After performing expression mapping the RNA samples to human genome version hg19 using *HISAT 2.1* and *featureCounts* 2.0.1 (features were combined using *Multi-Join* 1.1.1) at *Galaxy Server 2.10.8*, data were downloaded and loaded at *R* 4.0.3 for analyses using *Bioconductor* 2.5 and *DESeq2* 1.3. The counts data were first filtered to exclude genes with counts smaller than 10 when considering all six samples. The filtered data were then plotted in *R* using the *graphics* 4.0.3 package function *boxplot* (Fig. 1). As the raw counts were mainly near zero values, this generated some extreme outliers, especially for the fetal samples. To attempt to normalize the data, the raw counts values were transformed by *log2* conversion and by variance stabilization (performed using the function *varianceStabilizingTransformation* of the *DESeq2* package, which accounts for sample size. Although *log2* transforming generate fewer outliers, variance stabilization seemed to perform better, as this transformation produced outliers for all samples, which suggests that there are some genes that have a higher expression in all samples.

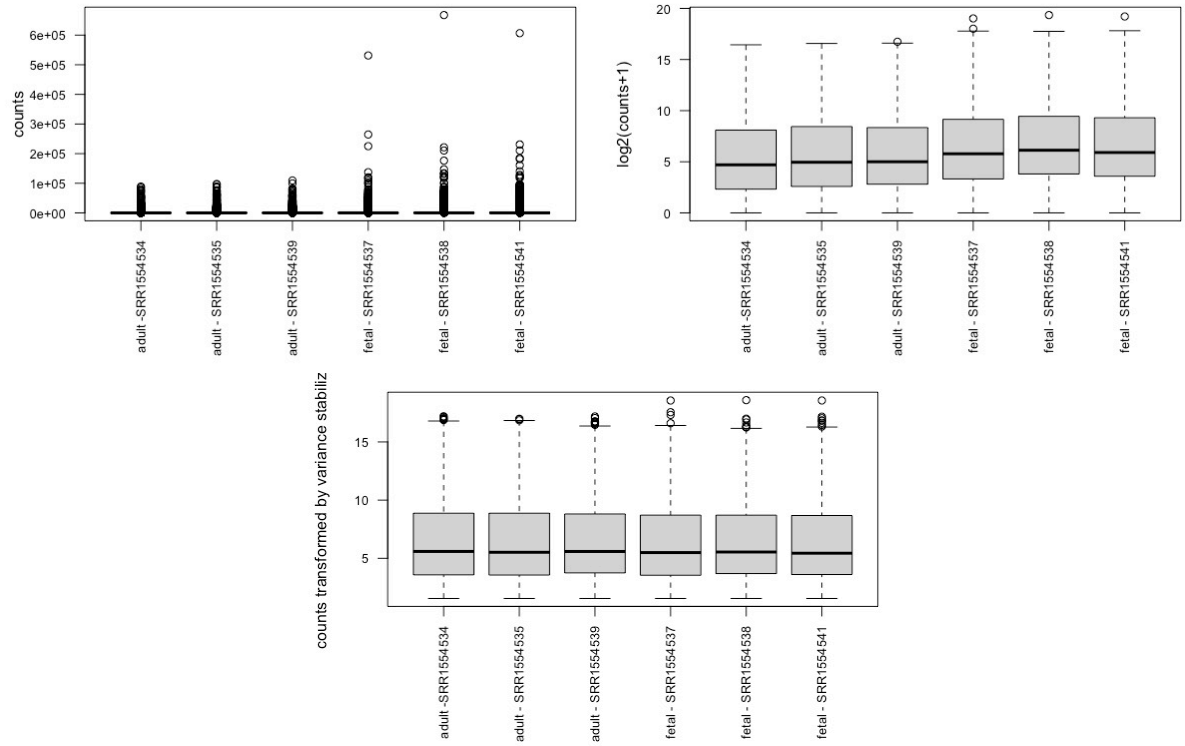

Fig. 1. Boxplots of the count data for samples. Data were filtered to exclude genes with expression count smaller than 10. Sample groups are indicated together with samples accession numbers. At top left: raw counts by samples. At top right: counts transformed by logarithimization. At bottom, center: counts transformed by variance stabilization.

As another step of the exploratory analysis, the counts data transformed by *log2* conversion and variance stabilization were used for further calculations using a principal components analysis (PCA) considering the sample age (fetal or adult) as groups of interest. There were virtually no differences in the PCA analysis of both transformation techniques when using the functions *prcomp* and *princomp*, with the use of the correlation parameter in the *princomp* function just scaling the data differently in relation to *prcomp*. Considering this, only the results for the *log2* transformation are presented (Fig. 2). The PC1 explained about 89% of the variance, and the PC2 about 5%. Age group in relation to PC2 easily differentiated the samples, with the adult samples presenting positive values and the fetal samples negatives values for the PC2. RNA integrity (RIN) and sex do not allowed a clear grouping of the samples when analyzed in relation to PC1 and PC2

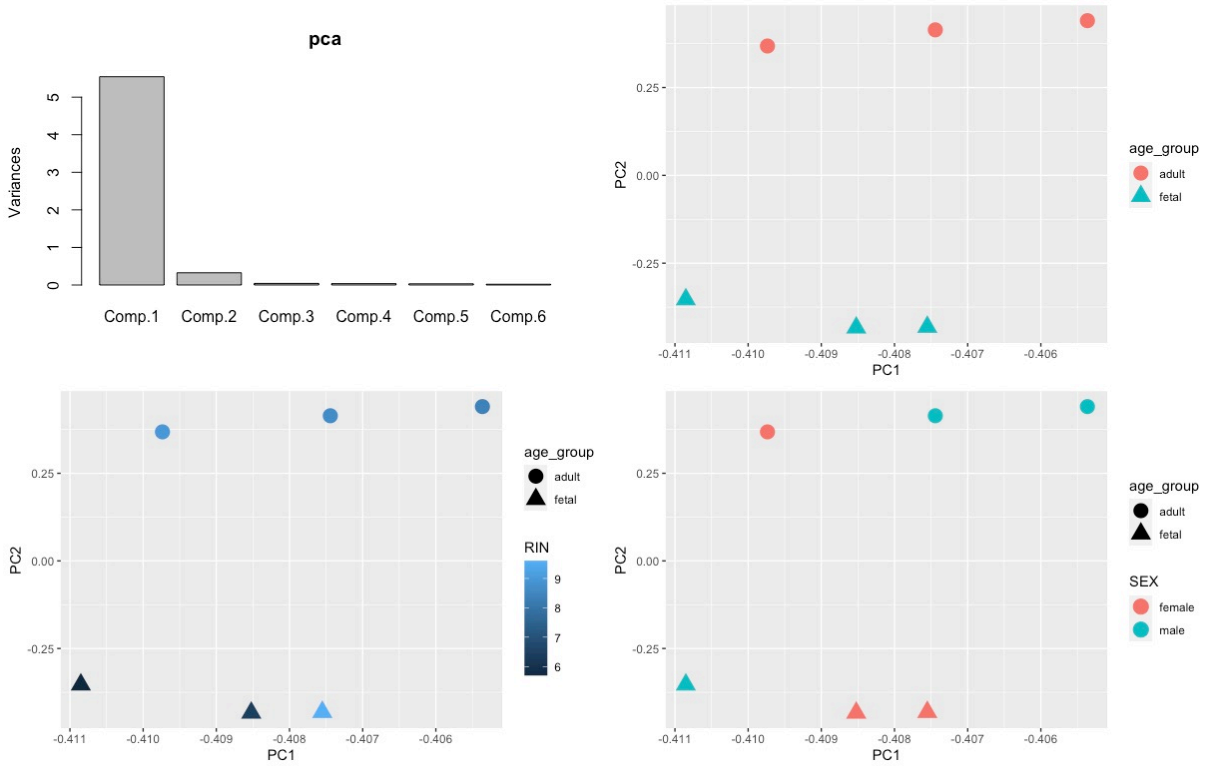

Fig. 2. Principal component analysis of counts by gene after filtering and log2 transformation. At top left: variance explained by each of the principal components recovered by the analysis. At top right: gene expression at fetal and adult samples in relation to PC1 and PC2. At bottom left: gene expression and RNA integrity (RIN) in relation to PC1 and PC2. At bottom right: gene expression and sex in relation to PC1 and PC2.