**Use of Inflammation mediators aspirated from synovial fluid to predict the outcome of ACL reconstruction surgery**

**Statistical analysis report**

Jenny Lee

[JennyLee.Stat@gmail.com](mailto:JennyLee.Stat@gmail.com)

March 18, 2019

1. Aim of this study

This study aims to determine the ability of a set of inflammatory mediators acquired at the acute phase of anterior cruciate ligament (ACL) tear injury to predict the progression of cartilage matrix degeneration over one year for subjects who underwent ACL reconstruction surgery.

1. Study Design and objectives
   1. Study design

A total of 18 Biomarker analytes aspirated from the synovial fluid of participants (N=25) at baseline (at the time of enrollment after acute injury and before surgery), were assessed. The participants were scanned for quantitative magnetic resonance imaging (qMRI) (T1rho) measurement three times at baseline, six months after the ACL reconstruction surgery, and one year after the surgery. Additional 10 participants with normal knees were recruited as a control cohort for image acquisition. Voxel-Based Relaxometry (VBR) analysis was performed to identify the size, severity, and location of clusters of voxels in cartilage compartments (medial femur (MF), lateral femur (LF), medial tibia (MT), lateral tibia (LT), patella (PAT), trochlea (TRO)) that showed higher level of qMRI measurements over a one-year period as compared to healthy controls.

2.2. Study objectives

* To discover and summarize the latent patterns among inflammation mediators using principal component analysis. (preliminary analysis)
* To determine whether baseline levels for a combination of biomarkers (derived from above) can predict the change in T1rho measurements in MF in one-year follow-up (preliminary analysis)
* To determine whether a combination of **baseline levels of biomarkers** (derived from above) is associated with **the volume of a cluster in MF** identified by VBR. (**primary analysis**)
* To determine whether the model based on features extracted from the VBR can explain the cartilage degeneration progress better than features based on mean global T1rho values. (secondary analysis)
* To determine whether combinations of baseline levels of biomarkers (derived from above) can predict volume of clusters identified by VBR in all compartments. (exploratory analysis)

1. Statistical Methods

3.1 Data quality control

* For details for the quality control of assays, please refer to Janet’s report.
* For each analyte and outcome, the anomalies and distribution were examined. If the distribution was not normally distributed, we performed the natural log transformation so that the transformed distribution approximates the normality.
* To control the possibly influential data points in this rather small study, the outliers that are located outside of +/- 1.58 \* IQR / sqrt(N) were assigned the value of the nearest non-outlier value (Winsorization).
* Each assay value was centered (subtracted by its mean value).
* All statistical analysis followed were conducted on the transformed data.

3.2. Data dimension reduction in analytes space – Principal component analysis (PCA)

Upon the initial observation of multicollinearity among analytes, we determined to use principal component analysis (PCA) to discover underlying relationships among predictor variables.

3.3. Multivariate analysis

Multivariate linear models were used to identify the combination of biomarkers, demographics, and clinical information that are associated to the progression of cartilage matrix degeneration. The results from PCA analysis above were used as predictors. The candidate models were: multivariate linear regression on principal components and ridge linear regression. The best model was chosen based on the model fit, robustness, and parsimony.

3.4. Software information

All analyses were performed in R (version 3.3.3) and Python (version 3.6.1).

1. Results

4.1. Data description

4.1.1. Summary of participants

A total of 25 study participants (Male: 15, Female: 10) with unilateral ACL injury were investigated. The age ranged from 18 to 56 years old and had a mean (SD) of 34.48 years old (12.32 years old). BMI ranged from 19.31 to 32.28 with a mean (SD) of 24.82 (3.36 ). The days since injury ranged from 3 to 33 days and had a mean (SD) of 17.40 days (7.92 days).

Additionally, four categorical variables provided by surgeons were included as predictors: presence of synovitis (0: absent, 1: present), presence of meniscal tear (0: no tear, 1: presence), presence of effusion (0: no effusion, 1: presence of effusion), presence of cartilage defect (0: absence, 1: presence).

Table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Effusion | Cartilage defects | Synovitis | Meniscal tear | N |
| 0 | 0 | 0 | 0 | 3 |
| 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 0 | 3 |
| 0 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 4 |
| 1 | 0 | 0 | 1 | 1 |
| 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 0 | 1 | 8 |

4.1.2. Summary of biomarker profiles

The summary statistics for the assay values of 18 analytes are shown below.

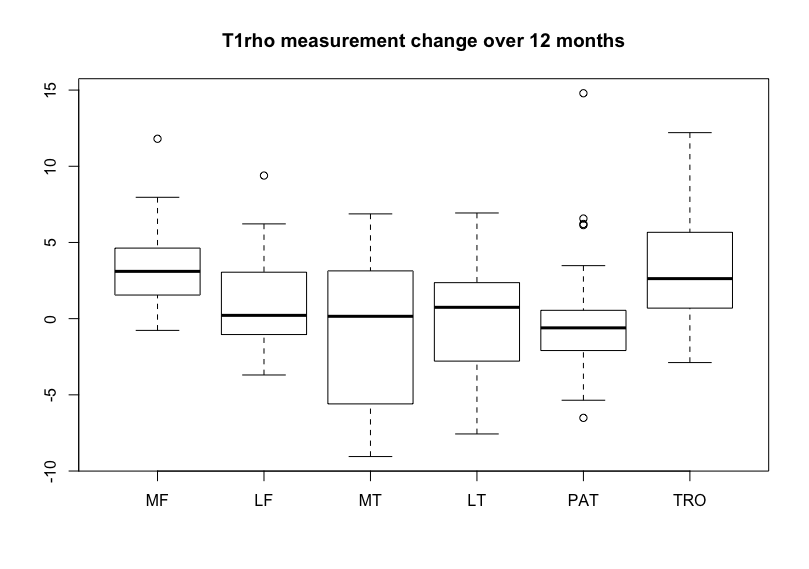
Table 2. Description of assay for 18 analytes (N = 25)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Min | 1Q | Median | 3Q | Max |
| IL1RA (pg /ml) | 37.51 | 235.44 | 656.71 | 2675.50 | 48986.97 |
| VEGF (pg/ml) | 69.62 | 241.43 | 426.99 | 624.66 | 1899.19 |
| IL1 (pg /ml) | 0.01 | 0.08 | 0.18 | 0.36 | 2.01 |
| IL6 (pg /ml) | 1.94 | 30.28 | 90.20 | 748.70 | 4728.82 |
| IL8 (pg /ml) | 12.38 | 30.98 | 57.37 | 114.22 | 973.77 |
| IL10 (pg /ml) | 0.06 | 0.31 | 0.65 | 1.21 | 1.92 |
| TNF (pg /ml) | 0.75 | 1.68 | 2.69 | 3.67 | 41.30 |
| MMP1 (pg /ml) | 64924 | 167049 | 365600 | 924605 | 1473227 |
| MMP3 (pg /ml) | 444564 | 1390551 | 1859172 | 5516018 | 11094058 |
| MMP9 (pg /ml) | 1812 | 5119 | 13134 | 21789 | 49839 |
| TIMP1 (pg /ml) | 525658 | 1597025 | 2256685 | 3520344 | 6173992 |
| COMP (ng /ml) | 8310 | 21103 | 27106 | 37445 | 55301 |
| sfCTXII (ng /ml) | 0.14 | 0.39 | 0.50 | 0.72 | 1.50 |
| sGAG (ug /ml) | 8.38 | 12.20 | 15.84 | 24.85 | 59.38 |
| ICAM-1 (ng /ml) | 117.20 | 149.10 | 189.00 | 235.80 | 374.30 |
| VCAM-1(ng /ml) | 431.00 | 550.00 | 663.60 | 792.80 | 1104.30 |
| MCP-1(pg /ml) | 195.90 | 332.30 | 480.70 | 685.90 | 1462.90 |
| TARC (pg /ml) | 13.69 | 22.50 | 31.36 | 44.92 | 166.45 |

4.1.3. qMRI T1rho measurements difference (baseline vs. 12 months follow-up)

The boxplot of the change of T1rho measurement over 12 months of period for each compartment is shown in Figure 1. While the median values of the change in T1rho measurements were greater than 0 in all six compartments, the distributions were spread wide and some participants showed lower T1rho measurements in one-year follow up.

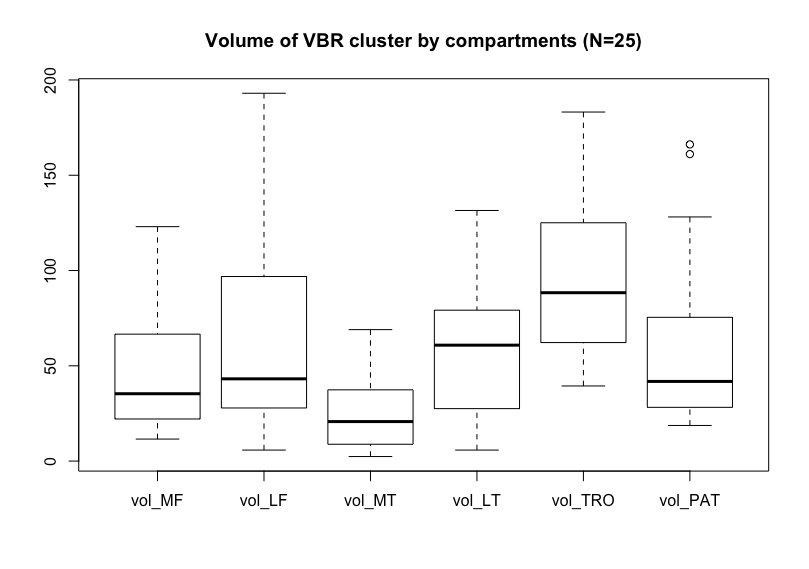
Figure 1. Boxplot of change in T1 rho measurements over 12 months period (N = 25)



4.1.4. Volume of VBR clusters

The boxplot of volumes of clusters in each compartment are compared in Figure 2.

**Figure 2. Boxplot of volumes of cluster in each compartment (N = 25)**

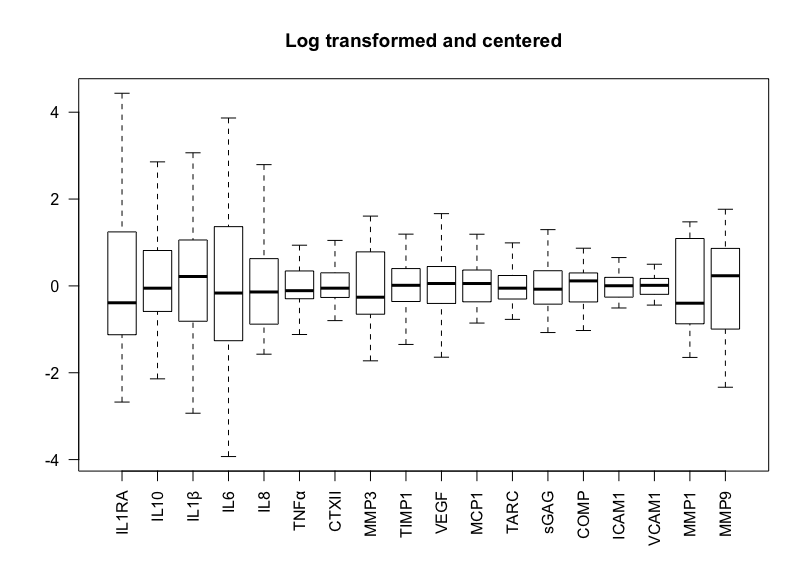


4.2. Data quality control and preprocessing

Among the total of 20 analytes initially given, we dropped two assays, GM-CSF and IL1-a, that showed the majority of the values below the lower level of detection (88% and 96%, respectively). A total of three samples of IL-1Beta were found the < LLOD and were extrapolated / imputed with 0.5 \* LLOD for each assay.

Figure 3 in Appendix shows the distribution of each analytes after natural log transformation and centering.

**Figure 3. A total of 18 analytes intensities after preprocessing**



4.3. Preliminary analysis

4.3.1. Principal component analysis

We performed PCA on the space of a total of 25 predictors (age, BMI, the days since injury, 18 biomarker assays in addition to 4 categorical data from surgeons). All predictors are centered by subtracting the mean of its values. The scree plot for the PCA analysis is shown in Figure 4. It seems that the variance levels off after 4-5 principal components. First four PCs account for 86.17% of total variance; first five PCs account for 89.42% of total variance.

Figure 4. Scree plot for PCA of a total of 18 biomarker assays

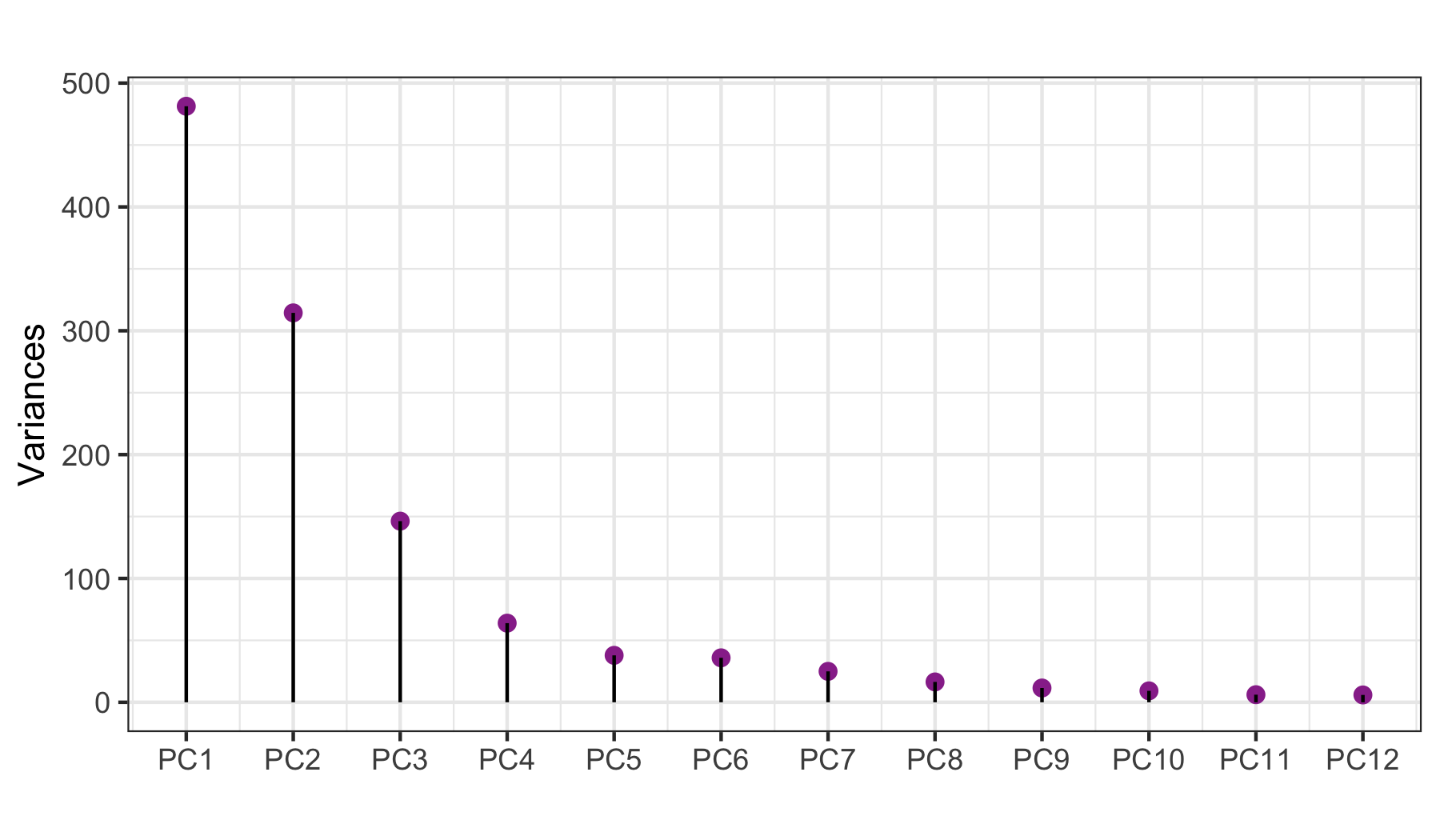
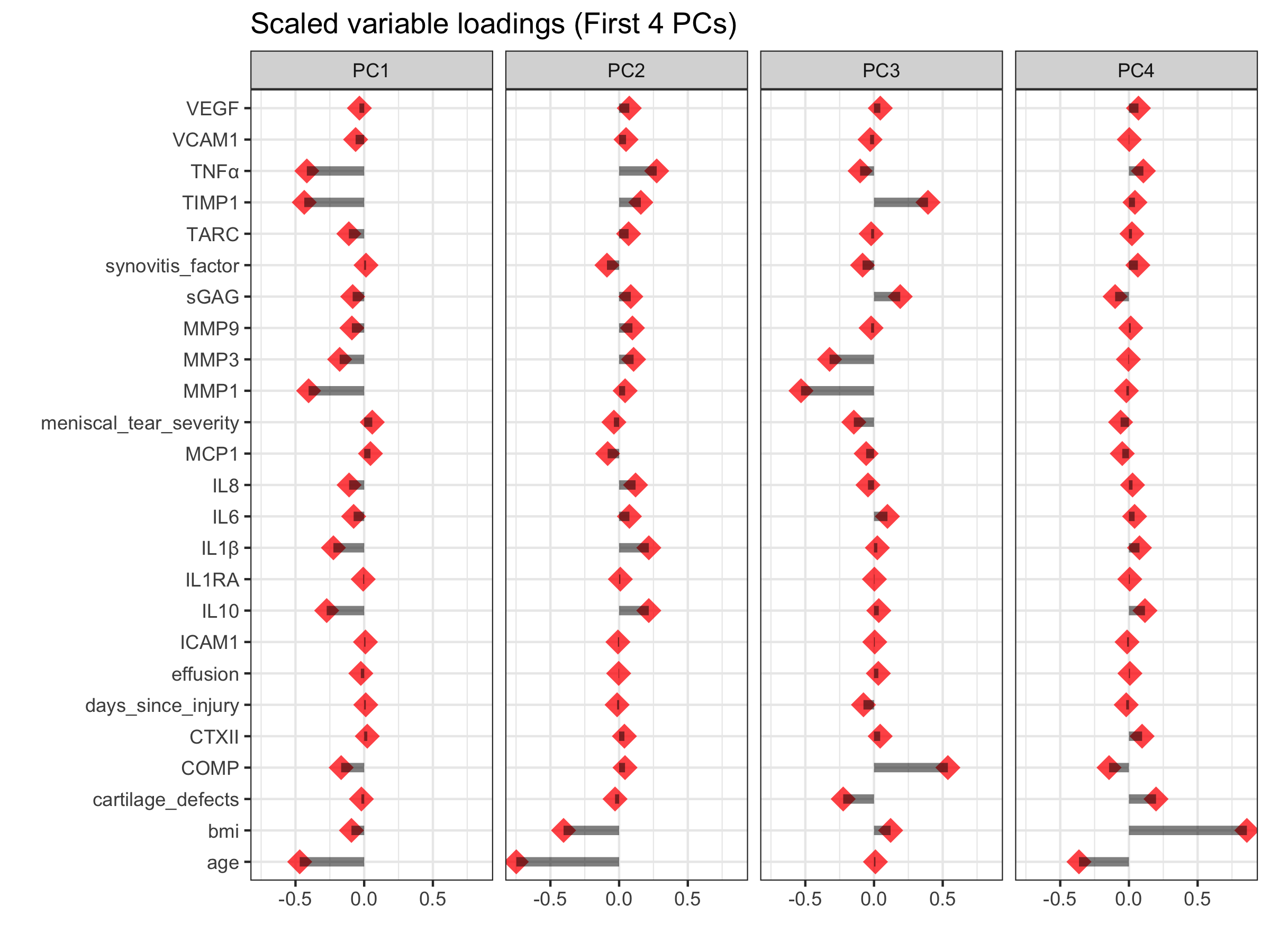


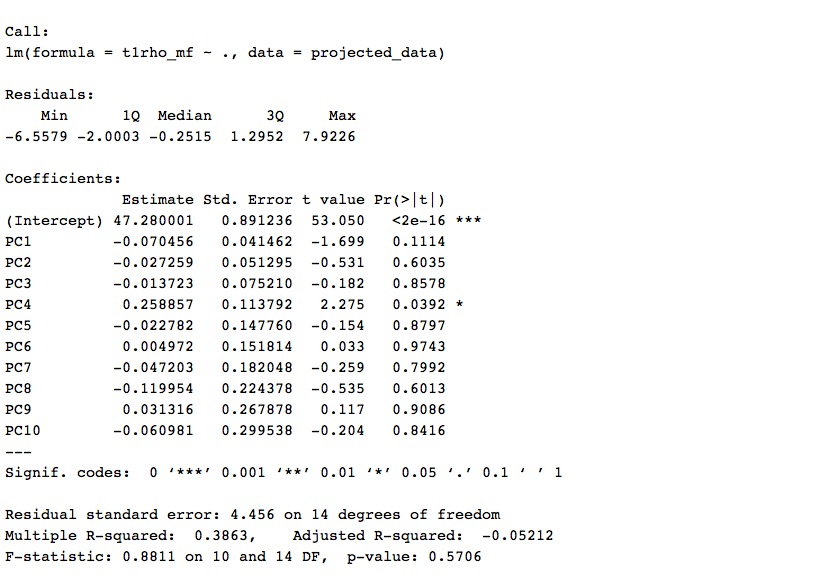
Figure 5. variable loadings plot of first five PCs from PCA



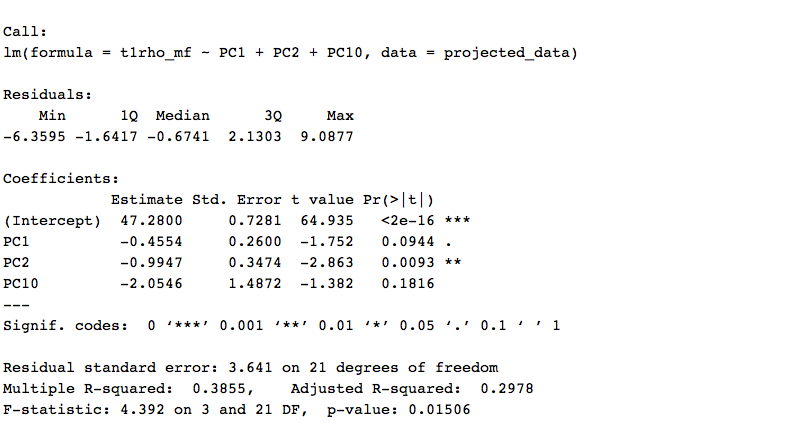
* The first PC aligns with MMP1. It separates participants with higher values of pro-inflammatory analytes and joint tissue analytes and those with lower values of those. There is a positive correlation between age and these analytes. (Note that these values are negatively projected on the first PC. For example, higher MMP1 is shown as lower PC1 loadings)
* The second principal component separates participants who are young and fit from the ones who are older and less fit.
* The third PC further separates within joint tissue metabolism analytes – COMP is positively correlated to the TIMP1 whereas they are negatively correlated to MMP1 and MMP3.
* PC4 separates participants who are younger with higher BMI from the ones older with lower BMI.
* Pairwise biplot of these PC loadings are in Appendix Figure 1.

4.4 Preliminary analysis – predicting T1rho with PCs as predictor variables

We fitted a multiple linear regression model using the T1rho measurement of medial femur in one-year follow up as outcome variable and first ten PCs as predictor variables. The details of summary of the model is shown below. The average value of predictor variables (recall that they are centered in the preprocessing step) were significant in predicting the T1rho measurement in MF in one-year follow up (p < 0.001). Only PC4, which separates younger and less fit subjects from older and more fit subjects was significant (p=0.0392)

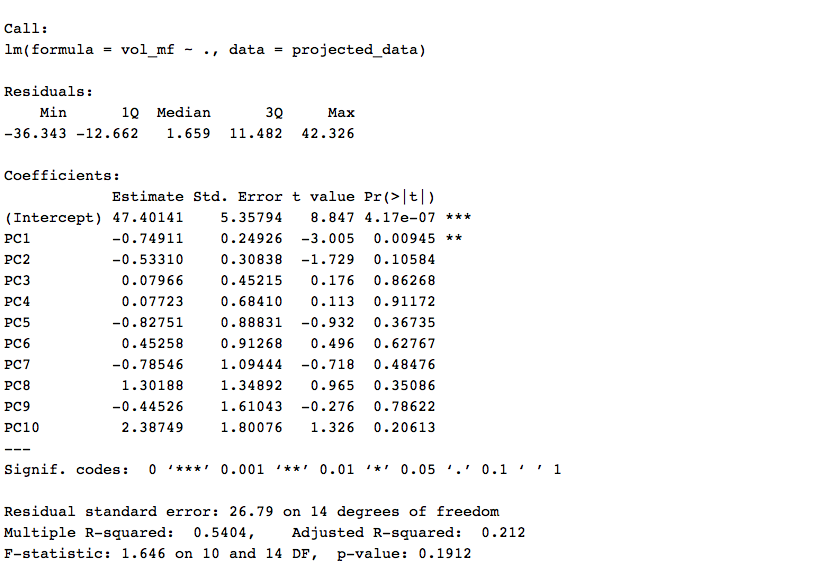


The final model is chosen by backward selection with AIC criteria. PC2 was negatively related to the T1rho measurement in global medial femur (t = -2.863, p = 0.0093)

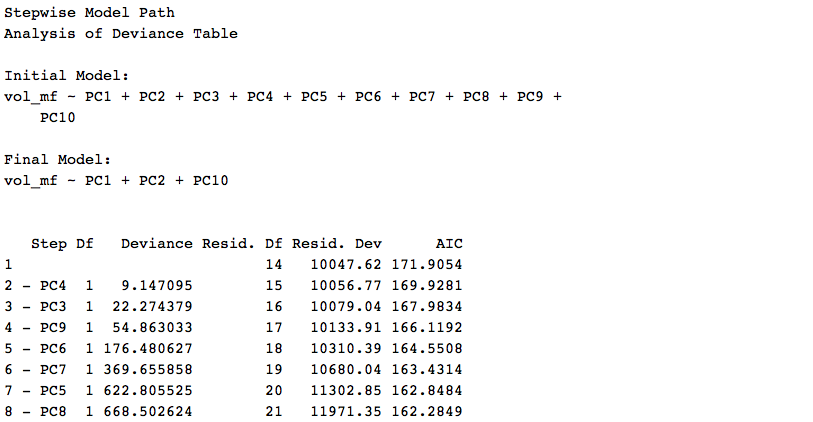


4.5. Primary analysis

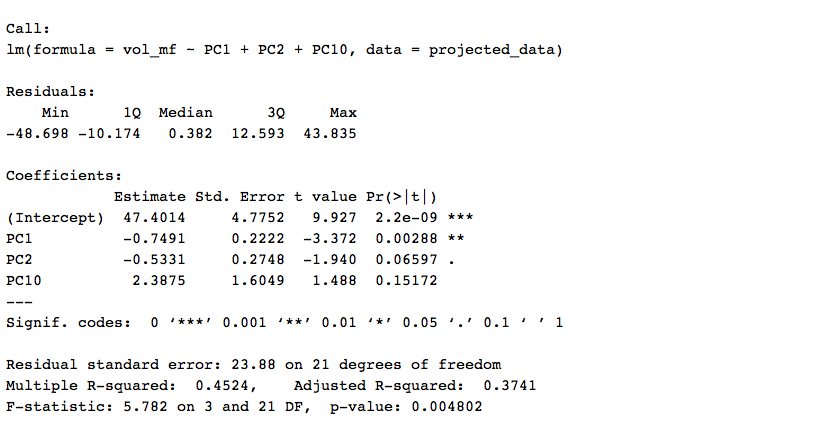
We fitted a multiple linear regression model using the volume of clusters in medial femur identified with VBR analysis as outcome variable and first ten principal components as predictor variables. Below shows the summary of the full model when first ten PCs were included. The average values of PCs were significant in predicting the average value of cluster volume in MF (p < 0.0001). The first PC – lower values of pro-inflammatory, joint tissue metabolism analytes, and younger age - were negatively related to the cluster volume in medial femur (t = -3.005, p = 0.0095). However, the F-test on the coefficients of the model failed to reject the null hypothesis (F=1.646, DF = (10, 14), p = 0.1912)



The final (parsimonious) model was chosen by AIC criteria in stepwise backward algorithm starting with 10 PCs. The summary of analysis of deviance table and final linear model are shown below.



The final model had a R2 of 45.24% (37.41% when adjusted) and PC1, PC2, and PC10 were included. (Interestingly, the same PCs were retained as the final model of T1rho measurements in global medial femur); volume of MF = 47.40 - 0.75 \* PC1 – 0.53 \* PC2 + 2.38 \* PC10. Details from the final model fit is shown below.



The diagnostic plot of the final model is shown in Figure 6.

The estimated coefficients were re-projected to the original predictor variable space. Details are shown below. The presence of meniscal tear at time of baseline, age, and BMI were positively related to the VBR cluster volumes in medial femur. The negative coefficients of IL10, though the confidence interval is wide, is a departure from other pro-inflammatory analytes (IL1Beta, IL6, IL8, TNF-alpha). Both chemokines, MCP1 and TARC, were negatively related to the cluster volume in MF. Among the joint tissue metabolism analytes, MMPs and TIMP were positively related to the volume; sGAG was very weakly negatively related to the volume; the Cis for coefficients of CTXII and COMP included zero.

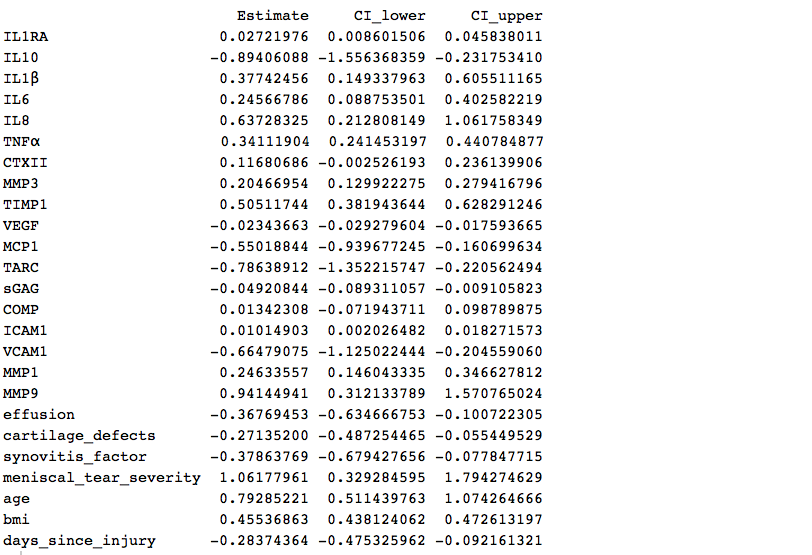
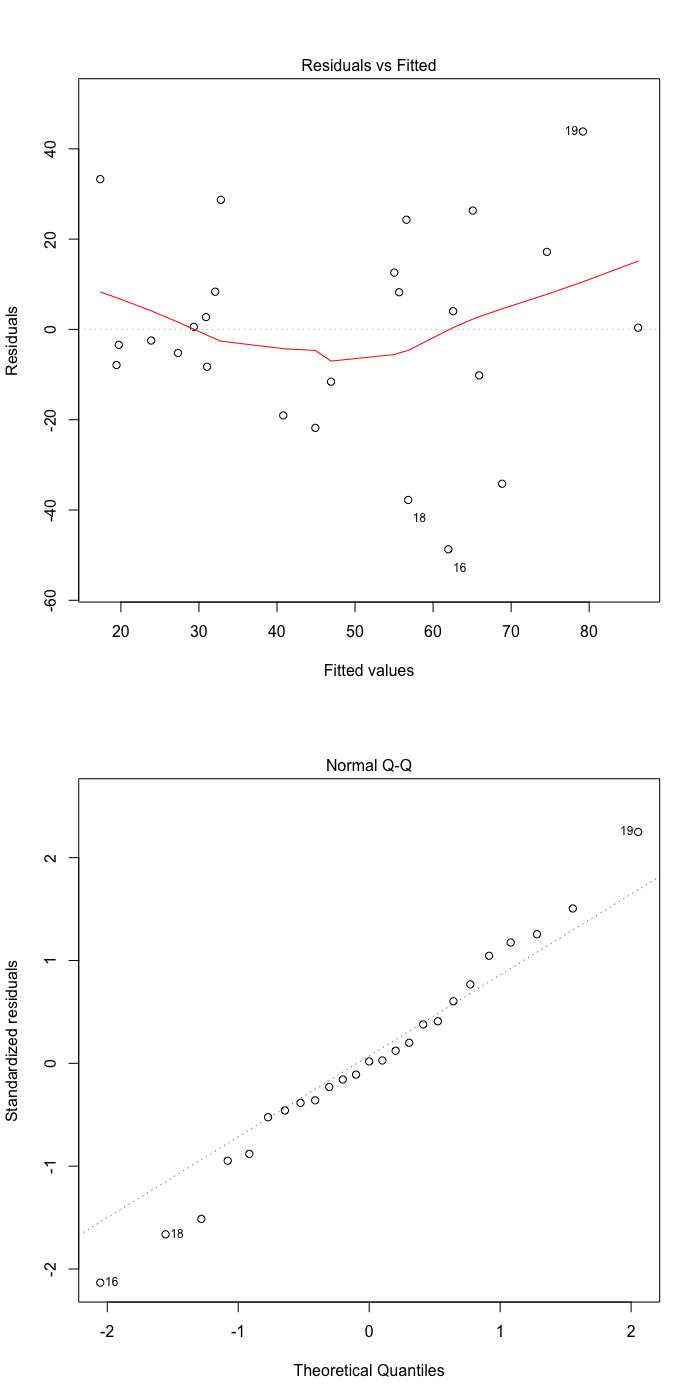
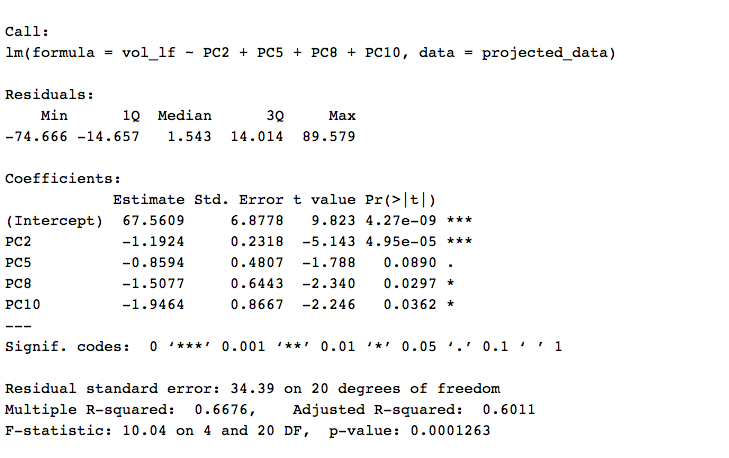


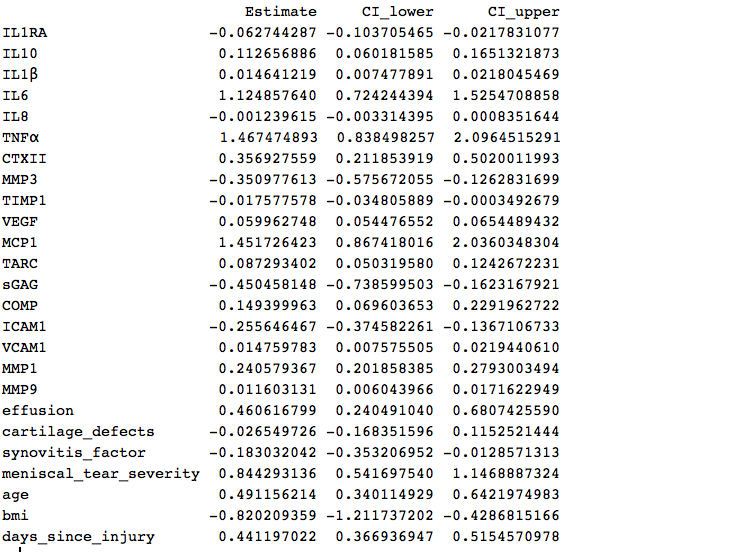
Figure 6. Diagnostic plot of final model



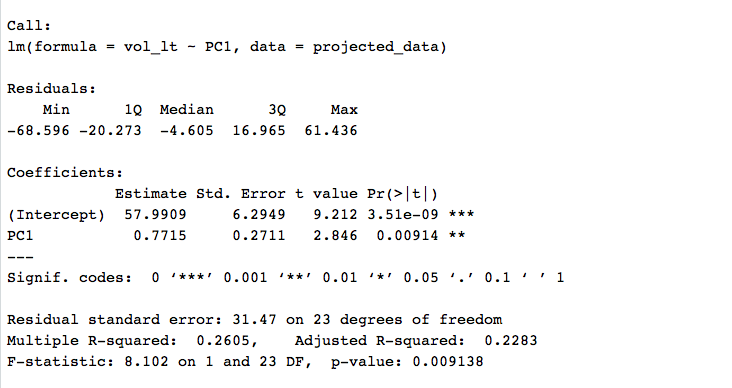
4.6. Exploratory analysis

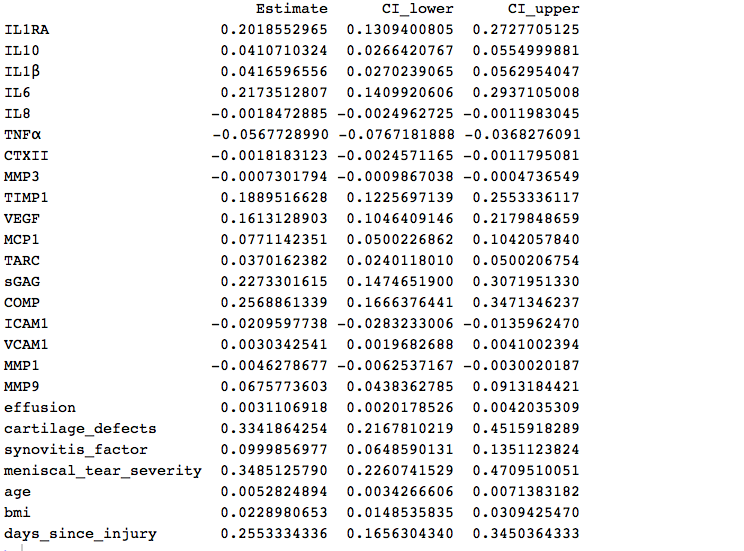
4.6.1. Predicting VBR cluster volumes in LF



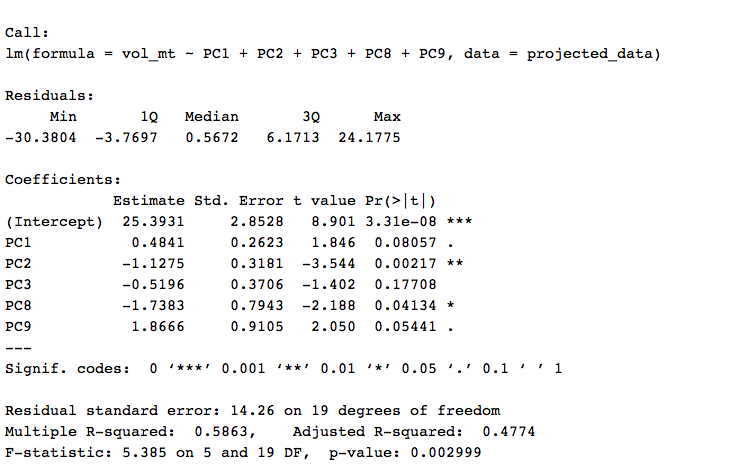


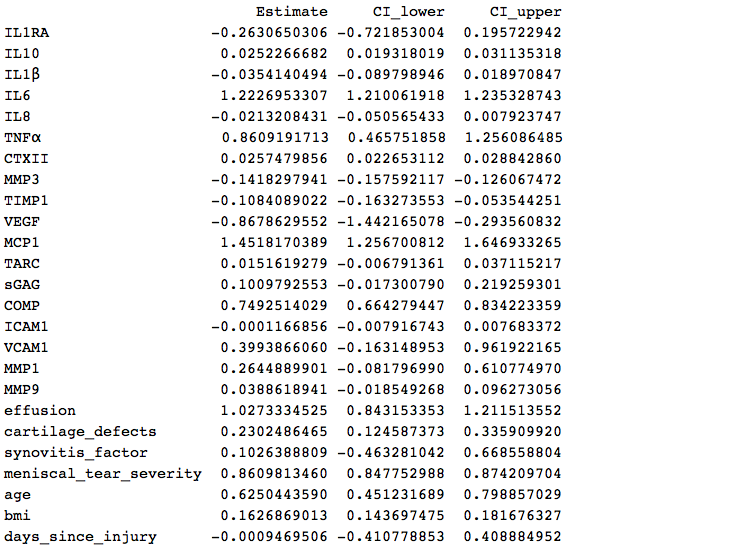
4.6.2. Predicting VBR cluster volumes in LT



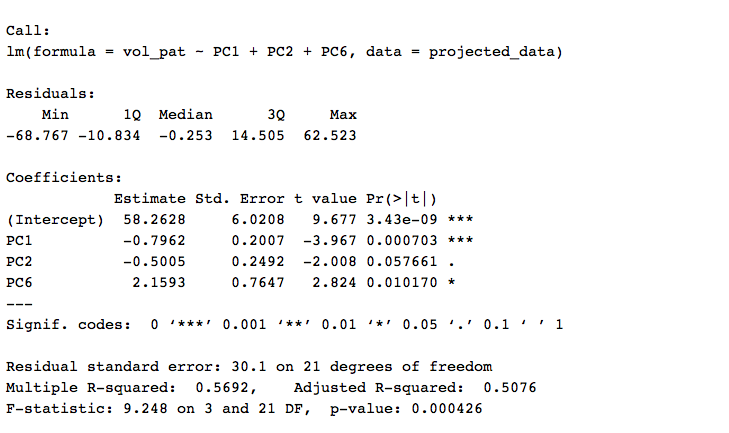


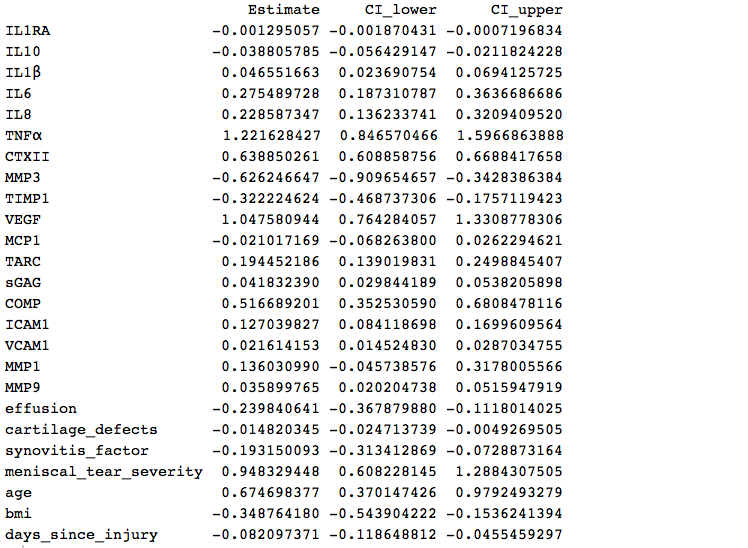
4.6.3. Predicting VBR cluster volumes in MT



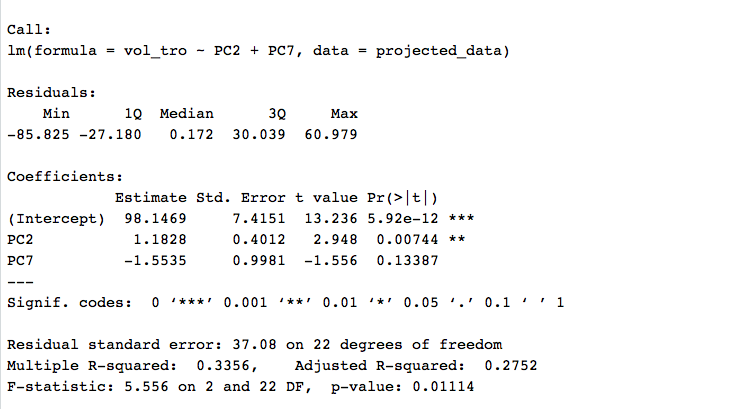


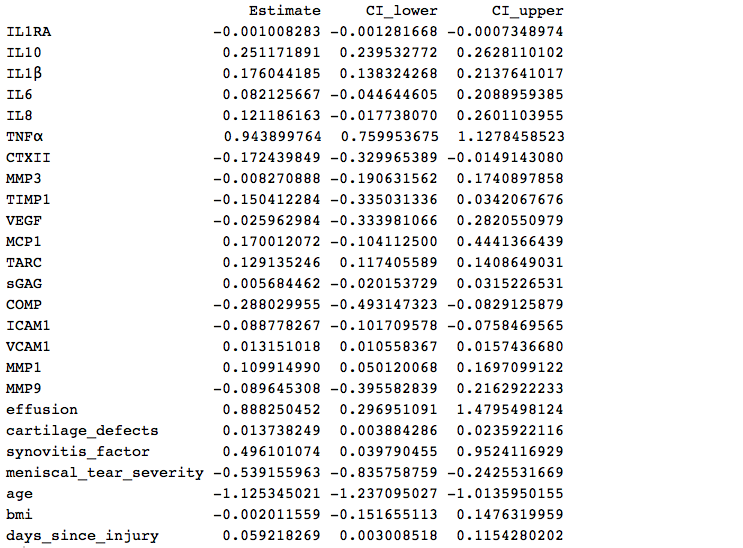
4.6.4. Predicting VBR cluster volumes in PAT





4.6.5. Predicting VBR cluster volumes in TRO





1. Appendix

Figure 1. Pairwise biplot of PCA

