

## **Supplemental Material**

### **Identifying early pulmonary arterial hypertension biomarkers in systemic sclerosis: Machine learning on proteomics from the DETECT cohort**

Yasmina Bauer<sup>1,2</sup>, Simon de Bernard<sup>3</sup>, Peter Hickey<sup>4,5</sup>, Karri Ballard<sup>6</sup>, Jeremy Cruz<sup>6</sup>, Peter Cornelisse<sup>2</sup>, Harbajan Chadha-Boreham<sup>7</sup>, Oliver Distler<sup>8</sup>, Daniel Rosenberg<sup>7</sup>, Martin Doelberg<sup>7</sup>, Sebastien Roux<sup>2</sup>, Oliver Nayler<sup>2</sup> and Allan Lawrie<sup>4</sup>

<sup>1</sup>Galapagos GmbH, Aeschengraben 20, CH-4051 Basel, Switzerland, <sup>2</sup>Idorsia Pharmaceuticals Ltd, Hegenheimermattweg 91, CH-4123 Allschwil, Switzerland, <sup>3</sup>AltraBio, 30 rue Pré-Gaudry, F-69007 Lyon, France, <sup>4</sup>Department of Infection, Immunity & Cardiovascular Disease, University of Sheffield, Medical School, Beech Hill Road, Sheffield, S10 2RX, UK, <sup>5</sup> Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF, <sup>6</sup> Myriad RBM, 3300 Duval Rd., #110, Austin, TX 78759, USA, <sup>7</sup> Actelion Pharmaceuticals Ltd, Gewerbestrasse 16, CH-4123 Allschwil, Switzerland, <sup>8</sup> University Hospital Zurich, Department of Rheumatology, Gloriastrasse 25, CH-8091 Zurich, Switzerland

### **Corresponding Author**

Allan Lawrie

IICD, Medical School

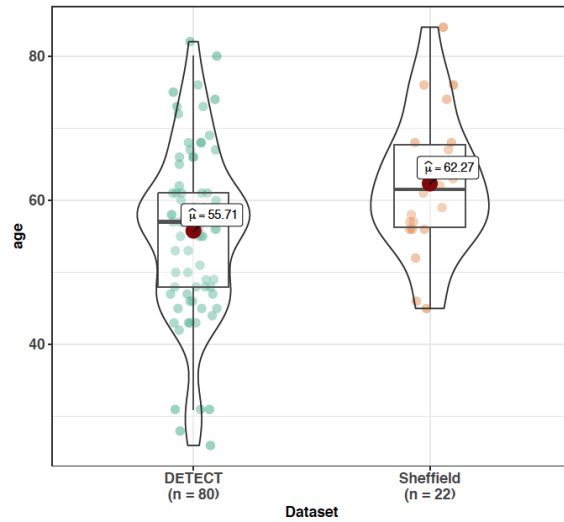
Beech Hill Road, Sheffield, S10 2RX

United Kingdom

[a.lawrie@sheffield.ac.uk](mailto:a.lawrie@sheffield.ac.uk) | Telephone: +44 (0) 114 215 9536

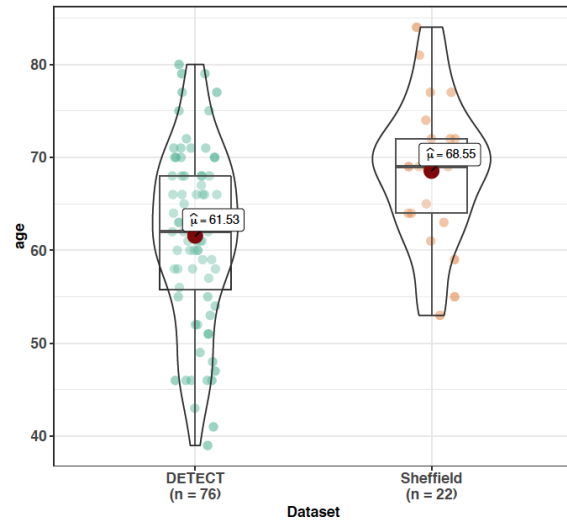
#### Status: non-PH

$\log_e(W_{Wilcoxon}) = 6.40$ ,  $p = 0.023$ ,  $\hat{\tau} = -0.23$ ,  $CI_{95\%} [-0.40, -0.06]$ ,  $n_{obs} = 102$



#### Status: PAH

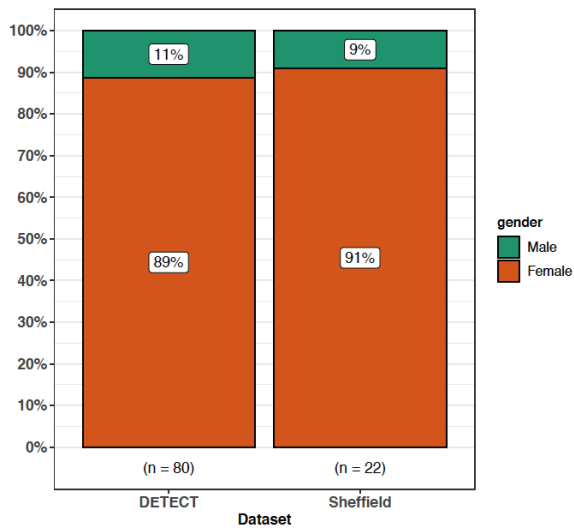
$\log_e(W_{Wilcoxon}) = 6.17$ ,  $p = 0.002$ ,  $\hat{\tau} = -0.31$ ,  $CI_{95\%} [-0.52, -0.12]$ ,  $n_{obs} = 98$



**Supplemental Figure S1:** Comparison of age across the DETECT Discovery and Sheffield Confirmatory Cohorts.

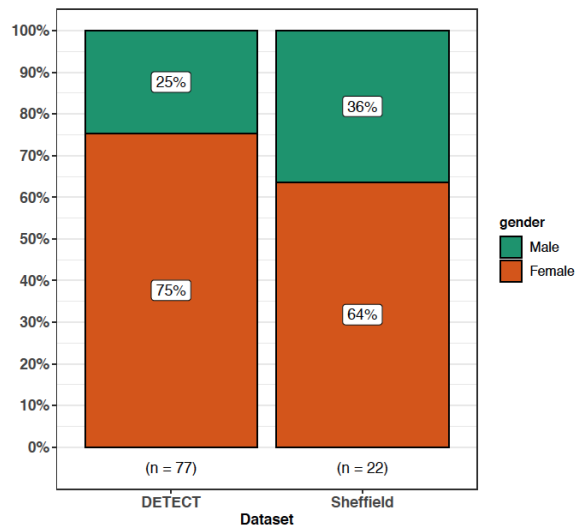
#### Status: non-PH

$\chi^2_{Pearson}(1) = 0.08$ ,  $p = 0.772$ ,  $\hat{V}_{Cramer} = 0.00$ ,  $CI_{95\%} [-0.15, 0.08]$ ,  $n_{obs} = 102$



#### Status: PAH

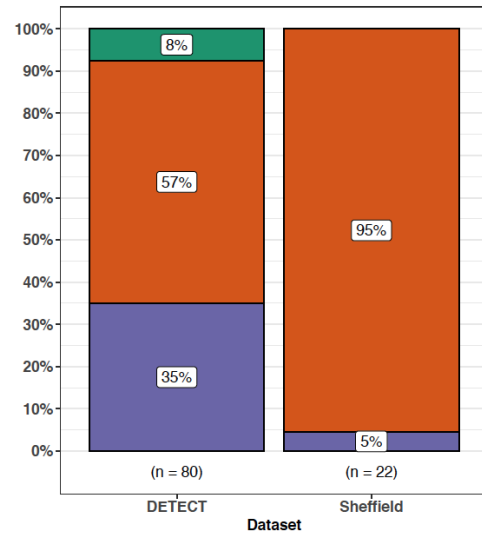
$\chi^2_{Pearson}(1) = 1.18$ ,  $p = 0.278$ ,  $\hat{V}_{Cramer} = 0.04$ ,  $CI_{95\%} [-0.18, 0.18]$ ,  $n_{obs} = 99$



**Supplemental Figure S2:** Comparison of sex across the DETECT Discovery and Sheffield Confirmatory Cohorts.

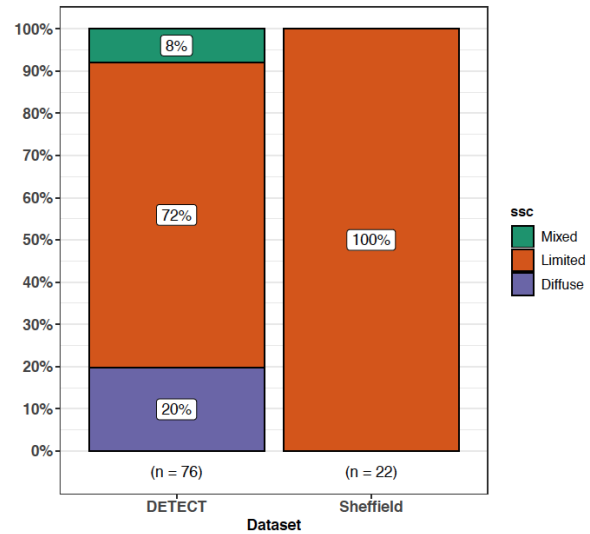
Status: non-PH

$\chi^2_{\text{Pearson}}(2) = 11.06, p = 0.004, \hat{V}_{\text{Cramer}} = 0.30, \text{CI}_{95\%} [0.16, 0.46], n_{\text{obs}} = 102$



Status: PAH

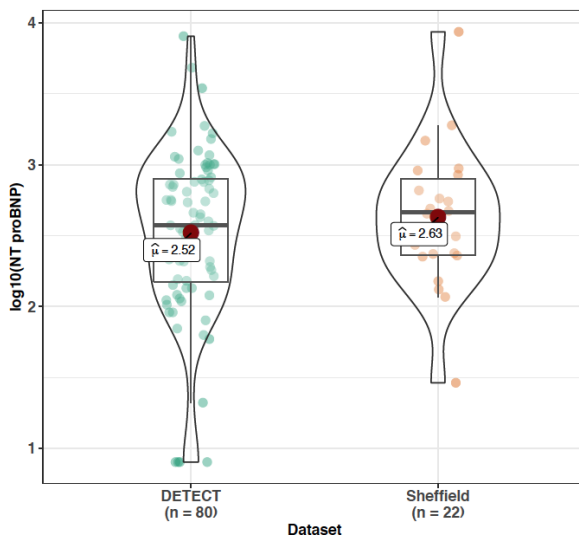
$\chi^2_{\text{Pearson}}(2) = 7.74, p = 0.021, \hat{V}_{\text{Cramer}} = 0.24, \text{CI}_{95\%} [0.14, 0.34], n_{\text{obs}} = 98$



**Supplemental Figure S3:** Comparison of systemic sclerosis type across the DETECT Discovery and Sheffield Confirmatory Cohorts.

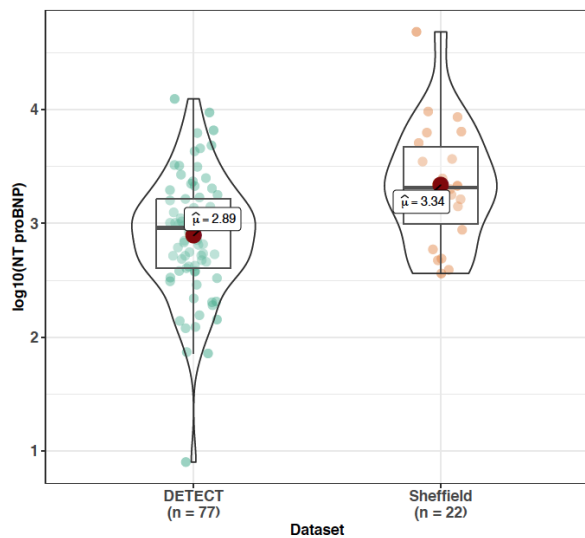
Status: non-PH

$\log_e(W_{\text{Wilcoxon}}) = 6.71, p = 0.625, \hat{\tau} = -0.05, \text{CI}_{95\%} [-0.21, 0.13], n_{\text{obs}} = 102$



Status: PAH

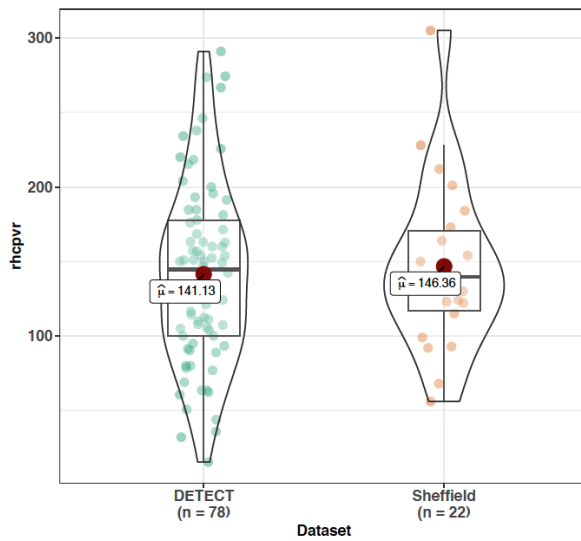
$\log_e(W_{\text{Wilcoxon}}) = 6.16, p = 0.002, \hat{\tau} = -0.32, \text{CI}_{95\%} [-0.52, -0.14], n_{\text{obs}} = 99$



**Supplemental Figure S4:** Comparison of NT-proBNP across the DETECT Discovery and Sheffield Confirmatory Cohorts.

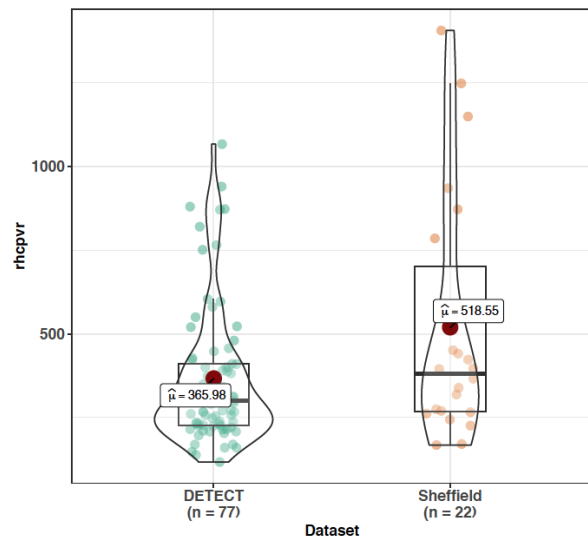
Status: non-PH

$\log_e(W_{\text{Wilcoxon}}) = 6.72$ ,  $p = 0.806$ ,  $\hat{\tau} = -0.02$ ,  $CI_{95\%} [-0.21, 0.16]$ ,  $n_{\text{obs}} = 100$



Status: PAH

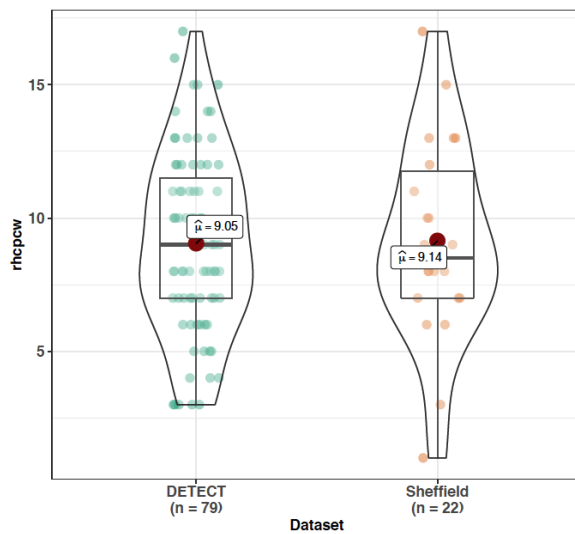
$\log_e(W_{\text{Wilcoxon}}) = 6.45$ ,  $p = 0.074$ ,  $\hat{\tau} = -0.18$ ,  $CI_{95\%} [-0.40, 0.02]$ ,  $n_{\text{obs}} = 99$



**Supplemental Figure S5:** Comparison of pulmonary vascular resistance across the DETECT Discovery and Sheffield Confirmatory Cohorts.

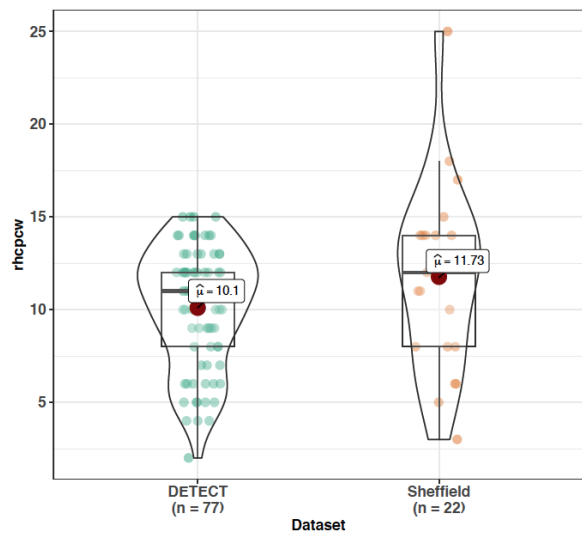
Status: non-PH

$\log_e(W_{\text{Wilcoxon}}) = 6.75$ ,  $p = 0.908$ ,  $\hat{\tau} = -0.01$ ,  $CI_{95\%} [-0.24, 0.19]$ ,  $n_{\text{obs}} = 101$



Status: PAH

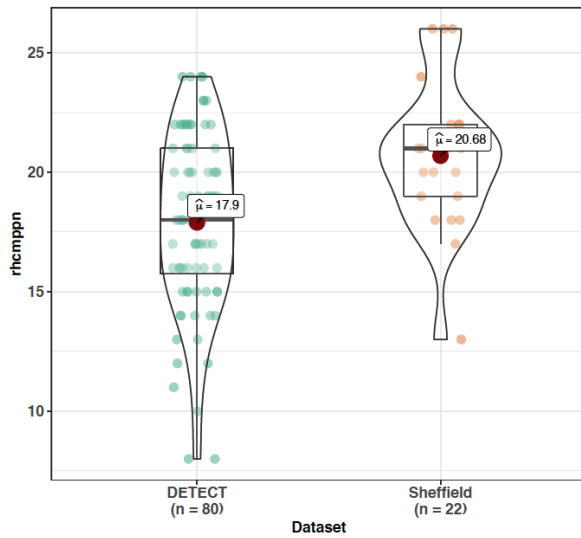
$\log_e(W_{\text{Wilcoxon}}) = 6.51$ ,  $p = 0.140$ ,  $\hat{\tau} = -0.15$ ,  $CI_{95\%} [-0.38, 0.09]$ ,  $n_{\text{obs}} = 99$



**Supplemental Figure S6:** Comparison of pulmonary arterial wedge pressure across the DETECT Discovery and Sheffield Confirmatory Cohorts.

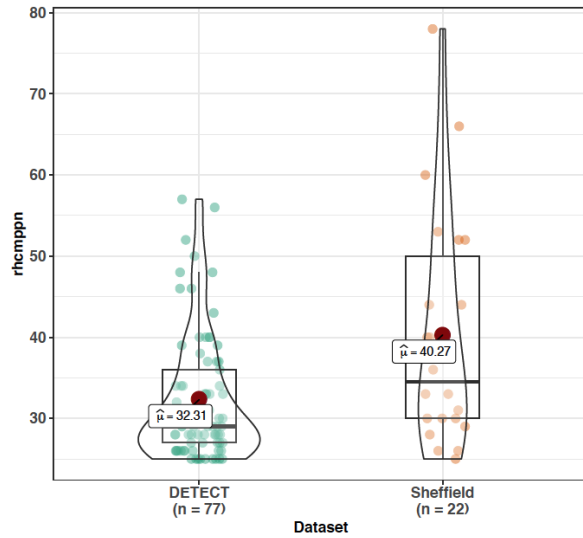
Status: non-PH

$\log_e(W_{\text{Wilcoxon}}) = 6.24$ ,  $p = 0.003$ ,  $\hat{\tau} = -0.30$ ,  $CI_{95\%} [-0.47, -0.13]$ ,  $n_{\text{obs}} = 102$



Status: PAH

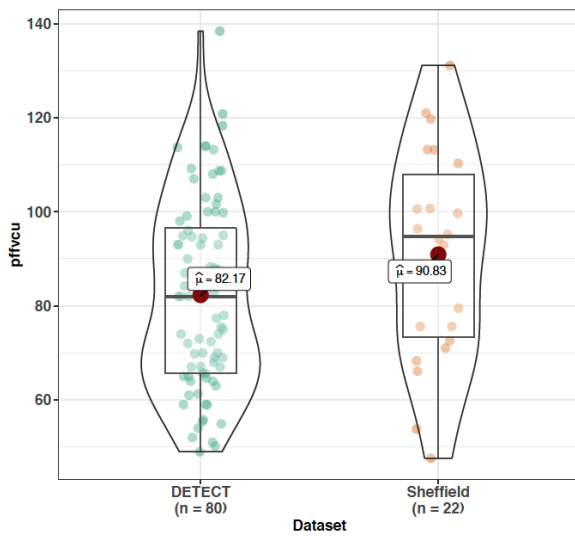
$\log_e(W_{\text{Wilcoxon}}) = 6.29$ ,  $p = 0.009$ ,  $\hat{\tau} = -0.26$ ,  $CI_{95\%} [-0.45, -0.08]$ ,  $n_{\text{obs}} = 99$



**Supplemental Figure S7:** Comparison of mean pulmonary artery pressure across the DETECT Discovery and Sheffield Confirmatory Cohorts.

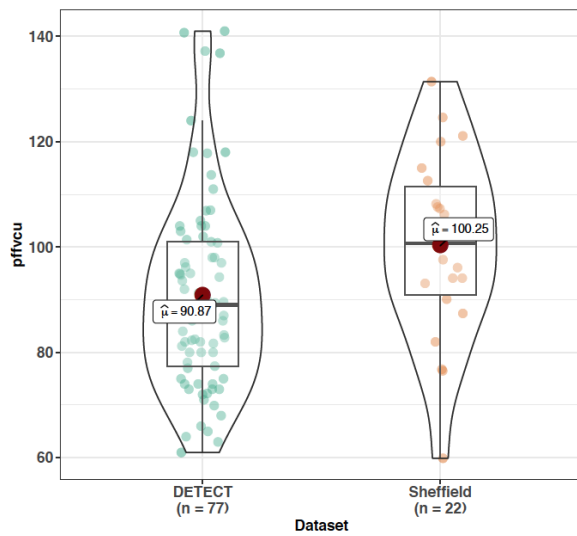
Status: non-PH

$\log_e(W_{\text{Wilcoxon}}) = 6.50$ ,  $p = 0.077$ ,  $\hat{\tau} = -0.18$ ,  $CI_{95\%} [-0.35, 0.00]$ ,  $n_{\text{obs}} = 102$



Status: PAH

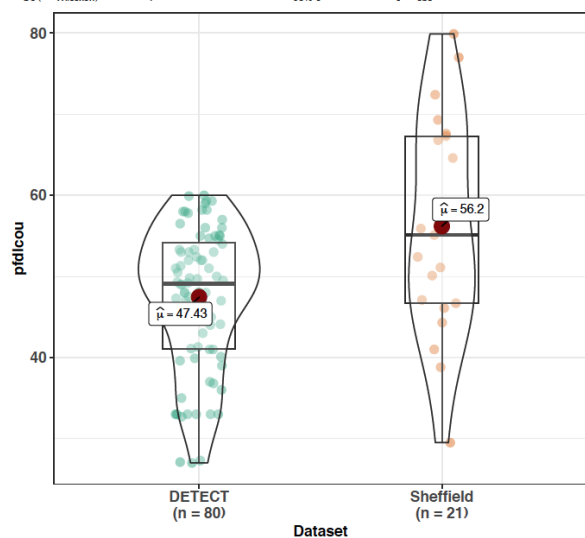
$\log_e(W_{\text{Wilcoxon}}) = 6.32$ ,  $p = 0.015$ ,  $\hat{\tau} = -0.25$ ,  $CI_{95\%} [-0.43, -0.06]$ ,  $n_{\text{obs}} = 99$



**Supplemental Figure S8:** Comparison of forced vital capacity across the DETECT Discovery and Sheffield Confirmatory Cohorts.

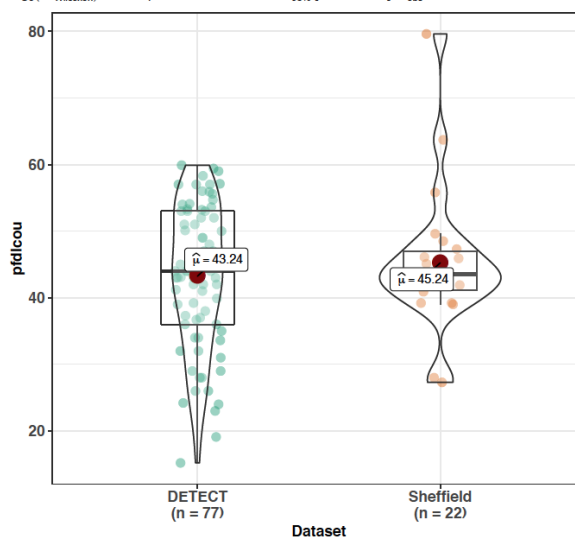
Status: non-PH

$\log_e(W_{\text{Wilcoxon}}) = 6.30$ ,  $p = 0.013$ ,  $\hat{\rho} = -0.25$ ,  $CI_{95\%} [-0.47, -0.07]$ ,  $n_{\text{obs}} = 101$

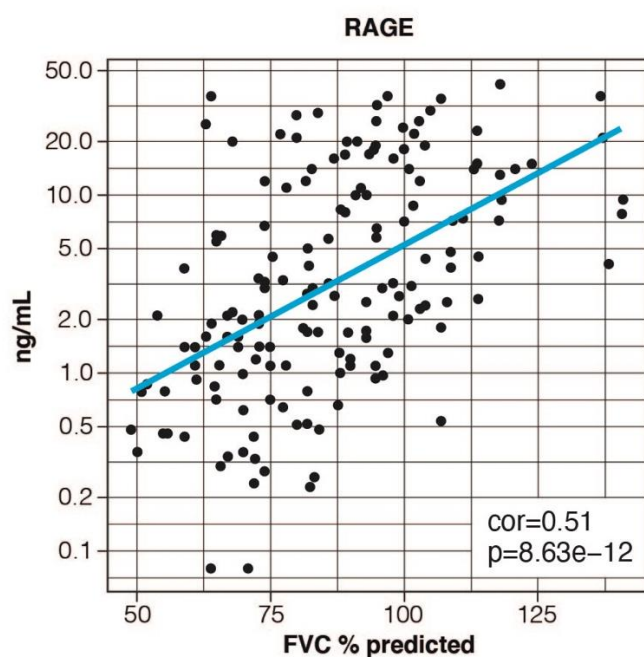


Status: PAH

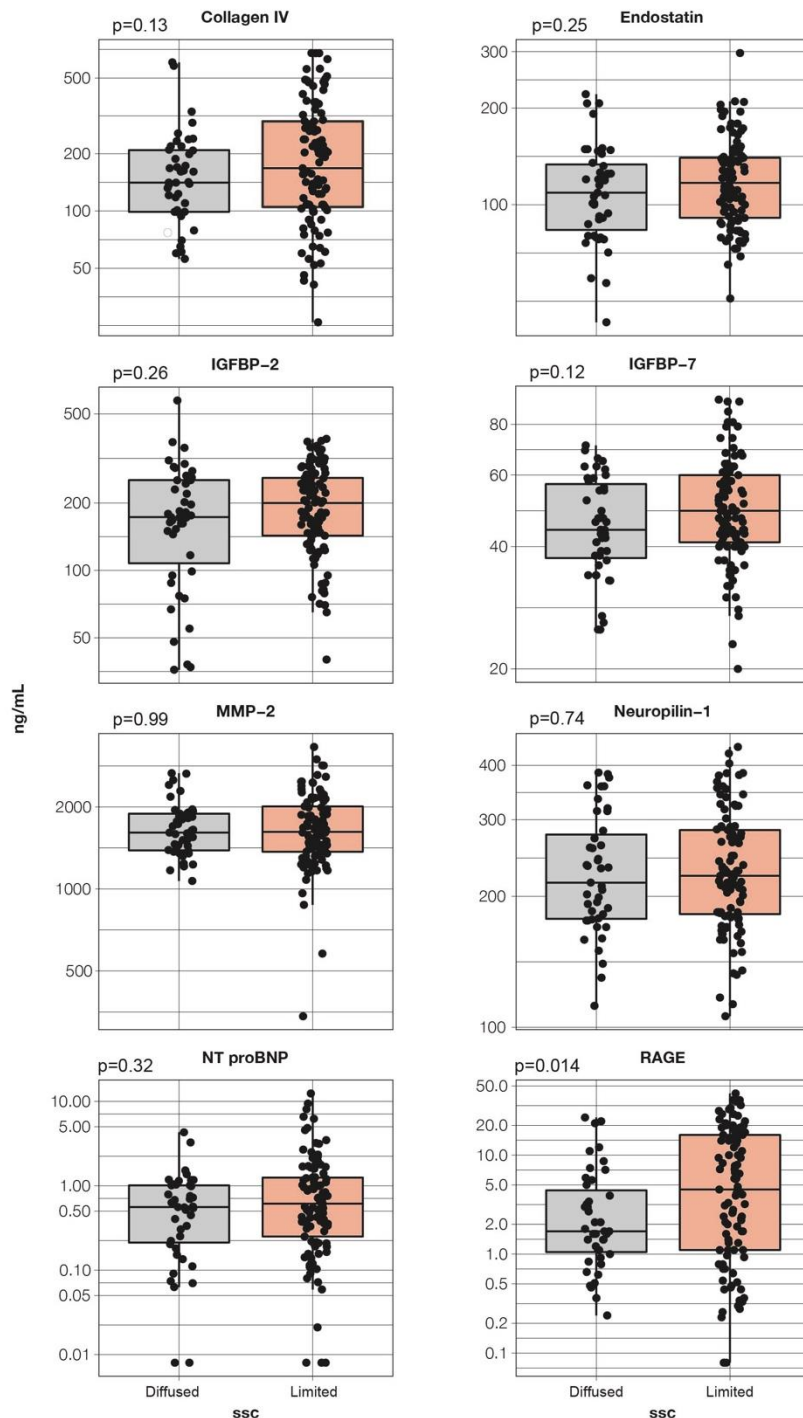
$\log_e(W_{\text{Wilcoxon}}) = 6.73$ ,  $p = 0.963$ ,  $\hat{\rho} = -0.01$ ,  $CI_{95\%} [-0.19, 0.16]$ ,  $n_{\text{obs}} = 99$



**Supplemental Figure S9:** Comparison of diffusion lung capacity for Carbon Monoxide across the DETECT Discovery and Sheffield Confirmatory Cohorts.



**Supplemental Figure S10:** Pearson correlation of RAGE protein levels with FVC % predicted in the DETECT cohort.



**Supplemental Figure S11:** Serum concentrations of the eight common analytes in DETECT serum patients with diffuse compared to limited SSc. Box and Whisker plots represent the interquartile range (box) with the line representing the median and whisker the full range of the data. Individual patient samples are represented by dots. The values in the top left corners represent the p-value of the Wilcoxon rank-sum test between the two patient groups.