Discovery of 11 candidate genes: Discovery study

(Lanigan *et al*. 2015, Moran *et al*. 2017)

1) ARACNe transcriptional network analysis of 2 breast cancer prognostic signatures identified 10 Master Transcriptional Regulator (MTR) genes: ATAD2, E2F1, E2F8, FOXM1, HMGB2, MYBL2, PTTG1, TCF19, UHRF1, ZNF367
2) Analysis of The Cancer Genome Atlas breast cancer dataset identified CDKN2A as a key tumor suppressor deregulated in tumors that had bypassed the cellular senescence checkpoint
3) ChIP analyses demonstrated that several MTRs bind to, and directly regulate, promoters of proliferation-associated genes
4) A subset of MTRs were demonstrated to be prognostic at mRNA level and protein level

Development of statistical algorithm for validation in the TransATAC cohort

(Barron *et al.* 2017, Loughman *et al.* 2018)

1) All 2,047 possible combinations of 10 MTRs + CDKN2A were analysed using Cox proportional hazards regression models of recurrence as a function of normalised gene expression
2) RT-qPCR assessment in training cohort (comprised of an expanded set of samples used in the discovery study above)
    - Hydrolysis probe gene expression assays for 10 MTRs, CDKN2A and three reference genes were designed and optimised
    - 225 samples with 37 distant recurrences from Malmö University Hospital were assessed
    - All 2,047 gene combinations were cross-validated 1,000 times x 2 folds
3) *In silico* analysis of publicly available array-based expression profiling data
    - 4 datasets (GSE3494, GSE6532, GSE21653, GSE20685) with 873 samples and 223 recurrences
    - All 2,047 gene combinations were fit (without cross-validation)
4) 3 best performing gene combinations were identified using hazard ratio, C-index, and consideration of over-fitting
5) Each of the 3 best performing gene combinations from 4) were refit to the RT-qPCR data from Malmö University Hospital. The molecular risk score for the gene combination is the Cox linear predictor
6) Each of the 3 molecular risk scores above were combined with clinical information (lymph node status, tumor size, tumor grade) and refit to the RT-qPCR data from Malmö University Hospital. The risk score is the Cox linear predictor
7) The 3 models were tested in new data sets – none of the new samples had been used in previous development
    - RT-qPCR data: 100 samples with 13 distant recurrences (Skåne University Hospital)
    - *In silico* analysis of array-based expression profiling data: 623 samples, 237 distant recurrences (Stockholm, GSE48091)
8) Final OncoMasTR algorithm scores: Fixed prior to validation in TransATAC
    - OncoMasTR scores were calculated blind to clinical data and recurrence status
    - OncoMasTR Molecular Score (OMm) is a linear combination of FOXM1, PTTG1, ZNF367. OMm is rescaled to range from 0 to 100
    - OMclin1 is a linear combination of OMm, lymph node status, tumor size and tumor grade. OMclin1 is rescaled to range from 0 to 10; 0 to < 5 is low risk, 5-10 is high risk
    - OncoMasTR Risk Score (OMclin2) is a nested version of OMclin1; a linear combination of OMm, lymph node status, and tumor size. Like OMclin1, OMclin2 is rescaled to range from 0 to 10; 0 to < 5 is low risk, 5-10 is high risk