Title

Longitudinal change in collagen degradation biomarkers in idiopathic pulmonary fibrosis: an analysis from the prospective, multicentre PROFILE study

Authors

RG Jenkins[1], JK Simpson[2], G Saini[1], JH Bentley[2], AM Russell[3], R Braybrooke[1], PL Molyneaux[4], TM McKeever[1], AU Wells[3], A Flynn[2], RB Hubbard[1], DJ Leeming[5], RP Marshall[2], MA Karsdal[5], PT Lukey[2], TM Maher[3,4]

Supplementary Table and Figure Legends

Tables

Marker	Meaning	Unit	LLOQ	ULOQ	Intra/Inter assay variability
C1M	Type I collagen degraded by MMP-2/9/13	ng/ml	21.2	421.0	≤ 13.7 % and ≤ 20.8 %
C3M	Type III collagen degraded by MMP-9	ng/ml	6.4	217.5	≤ 13.5 % and ≤14.5 %
C3A	Type III collagen degraded by ADAMTS-1/4/8	ng/ml	10.8	450.0	≤ 2.0 % and ≤ 2.5 %
C5M	Type V collagen degraded by MMP-2/9	ng/ml	17.3	2400	≤ 6.5 % and ≤ 7.5%
C6M	Type VI collagen degraded by MMP-2/9	ng/ml	5.2	267.6	≤ 6.2 % and ≤ 11.6%
ELM	Elastin degraded by MMP-12	ng/ml	4.5	391.1	≤ 8.2 % and ≤12.7 %
ELM2	Elastin degraded by MMP-9/12	ng/ml	5.4	207.4	≤ 15.5 % and ≤ 14.5 %
VICM	Citrullinated vimentin degraded by MMP-2/8	ng/ml	0.8	85.2	≤ 7.2 % and ≤15 %
BGM	Biglycan degraded by MMP-2/9	ng/ml	2.4	400	≤20.4 % and ≤ 20.9 %
CRPM	MMP degraded CRP-1/8	ng/ml	3.2	110.0	≤ 11.1 % and ≤20.8 %
P3NP	Type III collagen formation	ng/ml	4.0	120.2	≤ 8.2 % and ≤17.8 %

LLOQ- Lower Limit of Quantification, ULOQ- Upper Limit of Quantification

Supplementary Table 1. Assay performance characteristics and the upper and lower limits of quantification for each neoepitope.

	Brompton (n=91)	Nottingham (n=98)	All subjects (n=189)
Age (years)	67.7 (8.60)	72.2 (7.38)	70.1 (8.28)
Male sex	75 (82.4%)	74 (75.5%)	149 (78.8%)
Ethnicity			
European	80 (87.9%)	98 (100%)	178 (94.2%)
Asian	10 (11.0%)	0 (0%)	10 (5.3%)
Other	1 (1.1%)	0 (0%)	1 (0.5)
Ever smokers	63 (69.2%)	72 (73.5%) n=97	134 (70.9%) n=188
BMI (Kg/m^2)	28.1 (4.27)	28.6 (4.50)	28.4 (4.39)
Baseline lung function			
FVC %predicted	71.1 (19.4) n=87	83.1 (17.2) n=97	77.5 (19.2) n=184
FVC 60≤90 % predicted	47 (51.6)	62 (63.3)	109 (57.7)
DLco % predicted	40.5 (14.3) n=86	45 (15.5) n=90	42.8 (15.1) n=176
DLco 60≤90 % predicted	29 (31.9)	35 (35.7)	64 (33.9)
CPI	51.9 (12.9) n=86	46.8 (13.0) n=89	49.3 (13.2) n=175

Data are Mean (SD) or number (%) unless otherwise stated. BMI – Body mass index. FVC – forced vital capacity, DLco – diffusion capacity for carbon monoxide. CPI – Composite Physiological Index.

Supplementary Table 2. Comparison of baseline clinical characteristics between PROFILE subjects by recruitment centre (Brompton or Nottingham).

Variable	Statistic/Level	Discovery (N=20)	Validation (N=50)
Age	N	20	50
Age	Mean (SD)	56.6 (4.57)	63.6 (8.03)
Age	Median (Q1-Q3)	56.0 (52.5-61.0)	64.5 (57.0-68.0)
Age	Min - Max	50.0-64.0	50.0-82.0
Gender	Female (%)	4 (20.0)	10 (20.0)
Gender	Male (%)	16 (80.0)	40 (80.0)

Supplementary Table 3. Summary Statistics for age and gender within the control groups in the Discovery and Validation cohorts.

Model	Variable	Baseline/Gradient	Hazard Ratio (95% CI)	P value
1	**DLco**	BL	0.19 (0.09 ,0.40)	<.0001
	FVC	BL	0.76 (0.23 ,2.51)	0.66
2	**DLco**	BL	0.18 (0.09 ,0.36)	<.0001
	BGM	GR	1.07 (1.00 ,1.15)	0.048
3	**DLco**	BL	0.17 (0.09 ,0.34)	<.0001
	C1M	GR	1.01 (1.00 ,1.02)	0.012
4	**DLco**	BL	0.16 (0.08 ,0.32)	<.0001
	C3A	GR	1.05 (1.01 ,1.10)	0.016
5	**DLco**	BL	0.16 (0.08 ,0.32)	<.0001
	C3M	GR	1.10 (1.04 ,1.17)	0.0013
6	**DLco**	BL	0.18 (0.09 ,0.35)	<.0001
	C5M	GR	1.00 (1.00 ,1.00)	0.0036
7	**DLco**	BL	0.17 (0.09 ,0.33)	<.0001
	C6M	GR	1.04 (1.01 ,1.08)	0.014
8	**DLco**	BL	0.18 (0.09 ,0.35)	<.0001
	CRPM	GR	1.33 (1.10 ,1.60)	0.0034
9	**DLco**	BL	0.17 (0.09 ,0.34)	<.0001
	VICM	GR	1.01 (0.99 ,1.03)	0.20

Supplementary Table 4. The effects of the magnitude of 3 month change in neoepitopes when combined with baseline DLco in multiple linear regression. Baseline (BL) variables and longitudinal gradient (GR) variables are on different scales: lung function measures were transformed using a base 2 logarithm and the hazard ratio (HR) for FVC and DLco relates to the change in hazard when these measurements increase two-fold. The neoepitope gradients are on the original linear scale and so the hazard ratio for these variables relates to the change in hazard for every unit increase in the slope. All analyses are adjusted for age, site and smoking status.

	Gradient to Month 1		Gradient to Month 3		Gradient to Month 6	
Marker	**HR (95% CI)**	**P value**	**HR (95% CI)**	**P value**	**HR (95% CI)**	**P value**
BGM	1.007 (0.978, 1.037)	0.65	1.084 (1.030, 1.141)	0.0019	1.142 (1.041, 1.252)	0.0050
C1M	1.001 (0.996, 1.005)	0.74	1.010 (1.003, 1.017)	0.0039	1.012 (1.006, 1.018)	0.0002
C3A	1.006 (0.982, 1.031)	0.62	1.038 (0.989, 1.090)	0.13	1.039 (0.983, 1.099)	0.18
C3M	1.020 (0.975, 1.067)	0.39	1.106 (1.045, 1.170)	0.0005	1.112 (1.054, 1.174)	0.0001
C5M	1.001 (1.000, 1.003)	0.078	1.003 (1.001, 1.005)	0.0011	1.003 (1.001, 1.005)	0.0016
C6M	1.011 (0.997, 1.025)	0.14	1.042 (1.007, 1.078)	0.017	1.049 (1.016, 1.082)	0.0028
CRPM	1.145 (1.032, 1.271)	0.01	1.379 (1.164, 1.634)	0.0002	1.395 (1.178, 1.651)	0.0001
VICM	1.014 (0.999, 1.030)	0.072	1.015 (0.998, 1.032)	0.080	1.016 (1.001, 1.032)	0.043

Supplementary Table 5. The effects of the change in neoepitopes compared with overall survival.

The association between the gradient to month 1, 3 and 6 and survival in the validation cohort are presented. The gradients are on the original linear scale and so the hazard ratio for these variables relates to the change in hazard for every unit increase in the slope.

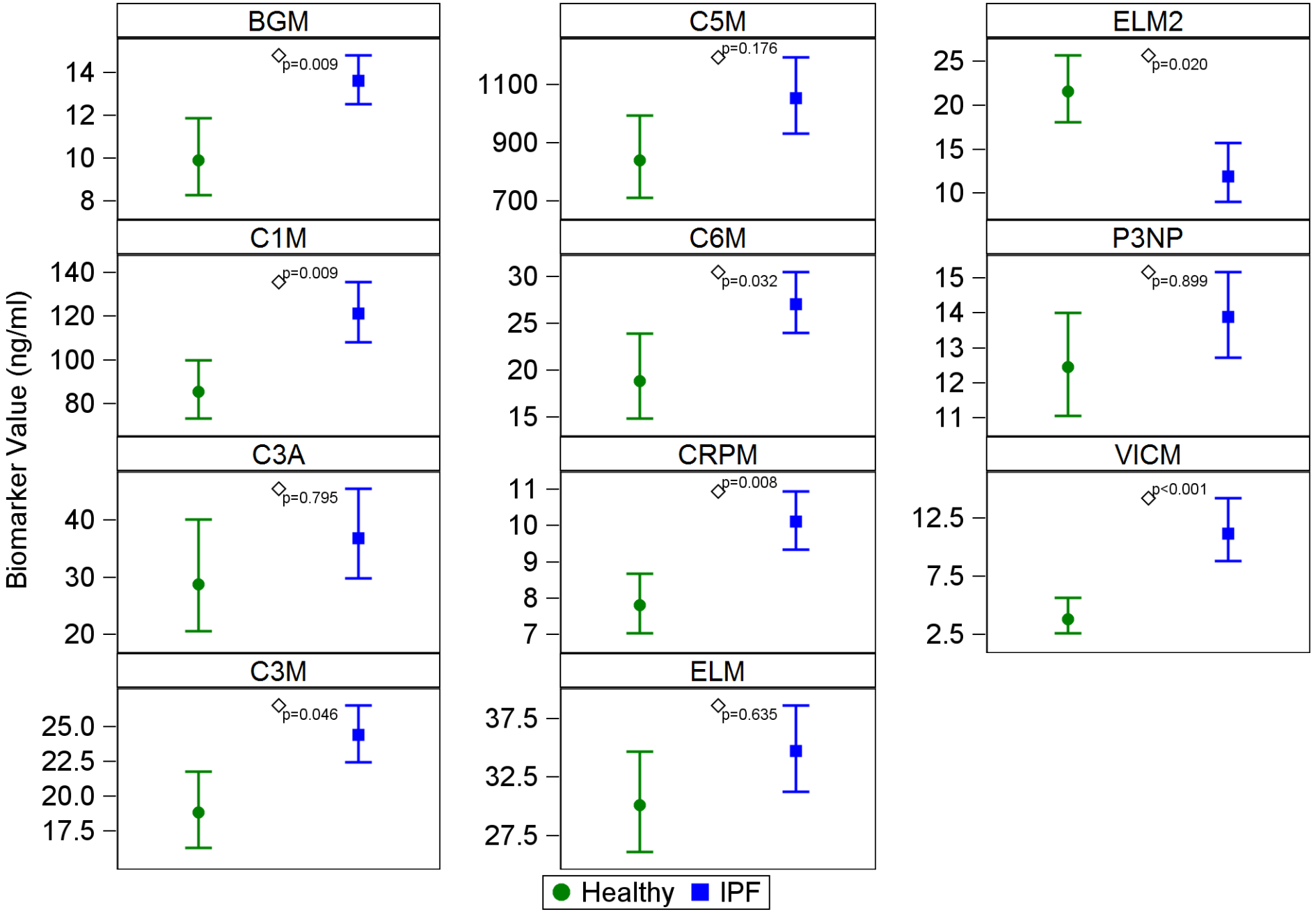
Figure Legends

Supplementary Figure 1: Baseline comparison of neoepitope levels in healthy controls and IPF subjects in the discovery cohort. Patients were recruited into the PROFILE study within six months of diagnosis (baseline) and serum samples stored for biomarker analysis. Biomarker data from the Discovery Cohort are presented as means (ng/ml) and 95% confidence intervals adjusted for age and gender. Numbers in each group are: 20 healthy subjects (gender matched only) represented by '●'; and 55 subjects with IPF represented by '■'. For ELM2 4% of values respectively were imputed. For all other epitopes the imputation rate was <1%. P values are provided where significant differences were observed between healthy subjects and IPF subjects.

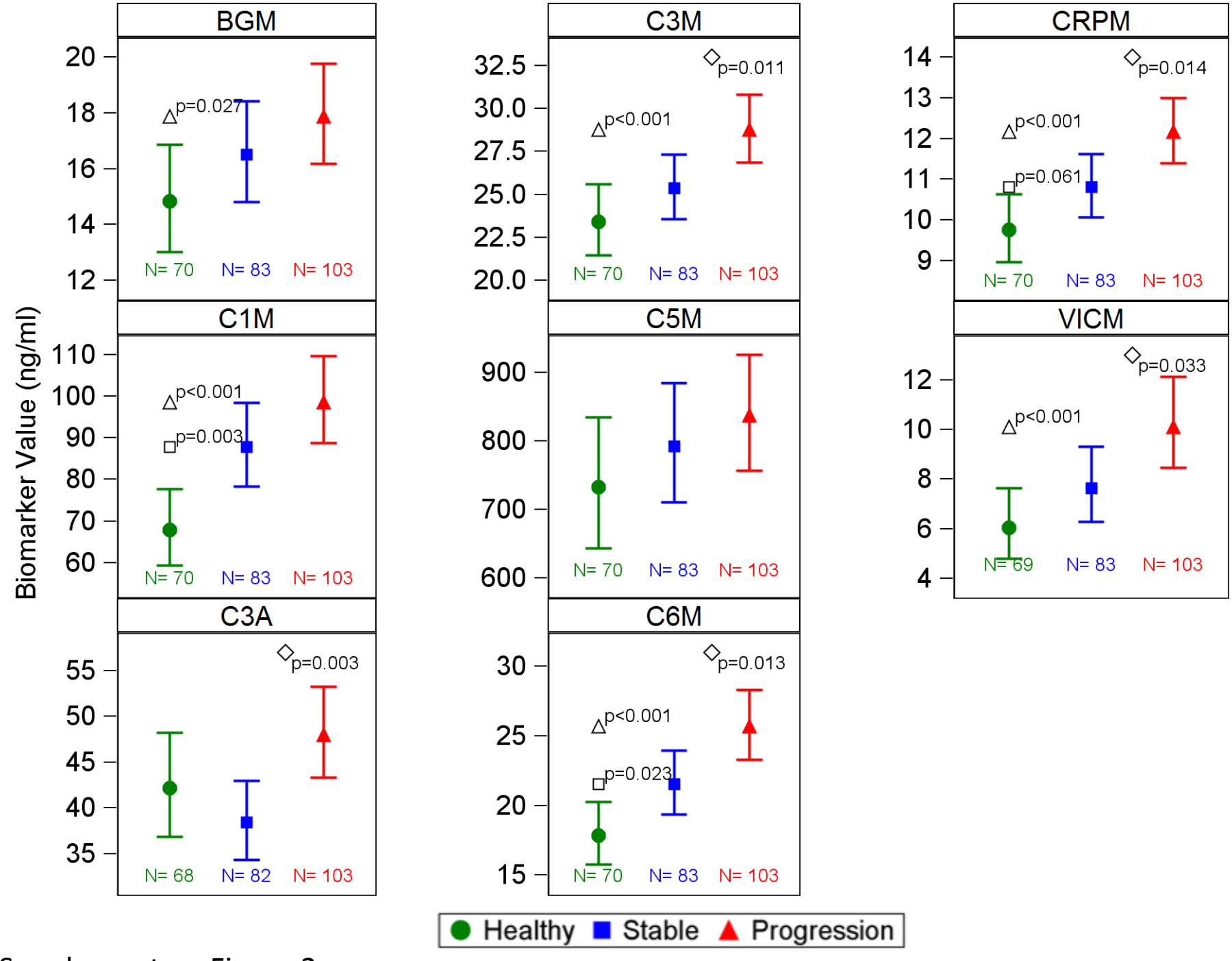
Supplementary Figure 2: Pooled Analysis from the discovery and validation cohort of baseline neoepitope levels in healthy controls and IPF subjects with stable and progressive disease. Combined numbers at baseline were 70 healthy subjects (●); 83 IPF subjects with stable disease (■) and 103 subjects with progressive IPF (▲). Three subjects for C3A and one for VICM were excluded due to missing baseline samples. Pooled biomarker data is presented as adjusted (least-squares) means (ng/ml) and 95% confidence intervals. These estimates were adjusted for age, cohort and age by cohort interaction. Disease progression was defined as all-cause mortality or ≥ 10% decline in FVC at 12 months. P values are presented where significant differences were observed between Healthy subjects and Stable IPF subjects (□); between healthy subjects and progressive IPF (Δ) and between stable IPF and progressive IPF (◊). P values are based on the pair-wise comparison between groups, adjusted for age, cohort and age by cohort interaction. The pooled analysis shows that there is a significant difference between levels of C1M and C6M in stable IPF patients and healthy controls. Levels of BGM, C1M, C3M, C6M, CRPM and VICM are significantly higher in progressive IPF patients compared to controls. Between progressive and stable IPF patients, levels of C3A, C3M, C6M, CRPM and VICM are significantly increased.

Supplementary Figure 3: Sensitivity analysis for the comparison of neoepitope levels in healthy controls and IPF subjects with stable and progressive disease at baseline and at 1, 3 and 6 months post baseline in the validation cohort. Disease progression was defined as all-cause mortality or ≥10% decline in FVC at 12 months. There were 16 cases without lung function data beyond baseline, so cases were adjudicated, following case note review, by the local principal investigator blinded to biomarker results. In the sensitivity analysis, these cases were removed leaving **5**0 IPF subjects with stable disease (■) and 65 subjects with progressive IPF (▲). Biomarker data represent means (ng/ml) and 95% confidence intervals, adjusted for age and gender. P values are provided where differences were observed between stable and progressive disease at a particular time point (◊). This sensitivity analysis supports the original analysis (Figure 3) by showing that levels of BGM, C1M, C3A, C3M, C6M and CRPM increase over time in the subjects with progressive, but not stable, disease.

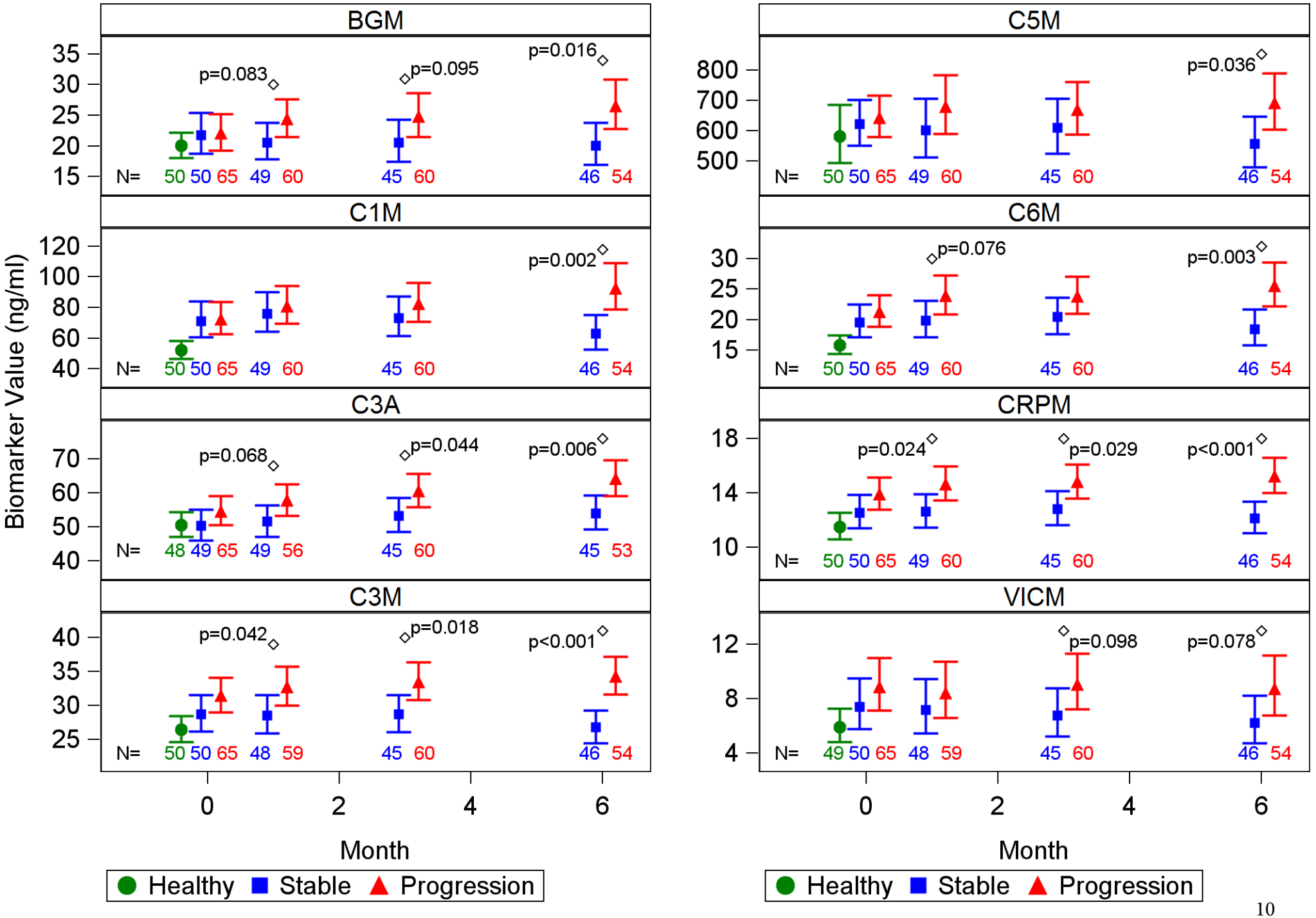
Supplementary Figure 4: Kaplan-Meier plot showing overall survival for the PROFILE IPF cohort. Curves are shown for the whole PROFILE IPF cohort of 189 subjects (——) and are broken down in to those for the discovery (·····) (n=55) and validation cohorts (·– – –) (n=134). There was no significant difference in survival between cohorts. Mortality data were available for all subjects and were collected from the NHS registry with a date of censoring of 1st October 2013.



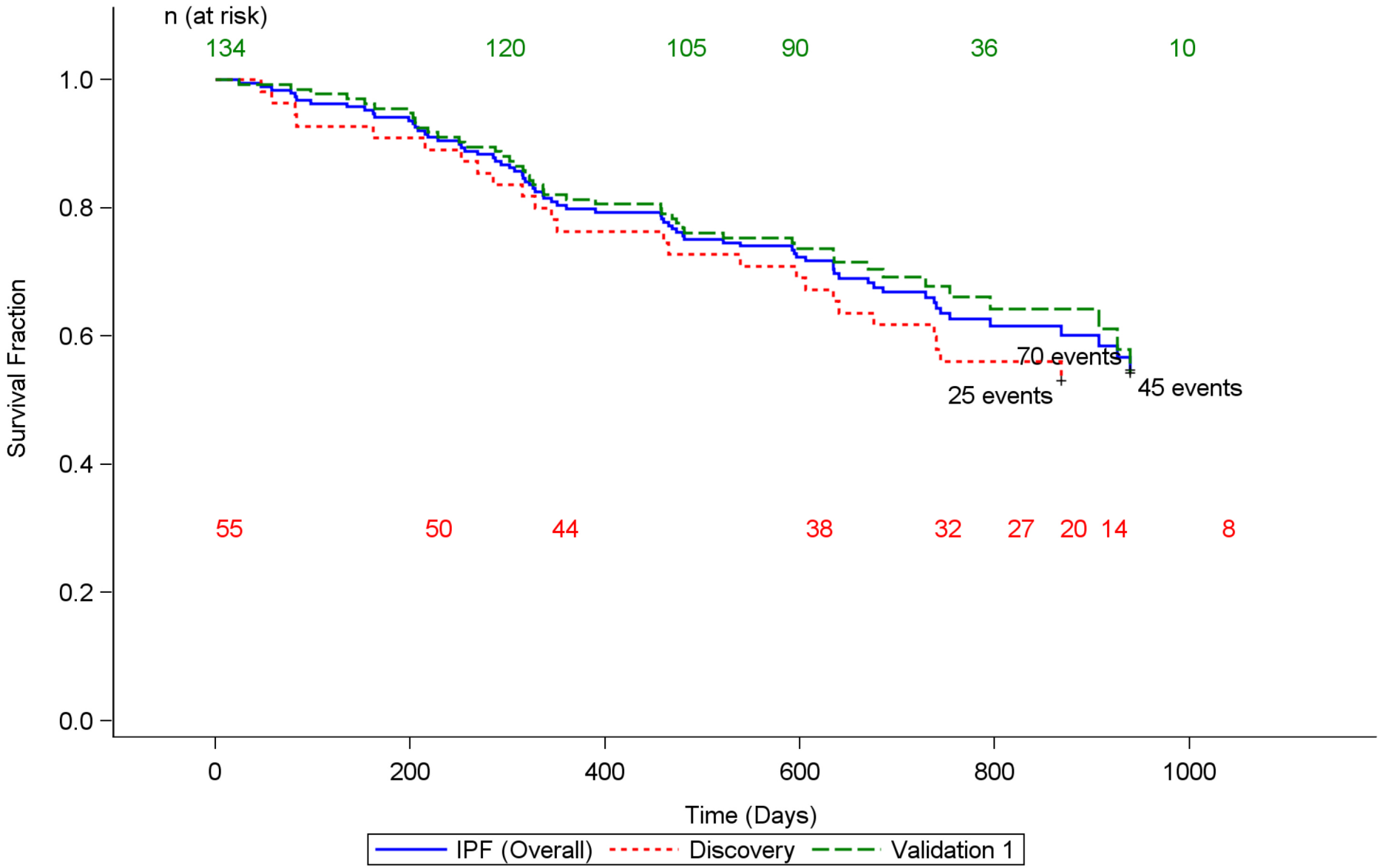
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3


n (at risk)
134
120
105
90
36
10
1.0
0.8
0.6
0.4
0.2
0.0
Survival Fraction
70 events
25 events
45 events
55
50
44
38
32
27
20
14
8
0
200
400
600
800
1000
Time (Days)
IPF (Overall)
Discovery
Validation 1