**Meeting 27-02-2018 Dragan, Zandra, Tim**

Discussed the results of the statistical analysis of the skin disease data set.

**Feature dimensionality**

* The greedy clustering was done as a way of reducing features
* A decision tree with also normal skin would be interesting
* The score was a validation score and should be made a testing score.

**Clustering genes**

* The scaling may have been done for all gene values together and must be checked for genes separately
* Clustering for pathways is possibly something that is interesting in this case too.

**Process/Cellular/Molecular relation**

* Define clearer what the three types of relations are (process, cellular, molecular)
* Check for the names of the genes that are used (may be database specific)
* Try to link it with other databases (GeneCards, KEGG)

**Overall**

* Try to better define all different aspects that are used. That way words as “relations”, “processes”, “significant” become clearer for the one that watches the presentation. Using figures for unknown processes and highlighting interesting parts would be beneficial for that.