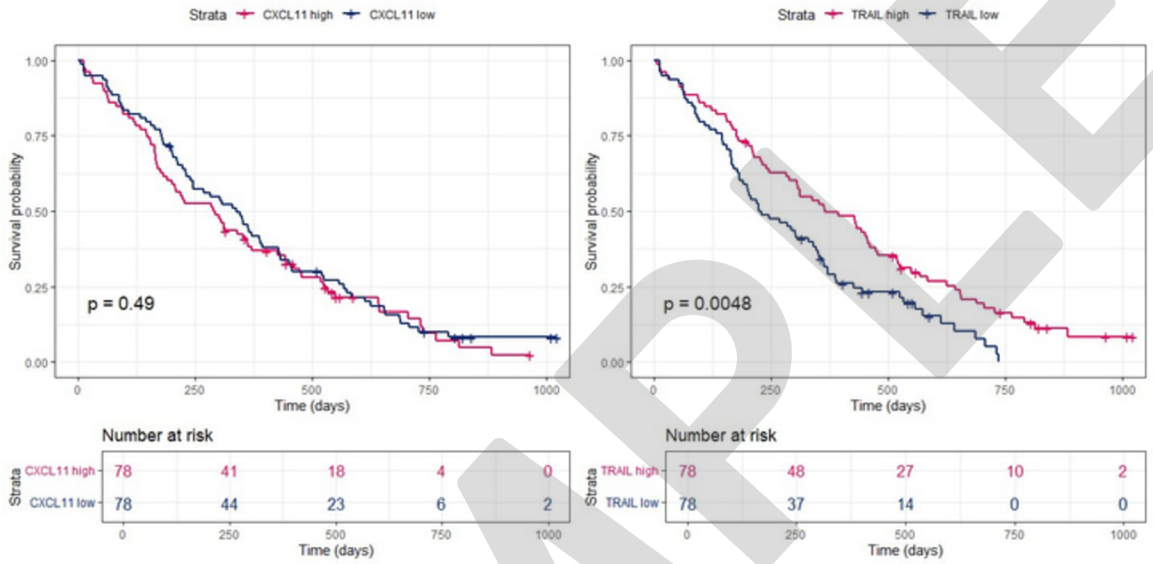

Figure 10: Representative Kaplan-Meier survival plots

8.4 Cox Regression:

Proteins significantly associated with survival were passed to a Cox multivariate regression model with provided metadata categories.

	estimate	std.error	statistic	p.value
TRAIL	-0.0578	0.02541	-2.2755	0.02287
age	0.01718	0.01045	1.64356	0.10027
sex	-0.5384	0.18645	-2.8879	0.00388
ph.ecog	0.37707	0.12294	3.06704	0.00216

Research use only

HEADQUARTER

Macrogen, Inc.
Laboratory, IT and Business Headquarter & Support Center
(08511) 1001, 10F, 254, Beotkkot-ro, Geumcheon-gu, Seoul, Republic of Korea (Gasan-dong, World Meridian 1)
Tel: +82-2-2180-7000
Email1: ngs@macrogen.com(Overseas)
Email2: ngskr@macrogen.com (Republic of Korea)
Web: www.macrogen.com
LIMS: dna.macrogen.com

SUBSIDIARY

Macrogen Europe
Laboratory, Business & Support Center
Meibergdreef 57, 1105 BA, Amsterdam, the Netherlands
Tel: +31-20-333-7563
Email: ngs@macrogen.eu

Macrogen Singapore
Laboratory, Business & Support Center
3 Biopolis Drive #05-18, Synapse, Singapore 138623
Tel: +65-6339-0927
Email: info-sg@macrogen.com

Psomagen (Macrogen USA)
Laboratory, Business & Support Center
1330 Piccard Drive, Suite 103, Rockville, MD 20850, United States
Tel: +1-301-251-1007
Email: inquiry@psomagen.com

Macrogen Japan
Laboratory, Business & Support Center
16F Time24 Building, 2-4-32 Aomi, Koto-ku, Tokyo 135-0064 JAPAN
Tel: +81-3-5962-1124
Email: ngs@macrogen-japan.co.jp

BRANCH

Macrogen Spain
Laboratory, Business & Support Center
Av. Sur del Aeropuerto de Barajas, 28, Office B-2, 28042 Madrid, Spain
Tel: +34-911-138-378
Email: info-spain@macrogen.com