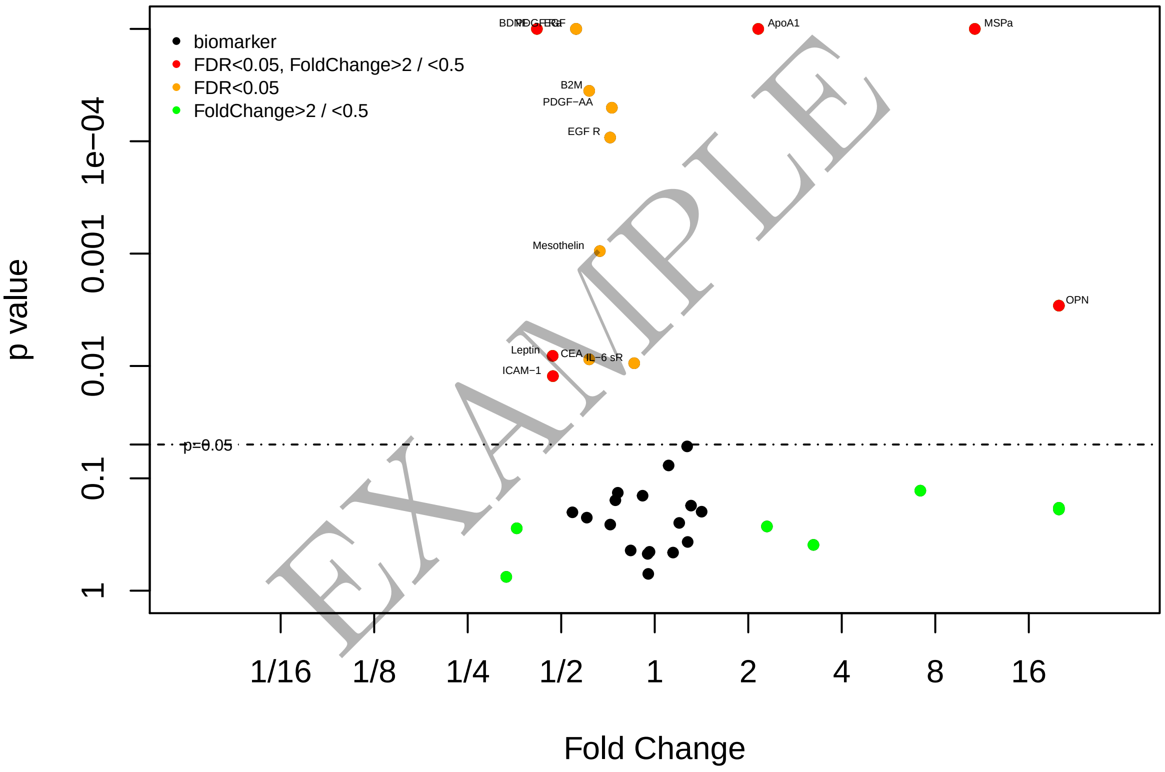
EXAMPLE REPORT
Biostatistics & Bioinformatics Services
"Differential Expression" Service

Volcano plot −− Patient vs. Control



Bioinformatics Team, RayBiotech
November 30, 2018

Contents

1 Introduction **2**

2 Methods **3**
- 2.1 Data filtration . . . 3
- 2.2 Comparison across groups . . . 3
- 2.3 Significance Analysis of Microarrays (SAM) . . . 3
- 2.4 Software . . . 3

3 Results **4**
- 3.1 Data filtration . . . 4
- 3.2 Comparison across groups . . . 4
- 3.3 Significance Analysis of Microarrays (SAM) . . . 6
- 3.4 Boxplots of differentially-expressed biomarkers . . . 6

References **14**

1 Introduction

The “Differential expression” service identifies differentially-expressed biomarkers between two groups using statistical methods like the Student’s t-test, Wilcoxon Rank-Sum, and Significance Analysis of Microarrays (SAM).

> *Need help understanding how the differential expression analyses are performed in layman’s terms? Please visit our* website.

Table 1: Number of samples in different groups

Group	# Sample
Control	80
Patient	80

Table 2: Sample list

Sample	Group	Sample	Group	Sample	Group	Sample	Group
SA001	Control	SA041	Control	SA101	Patient	SA141	Patient
SA002	Control	SA042	Control	SA102	Patient	SA142	Patient
SA003	Control	SA043	Control	SA103	Patient	SA143	Patient
SA004	Control	SA044	Control	SA104	Patient	SA144	Patient
SA005	Control	SA045	Control	SA105	Patient	SA145	Patient
SA006	Control	SA046	Control	SA106	Patient	SA146	Patient
SA007	Control	SA047	Control	SA107	Patient	SA147	Patient
SA008	Control	SA048	Control	SA108	Patient	SA148	Patient
SA009	Control	SA049	Control	SA109	Patient	SA149	Patient
SA010	Control	SA050	Control	SA110	Patient	SA150	Patient
SA011	Control	SA051	Control	SA111	Patient	SA151	Patient
SA012	Control	SA052	Control	SA112	Patient	SA152	Patient
SA013	Control	SA053	Control	SA113	Patient	SA153	Patient
SA014	Control	SA054	Control	SA114	Patient	SA154	Patient
SA015	Control	SA055	Control	SA115	Patient	SA155	Patient
SA016	Control	SA056	Control	SA116	Patient	SA156	Patient
SA017	Control	SA057	Control	SA117	Patient	SA157	Patient
SA018	Control	SA058	Control	SA118	Patient	SA158	Patient
SA019	Control	SA059	Control	SA119	Patient	SA159	Patient
SA020	Control	SA060	Control	SA120	Patient	SA160	Patient
SA021	Control	SA061	Control	SA121	Patient	SA161	Patient
SA022	Control	SA062	Control	SA122	Patient	SA162	Patient
SA023	Control	SA063	Control	SA123	Patient	SA163	Patient
SA024	Control	SA064	Control	SA124	Patient	SA164	Patient
SA025	Control	SA065	Control	SA125	Patient	SA165	Patient
SA026	Control	SA066	Control	SA126	Patient	SA166	Patient
SA027	Control	SA067	Control	SA127	Patient	SA167	Patient
SA028	Control	SA068	Control	SA128	Patient	SA168	Patient
SA029	Control	SA069	Control	SA129	Patient	SA169	Patient
SA030	Control	SA070	Control	SA130	Patient	SA170	Patient
SA031	Control	SA071	Control	SA131	Patient	SA171	Patient
SA032	Control	SA072	Control	SA132	Patient	SA172	Patient
SA033	Control	SA073	Control	SA133	Patient	SA173	Patient
SA034	Control	SA074	Control	SA134	Patient	SA174	Patient
SA035	Control	SA075	Control	SA135	Patient	SA175	Patient
SA036	Control	SA076	Control	SA136	Patient	SA176	Patient
SA037	Control	SA077	Control	SA137	Patient	SA177	Patient
SA038	Control	SA078	Control	SA138	Patient	SA178	Patient
SA039	Control	SA079	Control	SA139	Patient	SA179	Patient
SA040	Control	SA080	Control	SA140	Patient	SA180	Patient

2 Methods

2.1 Data filtration

Samples with missing data were excluded from the analysis. Biomarkers with no variation across all of the non-excluded samples (i.e., zero-variance) were also removed from the analysis.

2.2 Comparison across groups

Each biomarker was summarized by its mean and standard deviation, or median with minimum and maximum responses across the groups. The fold change between groups was calculated as the ratio of the Mean or Median. If the biomarkers met or did not meet normality criteria across two groups, the significance of expression difference was evaluated by the Student's t-test or Wilcoxon Rank-Sum, respectively. The significance of expression difference across three or more groups with normal or non-normal distributions was evaluated with ANOVA or the Kruskal–Wallis test. Biomarkers with FDR<0.05 were considered as differentially expressed.

2.3 Significance Analysis of Microarrays (SAM)

Significance Analysis of Microarrays (SAM) identifies differentially-expressed biomarkers or genes (Tusher, Tibshirani, and Chu 2001). Unlike the above-mentioned statistical tests that rely on p-value from theoretical probability distribution, SAM compares a t-like statistic d_i observed across groups with an estimate obtained from randomized permutations of the samples. The biomarkers with deviations larger than predefined Δ (default 1.2) were identified as differentially expressed.

2.4 Software

All the analyses were conducted with R programming language V3.5.1 (R Core Team 2017). The SAM analysis was implemented with R package *siggenes* 1.56.0 (Schwender 2012).

3 Results

3.1 Data filtration

Samples with missing data: None. No sample was excluded from analysis.

Biomarkers with zero-variance: None. No biomarker was excluded from the analysis.

3.2 Comparison across groups

Table 3: Comparison on 38 biomarkers between Control and Patient groups

biomarker	Control	Patient	FoldChange	statistic	p.value	FDR
MSPa	97.49(0, 4413.24)	1045.67(0, 10161.14)	10.73	Wilcoxon W = 739	0.0000	0.0000
ApoA1	219949.54(33001.97, 1695371.63)	473642.19(24268.22, 4581780.05)	2.15	Wilcoxon W = 1422	0.0000	0.0000
BDNF	6263.9(811.16, 23650.3)	2615.19(20.6, 20798.42)	0.42	Wilcoxon W = 4879	0.0000	0.0000
EGF	1103.53(52.07, 2814.99)	616.88(28.09, 2307.18)	0.56	Wilcoxon W = 4802	0.0000	0.0000
PDGF Ra	25916.74(3149.45, 88373.29)	14448.69(356.98, 79512.55)	0.56	Wilcoxon W = 4666	0.0000	0.0000
B2M	11058.82(1604.56, 52277.53)	6804.88(2062.12, 37383.07)	0.62	Wilcoxon W = 4412	0.0000	0.0002
PDGF-AA	216.43(63.31, 427.37)	157.51(26.34, 390.99)	0.73	Wilcoxon W = 4388.5	0.0001	0.0003
EGF R	4023.33(1781.16, 14473.82)	2891.12(601.3, 38140.09)	0.72	Wilcoxon W = 4346	0.0001	0.0004
Mesothelin	3649.12(315.67, 17467.94)	2428.82(333.13, 9076.65)	0.67	Wilcoxon W = 4169	0.0009	0.0040
OPN	5.6(0, 5072.88)	130.22(0, 33113.6)	23.27	Wilcoxon W = 2354	0.0029	0.0110
CEA	16759.42(438.83, 77759.09)	10314.14(326.83, 167759.7)	0.62	Wilcoxon W = 3969	0.0087	0.0276
IL-6 sR	806.29(360.37, 1492.8)	692.41(160.77, 1209.69)	0.86	Wilcoxon W = 3961	0.0095	0.0276
Leptin	3738.09(0, 184995.62)	1753.88(0, 132752.69)	0.47	Wilcoxon W = 3976	0.0081	0.0276
ICAM-1	36921.5(5426.15, 503784.53)	17360.62(582.07, 537584.1)	0.47	Wilcoxon W = 3934	0.0123	0.0334
CA125	1146.05(51.02, 21565.67)	1457.17(59.03, 169235.31)	1.27	Wilcoxon W = 2630	0.0520	0.1316
IL-6	47.66(0, 250.98)	52.82(10.21, 590.28)	1.11	Wilcoxon W = 2681	0.0768	0.1824
AgRP	11.92(0, 90.35)	9.07(0, 377.92)	0.76	Wilcoxon W = 3639	0.1344	0.2838
MCSF	758.21(0, 411657.18)	5429.97(0, 340268.58)	7.16	Wilcoxon W = 2771	0.1287	0.2838
TIMP-2	15204.26(80.49, 45626.02)	13897.5(1452.8, 52711.78)	0.91	Wilcoxon W = 3630	0.1427	0.2854
AFP	734.44(84.92, 4992.6)	548.67(63.9, 8138.3)	0.75	Wilcoxon W = 3615.5	0.1567	0.2977
IL-1 R6	266(0, 9241.89)	144.53(0, 12553.29)	0.54	Wilcoxon W = 3575	0.2006	0.3049
IL-8	67.77(0, 736.58)	88.7(0, 4338.76)	1.31	Wilcoxon W = 2802	0.1749	0.3049
MIF	3136.35(0, 94993.69)	4440.1(0, 355113.54)	1.42	Wilcoxon W = 2824.5	0.1981	0.3049
Prostasin	0(0, 19591)	258.75(0, 15765.43)	258746.00	Wilcoxon W = 2830	0.1831	0.3049
VEGF	1.08(0, 12636.64)	25.23(0, 376.12)	23.36	Wilcoxon W = 2832	0.1893	0.3049
transferrin	1119217.25(64388.85, 20162526)	676454.69(75036.27, 15786290.41)	0.60	Wilcoxon W = 3557	0.2238	0.3270
TIMP-4	401.47(48.24, 2527.24)	481.49(77.08, 2099.61)	1.20	Wilcoxon W = 2862	0.2494	0.3501
HE4	159.12(0, 694.62)	114.44(4.04, 1535.9)	0.72	Wilcoxon W = 3532	0.2579	0.3501
PDGF Rb	737.35(0, 303134.91)	1693.83(0, 280013.92)	2.30	Wilcoxon W = 2882	0.2681	0.3513
Prolactin	299275.3(0, 803723414.51)	107599.12(0, 1282915257.89)	0.36	Wilcoxon W = 3517.5	0.2785	0.3528
IL-2 Ra	3.87(0, 93.86)	4.94(0, 189.74)	1.28	Wilcoxon W = 2936.5	0.3684	0.4516
GROa	76.74(0, 14320.77)	248.85(0, 5916.37)	3.24	Wilcoxon W = 2954	0.3912	0.4646
CA15-3	4266.71(0, 63812.9)	4888.31(0, 293815.06)	1.15	Wilcoxon W = 2982	0.4579	0.4966
IGFBP-4	3416.62(0, 52073.53)	2858.45(0, 235380.99)	0.84	Wilcoxon W = 3427.5	0.4383	0.4966
IGFBP-3	52021.88±17408.19	50055.57±15437.32	0.96	t = 0.7559	0.4509	0.4966
Adiponectin/ACRP30	151624.34(17184.45, 302290.94)	143914.85(35376.37, 297628.98)	0.95	Wilcoxon W = 3412	0.4704	0.4966
CXCL16	1225.3(481.83, 3184.84)	1168.08(554.93, 4402.14)	0.95	Wilcoxon W = 3310	0.7086	0.7278
IFNa	798.87(0, 800584.3)	265.79(0, 668968.28)	0.33	Wilcoxon W = 3108	0.7534	0.7534

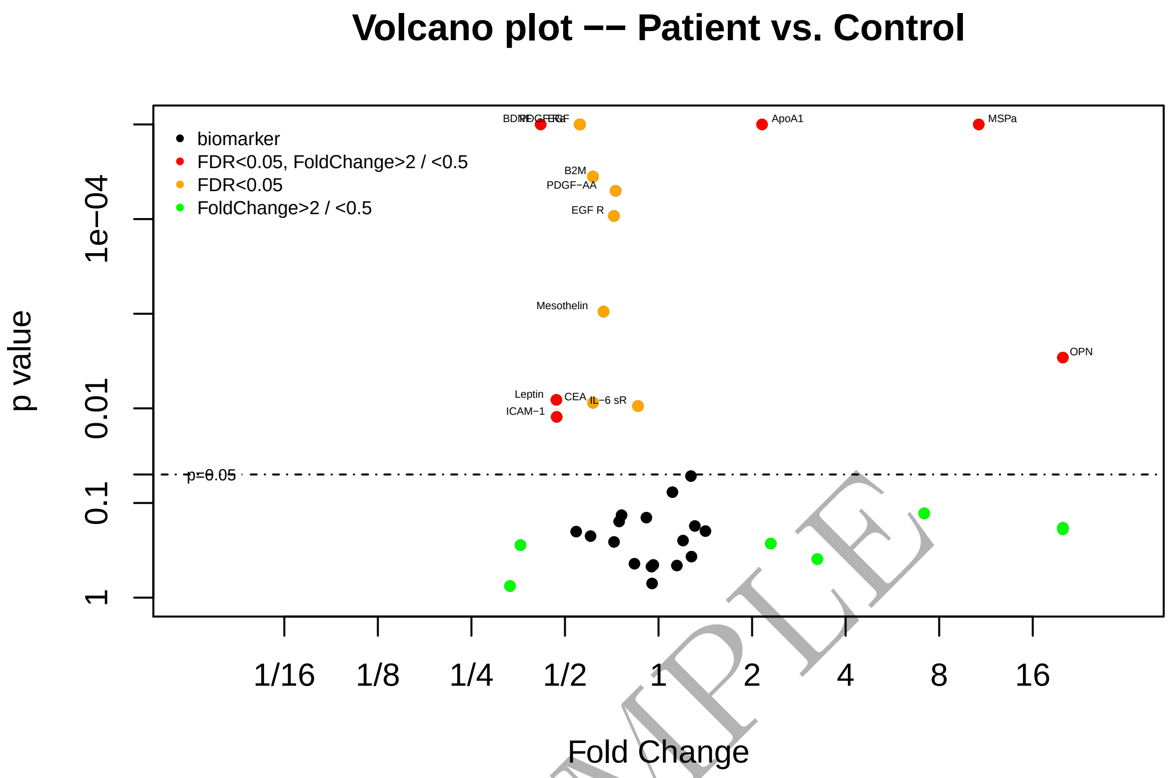


Figure 1: Volcano plot comparing the fold change and p-value of 38 biomarkers in which each point represents a biomarker

Table 4: 14 differentially-expressed biomarkers between Control and Patient groups (FDR<0.05)

biomarker	Control	Patient	FoldChange	statistic	p.value	FDR
MSPa	97.49(0, 4413.24)	1045.67(0, 10161.14)	10.73	Wilcoxon W = 739	0.0000	0.0000
ApoA1	219949.54(33001.97, 1695371.63)	473642.19(24268.22, 4581780.05)	2.15	Wilcoxon W = 1422	0.0000	0.0000
BDNF	6263.9(811.16, 23650.3)	2615.19(20.6, 20798.42)	0.42	Wilcoxon W = 4879	0.0000	0.0000
EGF	1103.53(52.07, 2814.99)	616.88(28.09, 2307.18)	0.56	Wilcoxon W = 4802	0.0000	0.0000
PDGF Ra	25916.74(3149.45, 88373.29)	14448.69(356.98, 79512.55)	0.56	Wilcoxon W = 4666	0.0000	0.0000
B2M	11058.82(1604.56, 52277.53)	6804.88(2062.12, 37383.07)	0.62	Wilcoxon W = 4412	0.0000	0.0002
PDGF-AA	216.43(63.31, 427.37)	157.51(26.34, 390.99)	0.73	Wilcoxon W = 4388.5	0.0001	0.0003
EGF R	4023.33(1781.16, 14473.82)	2891.12(601.3, 38140.09)	0.72	Wilcoxon W = 4346	0.0001	0.0004
Mesothelin	3649.12(315.67, 17467.94)	2428.82(333.13, 9076.65)	0.67	Wilcoxon W = 4169	0.0009	0.0040
OPN	5.6(0, 5072.88)	130.22(0, 33113.6)	23.27	Wilcoxon W = 2354	0.0029	0.0110
CEA	16759.42(438.83, 77759.09)	10314.14(326.83, 167759.7)	0.62	Wilcoxon W = 3969	0.0087	0.0276
IL-6 sR	806.29(360.37, 1492.8)	692.41(160.77, 1209.69)	0.86	Wilcoxon W = 3961	0.0095	0.0276
Leptin	3738.09(0, 184995.62)	1753.88(0, 132752.69)	0.47	Wilcoxon W = 3976	0.0081	0.0276
ICAM-1	36921.5(5426.15, 503784.53)	17360.62(582.07, 537584.1)	0.47	Wilcoxon W = 3934	0.0123	0.0334

3.3 Significance Analysis of Microarrays (SAM)

Table 5: 10 differentially-expressed biomarkers identified by SAM with delta=1.2

	Row	d.value	stdev	rawp	q.value	R.fold
MSPa	22	7.349387	297.45280	0.0000000	0.0000000	Inf
EGF	8	-5.598618	91.83085	0.0000000	0.0000000	0
BDNF	3	-5.139344	740.17929	0.0000000	0.0000000	0
ApoA1	36	5.129201	85751.28913	0.0000000	0.0000000	Inf
PDGF Ra	24	-4.800321	2370.32482	0.0000000	0.0000000	0
PDGF-AA	26	-4.402508	11.88127	0.0000000	0.0000000	0
B2M	31	-4.387495	1518.29295	0.0000000	0.0000000	0
CA125	4	3.225055	3129.63619	0.0005263	0.0000280	Inf
Mesothelin	20	-3.159356	392.06514	0.0005263	0.0000280	0
IL-6 sR	16	-2.870162	36.25126	0.0023684	0.0001133	0

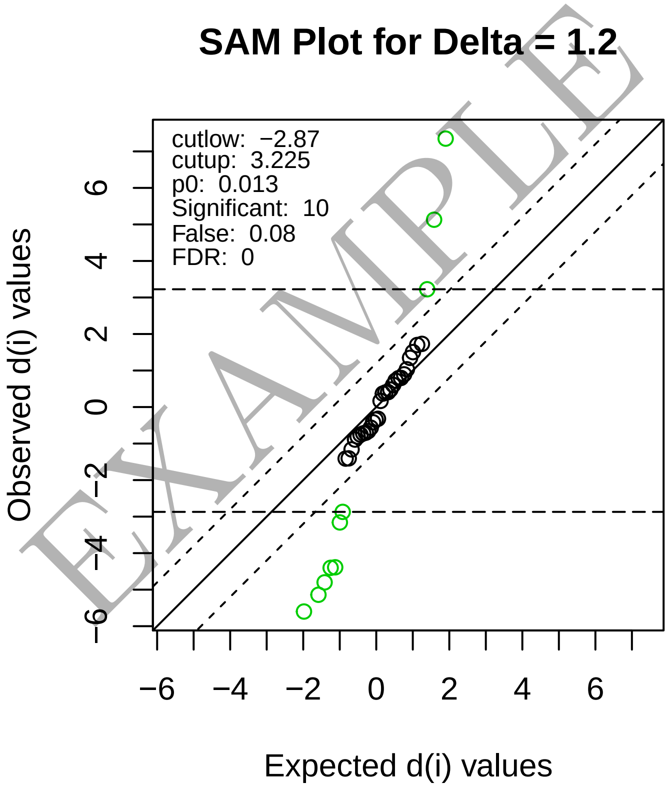


Figure 2: SAM plot for delta=1.2 of 38 biomarkers in two groups

3.4 Boxplots of differentially-expressed biomarkers

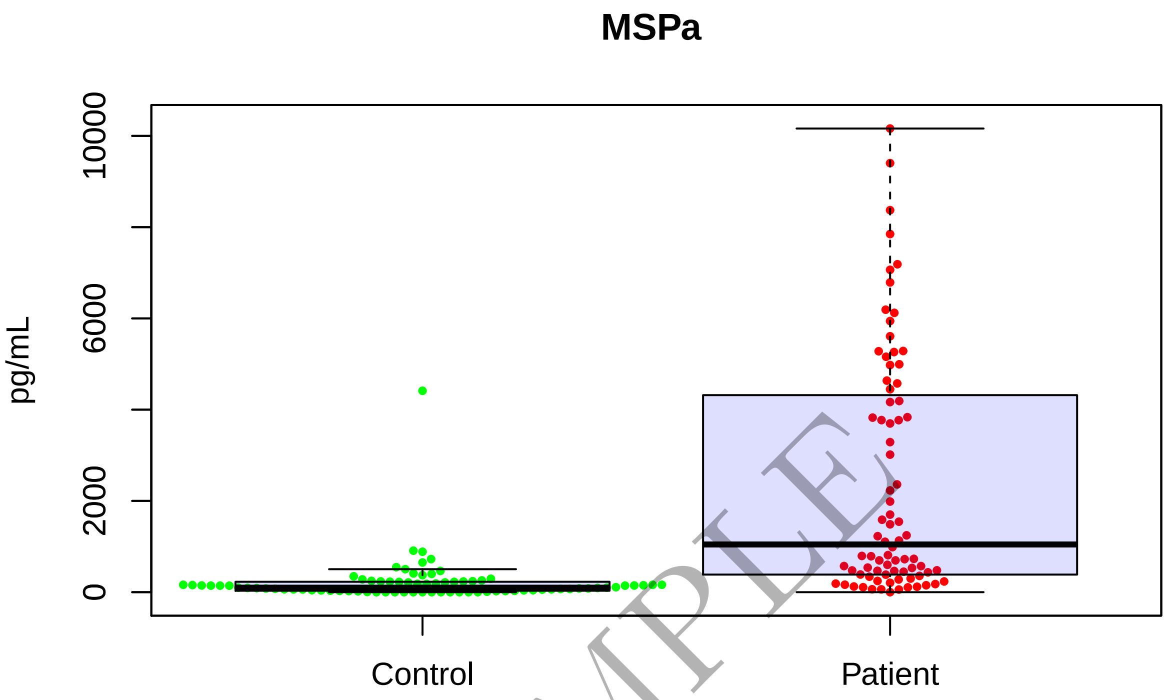


Figure 3: Boxplot of MSPa in Control and Patient groups

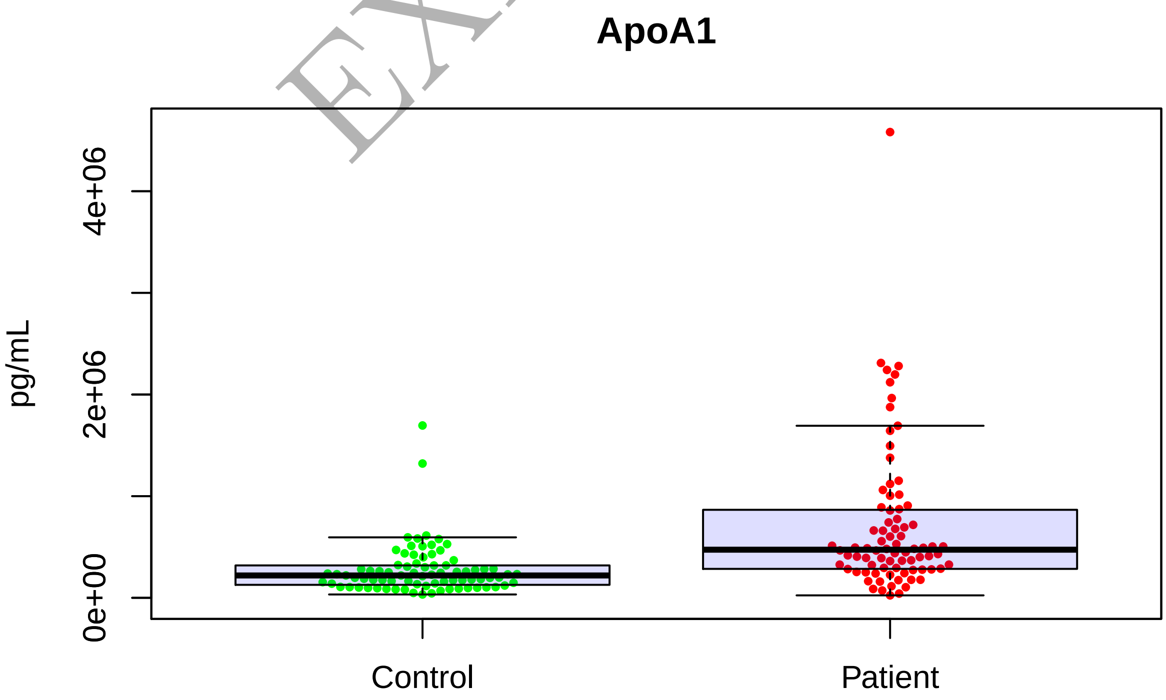


Figure 4: Boxplot of ApoA1 in Control and Patient groups

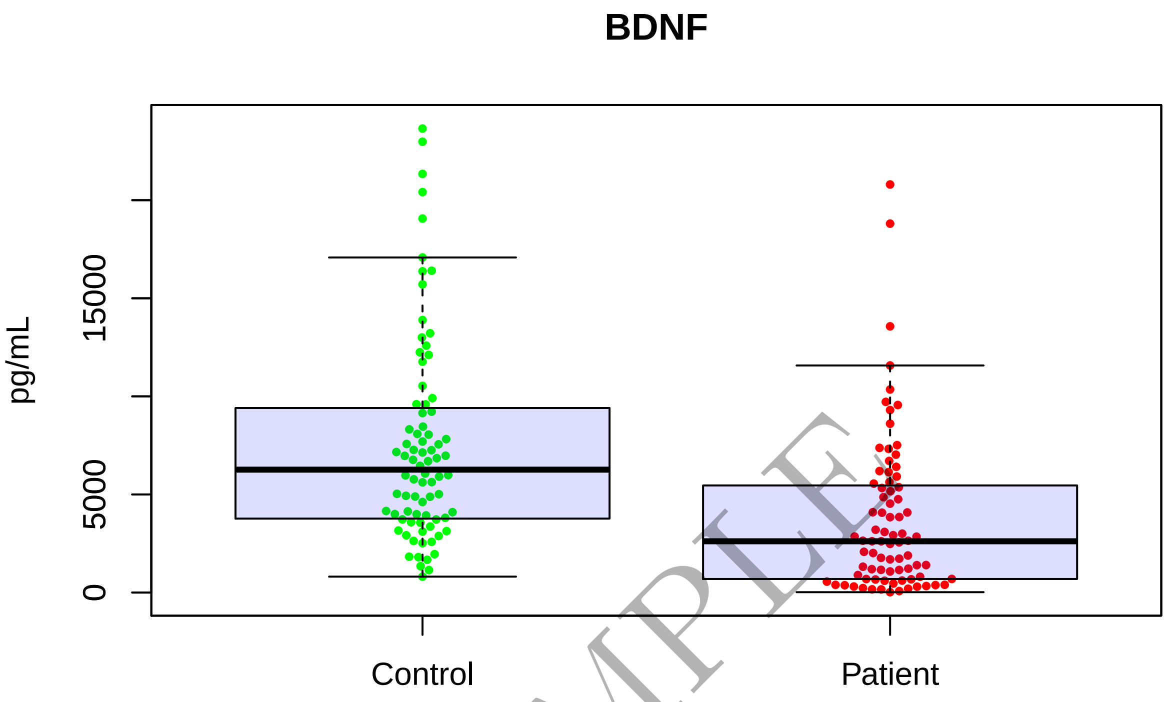


Figure 5: Boxplot of BDNF in Control and Patient groups

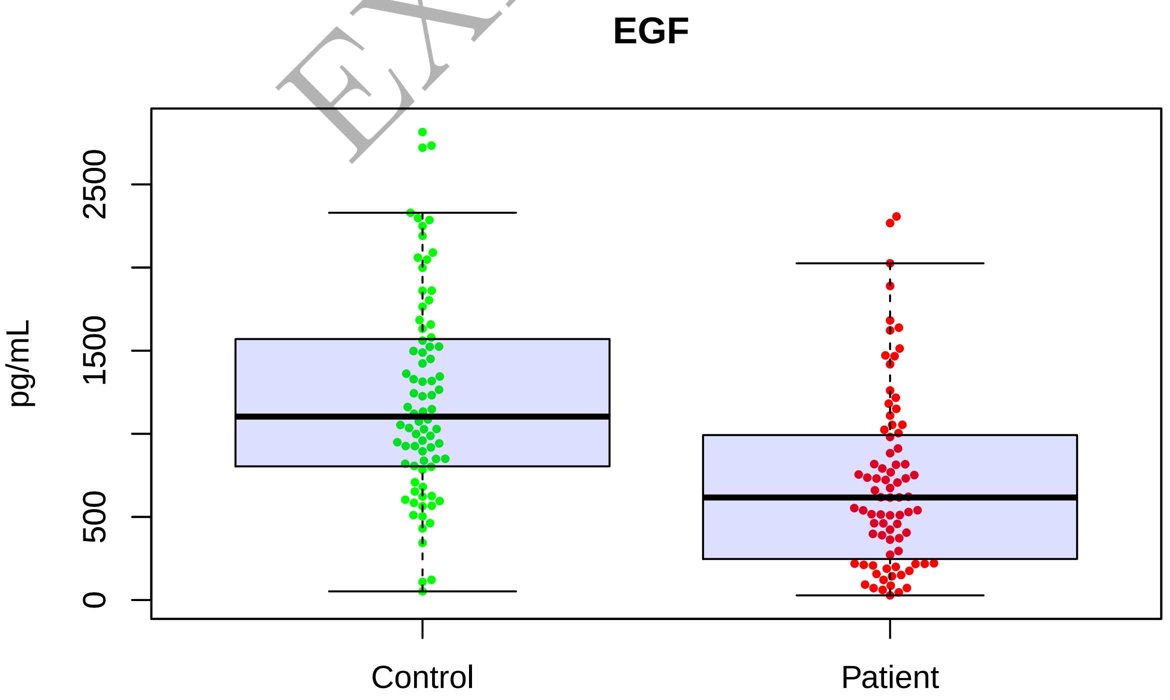


Figure 6: Boxplot of EGF in Control and Patient groups

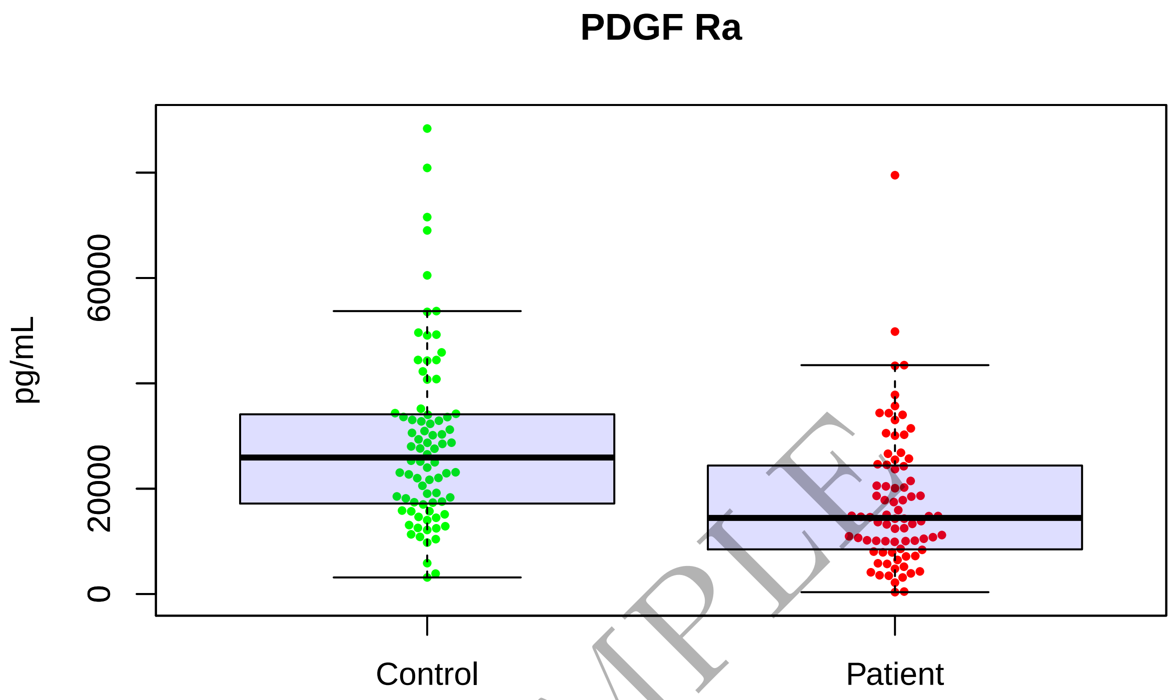


Figure 7: Boxplot of PDGF Ra in Control and Patient groups

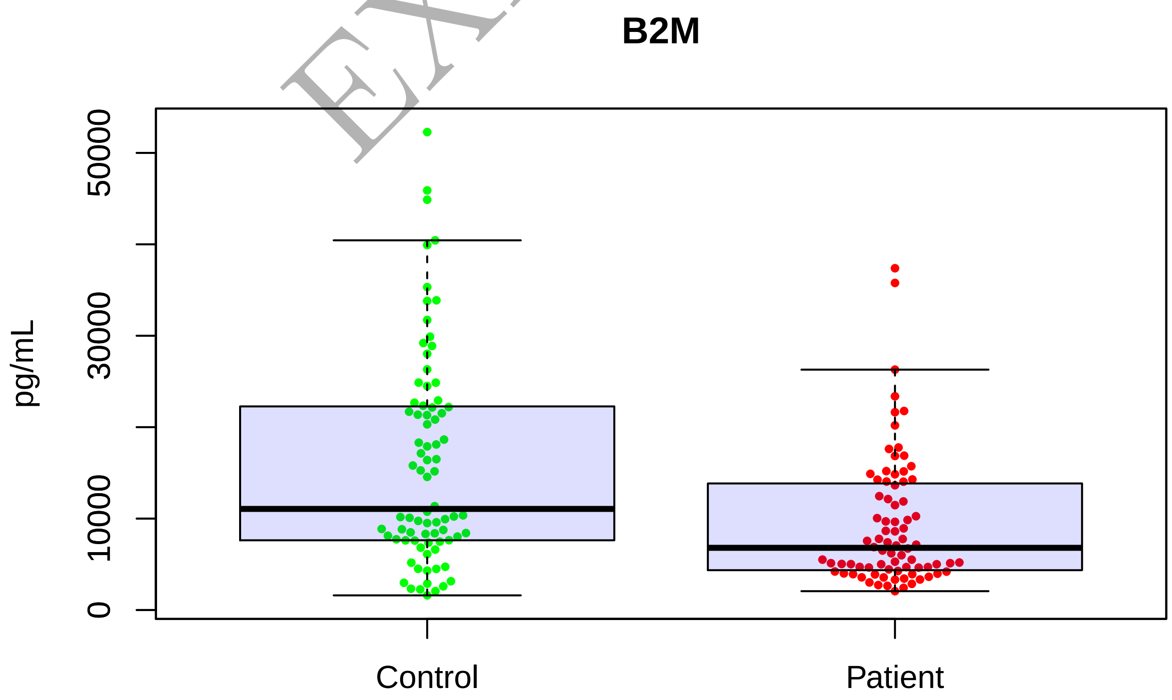


Figure 8: Boxplot of B2M in Control and Patient groups

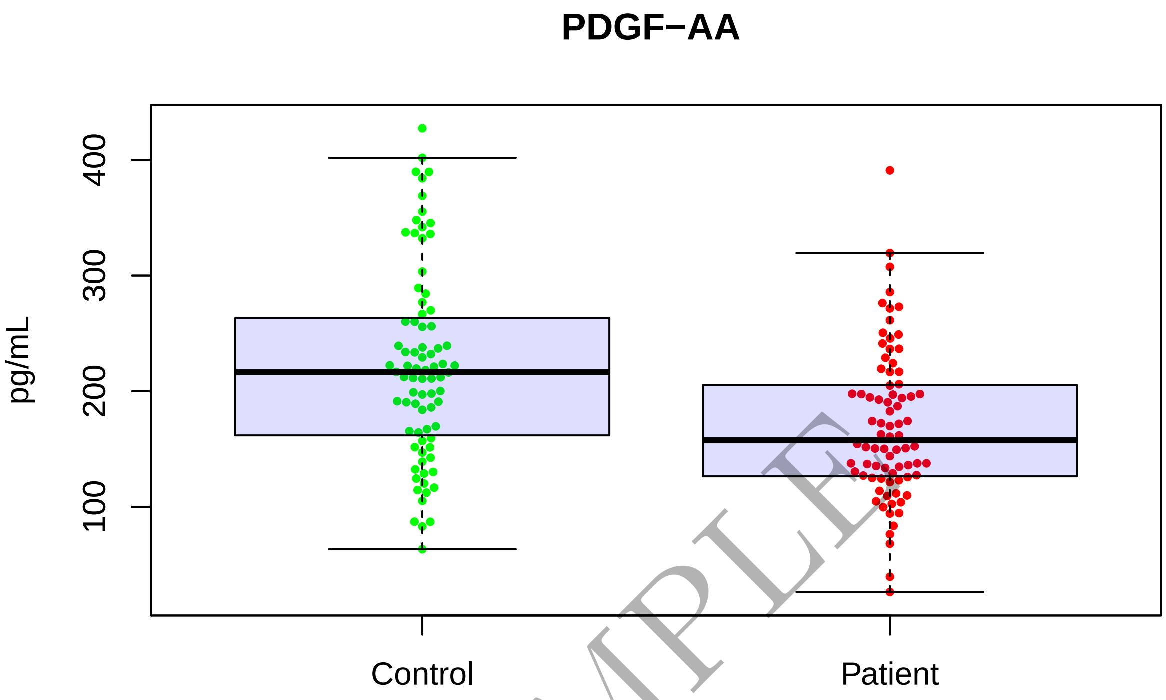


Figure 9: Boxplot of PDGF-AA in Control and Patient groups

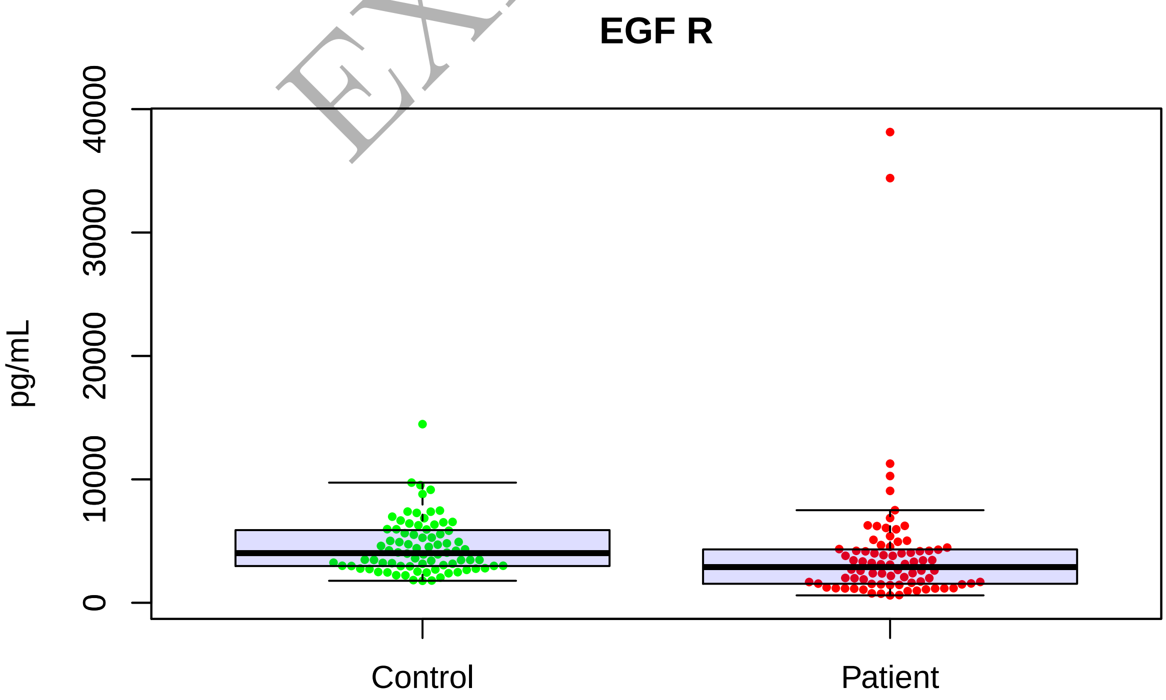


Figure 10: Boxplot of EGF R in Control and Patient groups

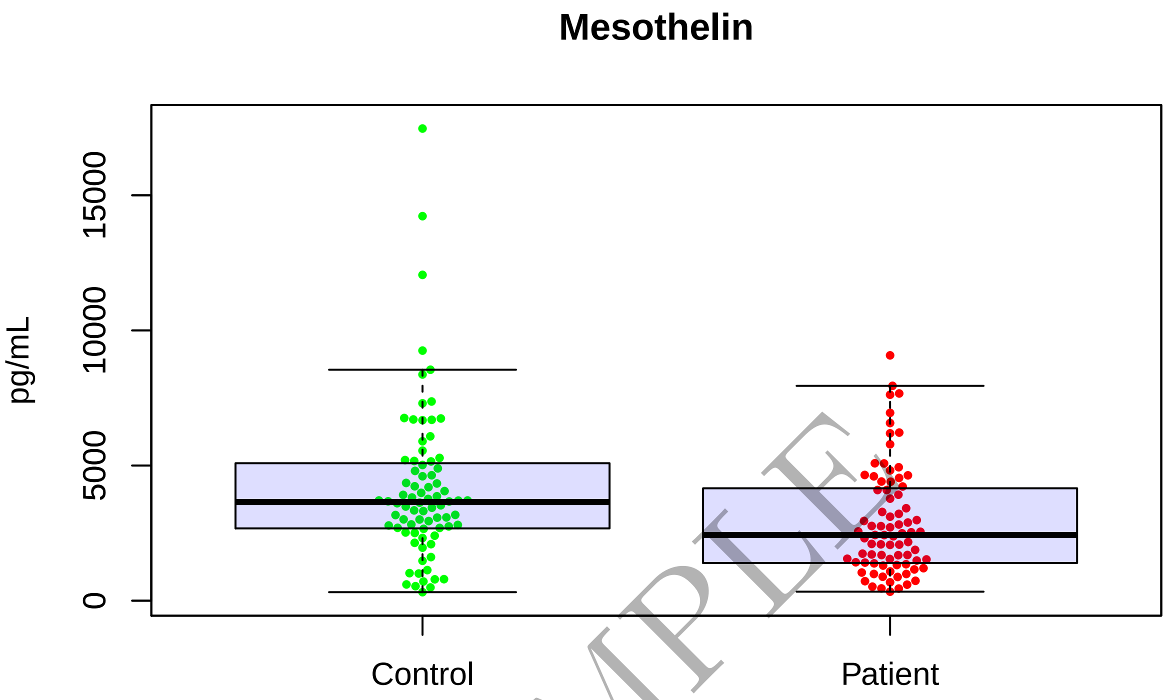


Figure 11: Boxplot of Mesothelin in Control and Patient groups

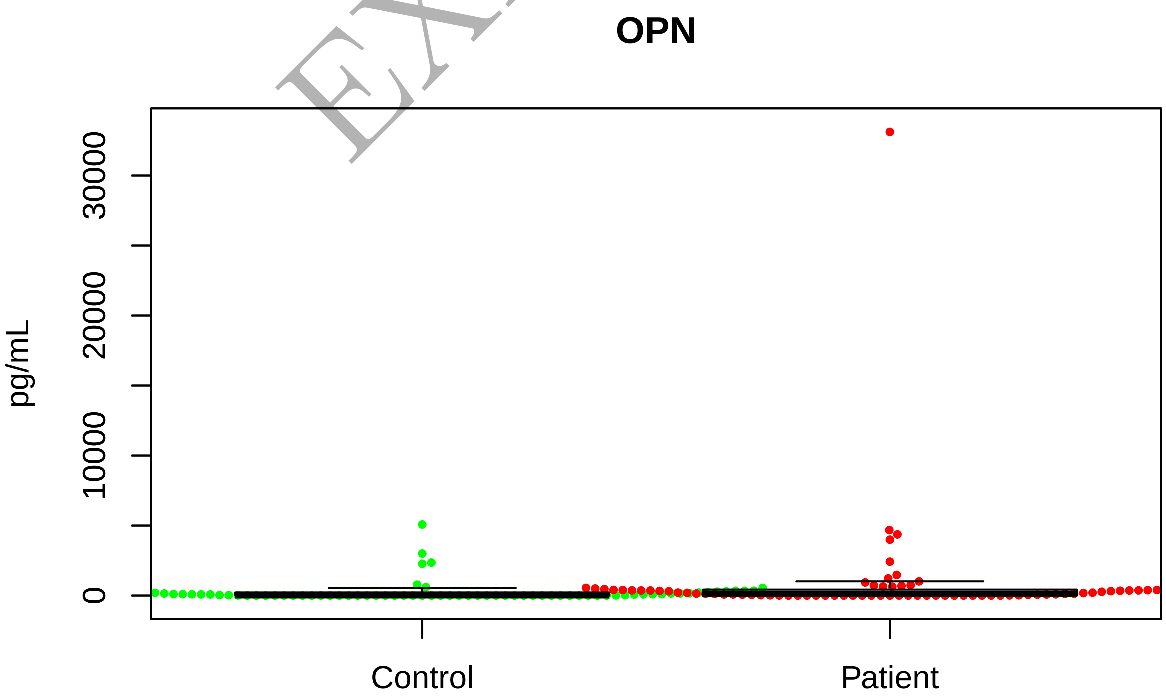


Figure 12: Boxplot of OPN in Control and Patient groups

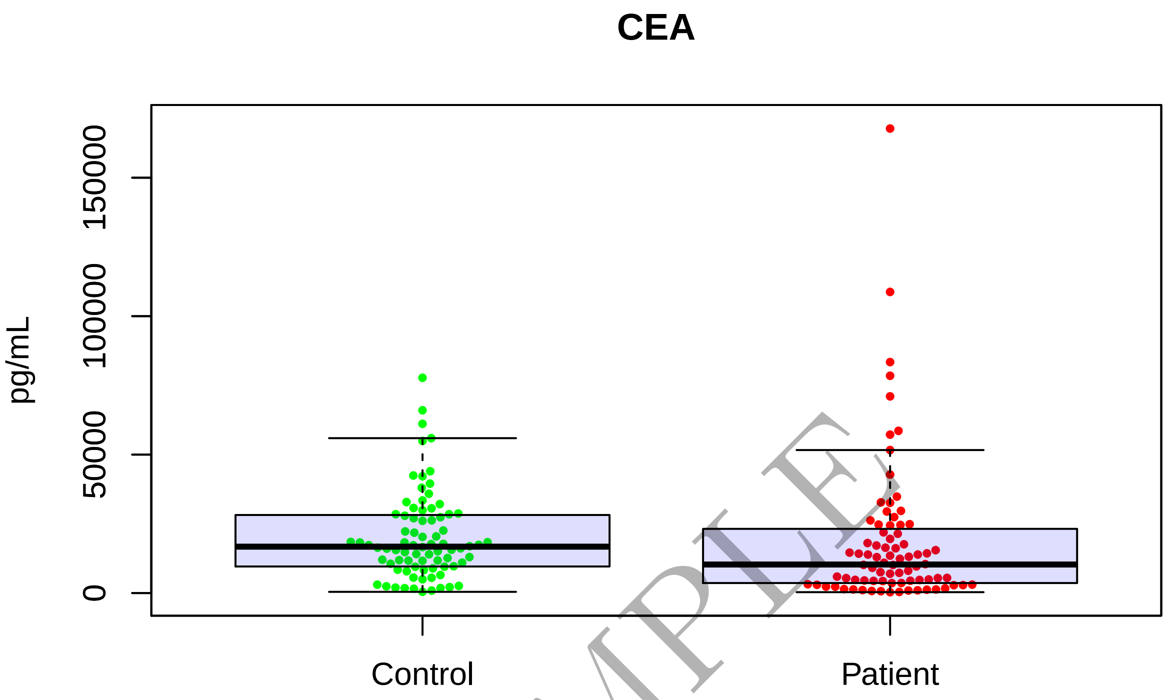


Figure 13: Boxplot of CEA in Control and Patient groups

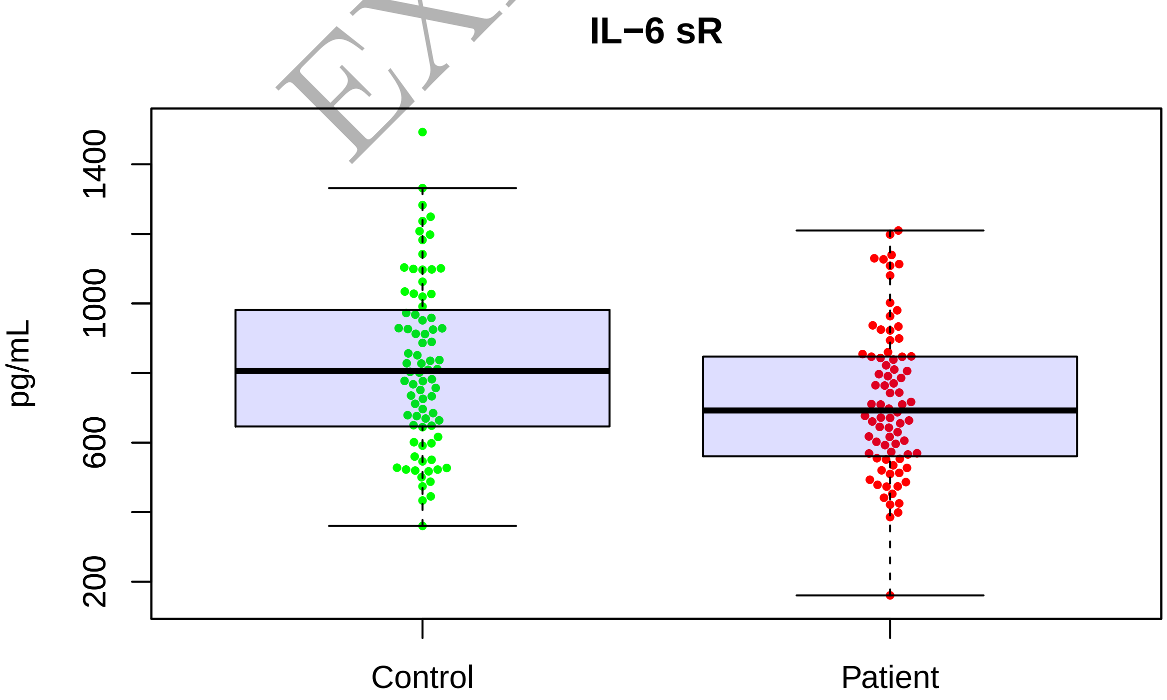


Figure 14: Boxplot of IL-6 sR in Control and Patient groups

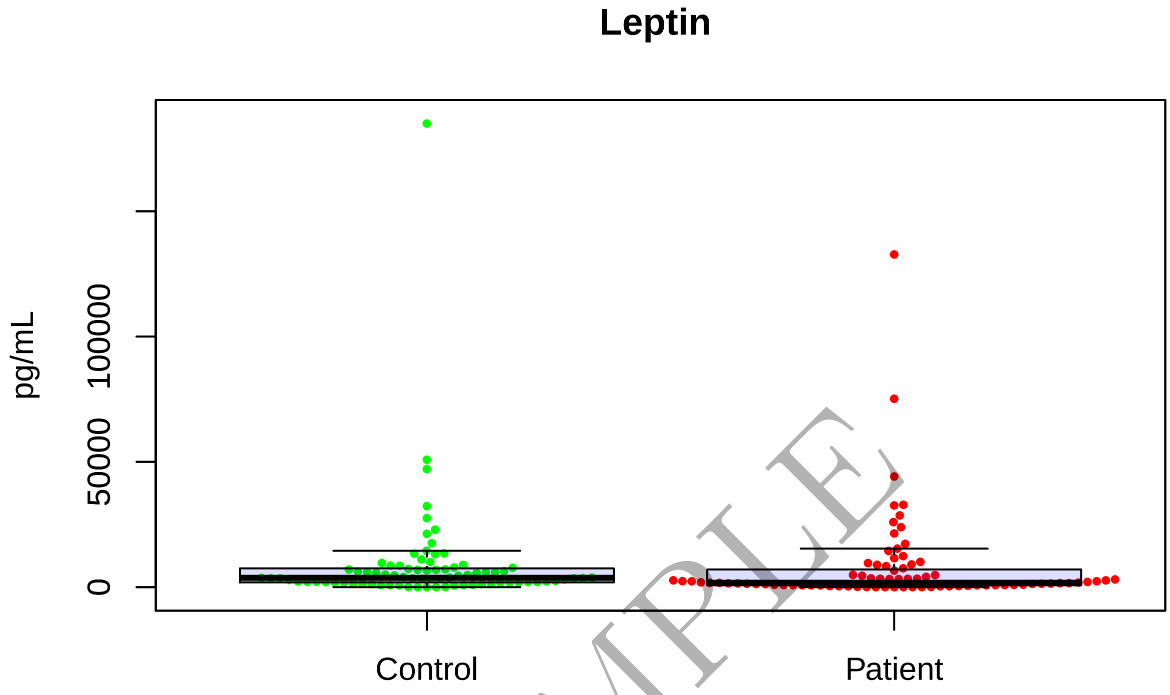


Figure 15: Boxplot of Leptin in Control and Patient groups

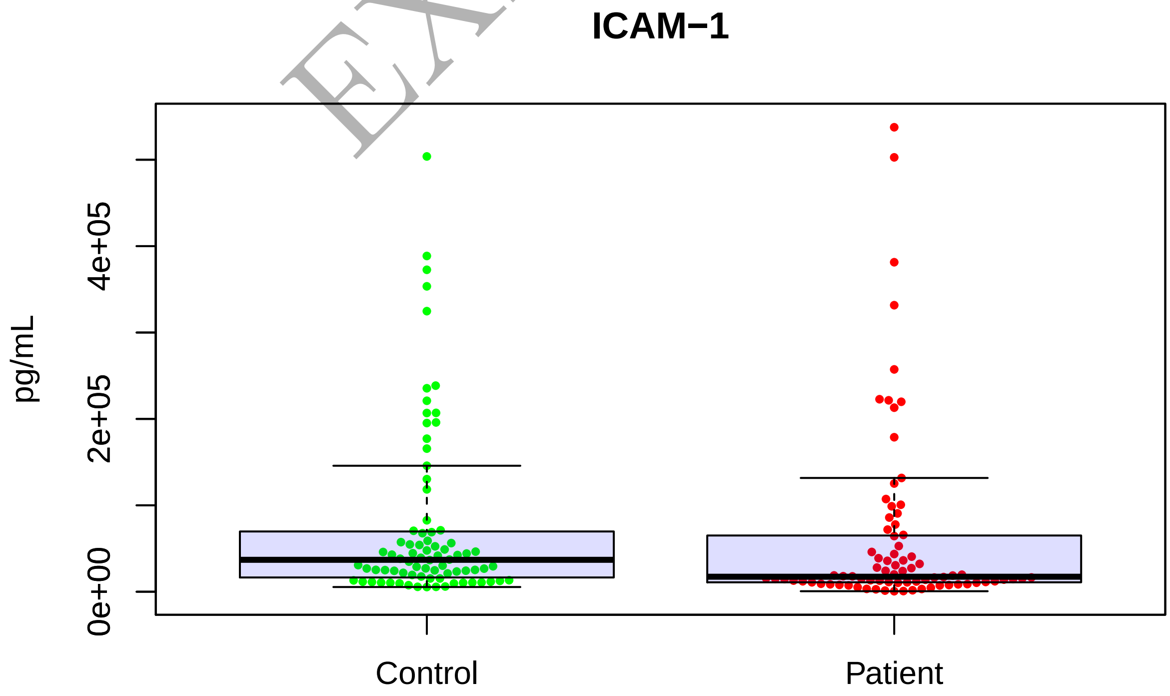


Figure 16: Boxplot of ICAM-1 in Control and Patient groups

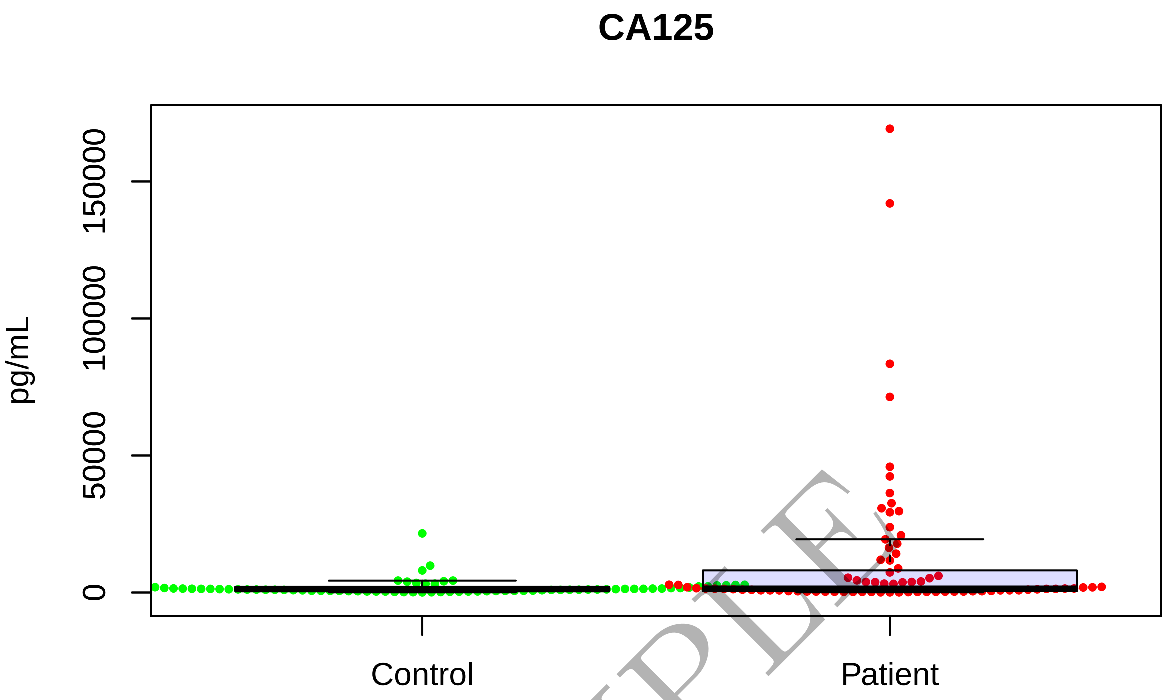


Figure 17: Boxplot of CA125 in Control and Patient groups

References

R Core Team. 2017. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schwender, Holger. 2012. *Siggenes: Multiple Testing Using Sam and Efron's Empirical Bayes Approaches.*

Tusher, G. V., R Tibshirani, and G Chu. 2001. "Significance Analysis of Microarrays Applied to the Ionizing Radiation Response." *Proceedings of the National Academy of Sciences* 98 (9): 5116–21.