endoDx_normalization_analysis

September 18, 2025

1 EndoDX initial analysis

With a working dataset of 9,061 proteins from 199 samples the data was filtered at 75%, 90%, and 100% detection thresholds, and LOWESS normalization was selected over other no normalization and mean-centered normalization for its correction of mean-variance dependency.

A subsequent PCA analysis confirmed the absence of significant batch effects in the processed data. Differential expression analysis was then conducted and the resulting biomarkers were validated through two robustness checks: an intersection of significant hits across the three detection thresholds and a sensitivity analysis comparing k-NN to Left-Censored imputation.

51 biomarkers that are robust to different imputation assumptions, including a set of 16 candidates found to be significant across all data filtering stringencies (to be evaluated if necessary).

1.0.1 Imports and environment setup

- Date of run: 2025-09-17
- Environment: python 3.13
- Packages required: pandas, numpy, sklearn, statsmodels, seaborn, matplotlib

1.1 Data Loading and Preprocessing

Load the EndoDx dataset with minimal preprocessing to establish our baseline. The data undergoes log2 transformation but no additional normalization at this stage.

```
Loading EndoDx dataset…
Dataset loaded successfully:
- Samples: 199
- Biomarkers: 9061
- Metadata variables: 26
- Groups: {1: 128, 0: 71}
```

1.1.1 Filter biomarker list by detection threshold

For this analysis, we will consider thresholds of 75%, 90% and 100%.

100% Detection Threshold

```
Filtering biomarkers with detection threshold: 100.0%
Original biomarkers: 9061
Filtered biomarkers ( 100.0% detection): 1453
```

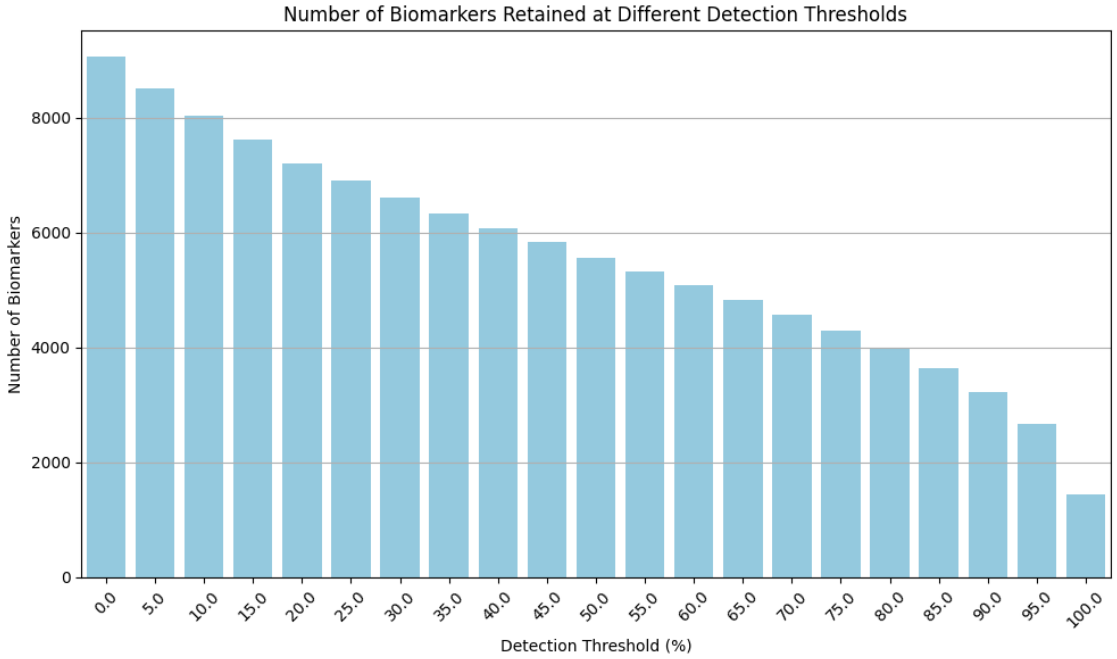
```
Removed biomarkers: 7608

Detection percentage distribution:
  Mean: 0.617
  Median: 0.709
  Min: 0.005
  Max: 1.000
```

90% Detection Threshold

```
Filtering biomarkers with detection threshold: 90.0%
Original biomarkers: 9061
Filtered biomarkers ( 90.0% detection): 3231
Removed biomarkers: 5830

Detection percentage distribution:
  Mean: 0.617
  Median: 0.709
  Min: 0.005
  Max: 1.000
```

75% Detection Threshold

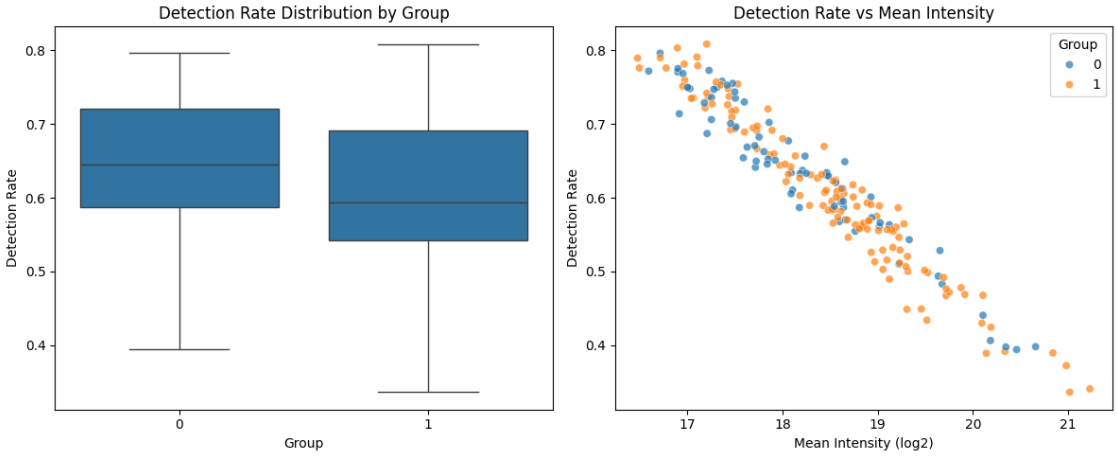
```
Filtering biomarkers with detection threshold: 75.0%
Original biomarkers: 9061
Filtered biomarkers ( 75.0% detection): 4293
Removed biomarkers: 4768

Detection percentage distribution:
  Mean: 0.617
  Median: 0.709
  Min: 0.005
  Max: 1.000
```



1.1.2 Detection summary

In the following plot we analyze the detection rates for samples of each condition and what happens with the intensities measured given the number of detections.

```
Creating sample-level detection quality analysis (boxplot + scatterplot)…
```




```
Sample Detection Quality Summary:
Mean detection rate: 0.617
Detection rate std: 0.106
```

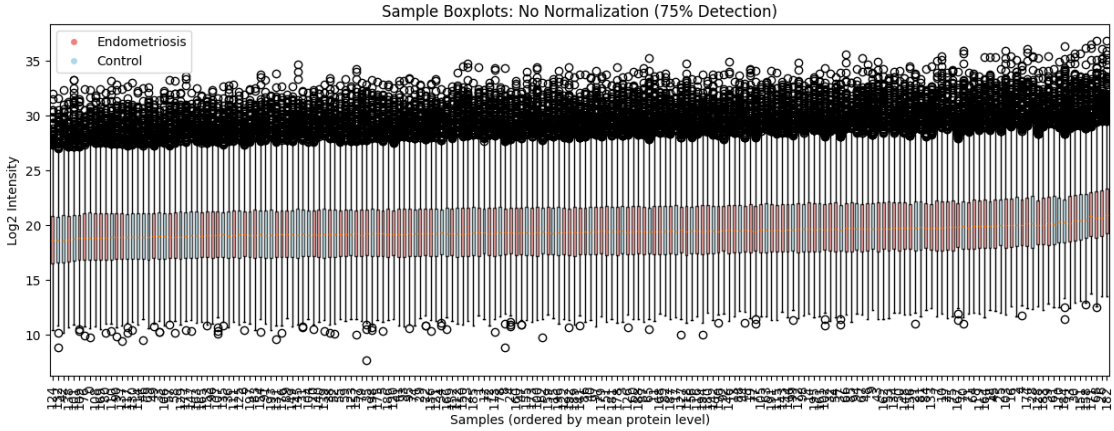
```
Min detection rate: 0.336
Max detection rate: 0.809

Detection rate by group:
        mean    std    min    max
Group
0      0.639  0.099  0.394  0.796
1      0.604  0.108  0.336  0.809
```

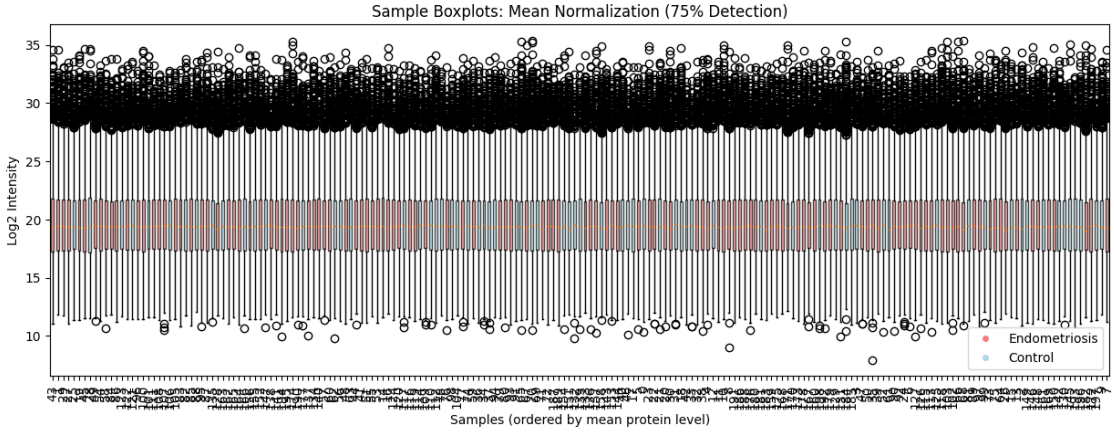
As can be observed, the mean intensity (in log scale) depends on the number of detections linearly and the mean intensities vary from 16 to 21. This indicates that a normalization strategy should be followed.

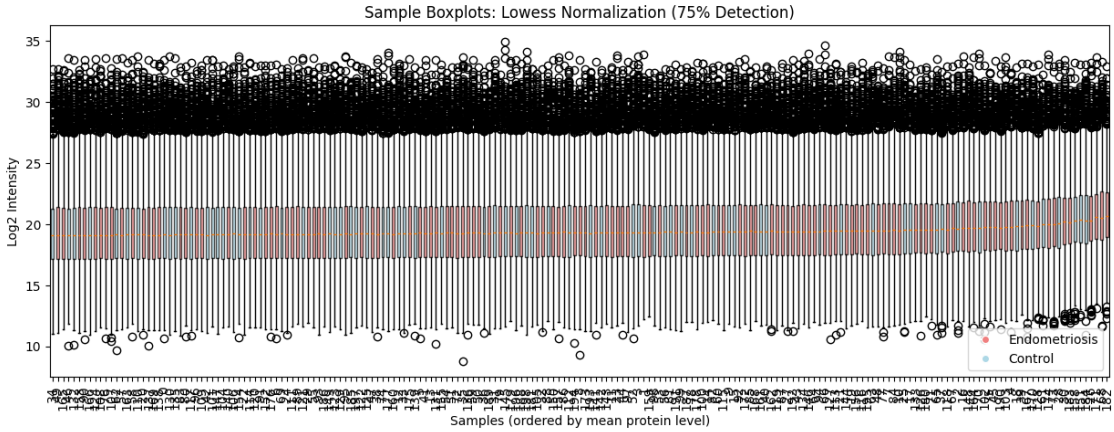
We observe also that detection rates don't vary according to group (Endometriosis=1, Control=0) and that samples with maximum detection rates detect around 80% of the biomarkers.

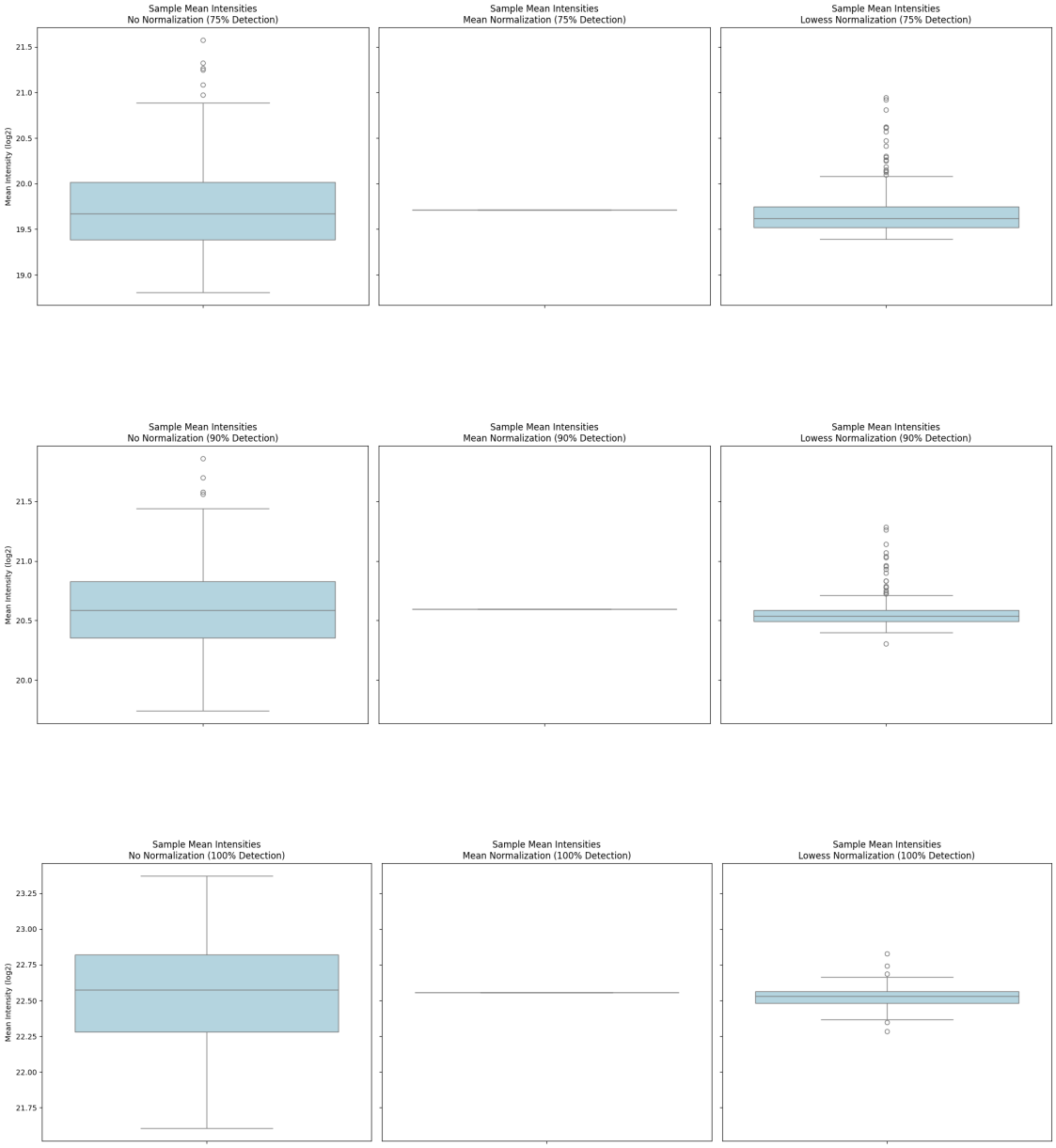
1.1.3 Load normalized matrices

```
Processing 75% detection threshold (4293 biomarkers):
   Loaded existing no_normalization
   Loaded existing mean_normalization
   Loaded existing lowess_normalization

Processing 90% detection threshold (3231 biomarkers):
   Loaded existing no_normalization
   Loaded existing mean_normalization
   Loaded existing lowess_normalization

Processing 100% detection threshold (1453 biomarkers):
   Loaded existing no_normalization
   Loaded existing mean_normalization
   Loaded existing lowess_normalization

============================================================
MATRIX LOADING SUMMARY
============================================================
no_normalization     |  75% detection: 199 samples, 4293 biomarkers
no_normalization     |  90% detection: 199 samples, 3231 biomarkers
no_normalization     | 100% detection: 199 samples, 9061 biomarkers
mean_normalization   |  75% detection: 199 samples, 4293 biomarkers
mean_normalization   |  90% detection: 199 samples, 3231 biomarkers
mean_normalization   | 100% detection: 199 samples, 1453 biomarkers
lowess_normalization |  75% detection: 199 samples, 4293 biomarkers
lowess_normalization |  90% detection: 199 samples, 3231 biomarkers
lowess_normalization | 100% detection: 199 samples, 1453 biomarkers
============================================================
All matrices ready for analysis!
```


Sample Boxplots: No Normalization (75% Detection)
Endometriosis
Control
Log2 Intensity
Samples (ordered by mean protein level)


Sample Boxplots: Mean Normalization (75% Detection)
Log2 Intensity
Endometriosis
Control
Samples (ordered by mean protein level)


Sample Boxplots: Lowess Normalization (75% Detection)
Log2 Intensity
Endometriosis
Control
Samples (ordered by mean protein level)



1.2 Measurment of normalization metrics

Here's a breakdown of the key metrics used to evaluate and compare normalization methods for mass spectrometry data.

1.2.1 Sample Mean CV

Measures the consistency of the overall intensity across all samples. In an ideal experiment, each sample should have a similar total amount of protein loaded. A high CV suggests that some samples are globally "brighter" or "dimmer" than others due to technical reasons (e.g., loading differences). The goal of normalization is to correct this.

1.2.2 Sample Median CV (Coefficient of Variation)

A more robust alternative to the Sample Mean CV. It serves the same purpose—to measure the consistency of overall sample intensity—but by using the median instead of the mean, it's less sensitive to a few extremely high-intensity outlier proteins within a sample.

1.2.3 Mean-Variance Correlation

Checks for a systematic bias common in mass spec data where a protein's variance increases with its abundance. This dependency violates the assumptions of many standard statistical tests (like the t-test), potentially leading to incorrect results. A good normalization method should break this relationship.

1.2.4 Pooled Within-Group CV (Bio-Signal Preservation)

This is a sanity check metric. Instead of measuring technical variation between samples, it measures the average variation inside your biological groups (e.g., the variation among all "Control" samples). This represents the real biological variability that you want to study.

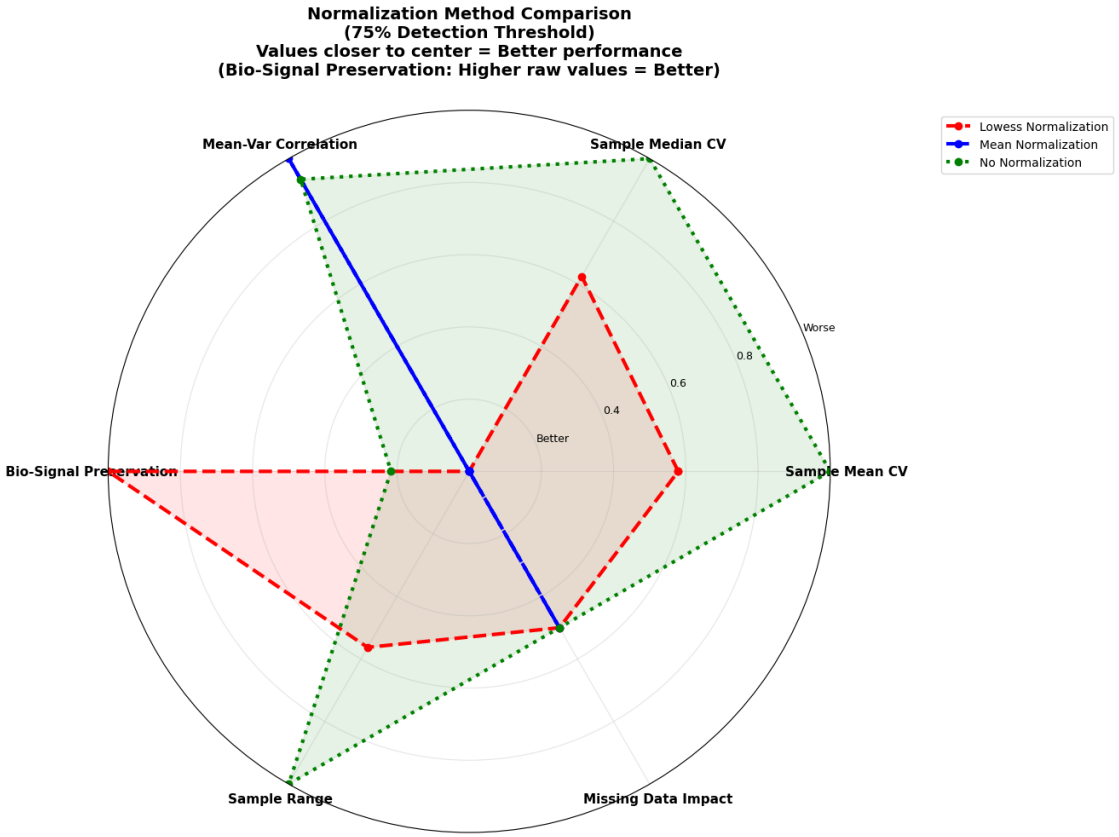
The dataset is divided into the biological groups (e.g., Control vs. Entrometriosis). Then for each protein, calculate its CV within the Control group. Do the same for the Endometriosis group. Then find the median of all protein CVs for the Control group and the median for the Disease group, and then average these two values.

The Bio-Signal value shown in the plot is min-max scaled, because the differences are very small for this metrics. Then the best performing pooled CV will be presented as 1, the worst with a 0 and the one in the middle with a value according to the distance with respect to the worse and the better.

1.2.5 Sample Range (Normalized)

This is a simple and intuitive metric for assessing outlier samples. It quantifies the difference between the "brightest" and "dimmest" samples in the dataset. A very large range suggests the presence of one or more samples that are technically very different from the rest.

```
Creating refined radar plot comparison of normalization methods…
Creating comprehensive normalization comparison report…

============================================================
ANALYZING DETECTION THRESHOLD: 75%
============================================================
Calculating normalization metrics for 75% detection threshold…
Found normalization methods: ['lowess_normalization', 'mean_normalization',
'no_normalization']
```



Raw Metrics for 75% Detection Threshold:

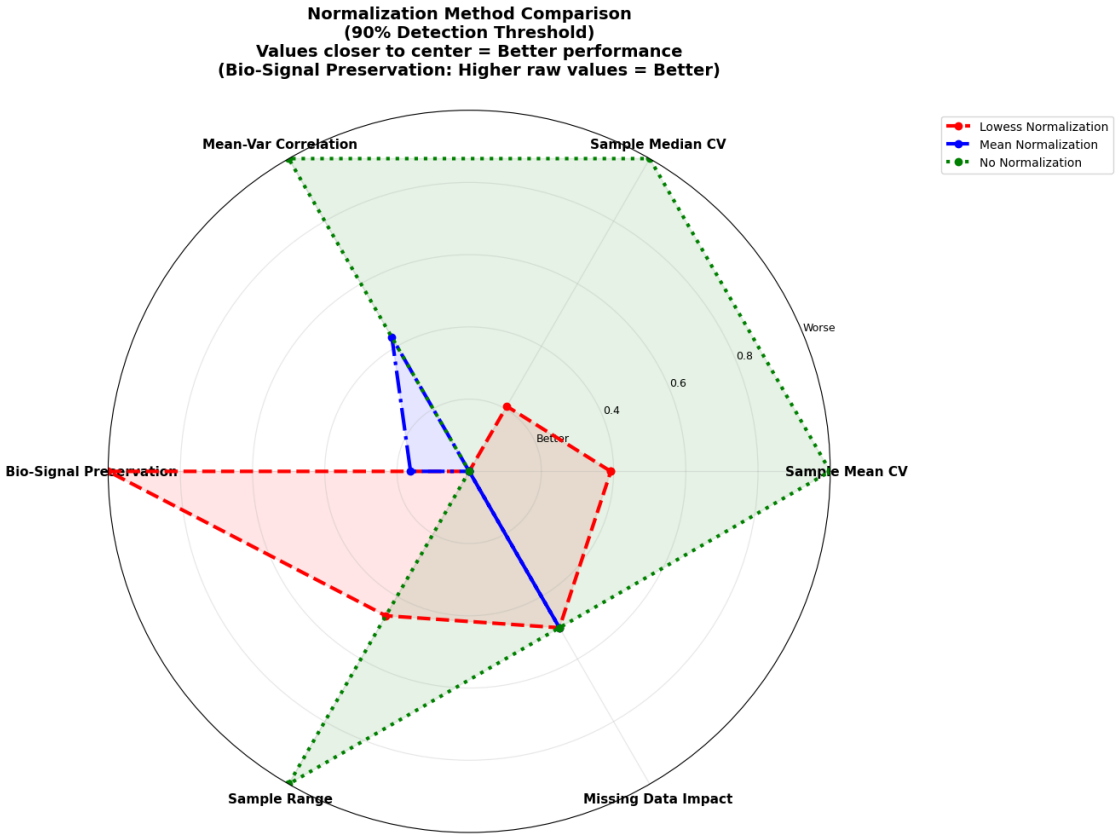
	sample_mean_cv	sample_median_cv	\
lowess_normalization	0.0144	0.0152	
mean_normalization	0.0000	0.0047	
no_normalization	0.0249	0.0215	

	mean_var_correlation_abs	pooled_within_group_cv	\
lowess_normalization	0.3001	0.0429	
mean_normalization	0.3499	0.0470	
no_normalization	0.3466	0.0461	

	sample_range_normalized	missing_data_penalty	\
lowess_normalization	0.0793	0.0563	
mean_normalization	0.0000	0.0563	
no_normalization	0.1409	0.0563	

	n_samples	n_proteins
lowess_normalization	199.0	4293.0
mean_normalization	199.0	4293.0
no_normalization	199.0	4293.0

```
Scaled Metrics for 75% Detection Threshold:

                     Sample Mean CV  Sample Median CV  Mean-Var Correlation  \
lowess_normalization          0.579             0.623                 0.000
mean_normalization            0.000             0.000                 1.000
no_normalization              1.000             1.000                 0.934

                     Bio-Signal Preservation  Sample Range  \
lowess_normalization                   1.000         0.563
mean_normalization                     0.000         0.000
no_normalization                       0.217         1.000

                     Missing Data Impact
lowess_normalization                 0.5
mean_normalization                   0.5
no_normalization                     0.5


============================================================
ANALYZING DETECTION THRESHOLD: 90%
============================================================
Calculating normalization metrics for 90% detection threshold…
Found normalization methods: ['lowess_normalization', 'mean_normalization',
'no_normalization']
```



Raw Metrics for 90% Detection Threshold:

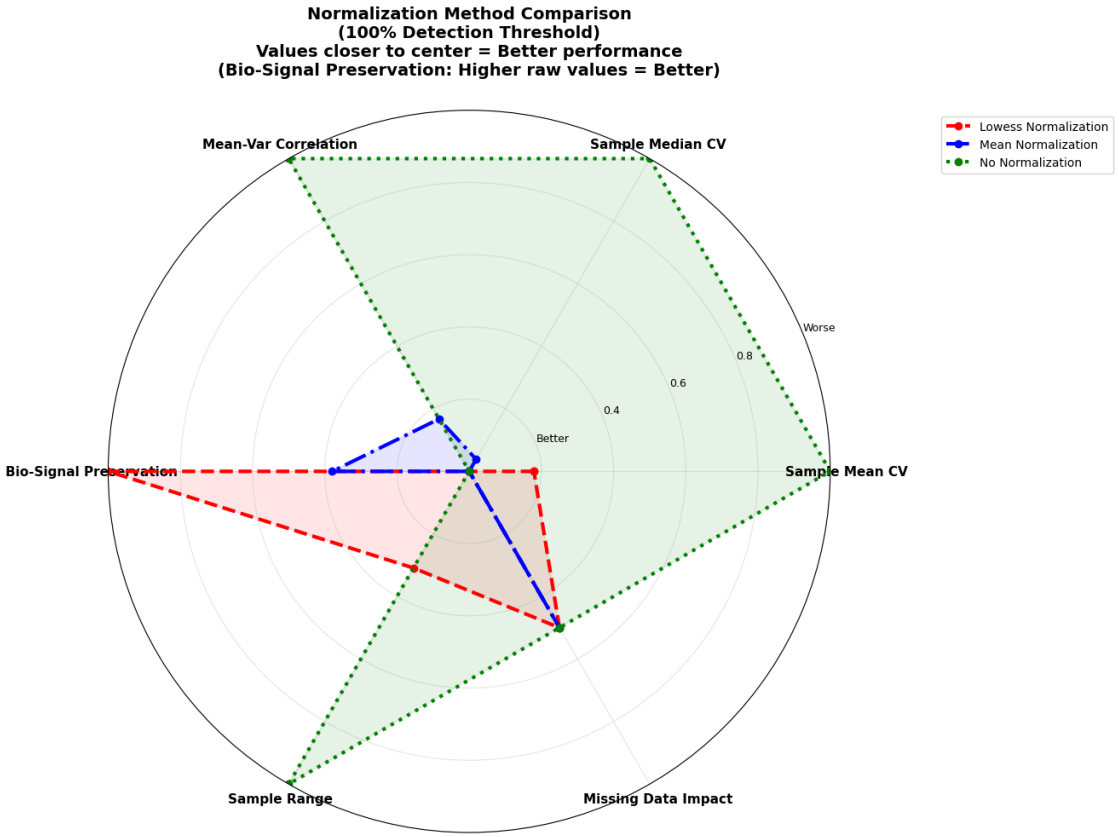
	sample_mean_cv	sample_median_cv	\
lowess_normalization	0.0069	0.0069	
mean_normalization	0.0000	0.0050	
no_normalization	0.0177	0.0144	

	mean_var_correlation_abs	pooled_within_group_cv	\
lowess_normalization	0.3039	0.0407	
mean_normalization	0.3296	0.0436	
no_normalization	0.3639	0.0441	

	sample_range_normalized	missing_data_penalty	\
lowess_normalization	0.0475	0.0195	
mean_normalization	0.0000	0.0195	
no_normalization	0.1030	0.0195	

	n_samples	n_proteins
lowess_normalization	199.0	3231.0
mean_normalization	199.0	3231.0
no_normalization	199.0	3231.0

```
Scaled Metrics for 90% Detection Threshold:

                     Sample Mean CV  Sample Median CV  Mean-Var Correlation  \
lowess_normalization          0.392             0.208                 0.000
mean_normalization            0.000             0.000                 0.428
no_normalization              1.000             1.000                 1.000

                     Bio-Signal Preservation  Sample Range  \
lowess_normalization                   1.000         0.462
mean_normalization                     0.162         0.000
no_normalization                       0.000         1.000

                     Missing Data Impact
lowess_normalization                 0.5
mean_normalization                   0.5
no_normalization                     0.5


============================================================
ANALYZING DETECTION THRESHOLD: 100%
============================================================
Calculating normalization metrics for 100% detection threshold…
Found normalization methods: ['lowess_normalization', 'mean_normalization',
'no_normalization']
```



Raw Metrics for 100% Detection Threshold:

	sample_mean_cv	sample_median_cv	\
lowess_normalization	0.0029	0.0033	
mean_normalization	0.0000	0.0038	
no_normalization	0.0161	0.0154	

	mean_var_correlation_abs	pooled_within_group_cv	\
lowess_normalization	0.3322	0.0358	
mean_normalization	0.3449	0.0388	
no_normalization	0.4077	0.0406	

	sample_range_normalized	missing_data_penalty	\
lowess_normalization	0.0242	0.0	
mean_normalization	0.0000	0.0	
no_normalization	0.0783	0.0	

	n_samples	n_proteins
lowess_normalization	199.0	1453.0
mean_normalization	199.0	1453.0
no_normalization	199.0	1453.0

```
Scaled Metrics for 100% Detection Threshold:

                      Sample Mean CV  Sample Median CV  Mean-Var Correlation  \
lowess_normalization           0.179             0.000                 0.000
mean_normalization             0.000             0.039                 0.167
no_normalization               1.000             1.000                 1.000

                      Bio-Signal Preservation  Sample Range  \
lowess_normalization                    1.000         0.309
mean_normalization                      0.379         0.000
no_normalization                        0.000         1.000

                      Missing Data Impact
lowess_normalization                  0.5
mean_normalization                    0.5
no_normalization                      0.5
```

1.2.6 Analysis of the results

Some of the following things are observed:

- Normalization is necessary: The no_normalization data consistently performs poorly on technical metrics (Sample Mean CV, Sample Range), confirming that a correction step is essential.
- As expected, mean_normalization always scores a "perfect" 0.000 on Sample Mean CV and Sample Range. This isn't a sign of superiority; it's a mathematical artifact of the method. It's a blunt tool that zeros out one specific error type.
- LOWESS is systematically better at correcting the Mean-Variance Correlation. At every threshold, it produces the lowest score on this metric. This is the most important factor, as correcting this systematic bias is critical for the validity of statistical tests we will use for biomarker discovery.
- Data shows that lowess_normalization results in the lowest raw pooled_within_group_cv. This means it reduces the internal biological variation slightly more than the other methods. However, the absolute differences in the raw values are very small. This minor reduction is an acceptable trade-off for the crucial benefit of fixing the mean-variance dependency.

1.2.7 Conclussion

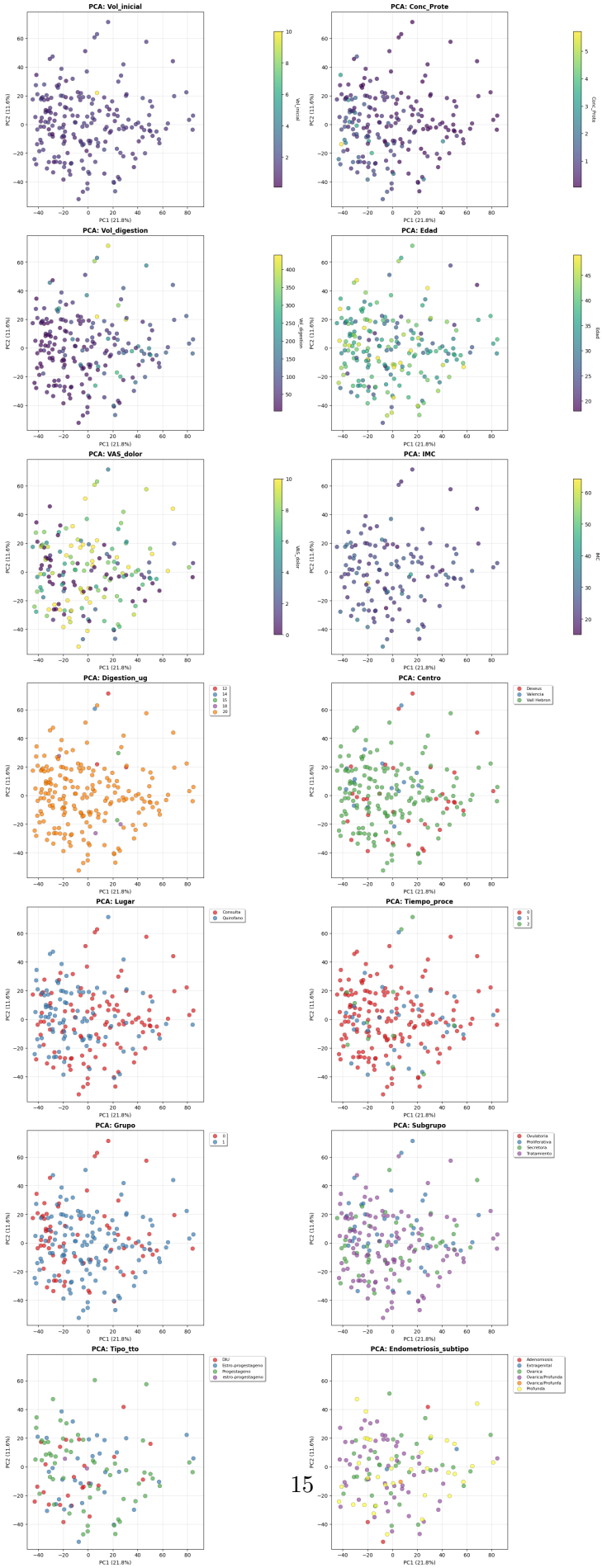
According to these metrics we should use **LOWESS NORMALIZATION**.

1.3 PCA analysis

We're coloring the samples in this plot by different metavariables to visually assess whether the primary source of variation is the biological condition of interest or a technical artifact.

To create a complete data matrix missing values were handled using KNN imputation. This estimates each missing value based on the expression profiles of the k most similar proteins present in the dataset. The advantage of it is the ability to preserve the underlying biological covariance structure by leveraging existing protein-protein correlations.

```
Numerical variables found: ['Vol_inicial', 'Conc_Prote', 'Vol_digestion',
'Edad', 'VAS_dolor', 'IMC']
Categorical variables found: ['Codigo_VHIR', 'Codigo_CRG', 'Digestion_ug',
'Centro', 'Lugar', 'Fecha_proce', 'Fecha_reco', 'Tiempo_proce', 'Subgrupo',
'Tipo_tto', 'Sintomatologia', 'Causa_infertilidad', 'Otras_enfermedades',
'Grupo', 'Endometriosis_subtipo', 'Previa_cito', 'Lubricante', 'Previa_TVUS',
'Hemolisis', 'Previa_cirugia']
```


PCA: Vol_inicial
PCA: Conc_Prote
PCA: Vol_digestion
PCA: Edad
PCA: VAS_dolor
PCA: IMC
PCA: Digestion_ug
PCA: Centro
PCA: Lugar
PCA: Tiempo_proce
PCA: Grupo
PCA: Subgrupo
PCA: Tipo_tto
PCA: Endometriosis_subtipo
PC1 (21.8%)
PC2 (11.6%)

The samples do not cluster by any variables like Centro or "tiempo de procesamiento", which is good and indicates that no corrections are needed to move to biomarker discovery.

1.4 Differential expression analysis

1.4.1 Comparison of Results Across Detection Thresholds

This analysis assesses the stability of biomarker discovery by comparing results from datasets filtered at different levels of data completeness prior to imputation. Each threshold represents a different trade-off between the number of proteins included and the amount of missing data tolerated.

- The 75% detection threshold is the most inclusive, retaining a large number of proteins for analysis. This maximizes the potential discovery pool but relies more heavily on the imputation of missing values.
- The 90% and 100% thresholds are progressively more stringent, analyzing only proteins that are very consistently detected. While this reduces the reliance on imputation and increases confidence in the raw data for each included protein, it also significantly narrows the scope of the analysis and may discard potentially important, less abundant biomarkers.

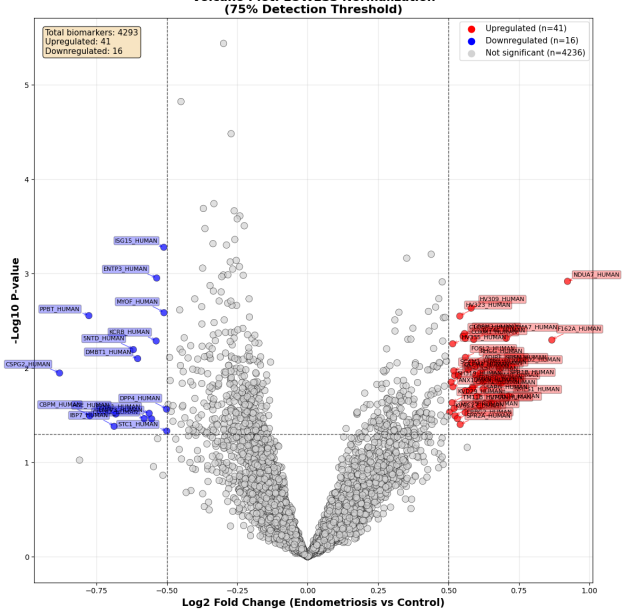
To create a complete data matrix suitable for analysis, missing values were imputed using the k-Nearest Neighbors (k-NN) method (below the analysis continue comparing this method with others).

By analyzing the intersection of significant biomarkers found at each threshold, we can infer the robustness of our findings. A biomarker that remains statistically significant across multiple filtering strategies is considered a high-confidence candidate, as its discovery is not dependent on a specific choice of data stringency. This consistency demonstrates a strong and stable biological signal.

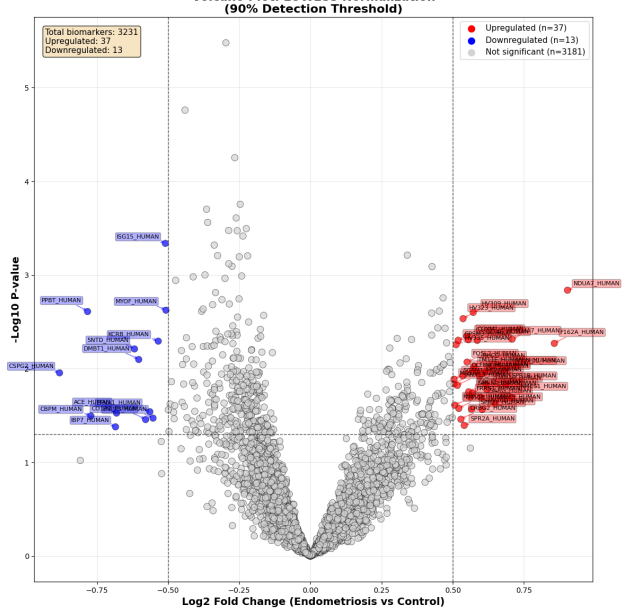
```
Creating enhanced volcano plots with ALL biomarker labels…
```

Differential Expression Analysis: Endometriosis vs Control
LOWESS Normalization - All Detection Thresholds

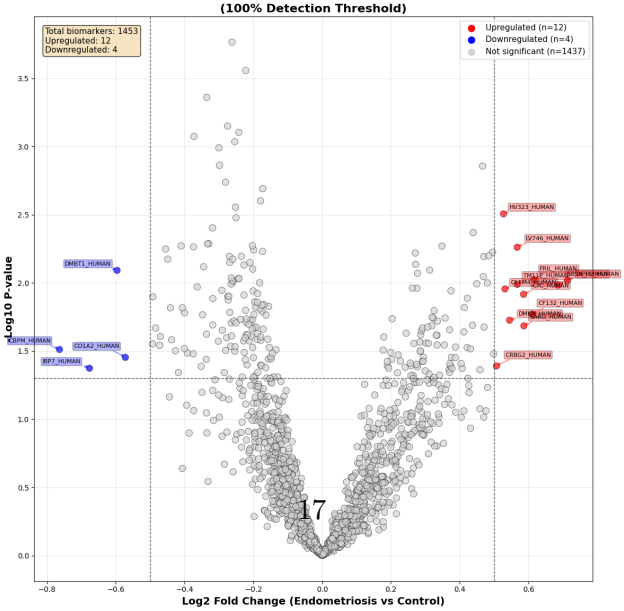
Volcano Plot: LOWESS Normalization
(75% Detection Threshold)



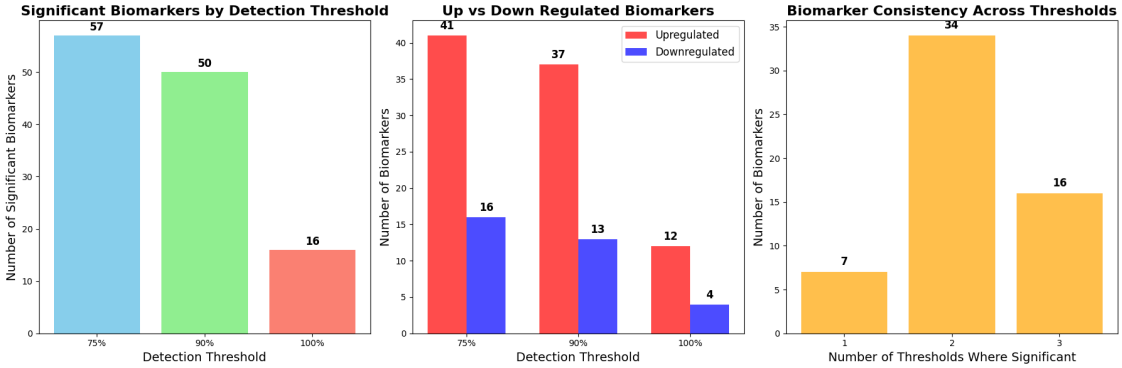
Volcano Plot: LOWESS Normalization
(90% Detection Threshold)



Volcano Plot: LOWESS Normalization
(100% Detection Threshold)



```
Creating biomarker intersection analysis plots…
```




```
================================================================================
DETAILED BIOMARKER INTERSECTION BREAKDOWN
================================================================================

 BIOMARKERS SIGNIFICANT AT ALL THREE THRESHOLDS (16 total):
   1. CBPM_HUMAN
   2. CF132_HUMAN
   3. CO1A2_HUMAN
   4. CRBG2_HUMAN
   5. CYC_HUMAN
   6. DMBT1_HUMAN
   7. DMKN_HUMAN
   8. FRIL_HUMAN
   9. HV323_HUMAN
  10. IBP7_HUMAN
  11. LV746_HUMAN
  12. LYPD2_HUMAN
  13. OLFM4_HUMAN
  14. SARG_HUMAN
  15. SBSN_HUMAN
  16. TM11E_HUMAN

 BIOMARKERS SIGNIFICANT AT 75% & 90% ONLY (34 total):
   1. 5NTD_HUMAN
   2. ACE_HUMAN
   3. ADIRF_HUMAN
   4. ANX13_HUMAN
   5. C1QBP_HUMAN
```

```
  6. COXM1_HUMAN
  7. CRCT1_HUMAN
  8. CSPG2_HUMAN
  9. EFNA1_HUMAN
 10. ENPP3_HUMAN
 11. F162A_HUMAN
 12. FOSL2_HUMAN
 13. FRRS1_HUMAN
 14. GPSM3_HUMAN
 15. HV309_HUMAN
 16. HV335_HUMAN
 17. HV70D_HUMAN
 18. ISG15_HUMAN
 19. KCRB_HUMAN
 20. KVD29_HUMAN
 21. MAST4_HUMAN
 22. MYOF_HUMAN
 23. NDUA7_HUMAN
 24. NMES1_HUMAN
 25. PPBT_HUMAN
 26. RHCG_HUMAN
 27. RL29_HUMAN
 28. SCAM1_HUMAN
 29. SPR1B_HUMAN
 30. SPR2A_HUMAN
 31. SPR2F_HUMAN
 32. TM11B_HUMAN
 33. TMA7_HUMAN
 34. TMM40_HUMAN

BIOMARKERS SIGNIFICANT AT 75% & 100% ONLY (0 total):
 None

BIOMARKERS SIGNIFICANT AT 90% & 100% ONLY (0 total):
 None

BIOMARKERS UNIQUE TO 75% THRESHOLD (7 total):
  1. DPP4_HUMAN
  2. ENTP3_HUMAN
  3. KV621_HUMAN
  4. LV537_HUMAN
  5. MIC27_HUMAN
  6. SH319_HUMAN
  7. STC1_HUMAN

BIOMARKERS UNIQUE TO 90% THRESHOLD (0 total):
 None
```

```
 BIOMARKERS UNIQUE TO 100% THRESHOLD (0 total):
  None

================================================================================
```

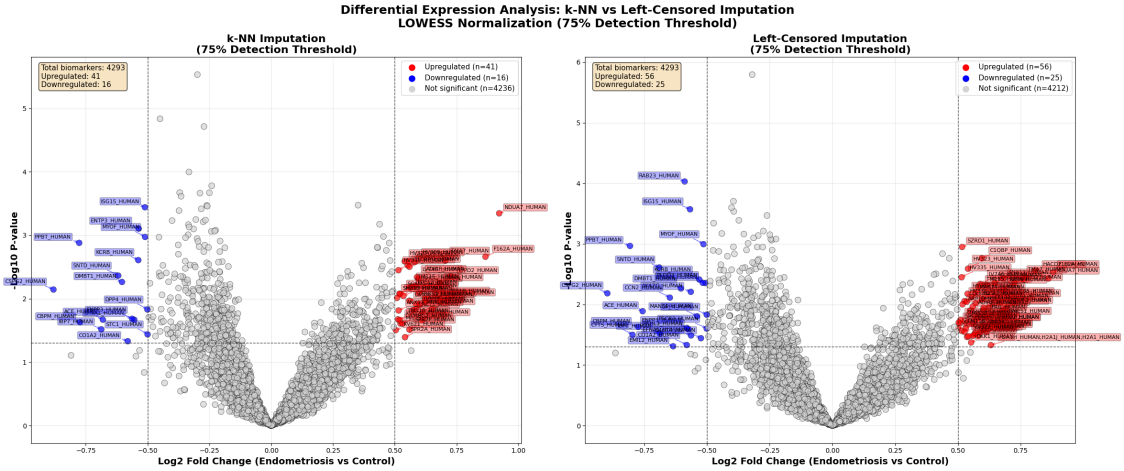
1.4.2 Comparison of Imputation Methods: k-NN vs. Left-Censored

This analysis compares two distinct imputation strategies to assess the robustness of biomarker discovery. The chosen methods operate on fundamentally different assumptions about why data is missing.

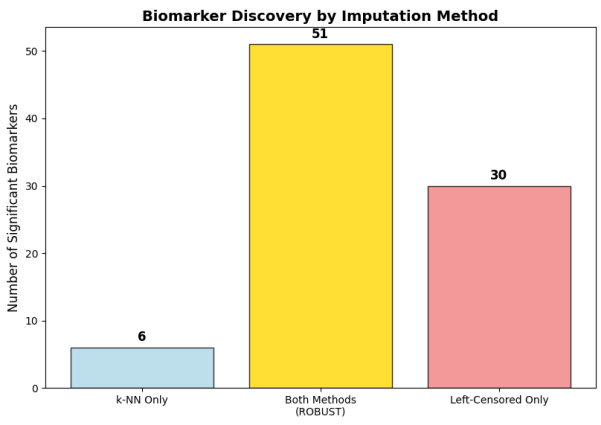
- k-Nearest Neighbors (k-NN) Imputation is a data-driven approach. It assumes that a missing protein value can be estimated from the k proteins with the most similar expression profiles across all samples. This method is good at preserving the natural covariance structure of the data, making it a strong choice for general-purpose imputation.
- Left-Censored Imputation is a biologically-motivated approach. It assumes that values are missing because their concentration was below the instrument's limit of detection. It therefore imputes missing data by sampling from a new, narrow distribution of low-intensity values, effectively treating them as present but low.

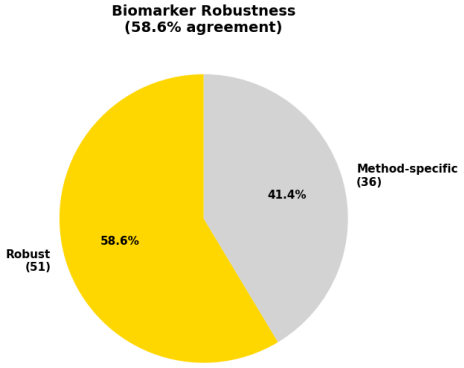
Analyzing the intersection of significant biomarkers identified from both analyses allows us to infer the robustness of our findings. A biomarker that is significant regardless of the imputation method used is a high-confidence candidate, as its discovery is not dependent on a single assumption about the nature of the missing data.

```
================================================================================
COMPARING IMPUTATION METHODS: k-NN vs Left-Censored
================================================================================
Original data shape: (199, 4319)
Number of biomarkers: 4293

1. Performing k-NN imputation…
2. Performing Left-Censored imputation…

Missing values before imputation: 48103
Missing values after k-NN: 0
Missing values after Left-Censored: 0

3. Performing differential expression analysis…
k-NN analysis completed: 4293 biomarkers analyzed
Left-Censored analysis completed: 4293 biomarkers analyzed
```




```
================================================================================
ROBUST BIOMARKER ANALYSIS
================================================================================
 SUMMARY STATISTICS:
   k-NN significant biomarkers: 57
   Left-Censored significant biomarkers: 81
   ROBUST biomarkers (both methods): 51
   k-NN only: 6
   Left-Censored only: 30
```






```
   Upregulated robust biomarkers: 37
   Downregulated robust biomarkers: 14

================================================================================
ROBUST BIOMARKERS SUMMARY TABLE
```

	Biomarker	k-NN_log2FC	k-NN_pval	LC_log2FC	LC_pval	Direction
0	5NTD_HUMAN	-0.619	0.004	-0.690	0.002	Down
1	ACE_HUMAN	-0.681	0.021	-0.756	0.013	Down
2	ADIRF_HUMAN	0.596	0.005	0.549	0.013	Up
3	ANX13_HUMAN	0.516	0.015	0.680	0.004	Up
4	C1QBP_HUMAN	0.552	0.003	0.595	0.002	Up
5	CBPM_HUMAN	-0.773	0.023	-0.773	0.023	Down
6	CF132_HUMAN	0.623	0.011	0.623	0.011	Up
7	CO1A2_HUMAN	-0.580	0.047	-0.580	0.047	Down
8	COXM1_HUMAN	0.559	0.003	0.647	0.004	Up
9	CRBG2_HUMAN	0.532	0.025	0.532	0.025	Up
10	CRCT1_HUMAN	0.574	0.015	0.669	0.009	Up
11	CSPG2_HUMAN	-0.881	0.007	-0.897	0.007	Down
12	CYC_HUMAN	0.596	0.008	0.596	0.008	Up
13	DMBT1_HUMAN	-0.603	0.005	-0.603	0.005	Down
14	DMKN_HUMAN	0.559	0.011	0.559	0.011	Up
15	DPP4_HUMAN	-0.501	0.015	-0.501	0.015	Down
16	EFNA1_HUMAN	-0.563	0.021	-0.563	0.033	Down
17	ENPP3_HUMAN	-0.554	0.021	-0.580	0.024	Down
18	F162A_HUMAN	0.867	0.002	0.879	0.003	Up
19	FOSL2_HUMAN	0.560	0.007	0.570	0.010	Up
20	FRIL_HUMAN	0.613	0.016	0.613	0.016	Up
21	FRRS1_HUMAN	0.586	0.012	0.601	0.027	Up
22	HV309_HUMAN	0.580	0.003	0.539	0.007	Up
23	HV323_HUMAN	0.540	0.003	0.540	0.003	Up
24	HV335_HUMAN	0.515	0.004	0.515	0.004	Up
25	HV70D_HUMAN	0.612	0.022	0.581	0.033	Up
26	IBP7_HUMAN	-0.687	0.030	-0.687	0.030	Down
27	ISG15_HUMAN	-0.511	0.000	-0.567	0.000	Down
28	KCRB_HUMAN	-0.538	0.002	-0.530	0.004	Down
29	KVD29_HUMAN	0.513	0.021	0.556	0.021	Up
30	LV746_HUMAN	0.590	0.005	0.590	0.005	Up
31	LYPD2_HUMAN	0.725	0.004	0.725	0.004	Up
32	MAST4_HUMAN	0.522	0.008	0.554	0.008	Up
33	MIC27_HUMAN	0.523	0.025	0.636	0.020	Up
34	MYOF_HUMAN	-0.510	0.001	-0.514	0.001	Down
35	NDUA7_HUMAN	0.923	0.000	0.861	0.004	Up
36	NMES1_HUMAN	0.713	0.011	0.671	0.019	Up
37	OLFM4_HUMAN	0.535	0.009	0.535	0.009	Up
38	PPBT_HUMAN	-0.777	0.001	-0.806	0.001	Down
39	RHCG_HUMAN	0.582	0.006	0.614	0.005	Up
40	RL29_HUMAN	0.652	0.015	0.692	0.012	Up
41	SARG_HUMAN	0.600	0.012	0.600	0.012	Up
42	SBSN_HUMAN	0.681	0.006	0.681	0.006	Up
43	SCAM1_HUMAN	0.519	0.008	0.508	0.020	Up
44	SPR1B_HUMAN	0.685	0.011	0.713	0.009	Up

45	SPR2A_HUMAN	0.542	0.040	0.588	0.029	Up
46	SPR2F_HUMAN	0.559	0.031	0.613	0.028	Up
47	TM11B_HUMAN	0.520	0.022	0.516	0.027	Up
48	TM11E_HUMAN	0.570	0.006	0.570	0.006	Up
49	TMA7_HUMAN	0.705	0.003	0.758	0.004	Up
50	TMM40_HUMAN	0.623	0.010	0.542	0.034	Up