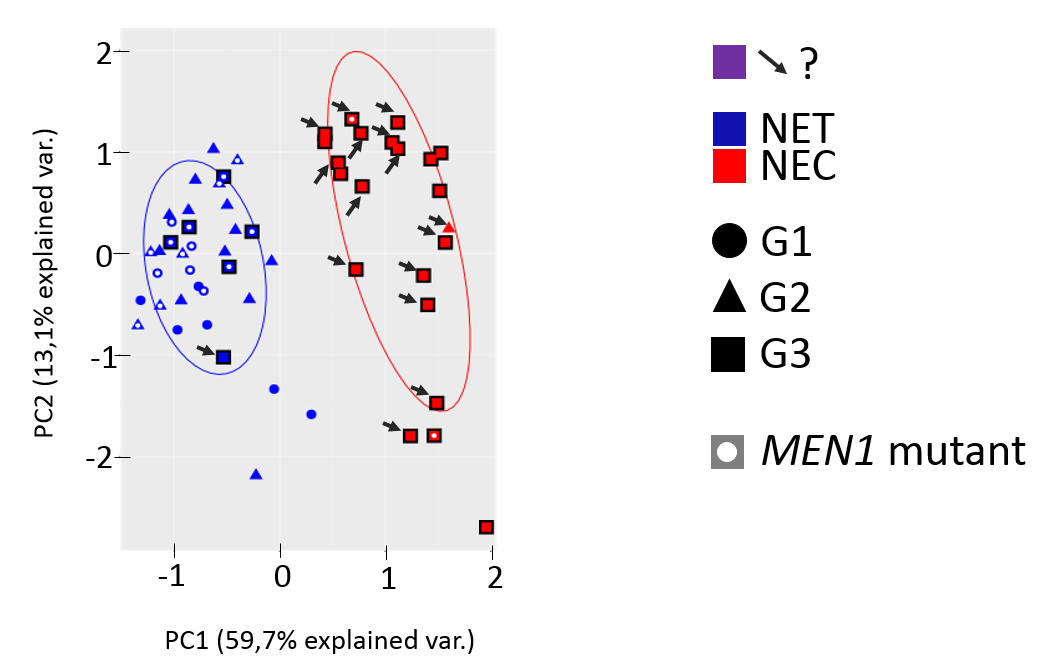
1. **PCA using (available) Sadanandam Genes – Discovery cohort**

* For the MASTER cohort (= validation cohort), Tincy could pick up 206 of the previously 207 used Sadanandam Genes; PRSS2 is missing => to keep it homogeneous, we should redo PCA with 206 genes for the Discovery cohort (=Scarpa + Grötzinger)
* We would like to identify the samples for which the pathologists disagree => colour the “ambiguous” samples purple instead of either red or blue. Which samples should be coloured red/blue/purple is given in the Excel file “Metadaten RNAseq Kollektiv\_ pathologisch morphologische Einordnung-1.xlsx”
* No need to include a legend, I will add it later when putting together the figures (font will probably have to be adapted, I suppose)

Aimed plot should look like this, but based on 206 instead of 207 genes, and all samples with and arrow should be purple:



1. **PCA using (available 206) Sadanandam Genes – Validation cohort**

* A PCA exactly like the one described above using the MASTER samples (using 206 Sadanandam genes); here, we only have the initial evaluation of the tumor, so it’s either red or blue according to the table; data are in Excel file “Heidelberg\_ gep\_nen\_data\_GEP\_NEN only\_.xlsx”

Grade: sheet “clinical GI only” column B

NEC/NET for G3: sheet “clinical GI only” column P

MEN1 mutations: sheet “snvs\_indels” columns G and H

1. **Do a PCA (206 Sadanandam genes) with the combined Discovery and Validation cohorts**

* This is to see how both cohorts behave with respect to each other
* To be able to tell apart if a sample came from the Discovery or the Validation cohort without losing any information (grade, MEN1 mutation), maybe use a lighter shade of blue and red for the Validation samples (or other suggestions?)