Imperial College London

Asthma subtyping and proteomic markers

10th May 2023

Siwei Wu, Fiona Rice, Seth Howes, Constance Popovic

Background

Imperial College London

- Asthma is most common chronic lung disease (262 million people worldwide in 2019)[1]
- Highly heterogeneous disease with multiple ways of subtyping
  - Immune – eosinophilic vs neutrophilic
  - Severity - mild, moderate, severe, life-threatening
  - Combination clusters - early-onset atopic and obese, noneosinophilic etc.[2]
- Omics-based approaches used in phenotyping other diseases
  - e.g. Mito-ML subtype of AML[3]
- Asthma proteomic endotyping not widely used in clinic as:
  - Nascent technology
  - Difficulty comparing different platforms[4]

Research aims

Imperial College London

For asthma (A, B, C) and control (D) cohorts

- **Aim 1** - identify proteomic markers associated with **asthma status**

For asthma cohorts (A, B C) only

- **Aim 2** - identify proteomic markers associated with **asthma immune subtypes**
- **Aim 3** - identify proteomic markers associated with **asthma severity**
    1. Diagnosed severity (mild/moderate vs. severe)
    2. Additional severity metrics (e.g. FEV1 percentage, OCS use)
    3. Derived severity score

Methods

Imperial College London

Aim 1: identify proteomic markers associated with asthma status (cohort A&B &C vs D)

- Univariate analysis: Logistic regression, BH adjusted
- Stability selection: Logistic LASSO

Aim 2: identify proteomic markers associated with asthma immune subtypes

- Univariate analysis: Multinomial regression, BH adjusted
- Stability selection: Multinomial LASSO

All models are adjusted for age, sex, race, BMI and smoking status

Methods

Imperial College London

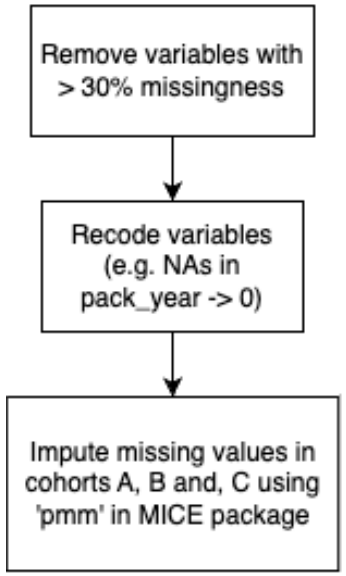
Aim 3: identify proteomic markers associated with asthma severity

1. **Diagnosed severity**: mild/moderate vs severe
   - Univariate analysis: logistic regression, BH adjusted
   - Stability selection: logistic LASSO
2. **10 severity metrics:** FEV1(%); FVC(%); Exacerbation per year; severe exacerbations per year; ACQ 5; OCS; ICU times; age of onset.
   - Stability selection LASSO (severity metric individually)
   - Stability selection sparse PLS (10 severity metrics combined)
3. **Derived severity score from severity metrics**
   - K-Means clustering(calibrated with silhouette score)
   - Univariate analysis: Multinomial logistic, BH adjusted
   - Stability selection: Multinomial logistic LASSO

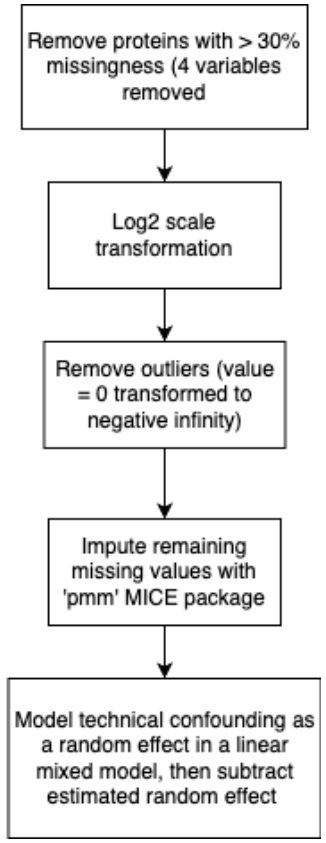
Data pre-processing

Imperial College London

Demographic & Clinical



Protein



- Eosinophil and neutrophil sputum percentage removed from demographic data
- IL-10, 1L-17, TSLP, and IL5 removed from protein data
- Denoised protein modelled as:
  - Denoised protein = protein – random effect (Plate ID) = Intercept + Residual
- Technical confounding modelled as[5]:
  - protein ~ (1|Plate ID)

Demographic summary

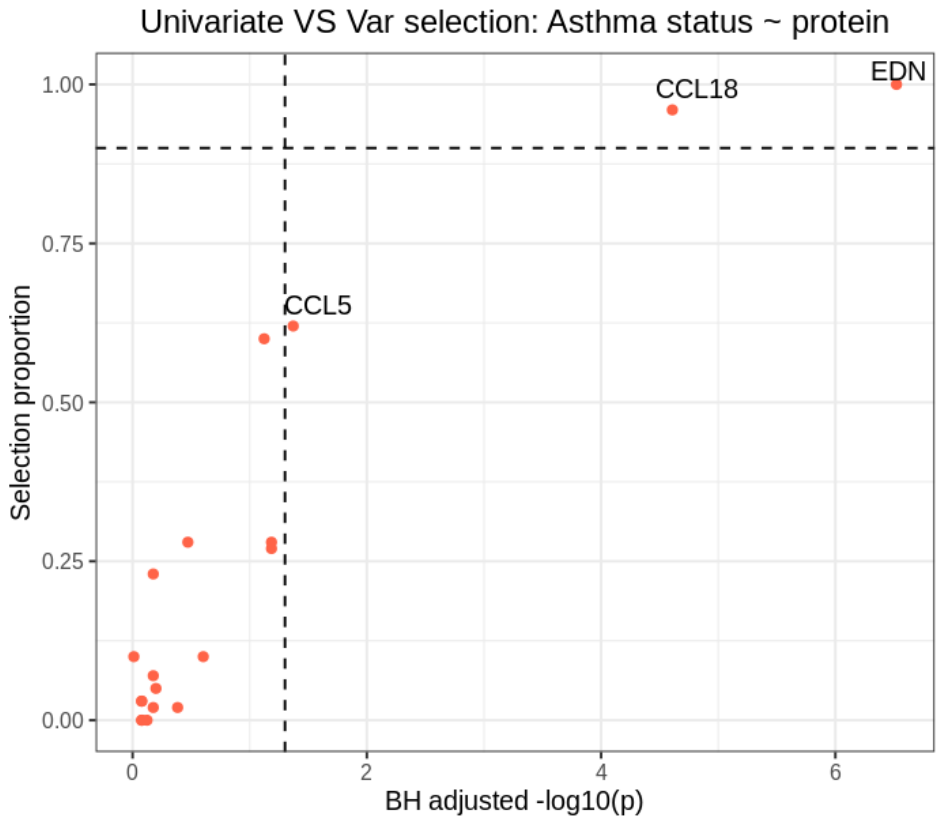
Imperial College London

	Mean (SD) or n (%)			
	Cohort A (n=279)	**Cohort B (n=102)**	**Cohort C (n=83)**	**Cohort D (n=93)**
Sex				
Female	186(66.7)	52(51.0)	41(49.4)	36(38.7)
Male	93(33.3)	50(49.0)	42(50.6)	57(61.3)
Age (years)	51.00 (14.27)	54.81 (11.31)	40.90 (15.44)	39.02 (13.65)
Race				
Others	23(8.2)	10(9.8)	8(9.6)	4(4.3)
White	256(91.8)	92(90.2)	75(90.4)	89(95.7)
BMI	29.11(6.37)	29.50(6.40)	25.55(4.43)	25.40(3.66)
Pack Years	0.41(1.22)	23.25(17.90)	0.42(1.25)	0.34(1.04)

Table 1: Demographic Characteristics of Participants

Aim 1: Proteomic markers associated with asthma status

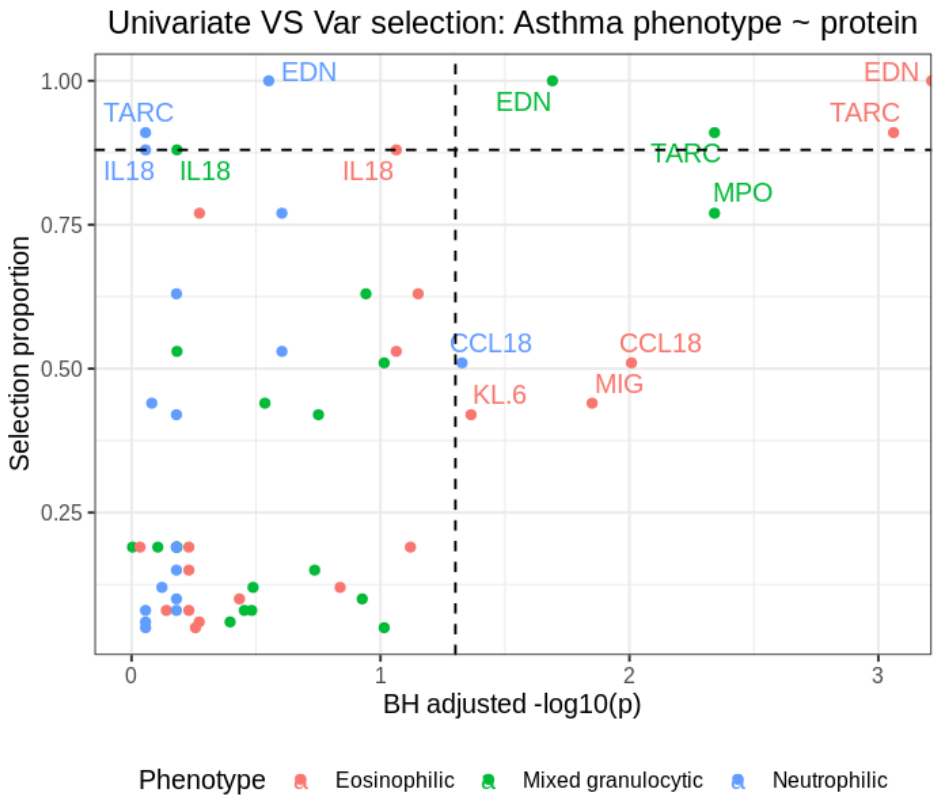
Imperial College London



- **CCL5** is significant in univariate analysis only
- No proteomic markers selected by stability lasso only
- **CCL18** and **EDN** selected by both univariate and stability selection methods

Aim 2: Proteomic markers associated with asthma immune phenotype

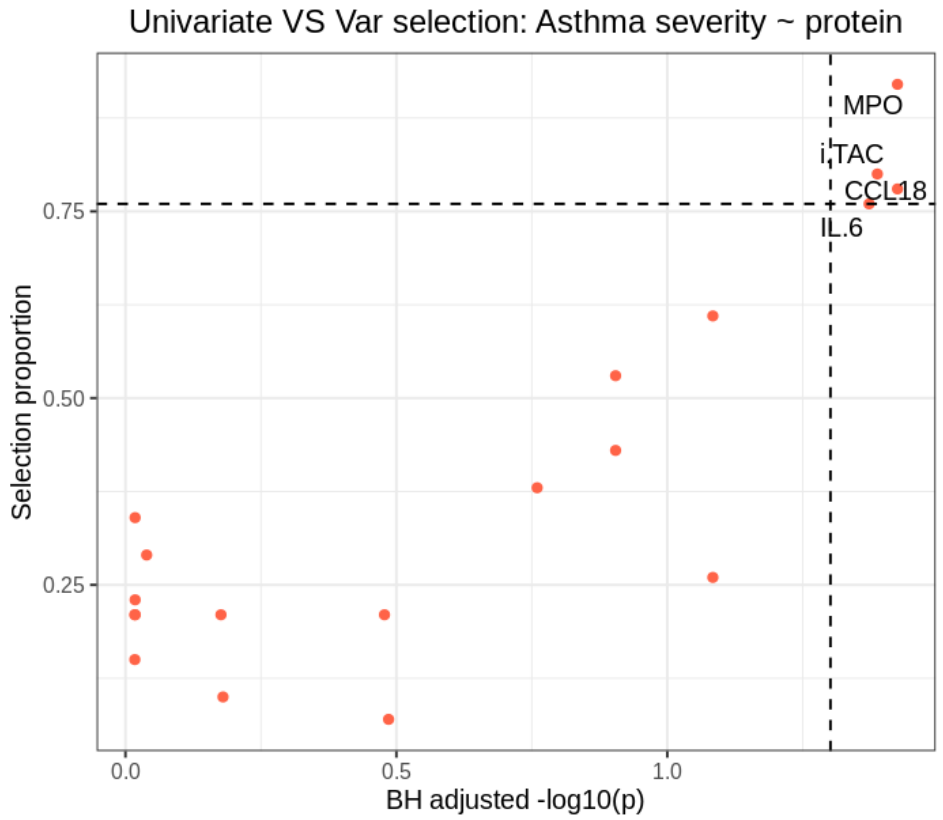
Imperial College London



- **EDN** and **TARC** are selected to be discriminative of different phenotypes
- **EDN**, **TARC** are also significant in univariate analysis for mixed granulocytic and eosinophilic subtypes
- More signals for mixed granulocytic

Aim 3.1: Proteomic markers associated with diagnosed asthma severity

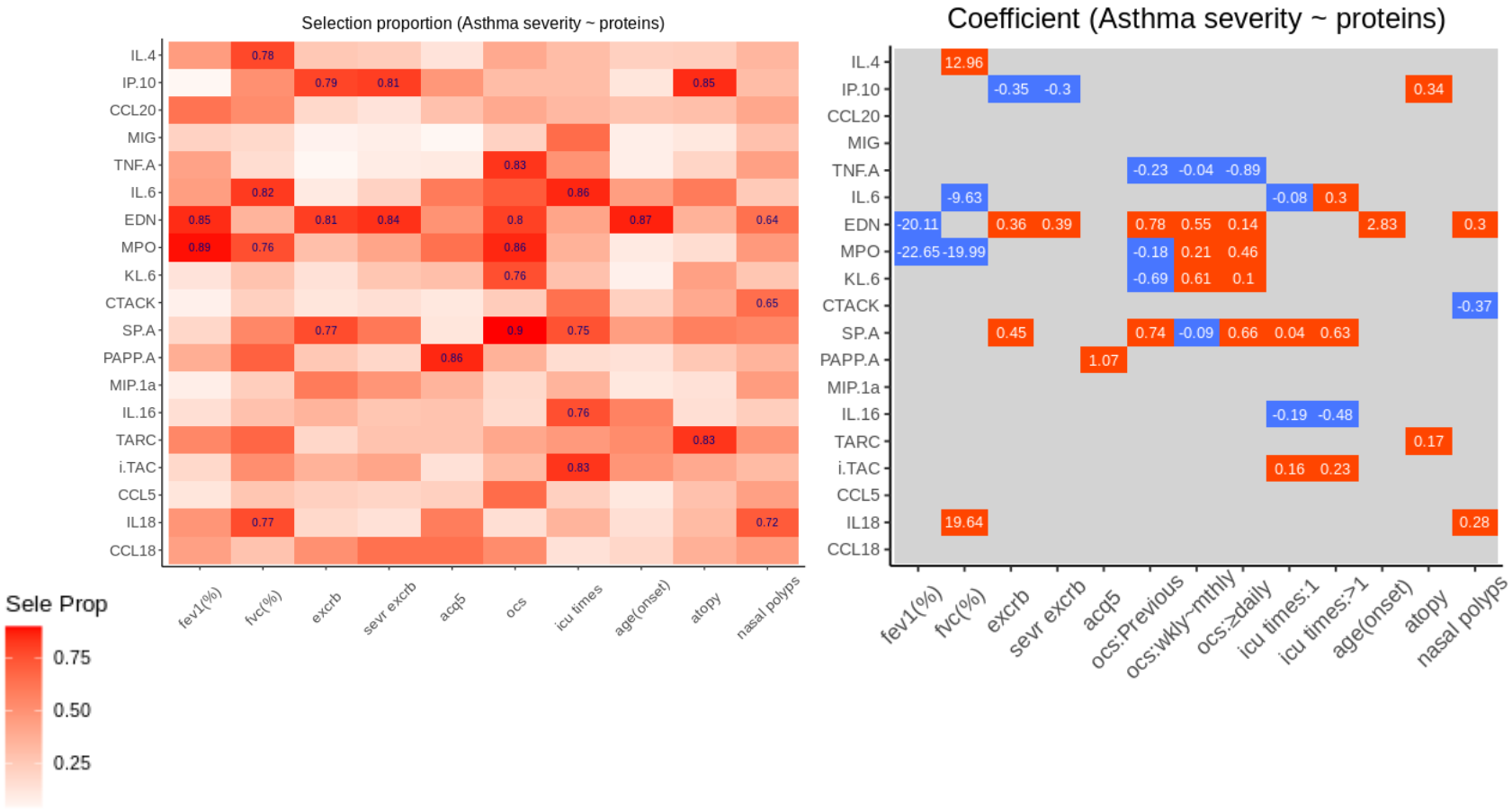
Imperial College London



- **MPO, ITAC**, **CCL18**, **IL-6** are associated with severe asthma compared to mild/moderate asthma
- No proteomic markers identified by univariate or stability selection analyses only

Aim 3: Stability selection lasso for 10 different clinical severity metrics

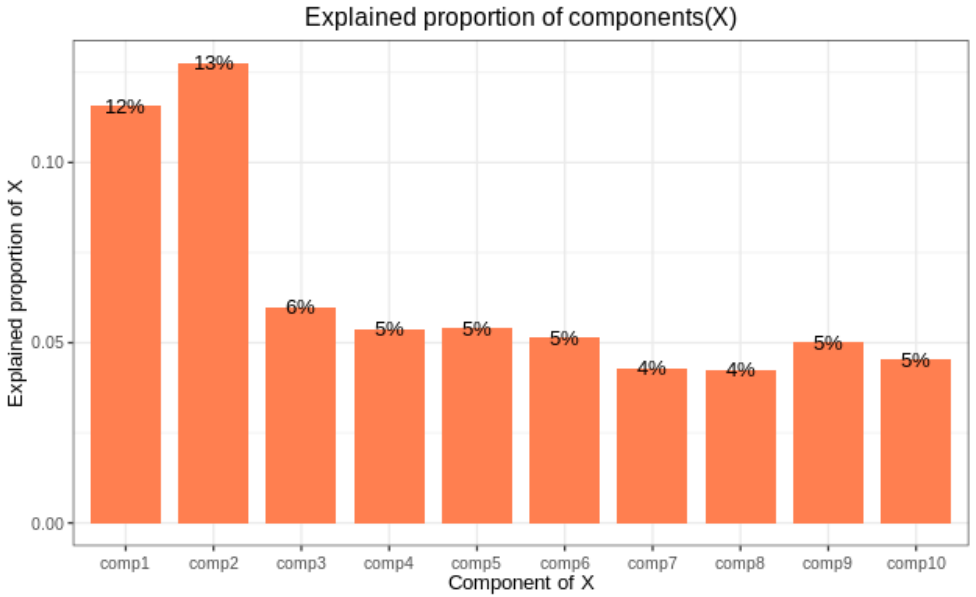
Imperial College London

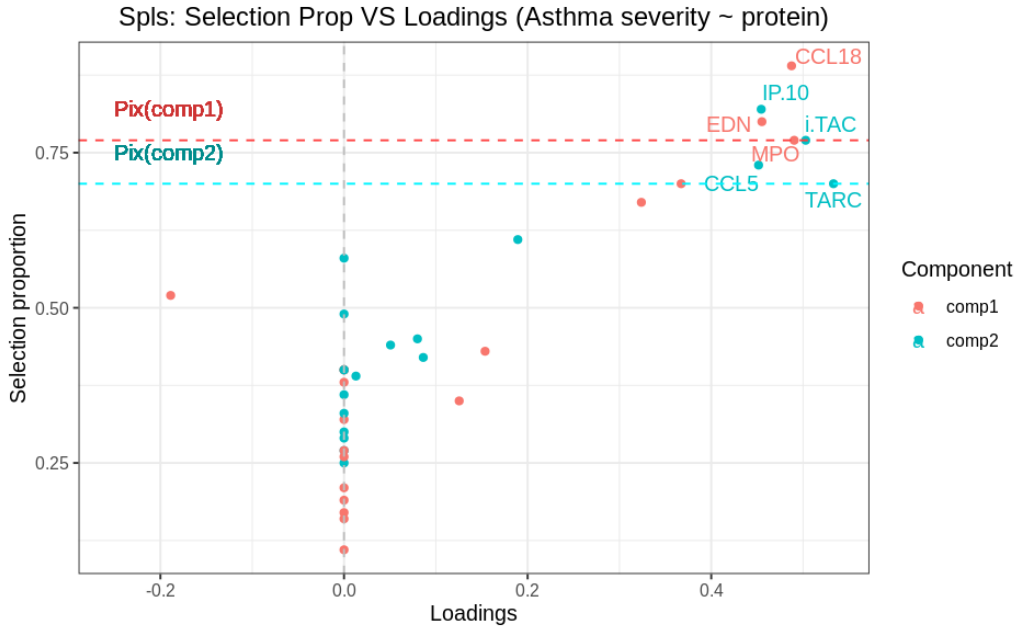


- All proteomic markers are associated with at least one severity metric
- **EDN** has the greatest number of associations with different metrics[6]
- **EDN** is positively associated with severe asthma

Aim 3.2: SPLS multivariate outcome, 10 different clinical severity metrics

Imperial College London



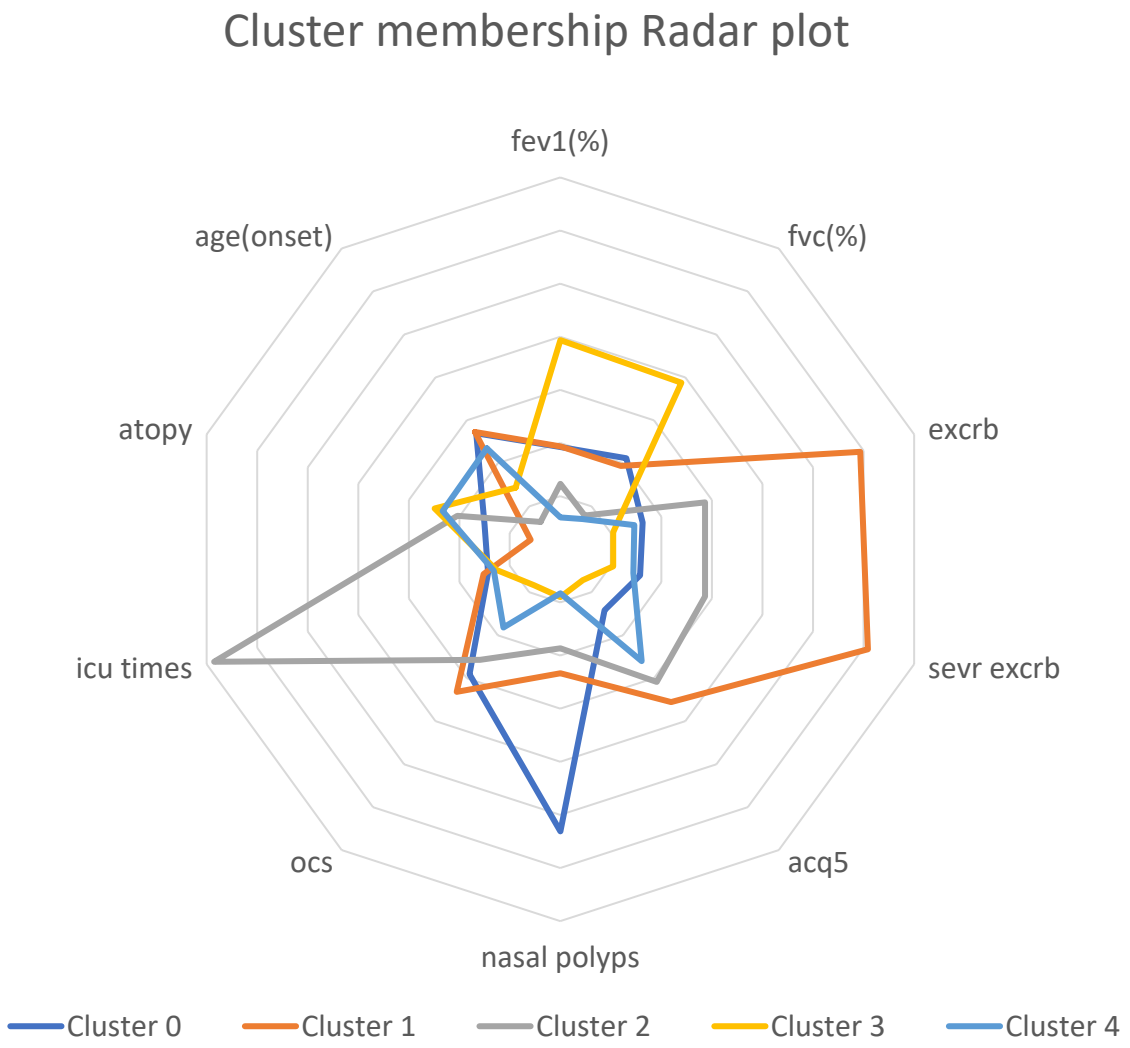


- First two components explain 25% of the variance between outcome loadings and independent variables

- **CCL-18**, **EDN,** and **MPO** are stably selected for component 1
- **IP-10**, **ITAC**, **CCL-5**, and **TARC** selected for component 2

Aim 3.3: Clinical severity score building

Imperial College London

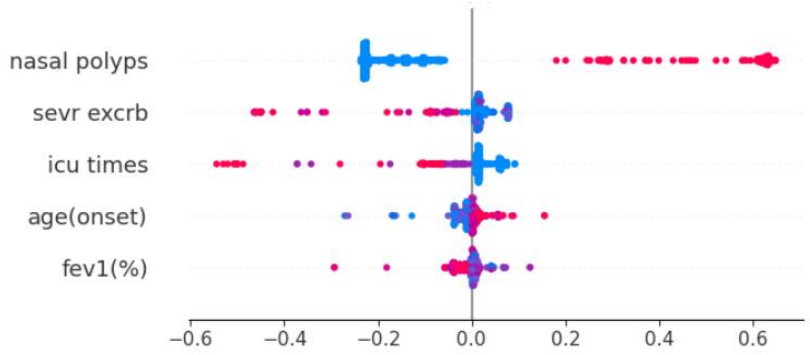


- K-Means clustering elbow point on WCSS is at **k=5 clusters**
- **Cluster 0** chosen as reference group as with highest nasal polyps
- Subtypes tend to be defined by 1 or 2 severity metrics, with other metrics similar between subgroups

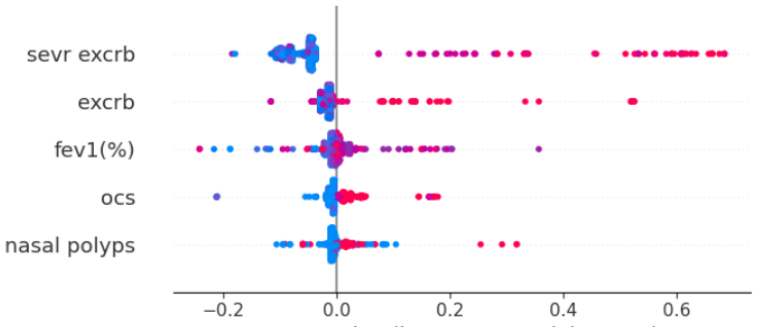
Aim 3.3: SHAP density plots for 5 clusters

Imperial College London

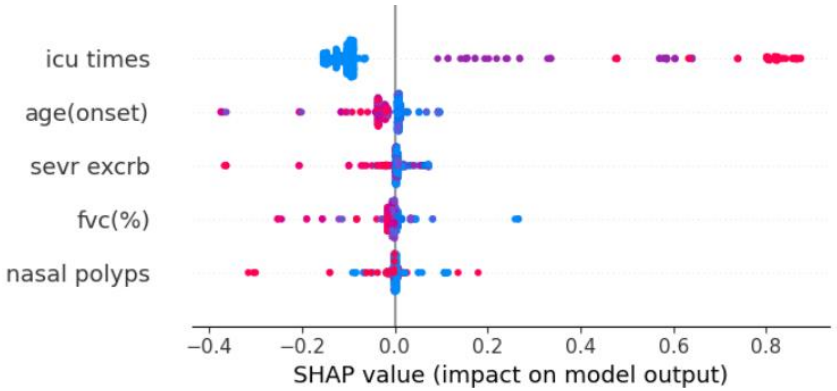
Cluster 0: Nasal polyps with fewer severe exacerbations



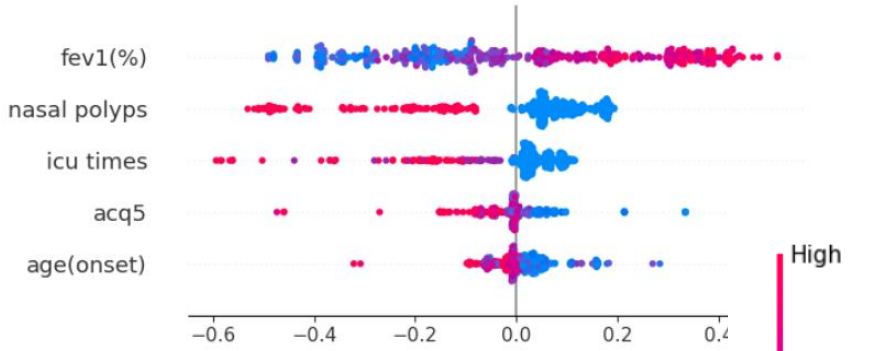
Cluster 1: High Severe exacerbations



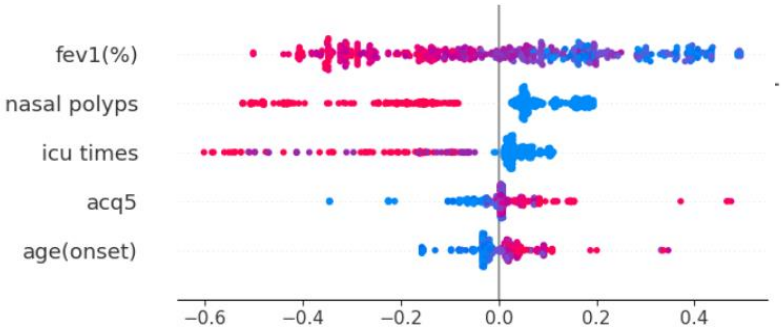
Cluster 2: High ICU times with low age of onset



Cluster 3: Highest FEV1 Well controlled asthma

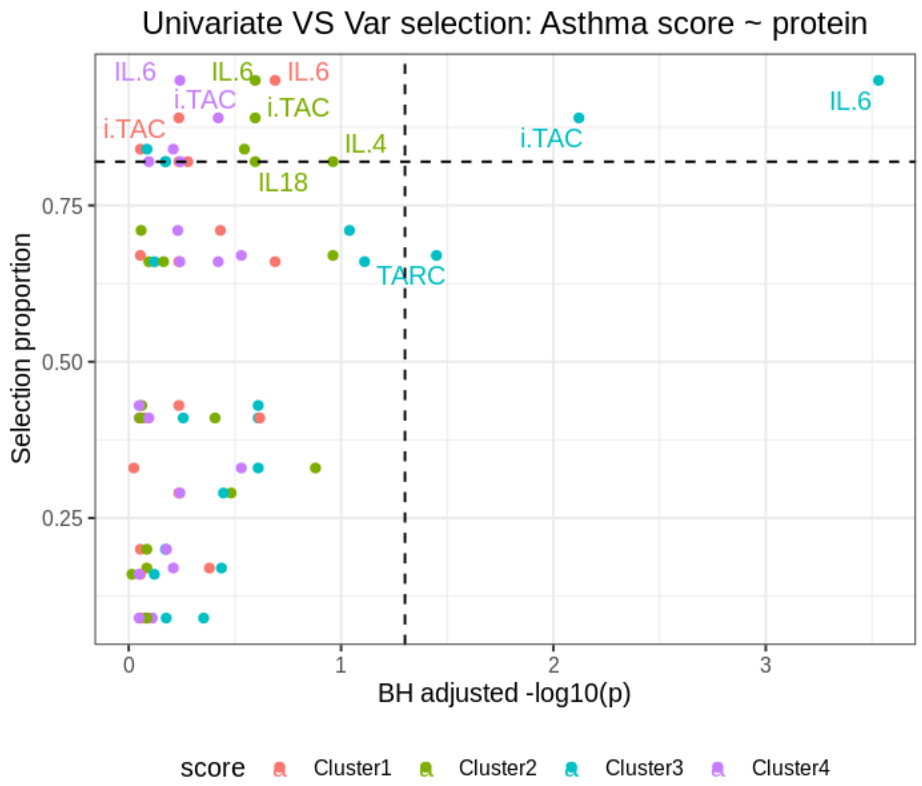


Cluster 4: Low FEV1(%) with few nasal polyps



Aim 3.3: Multinomial regression on 5 severity clusters

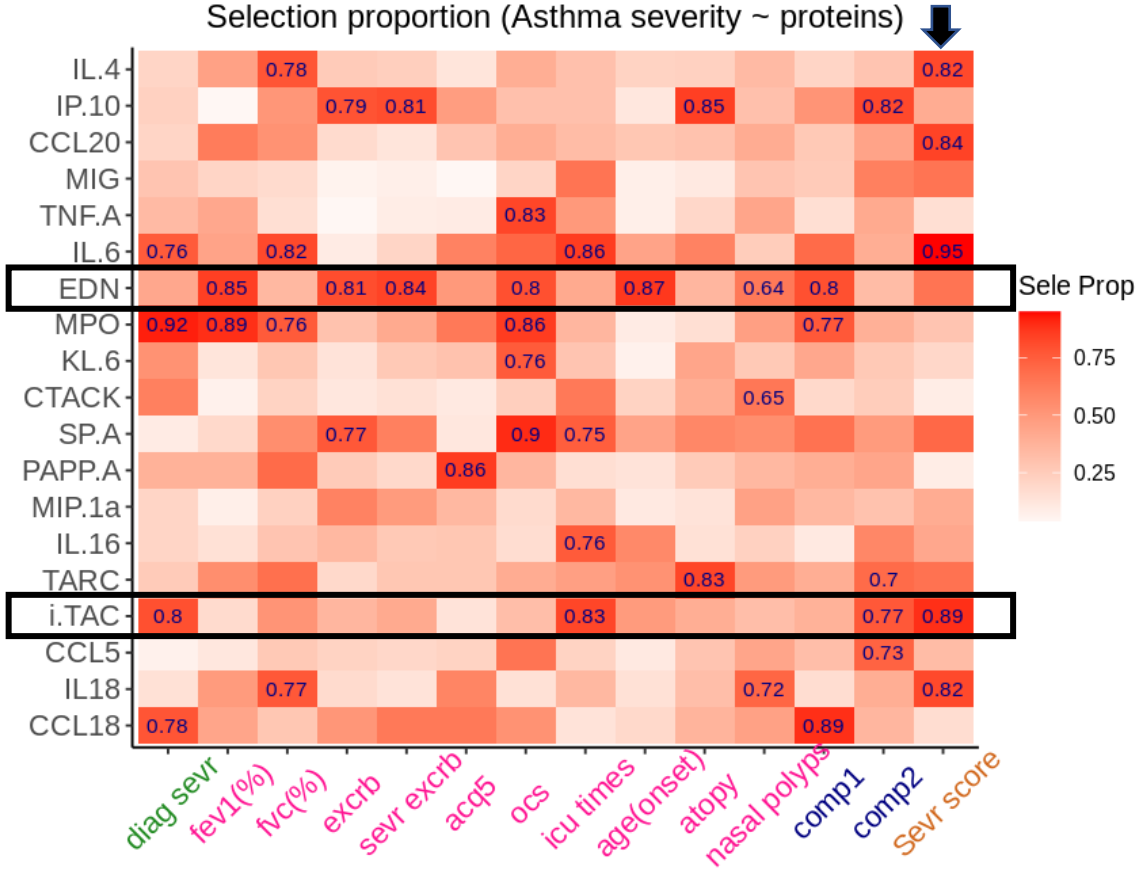
Imperial College London



- I-TAC and IL-6 are both stably selected and significant for **cluster 3**
- I-TAC and IL-6 are selected for every cluster, but not significant in all univariate analyses
- I-TAC, IL-6 and additionally IL.4, IL.18 were stably selected for cluster 2

Aim 3 Summary: Comparing proteins selected using different asthma severity classifications

Imperial College London



- **IL-18**, **IL-6**, **IL-4**, **I-TAC**, and **CCL20** are most important to discriminate different severity scores.
- **EDN** is predictive of individual metrics[7] but not representative of joint effect of different proteins
- **ITAC** is representative of joint effect of proteins

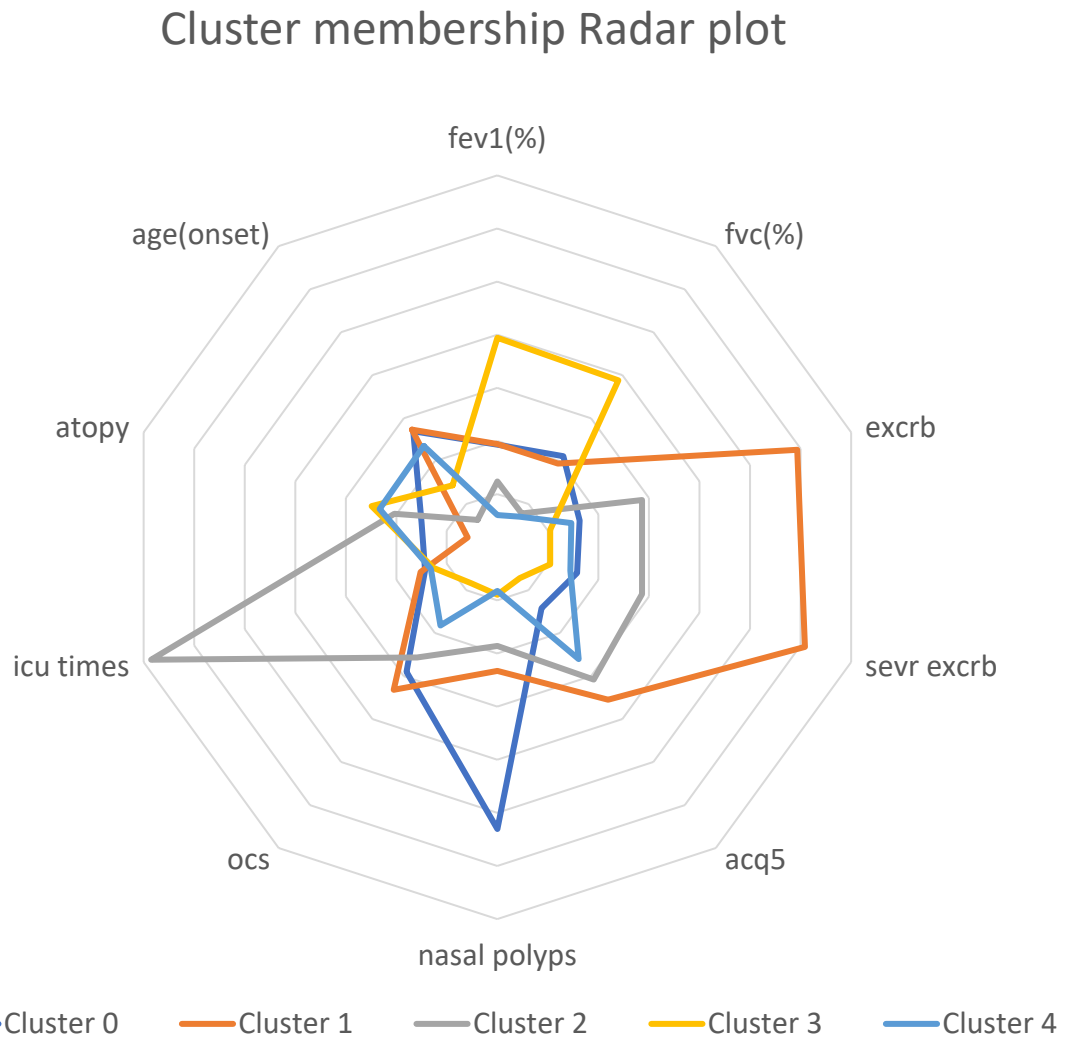
Conclusions

Imperial College London

- **CCL18** and **EDN** are indicators of asthma status
- **EDN** and **TARC** are associated with eosinophilic and mixed granulocytic asthma
- **5 different subtypes** capture participants with different clinical severity characteristics (e.g. high ICU times)
- **Cluster 4** is the only cluster associated with biomarkers both in univariate and stability selection methods (**IL-6** and **ITAC**)
- **IL-6** and **ITAC** are consistently across all analyses as being associated with severe asthma phenotypes

Discussion - severity score interpretation

Imperial College London



Cluster number	Cluster characteristics	Clinical interpretation
0	High nasal polyps	Chronic asthma
1	High normal and severe exacerbations	Poorly-controlled asthma, moderate FEV1, FVC
2	High ICU times, low FVC1	Severe asthma
3	High FEV1 and FVC	Well-controlled asthma
4	High ACQ5, very low FV1, FVC	Very severe asthma

Discussion – significant proteins

Imperial College London

- **IL-6** and **ITAC** have been identified as being associated with severe asthma in other studies[6]
- **IL-6** also associated with severe and critically ill COVID-19 cases[7]
- Potential for these factors to be used in assessing global asthma severity
- **Tocilizumab** (monoclonal antibody targeting **IL-6**) is currently licensed for use in treating severe COVID-19 and rheumatoid arthritis
- May be useful for treating patients with very severe (**cluster 2**) and very severe (**cluster 4**) asthma

Future work

Imperial College London

- Lack of exploration of non-linear relationships -> use SVM regression to capture these relationships
- Did not explore the group structure of the biomarkers (chemokines, cytokines, other immune) -> apply sg-PLS
- Carry out experimental assessment of the mechanistic role of significant biomarkers (e.g. **IL-6** and **I-TAC**) in asthma
- *In vivo* assessment of the efficacy of **IL-6** blockers

Imperial College London

Questions?

References

Imperial College London

1. GBD 2019 Diseases and Injuries Collaborators. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet, 396(10258), 1204-1222.
2. Fahy et al. (2015). Type 2 inflammation in asthma--present in most, absent in many. Nature Reviews Immunology, 15(1), 57-65.
3. Jayavelu et al.(2022). Oncogenic activation of FLT3 and c-KIT receptor tyrosine kinases promotes cancer stemness and resistance against the multikinase inhibitor dovitinib. Cancer Cell, 40(3), 301-317.
4. Ivanova et al. (2019). What did we learn from multiple omics studies in asthma? Allergy, 74(11), 2129-2145.
5. Vermeulen et al. (2018). Pre-diagnostic blood immune markers, incidence and progression of B-cell lymphoma and multiple myeloma: Univariate and functionally informed multivariate analyses. International Journal of Cancer, 143(6), 1335-1347.
6. Ramphul et al. (2022). Sputum biomarkers during acute severe asthma attacks in children-a case-control study. Acta Paediatrm 111(3), 620-627.
7. Gao et al. (2021). Risk factors for severe and critically ill COVID-19 patients: A review. Allergy, 76(2), 428-455.

Imperial College London

Appendix

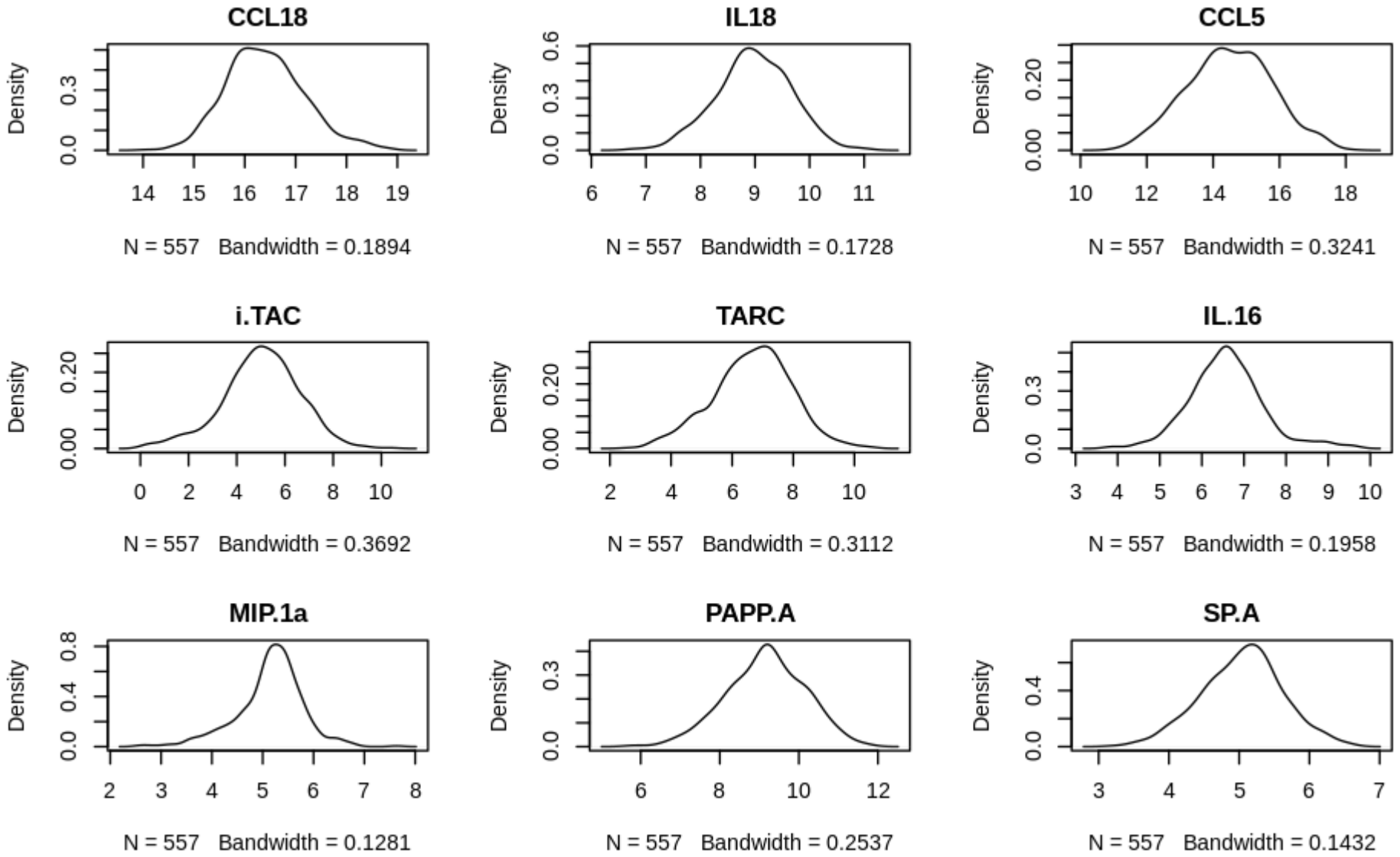


Figure A1: Distributions of selected Proteins across study population. Each subplot represents the distribution of one protein, showing its variance among the different cohorts. The plots provide insights into the heterogeneity of protein expression levels in the study population, thus informing the subsequent proteomic analyses.

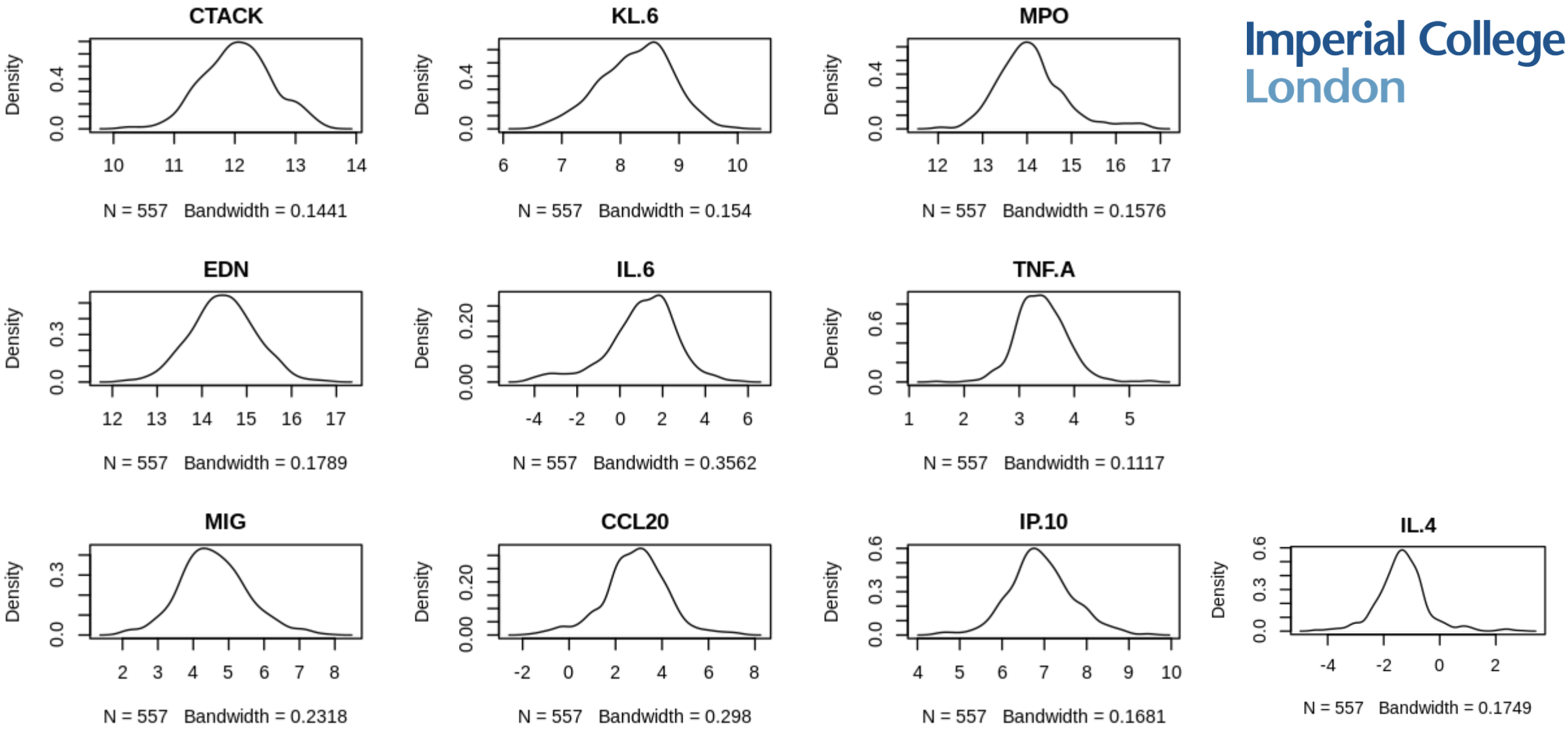


Figure A2: Distributions of selected Proteins across study population. Each subplot represents the distribution of one protein, showing its variance among the different cohorts. The plots provide insights into the heterogeneity of protein expression levels in the study population, thus informing the subsequent proteomic analyses.

Imperial College London

	CCL18	IL18	CCL5	i.TAC	TARC	IL.10	IL.16	MIP.1a	PAPP.A	IL17	TSLP	SP.A	CTACK	KL.6	MPO	EDN	IL.6	TNF.A	MIG	CCL20	IP.10	IL.4	IL.5
NA (Count)	2	2	2	1	1	1	1	1	5	2	7	1	1	2	2	2	1	1	1	1	1	1	3
NA (%)	0.36%	0.36%	0.36%	0.18%	0.18%	0.18%	0.18%	0.18%	0.90%	0.36%	1.26%	0.18%	0.18%	0.36%	0.36%	0.36%	0.18%	0.18%	0.18%	0.18%	0.18%	0.18%	0.54%
NA (Inf as NA) (Count)	2	2	3	20	17	421	1	1	6	266	231	1	1	2	2	2	60	1	1	57	1	8	306
NA (Inf as NA) (%)	0.36%	0.36%	0.54%	3.59%	3.05%	75.58%	0.18%	0.18%	1.08%	47.76%	41.47%	0.18%	0.18%	0.36%	0.36%	0.36%	10.77%	0.18%	0.18%	10.23%	0.18%	1.44%	54.94%
NA (Inf and outliners as NA) (count)	7	7	6	28	24	421	13	9	9	268	237	5	2	4	10	8	67	10	12	66	11	20	308
NA (Inf and outliners as NA) (%)	1.26%	1.26%	1.08%	5.03%	4.31%	75.58%	2.33%	1.62%	1.62%	48.11%	42.55%	0.90%	0.36%	0.72%	1.80%	1.44%	12.03%	1.80%	2.15%	11.85%	1.97%	3.59%	55.30%

Table A1: Summary of Missing Data in Protein Measurements. This table presents the counts and percentages of missing values in different protein measurements at various data preprocessing stages. Four proteins (IL-10, IL-17, TSLP, and IL5) with > 30% missingness are identified for removal, while missing values in other proteins are imputed. Each row corresponds to a specific protein, and each column represents a stage in the data cleaning process: initial data, after removing infinite values, and after further handling of outliers. This table provides an overview of the data quality and the effectiveness of the cleaning methods employed.

Imperial College London

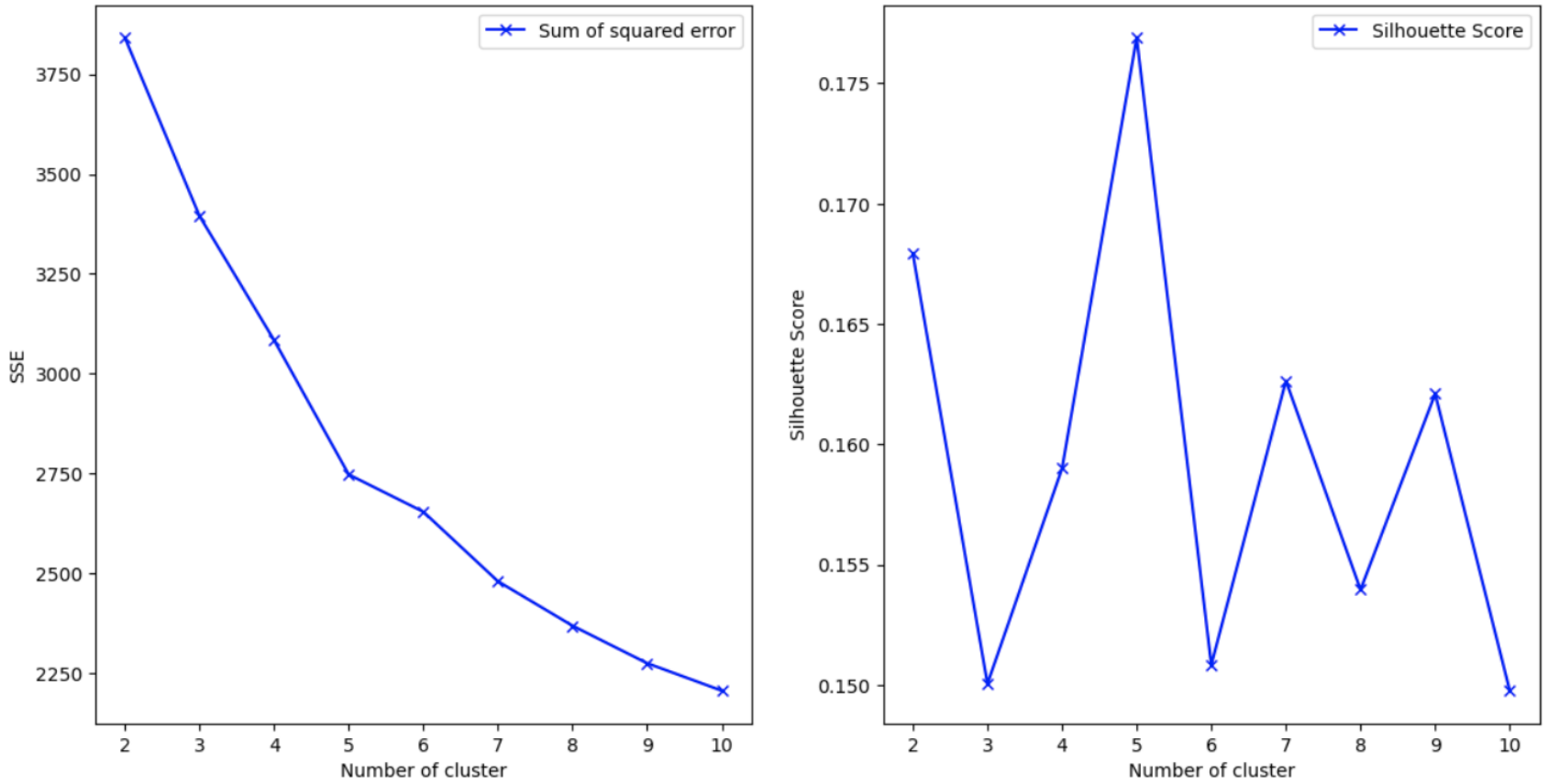


Figure A3: Elbow Method and Silhouette Score for Determining Optimal Cluster Number in Clinical Severity Score. The left plot employs the Elbow Method, showing the sum of squared distances from each point to its assigned centre, while the right plot presents the Silhouette Scores, reflecting the quality of clustering. Both methods aid in identifying the most appropriate number of clusters for the clinical severity score.

Imperial College London

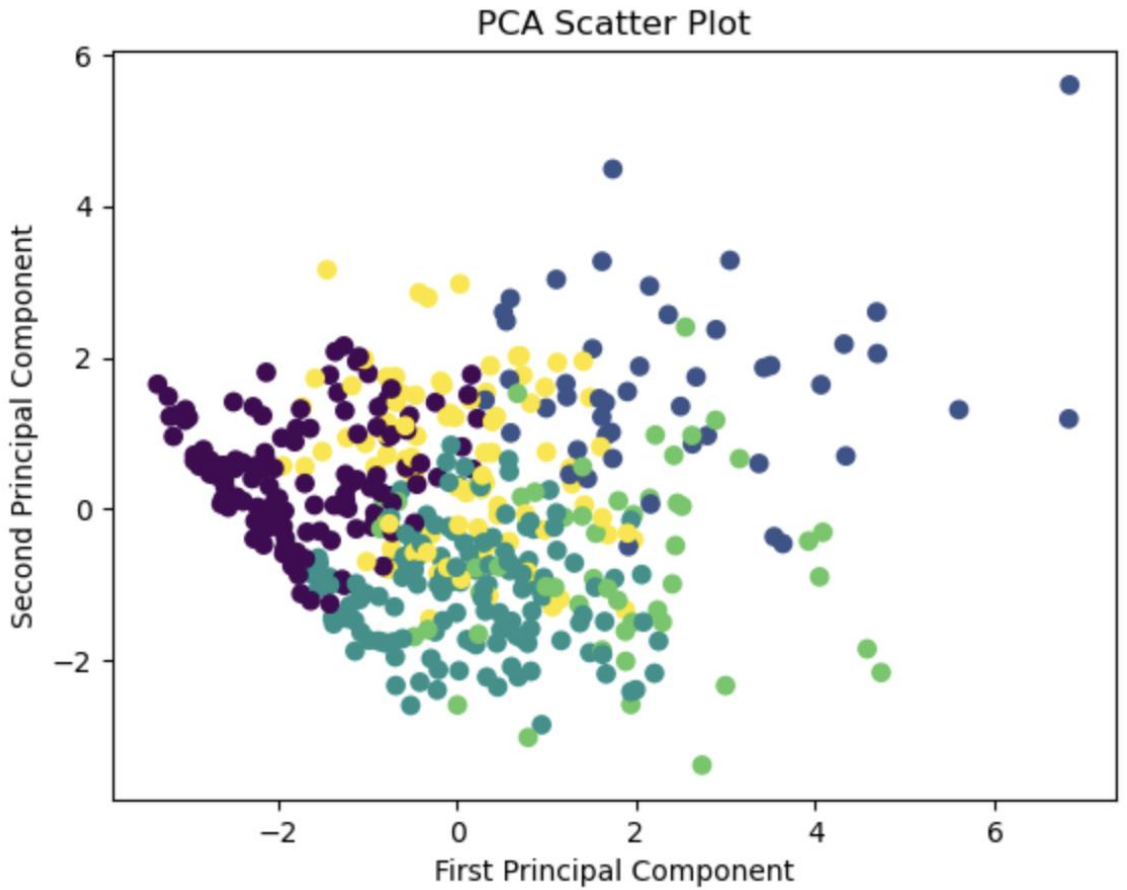


Figure A4: Principal Component Analysis (PCA) of Asthma Severity Clusters. This plot visualizes the distribution of asthma severity clusters in a reduced-dimensionality space. Each point corresponds to an individual patient, color-coded based on their cluster assignment. The axes represent the first and second principal components, which capture the largest sources of variation in the data.

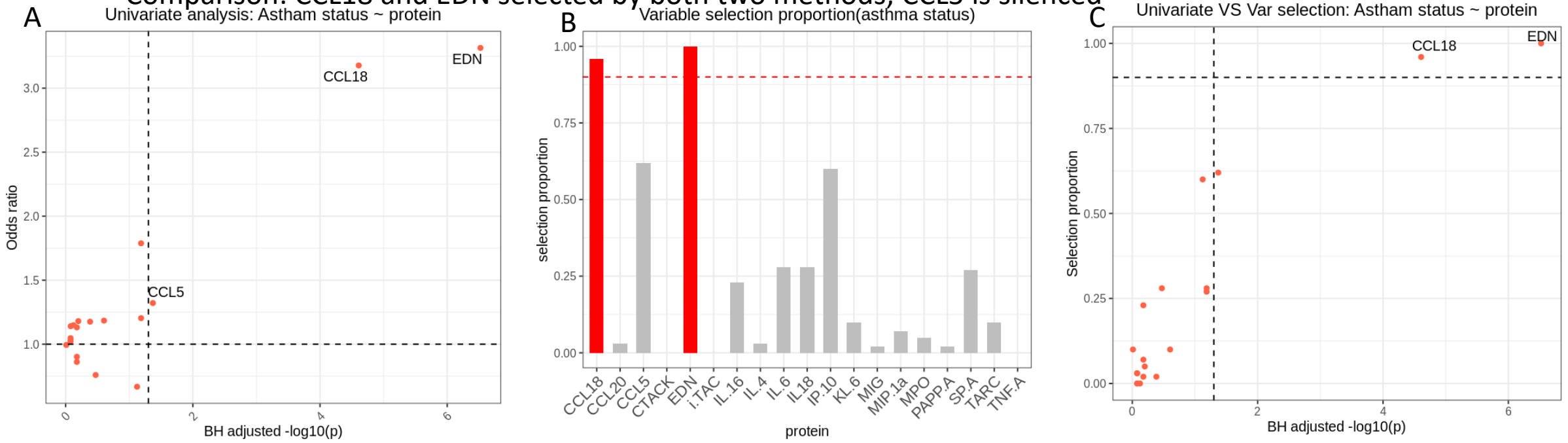
Immune cutoffs

Imperial College London

Paucigranulocytic: serum eosinophil < 300/ul, serum neutrophil < 7500/ul
Eosinophilic: serum eosinophil > 300/ul, serum neutrophil < 7500/ul
Neutrophilic: serum eosinophil < 300/ul, serum neutrophil > 7500/ul
Mixed granulocytic: serum eosinophil > 300/ul, serum neutrophil > 7500/ul

Asthma status ~ Proteomics

- Outcome: Asthma status: **asthma status (ABC vs D)**
- Univariate analysis:
- Stability logistic lasso:
- Comparison: CCL18 and EDN selected by both two methods; CCL5 is silenced

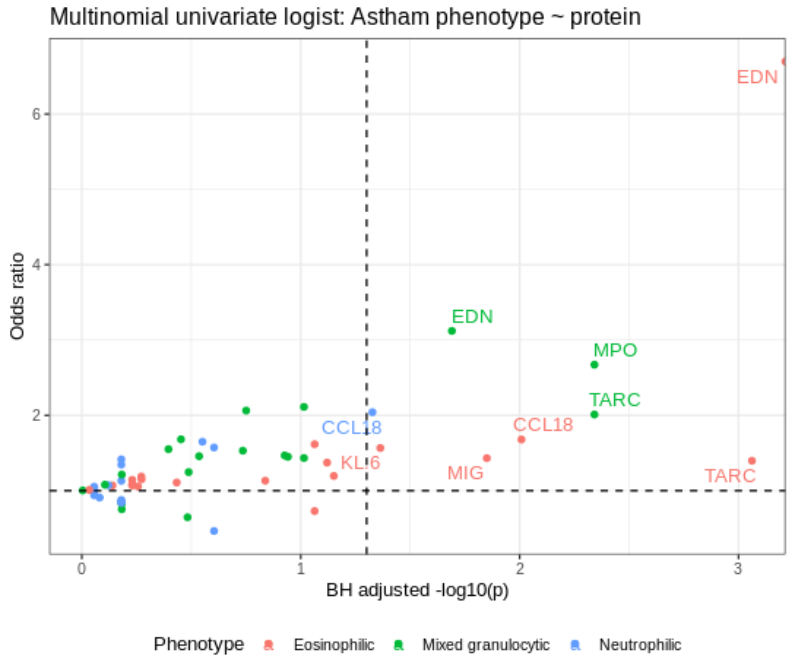


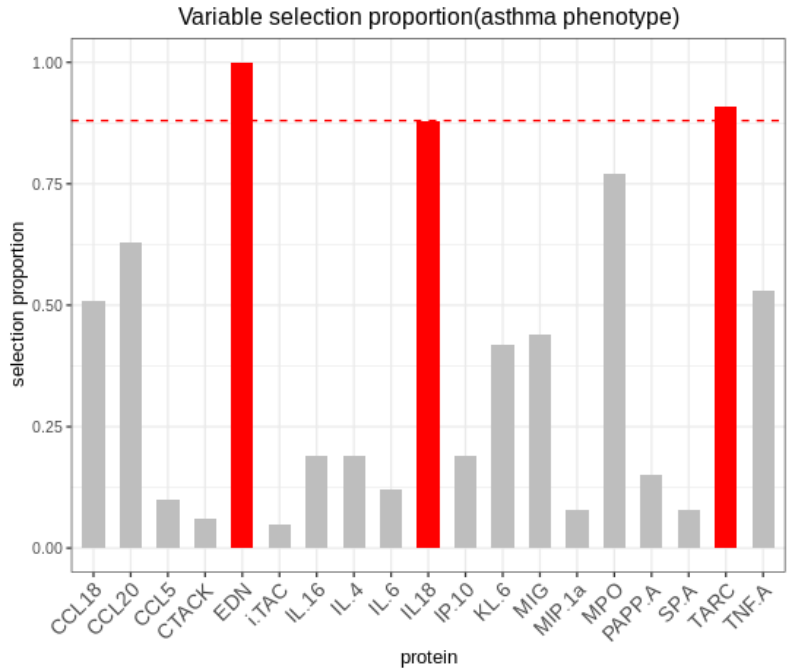
Asthma phenotype ~ Proteomics

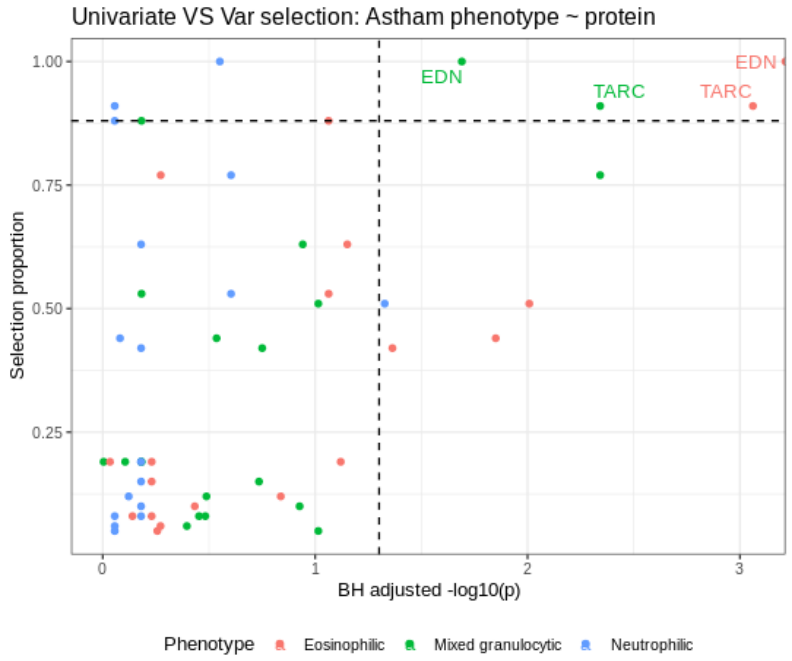
Imperial College London

Outcome:

- Univariate analysis:
- Stability logistic lasso:
- Comparison:



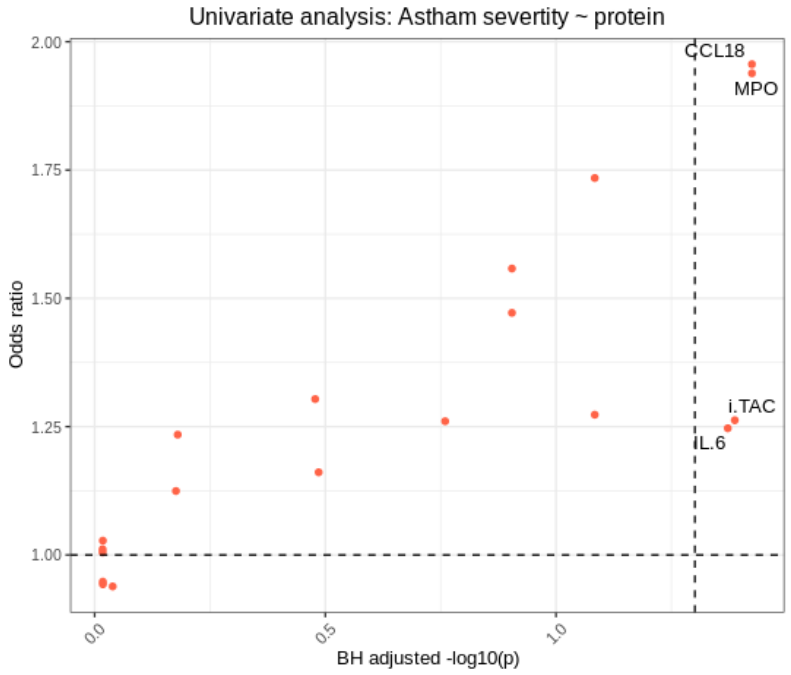


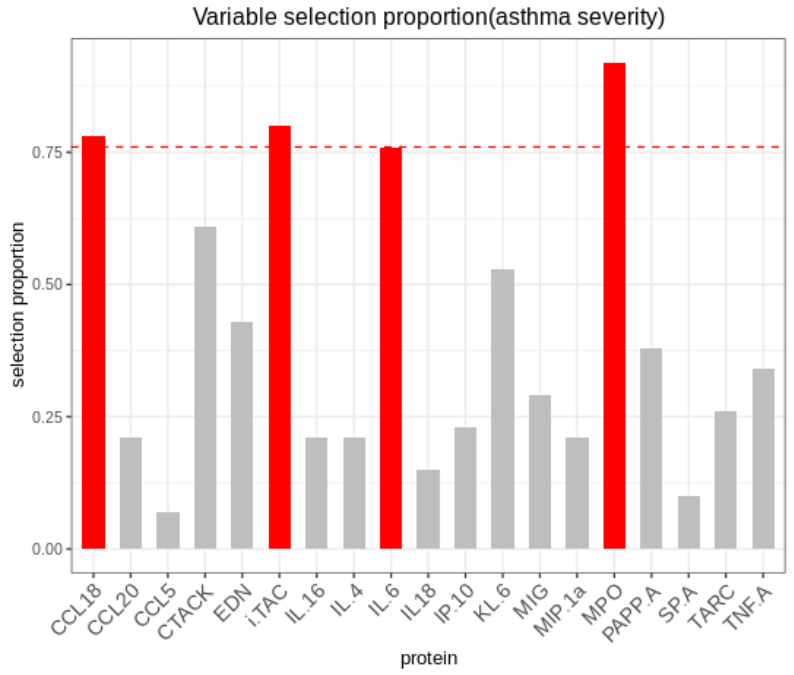


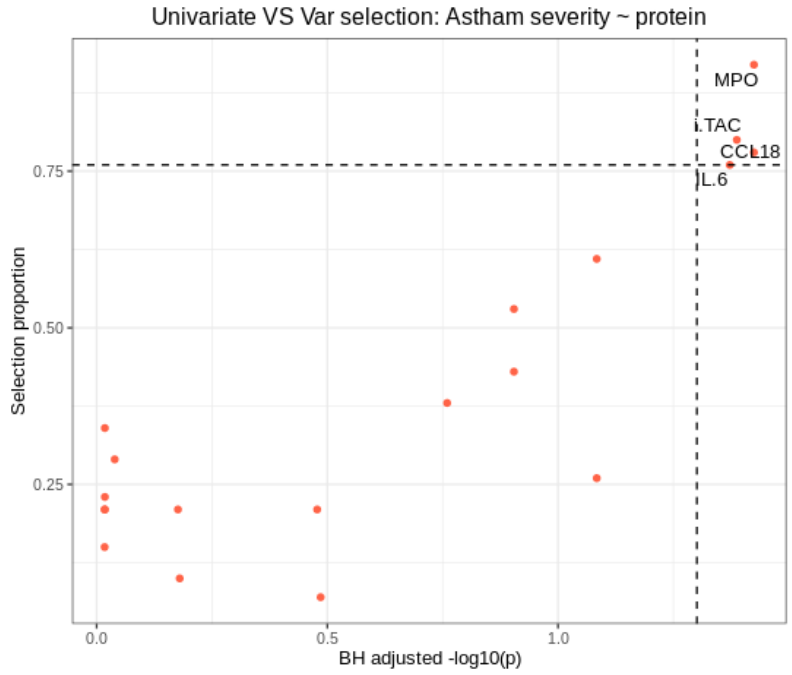
Diagnosed asthma severity ~ Proteomics

Outcome: Diagnosed severe asthma versus non-serve asthma (Cohort A & B versus C)

- Univariate analysis:
- Stability logistic lasso:
- Comparison:

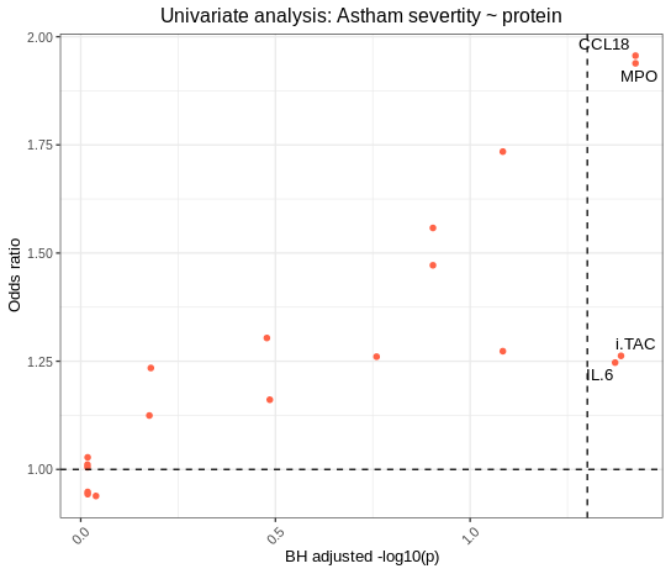


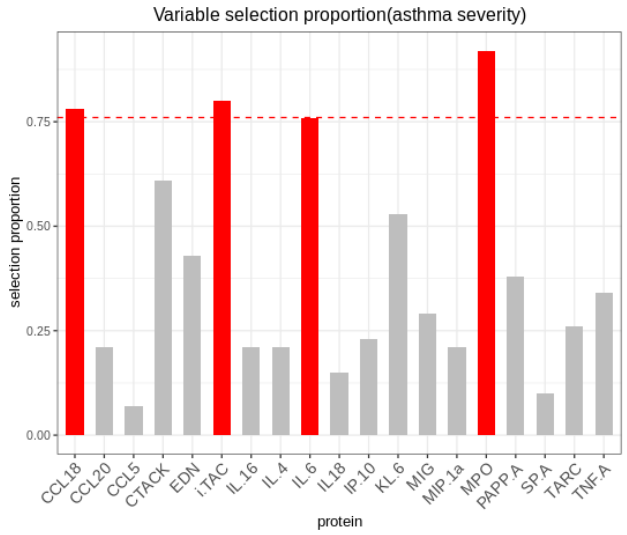


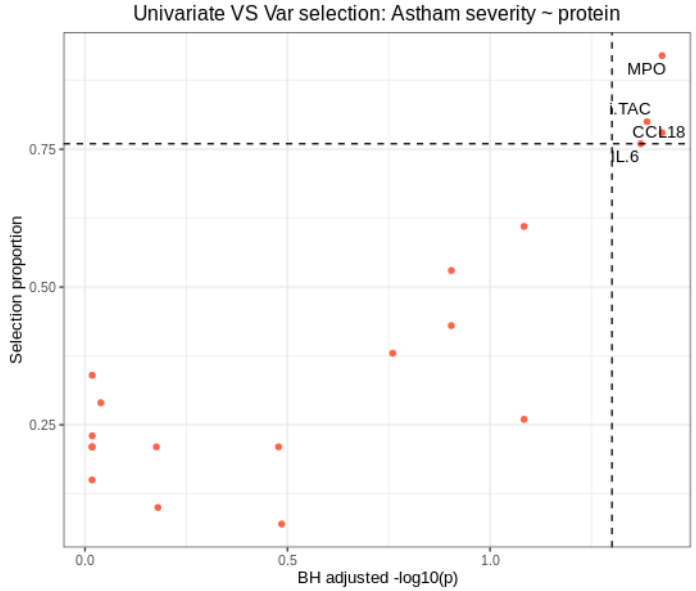


Diagnosed asthma severity ~ Proteomics

Imperial College London







Multinomial regression on 5 severity clusters

Imperial College London