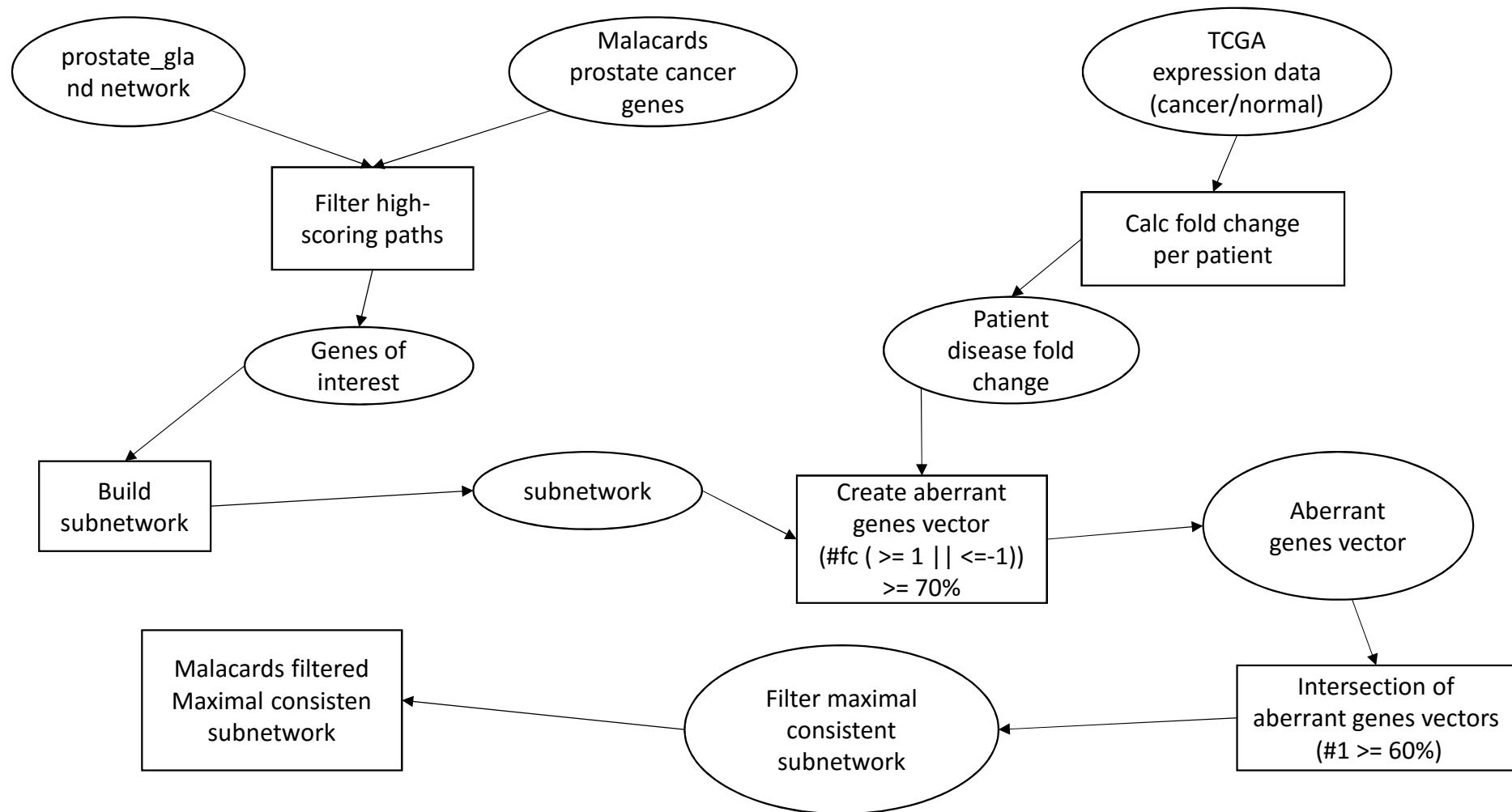# Prediciting Disease Lables Prostate Cancer

Julian Schmidt, Stefan Weber

25.07.2016

Approaches

- Malacards Filtered McSubnet Approach

- Z-Score Approach

- Accumulated Approach
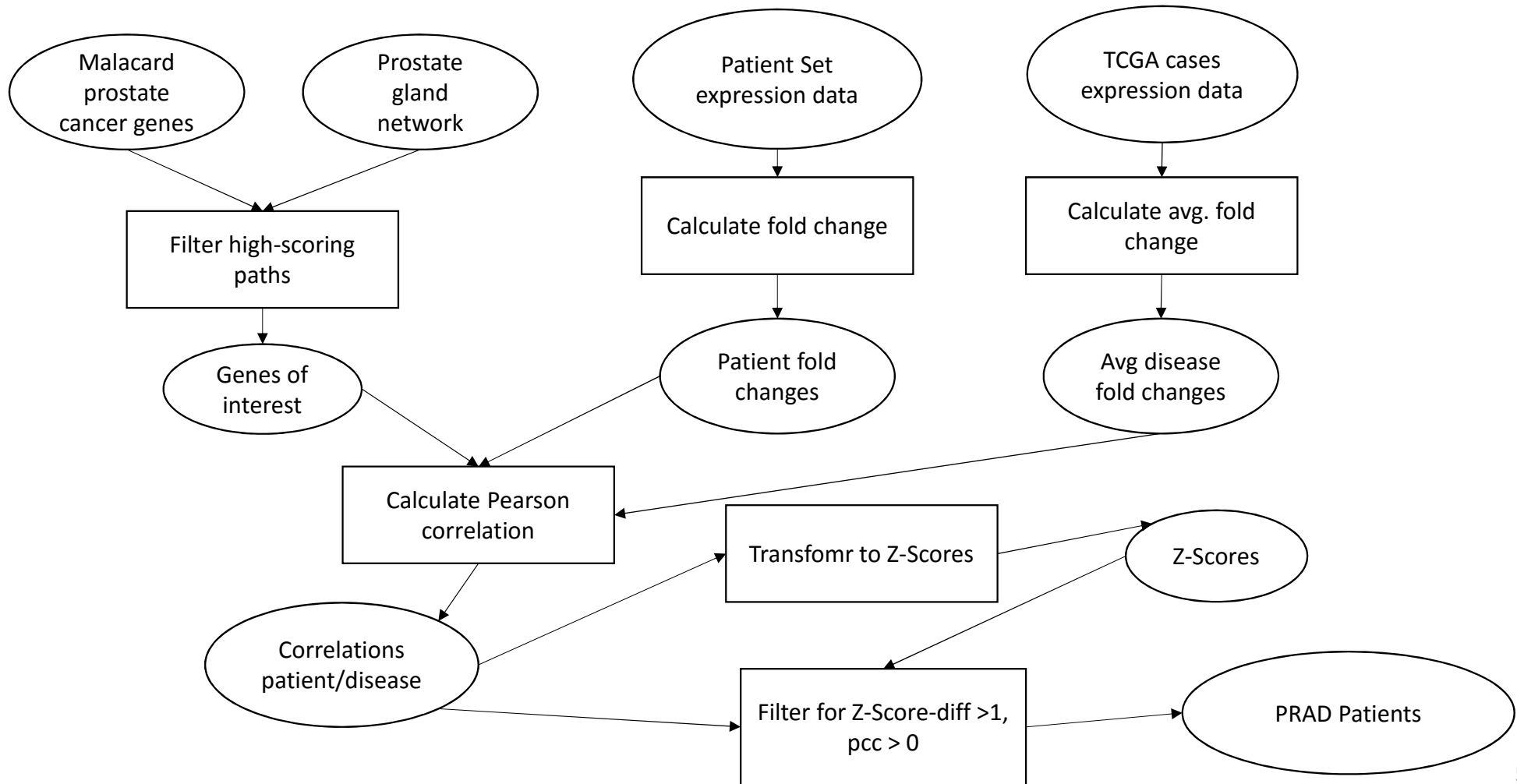
- Aberrant Neighbors Approach
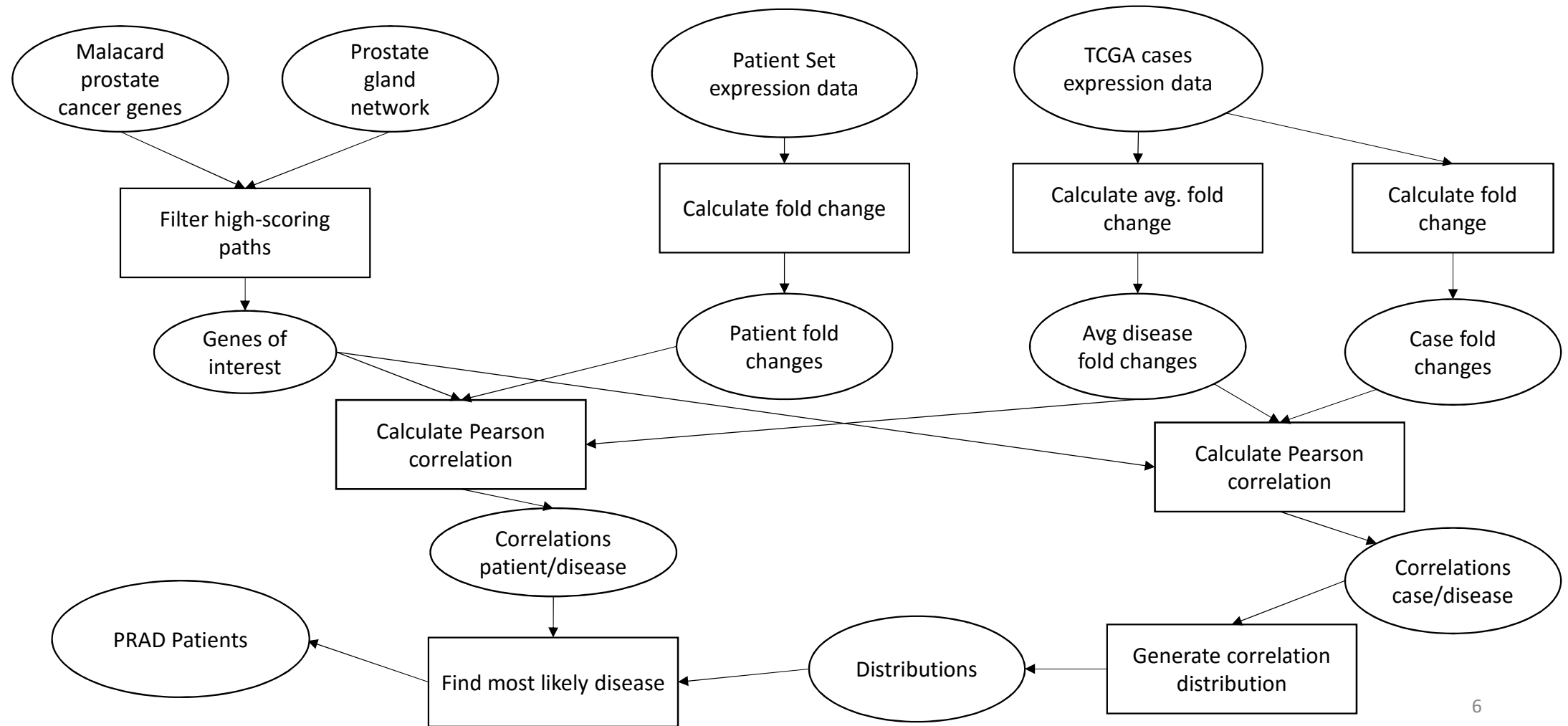
- SVM

Malacards Filtered McSubnet Approach

prostate_gland network

Malacards prostate cancer genes

TCGA expression data (cancer/normal)

Filter high-scoring paths

Calc fold change per patient

Genes of interest

Patient disease fold change

Build subnetwork

subnetwork

Create aberrant genes vector (#fc (>= 1 || <=-1)) >= 70%

Aberrant genes vector

Malacards filtered Maximal consisten subnetwork

Filter maximal consistent subnetwork

Intersection of aberrant genes vectors (#1 >= 60%)
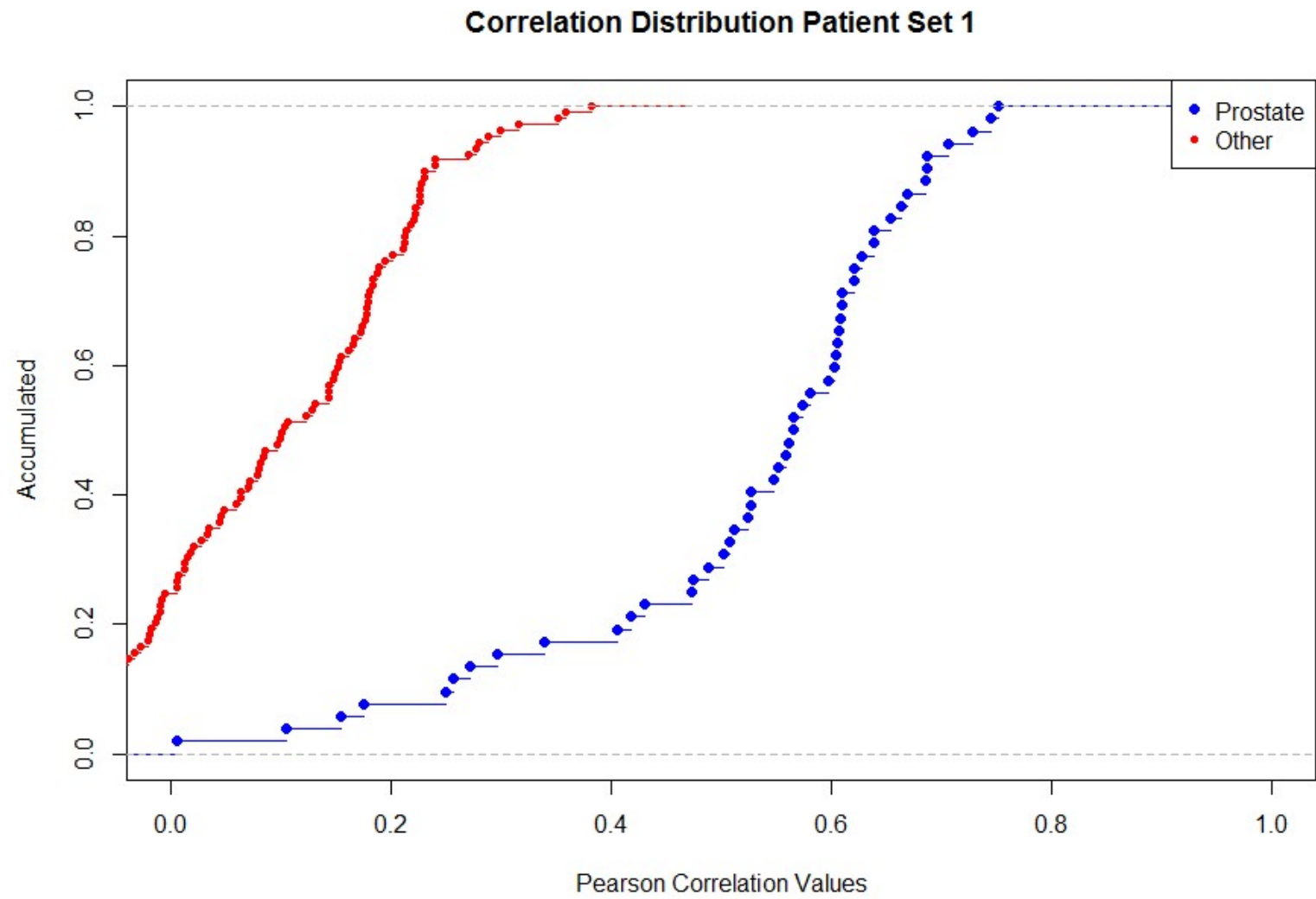
3

Malacards Filtered McSubnet Approach

- Limitations:
  - Subnetworks way to small to predict anything properly (max 5 genes)

Z-Score Approach
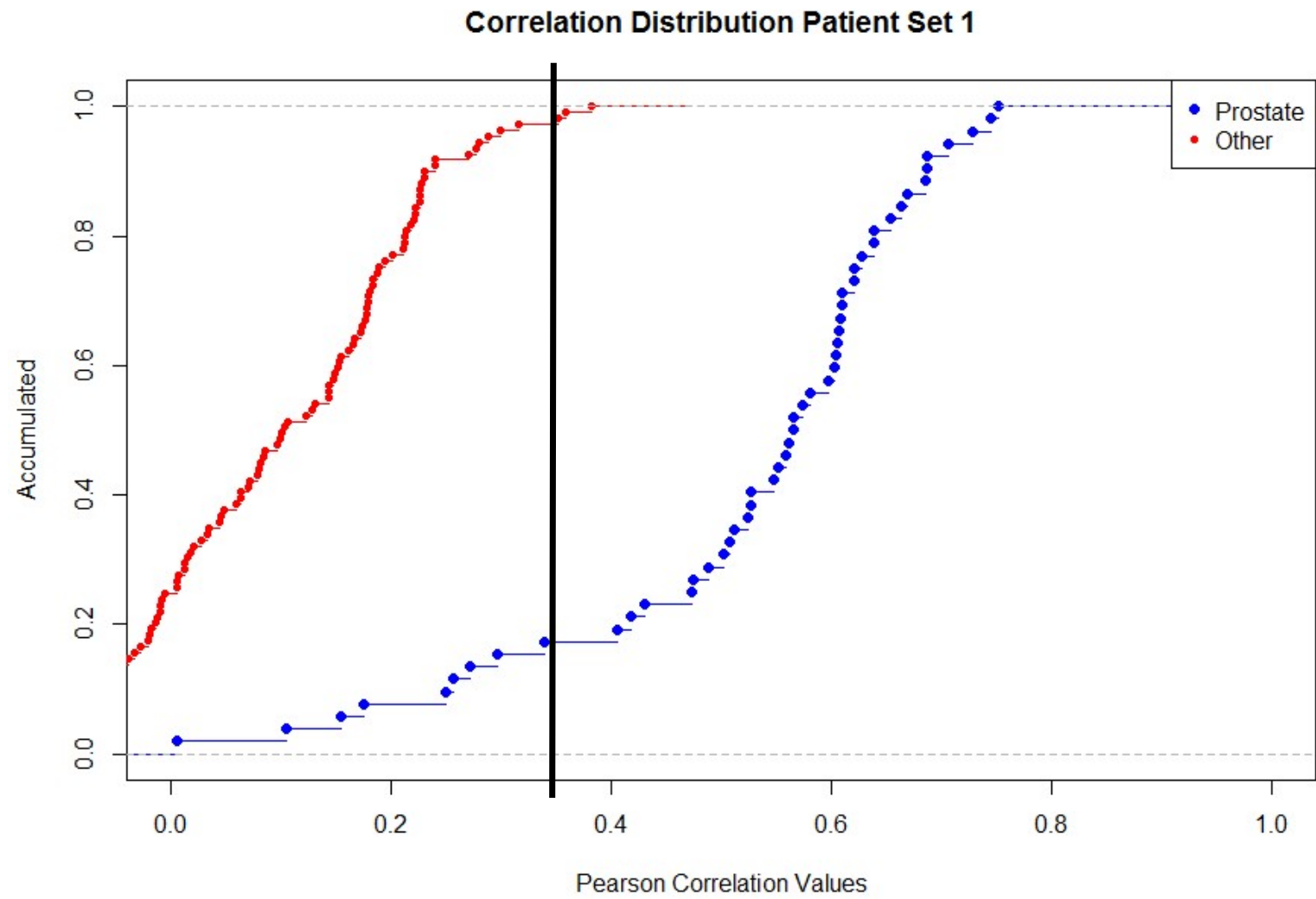

5

Accumulated Approach

Correlation Distribution Patient Set 1

Correlation Distribution Patient Set 1

Prostate
Other

Accumulated

Pearson Correlation Values

Validation PatientSet1

Approach	TP	TN	FN	FP
Z-Score	9	78	1	12
Accumulated	8	88	2	2

Prediction Aberrant neighbors

- Determine aberrant genes (aberrant: |log(FC)|>1)

- Score sample for average "association score" a of aberrant genes to genes of interest

- Scoring approaches:
  - Portion of genes with at least one edge to gene of interest with weight > threshold
  - $a_{n+1} = a_n + (1 - a_n) * \gamma(e)$ for $n \in \{0: \#genes\ of\ interest - 1\}, n_0 = 0$
  - $a_{n+1} = a_n + (1 - a_n) * \gamma(e)^2$ for $n \in \{0: \#genes\ of\ interest - 1\}, n_0 = 0$

Prediction SVM

Trainings data:

- 52 prostate (PRAD) cases
- 112 breast (BRCA) cases
- 58 thyroid (THCA) cases
- 49 lung (LUSC) cases
- (126 kidney (KIRC, KIRP, KICH))

Features

- Unique Prostate cancer Malacard genes
- Direct high scoring neighbors in GIANT prostate_gland network
- 117 Genes

Prediction SVM – PatientSet1

	Observed Prostate	Observed other
Predicted Prostate	10	4 (all KICH)
Predicted other	0	86

Lung, breast, thyroid as negative set

Prediction SVM – PatientSet1

	Observed Prostate	Observed other
Predicted Prostate	10	4 (all KICH)
Predicted other	0	86

Lung, breast, thyroid as negative set

	Observed Prostate	Observed other
Predicted Prostate	10	0
Predicted other	0	90

Lung, breast, thyroid, kidney as negative set

To Do

- Improve features selection
- Make MCSubnetwork Approach more efficient
  - With no malacards filter step
- Find maximal consistent subsets with bi-cliques