

SUPPLEMENTAL METHODS

Creating molecular scores for assigning continuous variables to phenotypes to reflect disease severity and detect changes in response to therapy.

A simple classifier validated for diagnosing symptomatic V30M FAP patients vs. asymptomatic V30M carriers does not reflect disease severity or change in response to therapy. Serial monitoring requires translation of the classifier's output to a continuous variable or molecular score based on the underlying gene expression derived from the classifier. Therefore, we developed a method for calculating molecular scores. Therefore we created molecular scores using the raw signal intensities from probesets that comprise a classifier. Simple addition of the classifier probeset signals would not reflect the score because signals from the upregulated genes and the signals from the downregulated genes would essentially cancel out each other so a more composite strategy needed to be used. So we separated the classifier into the upregulated and the downregulated genes and calculated the mean signal intensities. We then created the scores using the formula given below:

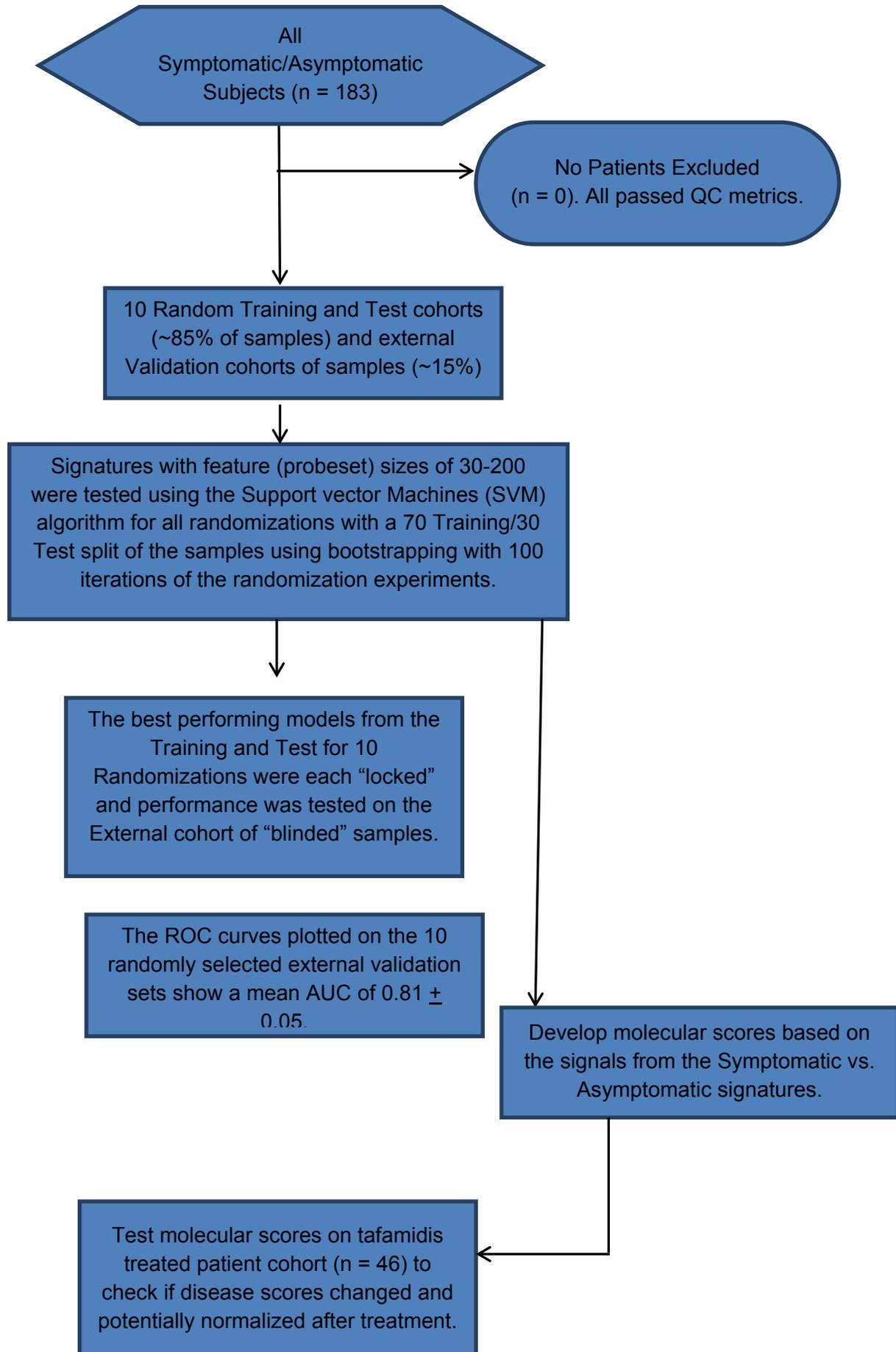
Raw Molecular Score = Mean signal intensities of upregulated probeset – Mean signal intensities of downregulated probeset, for each sample. To make the scoring system simpler and to remove confusion caused by negative and positive numbers we used a formula to scale the scores to a range of 1-100. The formula is as shown below:

$$\text{Scaled Score} = (\text{Raw Score} - (\text{MIN})) * (100 - 0) / (\text{MAX} - (\text{MIN})) + 0$$

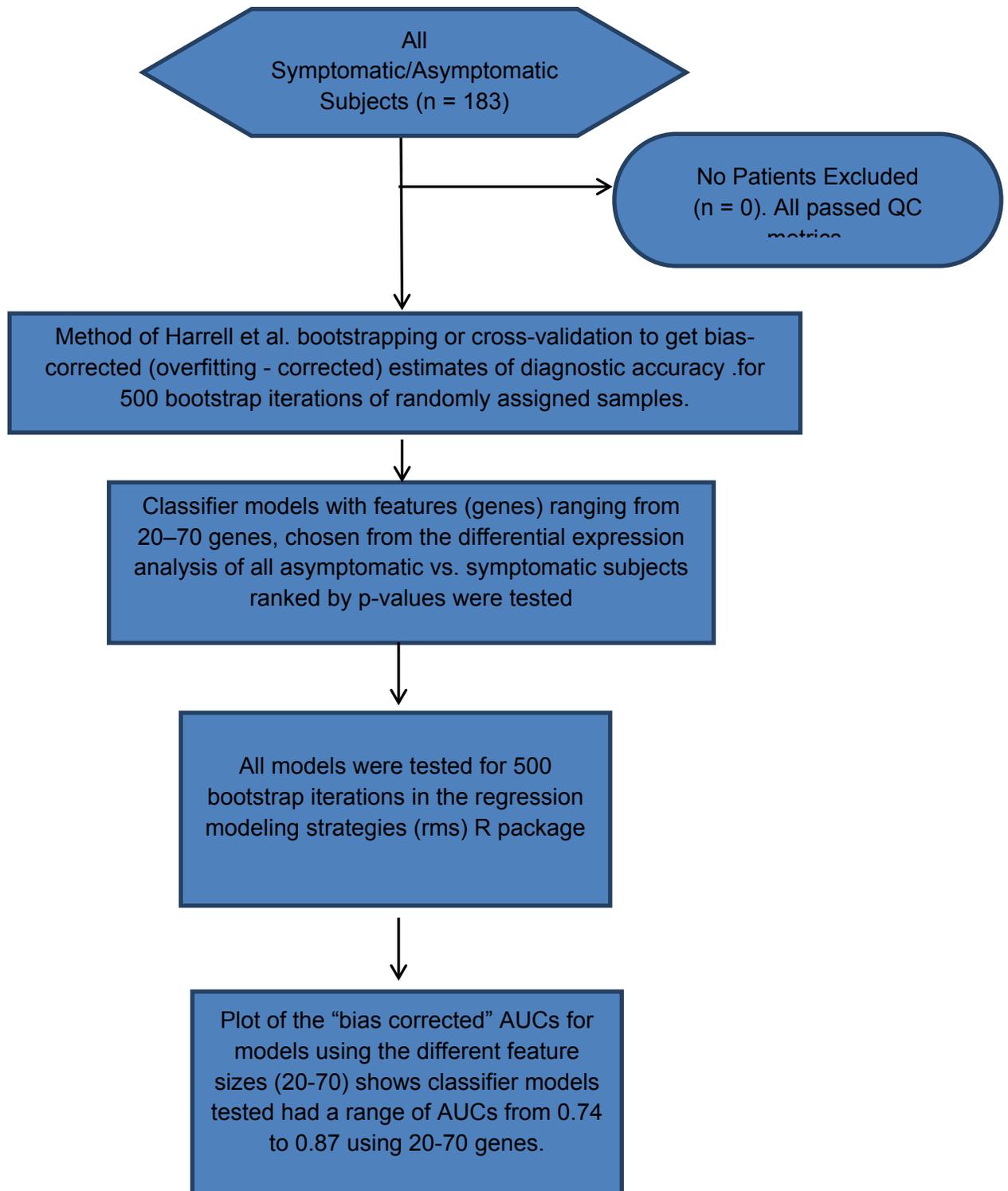
Where MIN = the minimum value of the range of scores and MAX = maximum value of the range of scores for the given classifier.

This adjusted value was then assigned as the molecular score for that given sample. These molecular scores correlated well with the dichotomous classification derived from the Support Vector Machine (SVM) algorithm.

Flowchart describing study design and diagnostic biomarker pipeline (Method 1)



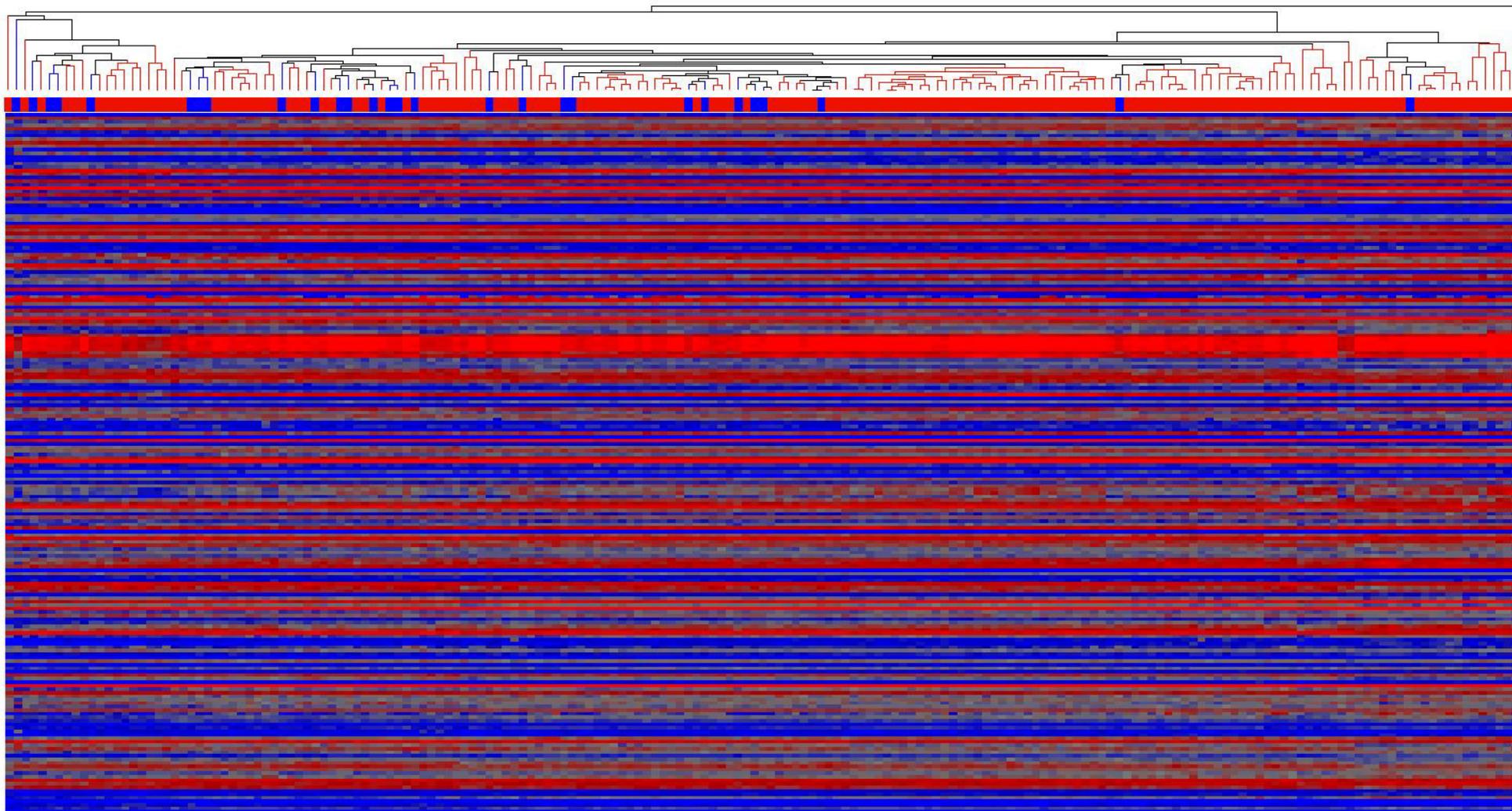
Flowchart describing study design and diagnostic biomarker pipeline (Method 2)



Clustering of FAP carriers based on Pain Medication:

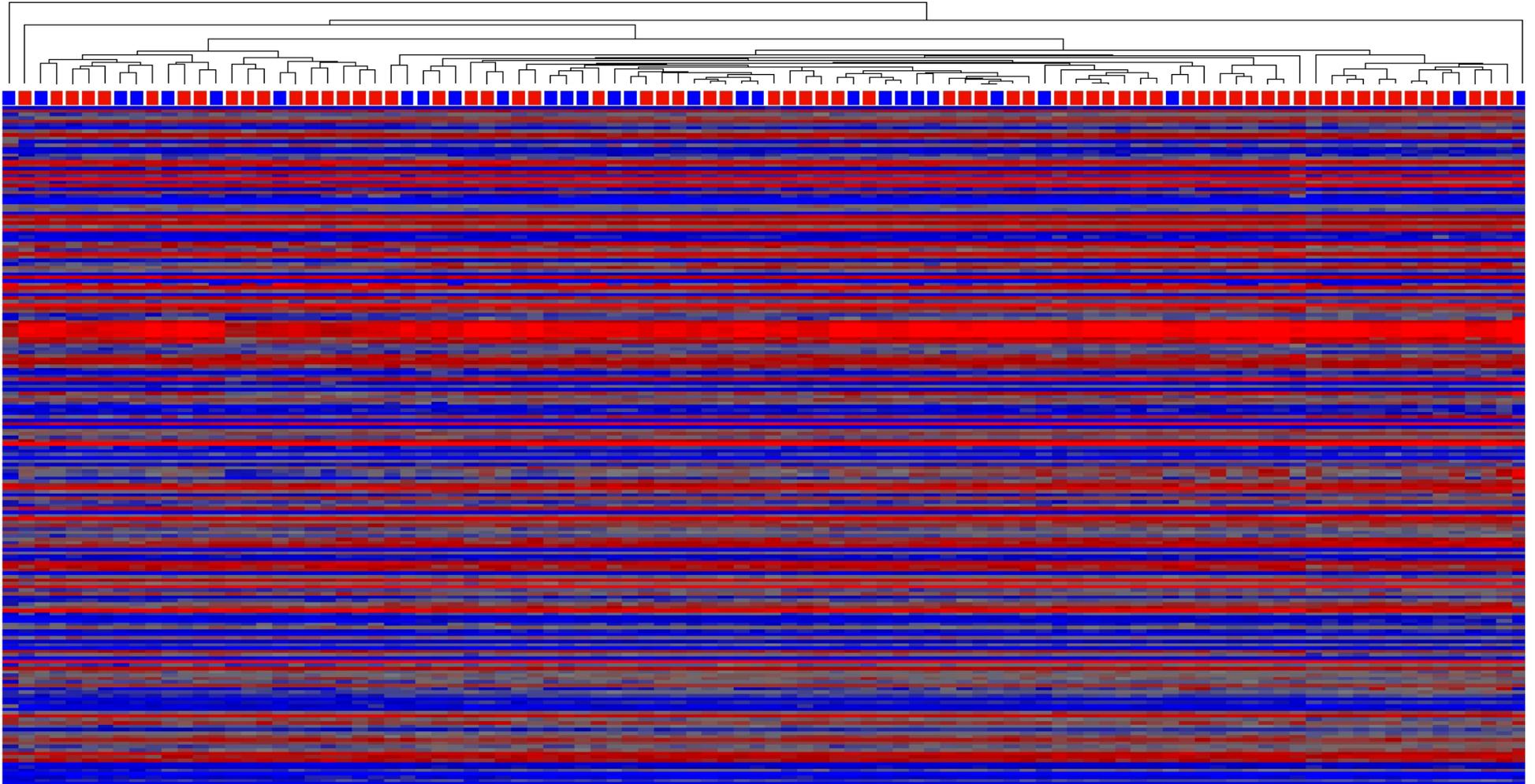
FAP patients were on pain medication (almost exclusively pregabalin or gabapentin) and none in the asymptomatic group were on these medications. We further explored this by asking the question whether our signatures are different between the symptomatic patients on medication and the ones that were not on any pain medication. We took the top 200 genes from our symptomatic vs asymptomatic signature and created a heatmap based on average linkage clustering with Euclidian distance of only the FAP subjects ($n = 183$). The results clearly show that the medication status did not distinguish the symptomatic patients by cluster both among Symptomatic vs. Asymptomatic FAP carriers (Figure 1) as well as among the Symptomatic FAP carriers (Figure 2). Furthermore of the 200 genes none were significantly differentially expressed ($FDR < 10\%$; lowest p -value = 0.02) between the symptomatic patients on pain medications vs. those who were not on any pain medication. This suggests that the pain medications did not influence the gene expression signatures.

Figure 1: Average Linkage clustering of 183 symptomatic and asymptomatic FAP carriers by Euclidian distance based on the pain medication clinical variable



The color of the cells indicates the spectrum of signal intensities from high (red) to low (blue) intensity genes. The dendrogram color indicates patients on pain medication (blue) and those who are not on pain medication (red)

Figure 1: Average Linkage clustering of 96 symptomatic FAP carriers by Euclidian distance based on the pain medication clinical variable



The color of the cells indicates the spectrum of signal intensities from high (red) to low (blue) intensity genes. The dendrogram color indicates patients on pain medication (blue) and those who are not on pain medication (red)