

# ms\_6\_samples\_analysis

April 2, 2025

## 1 Mass spec 6 samples analysis

### 1.0.1 Imports and environment setup

- Date of run: 2025-03-02
- Environment: python 3.12
- Packages required: pandas, numpy, seaborn, matplotlib

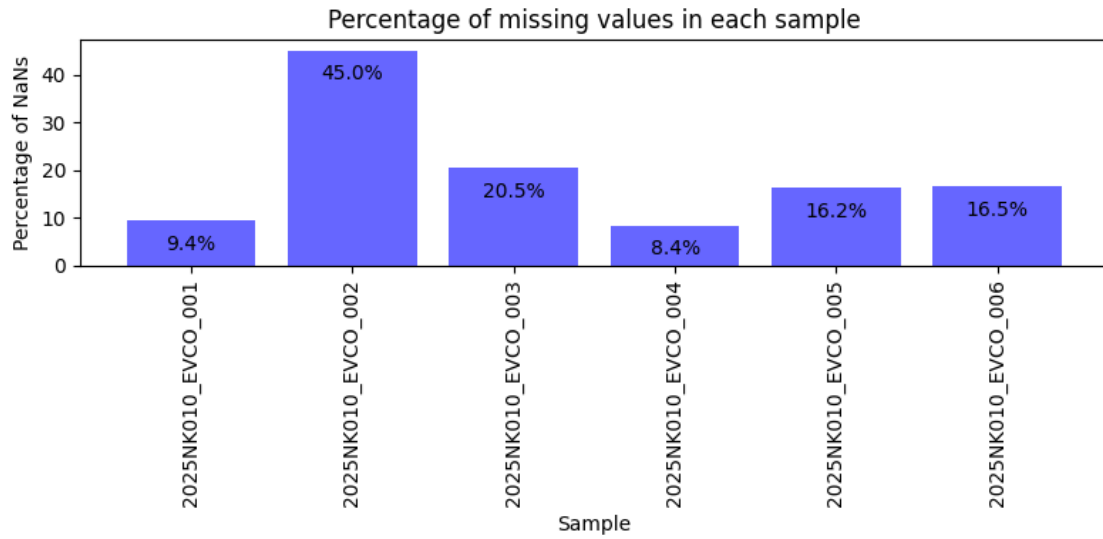
## 2 Data Loading

Both the MS and ELISA+metadata matrices were uploaded and some transformations to clean column names are done.

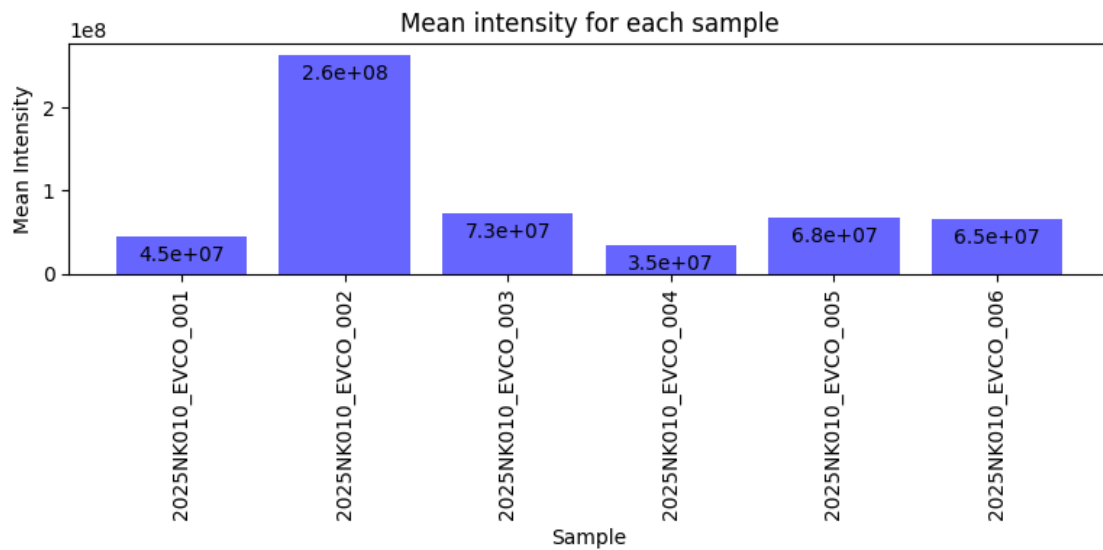
### 2.1 Samples analysis

We will analyze in MS the number of missing values and identities distributions.

```
2025NK010_EVCO_001    667
2025NK010_EVCO_002    3191
2025NK010_EVCO_003    1452
2025NK010_EVCO_004     594
2025NK010_EVCO_005    1152
2025NK010_EVCO_006    1173
dtype: int64
```

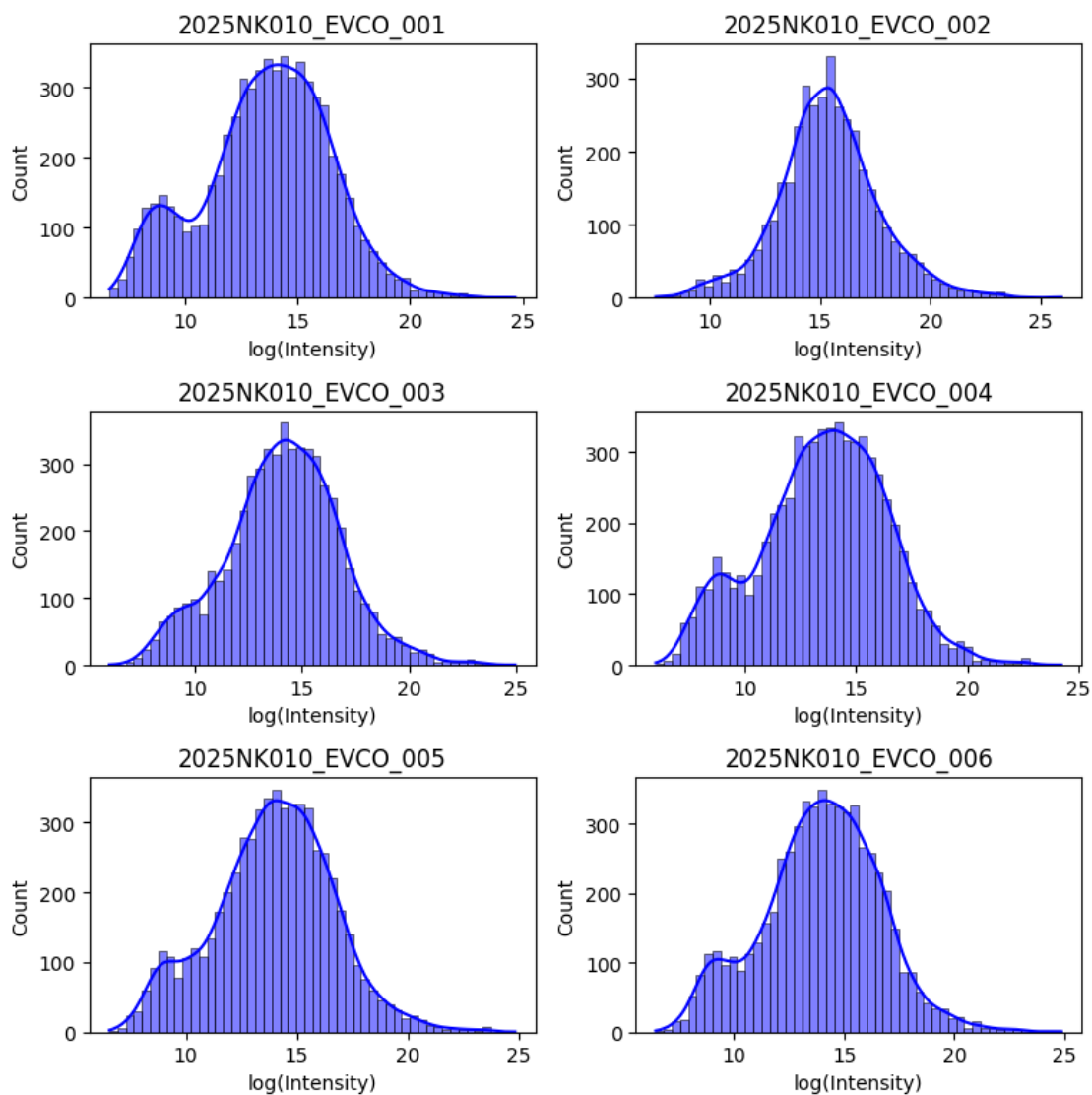


As can be observed, the number of missing values in the sample 2 is much higher than the rest.



The mean of intensities correlate with the percentage of missing values, are these two things related?

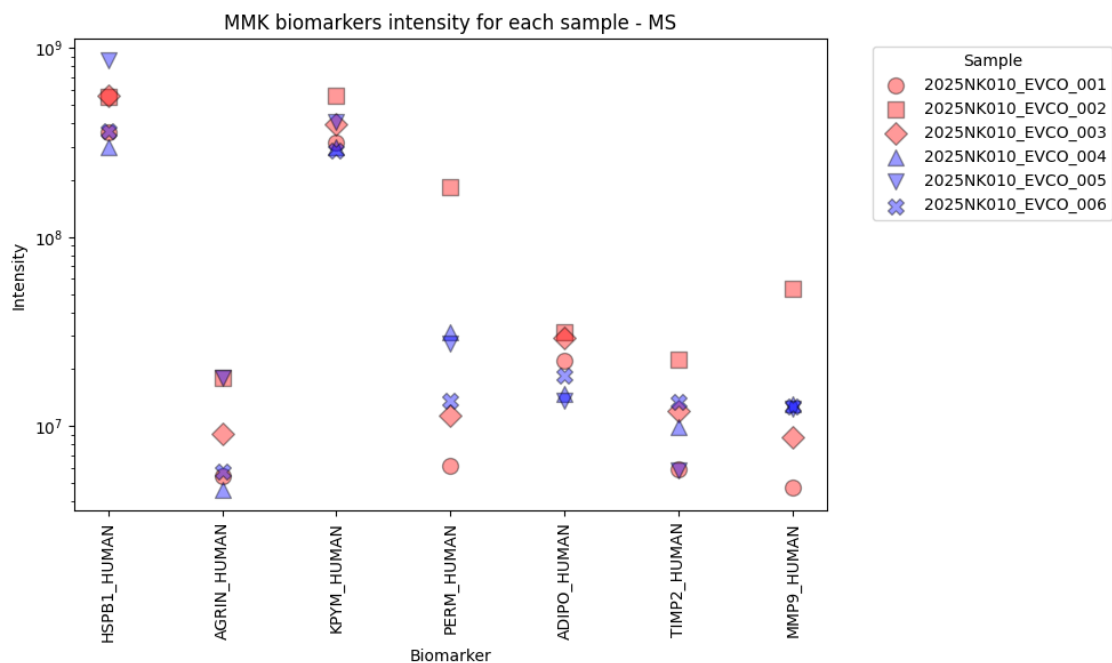
### 2.1.1 Distributions of intensities, per sample



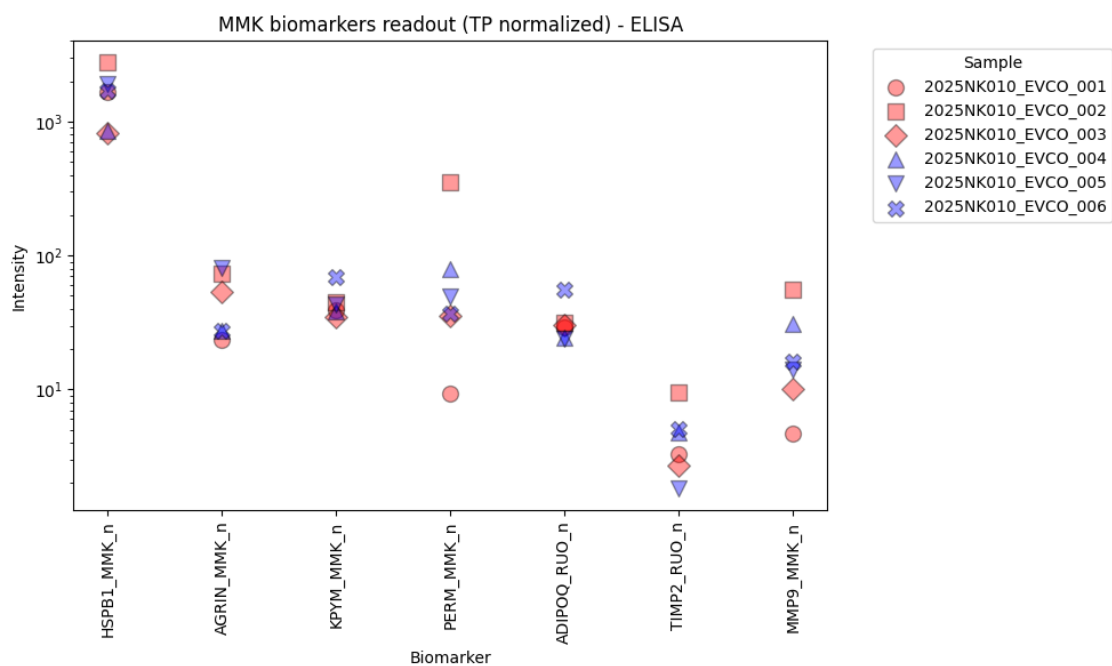
Notably, in the sample 2, the distribution lacks the enriched left tail of low intensity proteins.

### 2.2 What happens for the interest biomarkers?

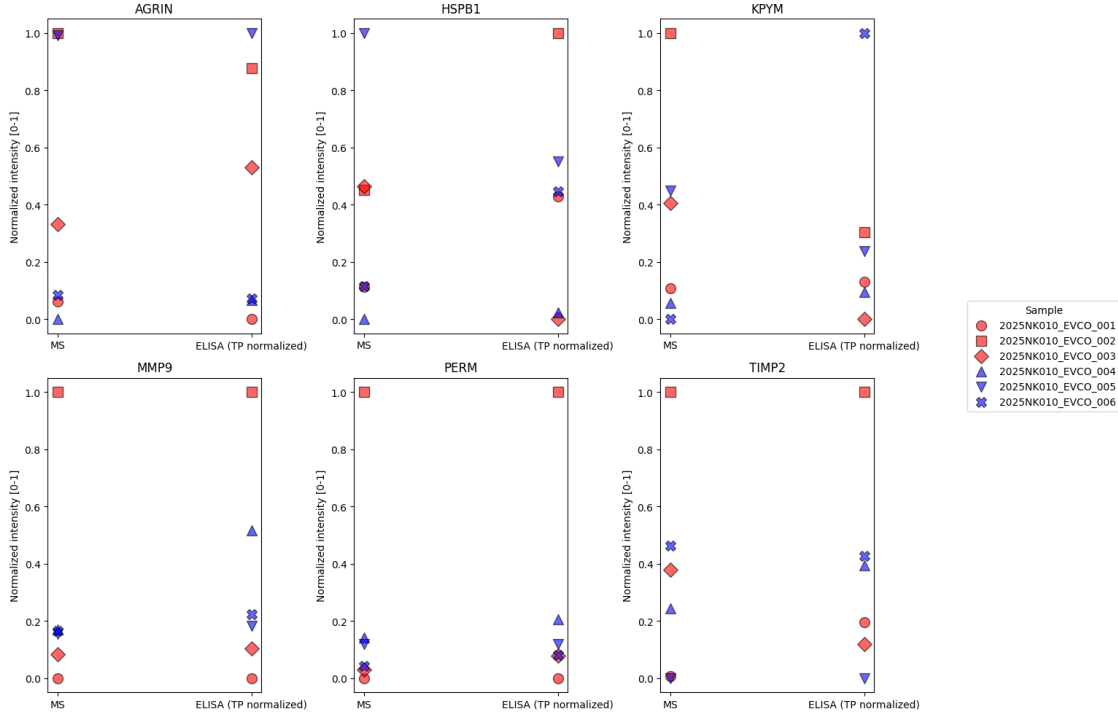
Let's analyze the distributions of intensities for the MMK interest biomarkers over the samples, in MS.



And now let's do the same with the ELISA results we have obtained (normalized by total protein).



Now, let's do the same but comparing the MS and ELISA for each biomarker individually. Since they are expressed in different scales/units, we need to normalize their intensities to compare them.



## 2.3 Fold changes

A fold change greater than 1 indicates that the biomarker is upregulated in cancer, meaning it shows higher intensity in cancer samples. Conversely, a fold change less than 1 suggests downregulation in cancer, i.e., the biomarker is less abundant in cancer compared to benign samples. Fold change values are always positive since they represent ratios. However, to make interpretation more intuitive, log2-transformed fold changes are often used: in this scale, positive values indicate upregulation, negative values indicate downregulation, and zero means no change.

### 2.3.1 Interest biomarkers fold change (MS vs ELISA)

|   | Biomarker | MS_FoldChange | ELISA_FoldChange | Same sign fold change? |
|---|-----------|---------------|------------------|------------------------|
| 0 | AGRN      | -0.201        | -0.156           | Yes                    |
| 1 | HSPB1     | 0.047         | -0.238           | No                     |
| 2 | KPYM      | -0.351        | 0.331            | No                     |
| 3 | MMP9      | -0.815        | -0.203           | Yes                    |
| 4 | PERM      | -1.475        | -1.256           | Yes                    |
| 5 | TIMP2     | -0.470        | -0.401           | Yes                    |

### 2.3.2 Highest cancer upregulated fold changes

|      | Biomarker   | Mean_Benign | Mean_Cancer  | Log2_Fold_Change |
|------|-------------|-------------|--------------|------------------|
| 2407 | LIMK2_HUMAN | 2627.990    | 14900000.000 | 12.469           |
| 1752 | PTN3_HUMAN  | 1885.490    | 1716277.185  | 9.830            |
| 1192 | IGF1R_HUMAN | 1799.070    | 676608.960   | 8.555            |
| 3704 | KAZRN_HUMAN | 3269.570    | 541000.000   | 7.370            |
| 4463 | ARH40_HUMAN | 5774.810    | 717947.410   | 6.958            |
| 169  | S35U4_HUMAN | 6688.690    | 639869.000   | 6.580            |
| 2502 | RAB38_HUMAN | 3403.990    | 312997.000   | 6.523            |
| 448  | CAH12_HUMAN | 1287.460    | 111642.000   | 6.438            |
| 2196 | PCP4_HUMAN  | 2394.330    | 168667.000   | 6.138            |
| 4687 | YIPF5_HUMAN | 5885.195    | 388639.675   | 6.045            |
| 6382 | WAC2C_HUMAN | 3687.120    | 188964.000   | 5.679            |
| 4100 | RASL3_HUMAN | 3155.290    | 153518.927   | 5.605            |
| 5400 | SHRPN_HUMAN | 122101.900  | 4733998.200  | 5.277            |
| 6102 | CDC23_HUMAN | 15123.300   | 338560.333   | 4.485            |
| 5823 | NDUBB_HUMAN | 8373.830    | 155451.200   | 4.214            |
| 3885 | RHG27_HUMAN | 22251.900   | 406093.233   | 4.190            |
| 6439 | WIF1_HUMAN  | 492427.000  | 8010000.000  | 4.024            |
| 1924 | S10A3_HUMAN | 161866.000  | 2460000.000  | 3.926            |
| 4783 | IP3KC_HUMAN | 4488.950    | 53377.237    | 3.572            |
| 3491 | KIZ_HUMAN   | 234115.000  | 2670000.000  | 3.512            |

### 2.3.3 Highest benign upregulated fold changes

|      | Biomarker   | Mean_Benign   | Mean_Cancer | Log2_Fold_Change |
|------|-------------|---------------|-------------|------------------|
| 3001 | SEC20_HUMAN | 303020.000    | 8159.930    | -5.215           |
| 2076 | PACR_HUMAN  | 23700000.000  | 623467.500  | -5.248           |
| 3056 | SG2A2_HUMAN | 222576666.667 | 5700000.000 | -5.287           |
| 4253 | BPIB2_HUMAN | 18776038.000  | 478888.667  | -5.293           |
| 3782 | ANM9_HUMAN  | 364311.795    | 8798.290    | -5.372           |
| 4868 | DIRA2_HUMAN | 81695.800     | 1914.510    | -5.415           |
| 2326 | SCNNB_HUMAN | 1030000.000   | 22421.400   | -5.522           |
| 2117 | GPDM_HUMAN  | 250142.000    | 3705.535    | -6.077           |
| 1507 | H12_HUMAN   | 5109038.000   | 67950.200   | -6.232           |
| 4984 | SG3A1_HUMAN | 59045316.333  | 671252.500  | -6.459           |
| 4443 | DMXL2_HUMAN | 353762.000    | 3024.990    | -6.870           |
| 6292 | SIK3_HUMAN  | 1230316.000   | 8246.617    | -7.221           |
| 6461 | AUP1_HUMAN  | 507758.500    | 3387.385    | -7.228           |
| 2847 | TAP2_HUMAN  | 5600000.000   | 35118.433   | -7.317           |
| 4692 | OSBP2_HUMAN | 4872829.600   | 16459.210   | -8.210           |
| 1963 | CBS_HUMAN   | 850742.500    | 2441.550    | -8.445           |
| 1386 | KCRU_HUMAN  | 1680000.000   | 4485.350    | -8.549           |
| 2419 | ATX1_HUMAN  | 3460000.000   | 9231.020    | -8.550           |
| 5883 | IF2B1_HUMAN | 1730000.000   | 1514.790    | -10.157          |
| 2997 | NFAC3_HUMAN | 2320000.000   | 1818.630    | -10.317          |

All the presented biomarkers have no NaNs, measurements were obtained for all the displayed proteins.