SUPPLEMENTARY RESULTS

Title: Biomarker Signatures for Progressive Idiopathic Pulmonary Fibrosis

Britt Clynick[1,2,3]#, Tamera J Corte[1,4,5]#, Helen E Jo[1,4,5], Iain Stewart[6], Ian N Glaspole[7,8], Christopher Grainge[9,10], Toby M Maher[11], Vidya Navaratnam[6,12], Richard Hubbard[6], Peter MA Hopkins[13,14], Paul N Reynolds[15,16], Sally Chapman[16], Christopher Zappala[13], Gregory J Keir[13], Wendy A Cooper[4,5,17], Annabelle M Mahar[5], Samantha Ellis[8], Nicole Goh[18,19], Emma De Jong[2,3], Lilian Cha[2,3], Dino BA Tan[2,3], Lucy Leigh[9,20], Christopher Oldmeadow[9,20], E Haydn Walters[1,8,21,22,23], R Gisli Jenkins[6], Yuben Moodley[1,2,3,24].

Corresponding author:

Yuben Moodley

Institute for Respiratory Health Inc,

E Block, Sir Charles Gairdner Hospital

Nedlands, Western Australia 6009 AUSTRALIA

Email: yuben.moodley@uwa.edu.au

Phone: +61 414 383 338

<u>Table S1 – Biomarkers selected for ELISA and Proteomic analysis in the AIPFR cohort.</u> Each ELISA biomarker is grouped according to its potential role in IPF pathogenesis, and the proteomic analytes are listed.

Biomarkers	Abbreviations
ELISA	
Inflammation	
C-reactive protein	CRP
Intercellular adhesion molecule-1	ICAM1
Vascular cell adhesion molecule-1	VCAM1
Extracellular newly identified receptor for advanced glycation end-products binding protein	ENRAGE
Fibrosis	
Osteopontin	OPN
Periostin	POSTN
Matrix metallopeptidase 7	MMP7
Fibulin-1	FBLN1
Chemokines	
C-X-C motif chemokine ligand 13	CXCL13
C-C motif chemokine ligand 18	CCL18
Lung physiology	
Surfactant protein-A	SPA
Surfactant protein-D	SPD
PROTEOMICS	
Lumican	LUM
N-acetylmuramoyl-L-alanine amidase	PGRP2
Beta-2-glycoprotein 1	B2GPI
Complement component C9	CO9

Fibronectin	FN
Inter-alpha-trypsin inhibitor heavy chain H1	ITIH1
Leucine-rich alpha-2-glycoprotein	A2GL
Apolipoprotein E	APOE
Gelsolin	GELS
Pigment epithelium-derived factor	PEDF
Alpha-1-antichymotrypsin	AACT
Apolipoprotein C-I	APOCI
Apolipoprotein C-II	APOCII
Apolipoprotein C-III	APOCIII
Serum amyloid A-4 protein	SAA4

Figure S1 – A comparison of unadjusted estimates of progression index with disease progression. Adjusted estimates as reported (FVC, age, gender, BMI) and adjusted sensitivity estimates for DLCO (DLCO, age, gender, BMI), demonstrating comparable findings across TLF, AIPFR and pooled in multilevel analysis.
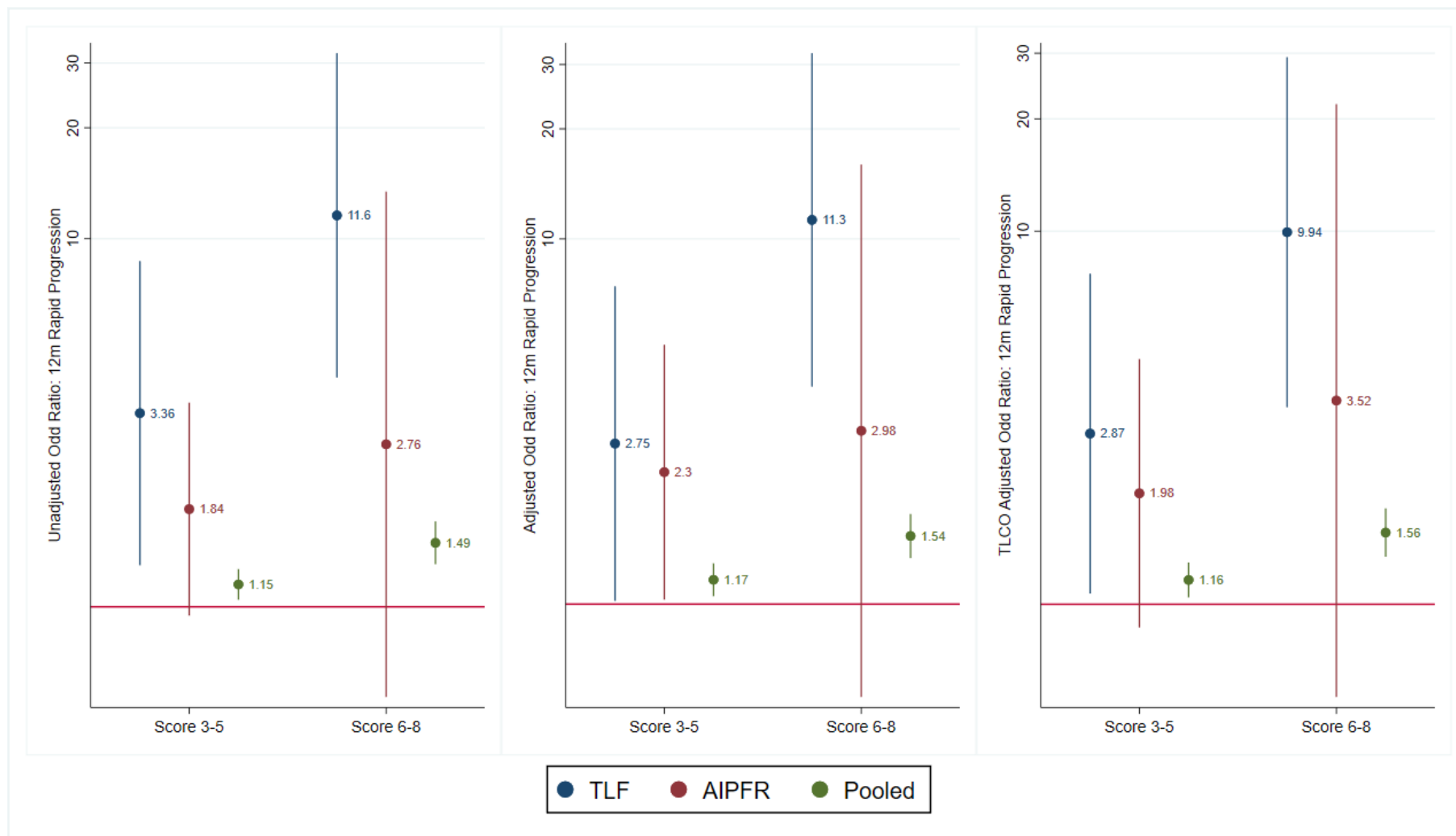
Figure S2 – CONSORT diagram.

Prospective Assessment | Retrospective Assessment

Enrolment / Assessed for eligibility

AIPFR *n* = 217 → Excluded – no serum available; only one lung function test datapoint *n* = 28

Had biomarker data with at least two lung function tests available *n* = 189

TLF *n* = 211 → Excluded – no serum available; only one lung function test datapoint *n* = 6

Had biomarker data with at least two lung function tests available *n* = 205

PROFILE *n* = 511 → Excluded – no specific biomarker data available ; only one lung function test datapoint *n* = 389

Had MMP7, ICAM1 and OPN data with at least two lung function tests available *n* = 122

Analysis

Identification of Candidate Prognostic Biomarkers: Multivariable adjustments for demographic and physiological parameters: age, gender, smoking status, and baseline FVC.

Development of a Progression Index. All progression index models were adjusted for age, gender, baseline FVC, and BMI.

Validation of a Progression Index. All progression index models were adjusted for age, gender, baseline FVC, and BMI.
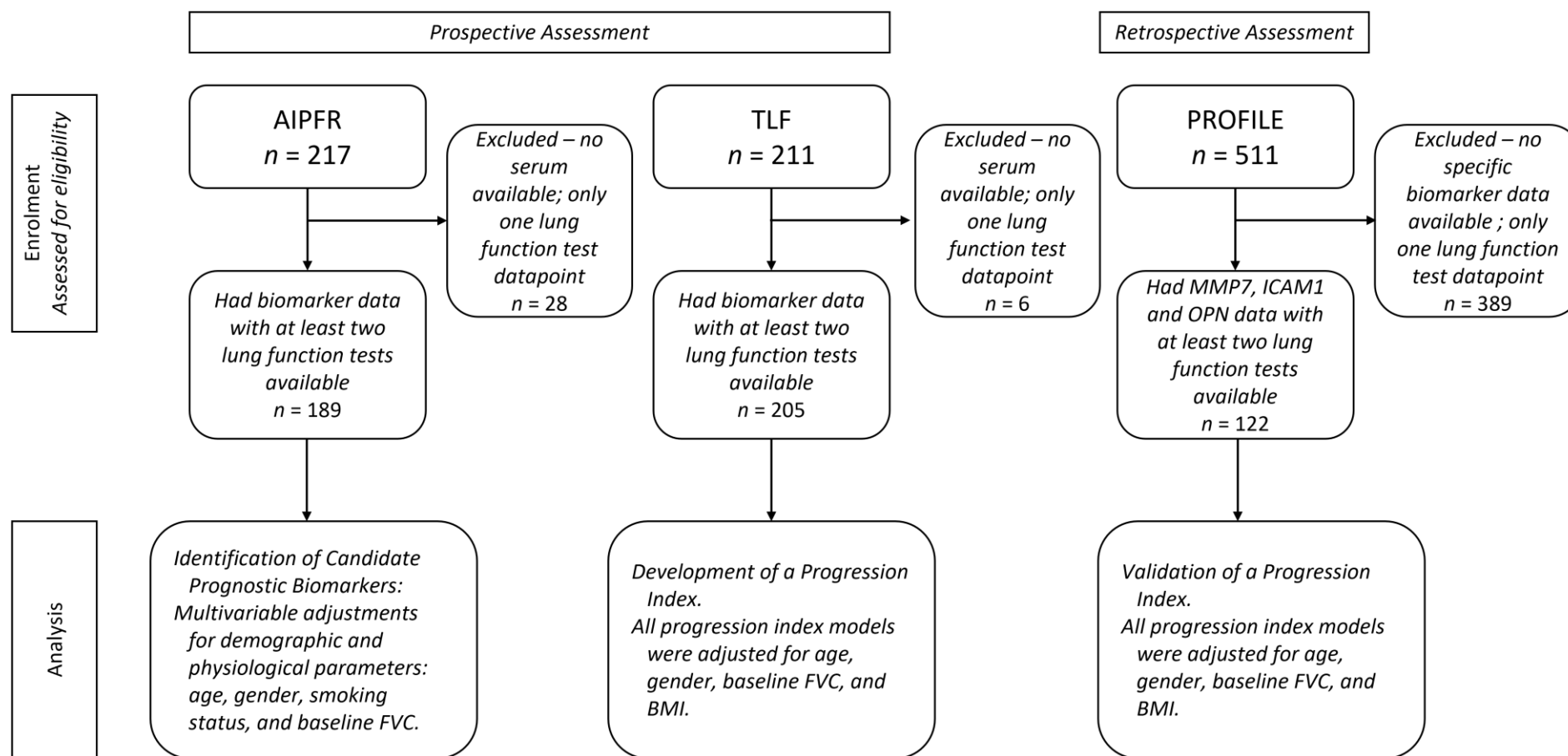
Table S2 – Progression at 12 months outcome predicted by the LASSO model for biomarkers in AIPFR. All coefficients are on the log-odds scale. Biomarkers that were shrunk to zero ≤ 5% of the repeated model are indicated in bold. Area under the curve (AUC) (95%CI) from empirical ROC curves for ELISA and proteomics biomarkers >0.6 were more predictive of disease progression.

Biomarker	n shrunk to 0	Mean (SD) of non-zero coefficients	AUC (95% CI)
ELISA			
SPA	76	0·00 (0·00)	0·49 (0·40, 0·59)
FBLN1	58	-0·00 (0·00)	0·53 (0·43, 0·63)
VCAM1	**3**	**-0·00 (0·00)**	**0·55 (0·45, 0·65)**
CRP	43	0·00 (0·00)	0·57 (0·47, 0·67)
SPD	**1**	**0·01 (0·00)**	**0·60 (0·51, 0·70)**
ENRAGE	**1**	**0·00 (0·00)**	**0·60 (0·50, 0·70)**
ICAM1	**2**	**0·00 (0·00)**	**0·62 (0·54, 0·71)**
MMP7	76	-0·00 (0·00)	0·62 (0·53, 0·70)
OPN	**0**	**-1·22 (0·20)**	**0·63 (0·53, 0·73)**
CXCL13	31	0·00 (0·00)	0·65 (0·56, 0·74)
POSTN	**0**	**0·01 (0·00)**	**0·65 (0·56, 0·74)**
PROTEOMICS			
FN	100	0·00 (0·00)	0·47 (0·36, 0·58)
LUM	47	0·05 (0·08)	0·51 (0·39, 0·62)
APOCI	100	0·00 (0·00)	0·52 (0·41, 0·63)
APOCII	100	0·00 (0·00)	0·53 (0·42, 0·64)
CO9	97	0·00 (0·02)	0·54 (0·44, 0·65)
AACT	20	0·27 (0·22)	0·55 (0·43, 0·66)
APOCIII	20	0·12 (0·11)	0·55 (0·44, 0·66)

SAA4	**4**	**-1·02 (0·41)**	**0·55 (0·44, 0·66)**
APOE	**1**	**1·62 (0·45)**	**0·55 (0·44, 0·67)**
B2GPI	100	0·00 (0·00)	0·59 (0·48, 0·69)
ITIH1	5	-0·49 (0·24)	0·62 (0·51, 0·73)
A2GL	**0**	**0·71 (0·10)**	**0·62 (0·52, 0·72)**
PEDF	**4**	**-1·12 (0·45)**	**0·62 (0·52, 0·72)**
PGRP2	97	0·02 (0·13)	0·63 (0·52, 0·73)
GELS	**0**	**-2·44 (0·21)**	**0·66 (0·55, 0·76)**

Abbreviations: SD, standard deviation; AUC, area under the curve; CI, confidence interval.

Table S3 – Univariable analyses of the ELISA and Proteomics biomarkers associated with PFS and mortality in the AIPFR cohort. All outcomes were regressed on each ELISA biomarker in separate univariable Cox proportional hazards models (using SAS PHREG). P-values which remain significant after adjustment have been indicated in bold.

Biomarker	Progression-free Survival			Mortality		
	n	Hazard Ratio (95% CI)	p-value	n	Hazard Ratio (95% CI)	p-value
ELISA						
CCL18	191	1·003 (0·994,1·013)	0·5025	191	1·002 (0·991,1·014)	0·7117
CRP	190	1·009 (0·987,1·032)	0·4340	190	1·013 (0·989,1·038)	0·3013
CXCL13	191	1·000 (0·999,1·001)	0·9200	191	1·000 (0·999,1·001)	0·7853
ENRAGE	**189**	**1·002 (1·000,1·004)**	**0·0125**	**189**	**1·003 (1·001,1·005)**	**0·0058**
FBLN1	190	1·001 (0·997,1·004)	0·7326	190	0·998 (0·993,1·003)	0·3602
ICAM1	**191**	**1·002 (1·000,1·004)**	**0·0125**	191	1·001 (0·999,1·003)	0·4478
MMP7	191	1·000 (1·000,1·000)	0·1727	191	1·000 (1·000,1·000)	0·3914
MUC1	49	1·006 (0·959,1·055)	0·8116	49	1·009 (0·960,1·060)	0·7191
OPN	**191**	**1·021 (1·012,1·030)**	**<0·0001**	**191**	**1·024 (1·014,1·034)**	**<0·0001**
POSTN	**189**	**1·012 (1·004,1·020)**	**0·0046**	**189**	**1·011 (1·000,1·021)**	**0·0450**
SPA	190	1·000 (0·996,1·003)	0·8965	190	1·001 (0·997,1·005)	0·6821
SPD	**184**	**1·026 (1·012,1·039)**	**0·0001**	184	1·016 (0·999,1·033)	0·0639
VCAM1	190	1·000 (1·000,1·000)	0·6137	190	1·000 (1·000,1·000)	0·8697
PROTEOMICS						
LUM	157	1·354 (0·888,2·064)	0·1594	157	1·150 (0·659,2·004)	0·6233
PGRP2	157	0·918 (0·258,3·263)	0·8945	157	0·353 (0·055,2·256)	0·2710
B2GPI	157	1·041 (0·856,1·267)	0·6855	157	0·805 (0·528,1·226)	0·3123
CO9	**157**	**1·349 (1·010,1·802)**	**0·0424**	157	1·215 (0·823,1·793)	0·3281

FN	157	0·599 (0·153,2·348)	0·4623	157	0·819 (0·184,3·656)	0·7940
ITIH1	157	1·080 (0·810,1·440)	0·6006	157	0·610 (0·283,1·315)	0·2070
A2GL	**157**	**1·215 (1·023,1·443)**	**0·0266**	157	1·181 (0·965,1·446)	0·1074
APOE	**157**	**1·421 (1·022,1·975)**	**0·0366**	157	1·264 (0·813,1·965)	0·2977
GELS	157	0·714 (0·248,2·057)	0·5330	157	0·216 (0·048,0·977)	0·0465
PEDF	157	0·938 (0·405,2·172)	0·8814	157	0·486 (0·141,1·681)	0·2547
AACT	**157**	**1·611 (1·213,2·140)**	**0·0010**	**157**	**1·539 (1·093,2·167)**	**0·0135**
APOCI	137	1·600 (0·553,4·631)	0·3856	137	1·545 (0·478,4·996)	0·4671
APOCII	151	1·377 (0·442,4·289)	0·5812	151	1·702 (0·519,5·578)	0·3797
APOCIII	144	1·360 (0·742,2·494)	0·3201	144	1·580 (0·863,2·893)	0·1384
SAA4	154	1·094 (0·553,2·163)	0·7968	154	0·853 (0·356,2·045)	0·7217

Abbreviations: CI, confidence interval.

Table S4 – Area under the curve (95%CI) of continuous and tertiles from empirical ROC curves for ELISA and Proteomics biomarkers predicting disease progression in TLF.

Biomarker	AUROC	95%CI	p-value*	Tert 1	Tert 2	Tert 3	AUROC	95%CI	p-value*
ELISA									
SPD	0·56	(0·47; 0·64)	0·7822	19·70	38·06	78·82	0·55	(0·47; 0·63)	0·7173
OPN	0·61	(0·54 ;0·69)	0·4616	16·23	24·96	36·43	0·61	(0·54; 0·69)	0·4652
MMP7	0·64	(0·56; 0·71)	0·3097	1·87	2·98	4·43	0·63	(0·56; 0·70)	0·3589
ICAM1	0·67	(0·60; 0·75)	0·0852	218·09	347·58	619·40	0·64	(0·60; 0·72)	0·2042
POSTN	0·67	(0·59; 0·75)	0·0865	27·76	44·62	80·97	0·66	(0·59; 0·74)	0·0948
PROTEOMICS									
PGRP2	0·55	(0·46; 0·63)	0·6243	1·83	2·65	3·94	0·54	(0·46; 0·62)	0·5575
CO9	0·52	(0·44; 0·60)	0·3615	5·61	8·14	12·11	0·51	(0·43; 0·59)	0·2629
A2GL	0·52	(0·43; 0·60)	0·3061	3·83	6·03	9·76	0·52	(0·44; 0·61)	0·3671
APOE	0·49	(0·41; 0·58)	0·2003	1·01	1·62	2·46	0·51	(0·43; 0·59)	0·3073
GELS	0·56	(0·48; 0·64)	0·8110	1·01	1·51	2·41	0·55	(0·47; 0·63)	0·6646
PEDF	0·54	(0·46; 0·63)	0·6140	1·52	2·29	3·04	0·55	(0·47; 0·62)	0·6571
AACT	0·55	(0·46; 0·63)	0·6554	4·46	6·95	11·75	0·55	(0·47; 0·63)	0·7406
SAA4	0·55	(0·47; 0·64)	0·7697	0·40	0·65	1·04	0·56	(0·48; 0·64)	0·8533

*Median tertile value: ELISA as ng/ml, Proteomics as 180 ratio. *p-value for difference from GAP AUROC.*

Table S5 – Odds ratio of tertiles for disease progression in TLF. Tertile 3 versus Tertile 1. Significant values are indicated in bold. * includes odds ratios adjusted for age, gender, ppFVC and BMI

Biomarker	OR	95%CI	p-value	OR*	95%CI	p-value
ELISA						
SPD	1·63	(0·79; 3·35)	0·185	1.11	(0.65;2.98)	0.399
OPN	3·13	(1·42; 6·90)	**0·005**	3.40	(1.41;8.24)	**0.007**
MMP7	3·89	(1·69; 8·91)	**0·001**	4.05	(1.70;9.61)	**0.002**
ICAM1	4·16	(1·86; 9·31)	**0·001**	5.33	(2.21;12.84)	**<0.001**
POSTN	4·57	(2·12; 9·87)	**<0·001**	4.04	(1.73; 9.44)	**0.001**
PROTEOMICS						
PGRP2	1·41	(0·70; 2·88)	0·338	0.87	(0.39;1.95)	0.734
CO9	1·09	(0·53; 2·28)	0·809	1.22	(0.57;2.63)	0.604
A2GL	1·24	(0·61; 2·50)	0·551	1.06	(0.50;2.23)	0.883
APOE	1·09	(0·53; 2·25)	0·546	1.13	(0.53;2.42)	0.754
GELS	1·53	(0·74; 3·16)	0·250	1.55	(0.71;3.35)	0.268
PEDF	1·56	(0·74; 3·29)	0·240	1.68	(0.75;3.80)	0.209
AACT	1·63	(0·79; 3·35)	0·185	1.70	(0.79;3.69)	0.177
SAA4	1·69	(0·84; 3·39)	0·142	1.84	(0.87;3.89)	0.109

*adjusted for age, gender, ppFVC, BMI

<u>Table S6 – Biomarker differences between progressor and stable groups in PROFILE.</u>
Only significant biomarkers identified in TLF have been displayed. Absolute values, median
(IQR), non-parametric test· Significant values are indicated in bold.

Biomarker	Unit	Stable median (IQR)	Progressive median (IQR)	p-value
PROFILE Cohort				
OPN	ng/ml	110 (63·5 – 150·0)	120 (79·0 – 190·0)	0·1204
MMP7	ng/ml	0·91 (0·70 – 1·40)	1·1 (0·85 – 1·50)	**0·0331**
ICAM1	ng/ml	208 (159·0 – 276·5)	241 (195·0 – 314·0)	**0·0352**
POSTN	ng/ml	NO OBSERVATIONS		

Note: proteomic unit is ratio of corrected unlabelled peak area/180 peak area.
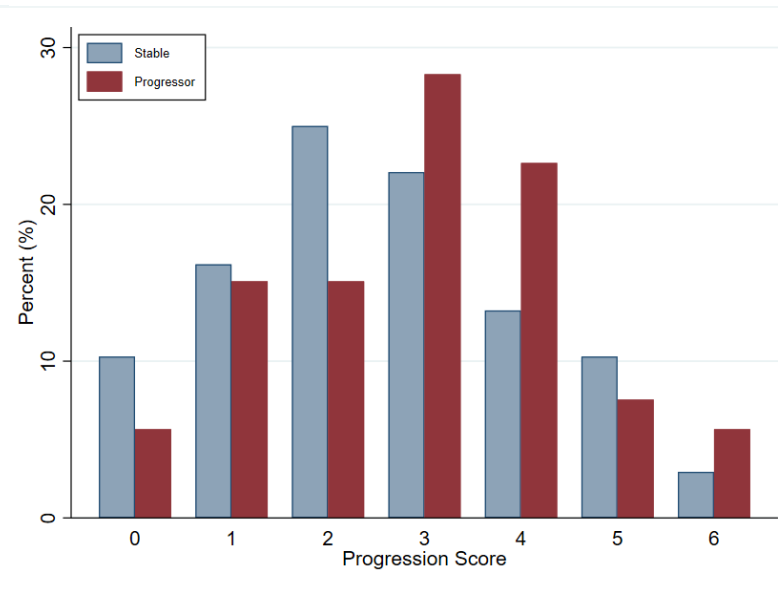*Abbreviations: IQR, interquartile range.*

Table S7 – Progression Index values percentage distribution across all cohorts. Index created with four biomarkers based on tertiles: score 0,1,2 depending on T1, T2, T3 for each of four biomarkers (min 0, max 8). Number of people with positive biomarker per index value of 0-8.

Index	TLF			AIPFR			PROFILE		
	All n=205 (%)	**Stable n=138 (%)**	**Progressor n=67 (%)**	**All n=189 (%)**	**Stable n=136 (%)**	**Progressor n=53 (%)**	**All n=205 (%)**	**Stable n=138 (%)**	**Progressor n=67 (%)**
0	12 (5·9)	11 (8·0)	1 (1·5)	17 (9)	14 (10·3)	3 (5·7)	4 (3·3)	3 (4·4)	1 (1·9)
1	13 (6·3)	12 (8·7)	1 (1·5)	30 (15·9)	22 (16·2)	8 (15·1)	0	0	0
2	28 (13·7)	24 (17·4)	4 (6·0)	42 (22·2)	34 (25)	8 (15·1)	85 (69·7)	53 (77·9)	32 (59·3)
3	35 (17·1)	28 (20·3)	7 (10·5)	45 (23·8)	30 (22·1)	15 (28·3)	25 (20·5)	10 (14·7)	15 (27·8)
4	34 (16·6)	23 (16·7)	11 (16·4)	30 (15·9)	18 (13·2)	12 (22·6)	7 (5·7)	1 (1·5)	6 (11·1)
5	31 (15·1)	19 (13·8)	12 (17·9)	18 (9·5)	14 (10·3)	4 (7·5)	1 (0·8)	1 (1·5)	0
6	21 (10·2)	7 (5·1)	14 (20·9)	7 (3·7)	4 (2·9)	3 (5·7)	0	0	0
7	23 (11·2)	10 (7·3)	13 (19·4)	0	0	0	0	0	0
8	8 (3·9)	4 (2·9)	4 (6·0)	0	0	0	0	0	0

TLF Cohort

AIPFR Cohort
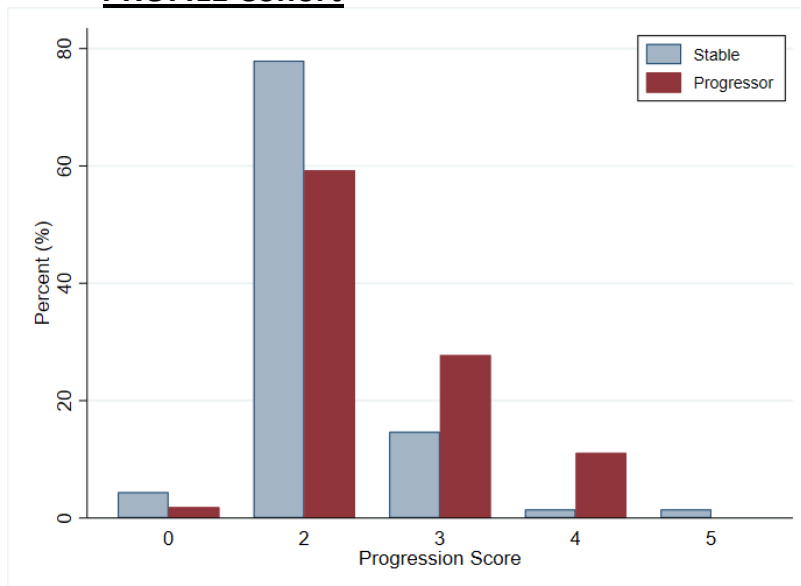
PROFILE Cohort

Figure S3 – Progression Score value percentage distribution across all 3 cohorts. Score created with four biomarkers (MMP7, ICAM1, OPN and POSTN) based on tertiles: score 0,1,2 depending on T1, T2, T3 for each of four biomarkers (min 0, max 8). Percentage of participants with progression scores between 0-8 have been totalled across all cohorts.

Table S8 – Adjusted associations with outcome in PROFILE. Prediction of progression at 12-months and mortality. Significant p-values are indicated in bold. Note: participants missing biomarker value for POSTN and no observations for PS 6-8.

Score	n	OR*	95%CI	p-value	HR mort*	95%CI	p-value
PROFILE							
0-2	89	1			1		
3-5	33	2·06	(0·59; 7·24)	0·258	2·23	(1·02; 4·85)	**0·043**
6-8	-	No observations			No observations		

Adjusted for Age, gender, baseline ppFVC, BMI. Modelled with robust standard errors.

Abbreviations: OR, odds ratio; CI, confidence interval; HR, hazard ratio; mort, mortality; PFS,

progression-free survival.