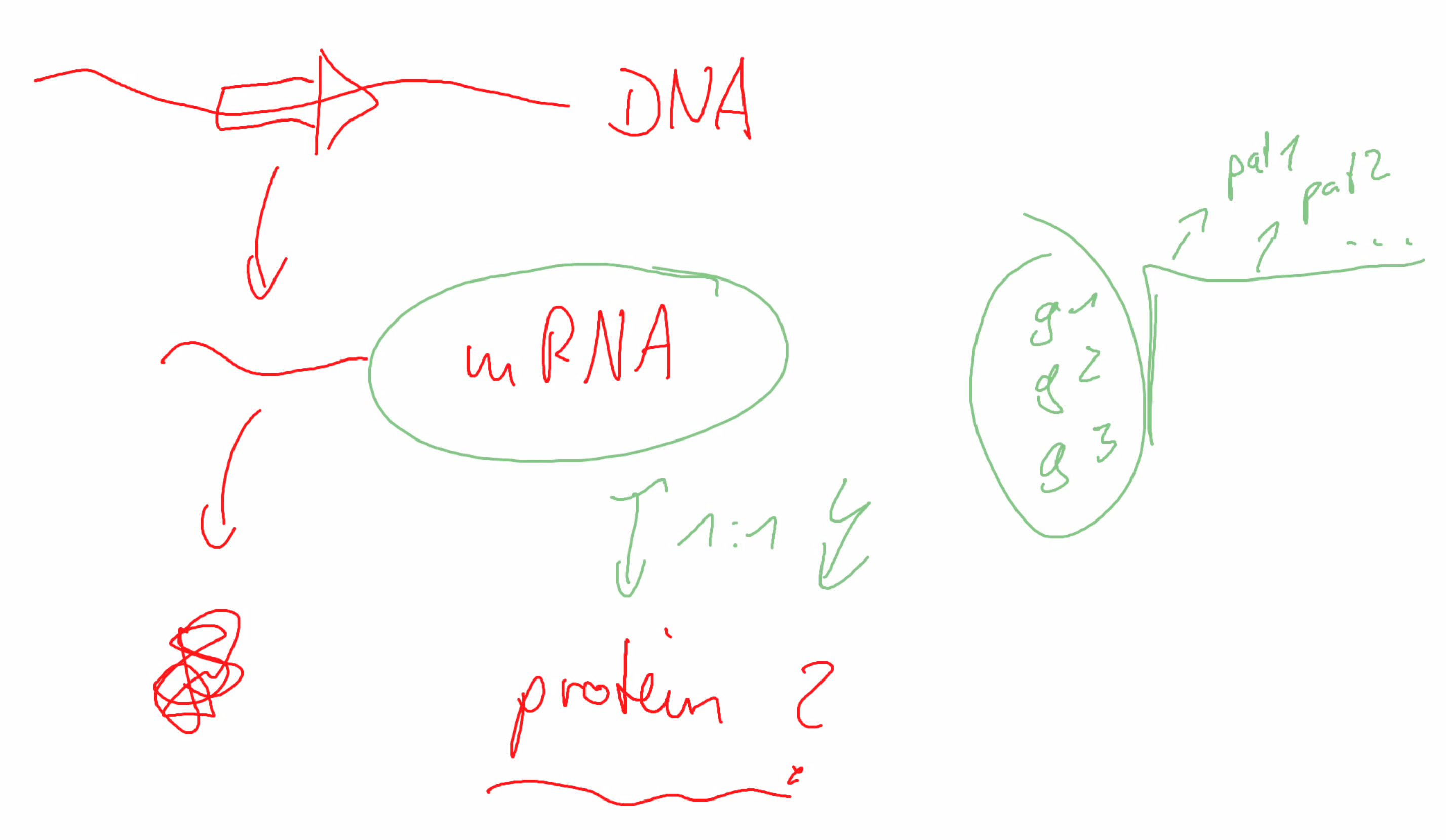
DNA->mRNA->protein



Relevancy of gene types.

Filter out males

PAM50 microarrays

# Initially



Filter data by outliers and invalid.

Create project description by what we want to do (Deadline first week of February).

Create plan and thoughts on what to do.

Read material given by Richard.

# First part

RNAseq data

Use RNAseq to predict PAM50 label

PAM50 has 5 groups, so this is supervised.

Use interpretable machine learning (decision tree’s etc).

# Second part

Ignore PAM50 and use unsupervised.

Invent new PAM50 and designate amount of groups.

Rate PAM50 vs ours.

Does it make sense to switch from PAM50?

Survival years biased by age and preexisting conditions.