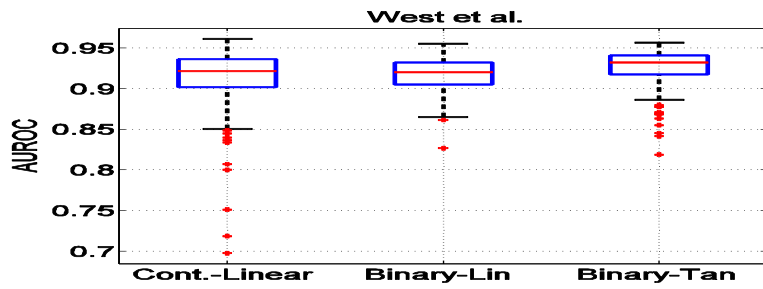
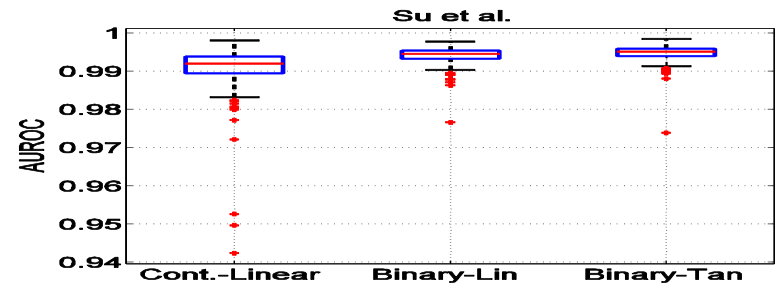


Reducing variability caused by the algorithmic choice of pre- processing raw microarray data

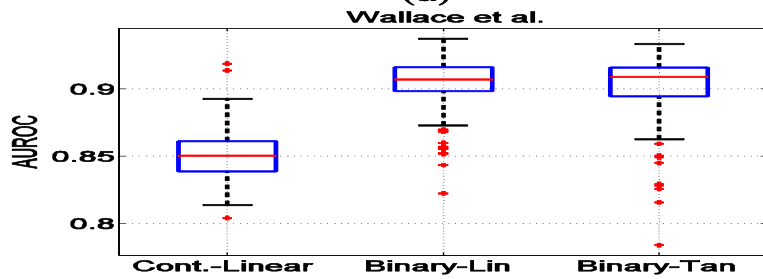
Salih Tuna and Mahesan Niranjan
Wellcome Trust Sanger Institute and
University of Southampton



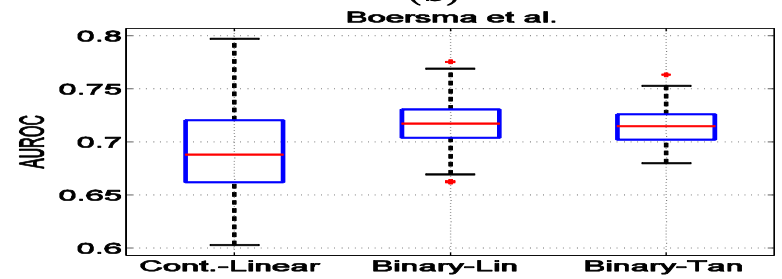
(a)



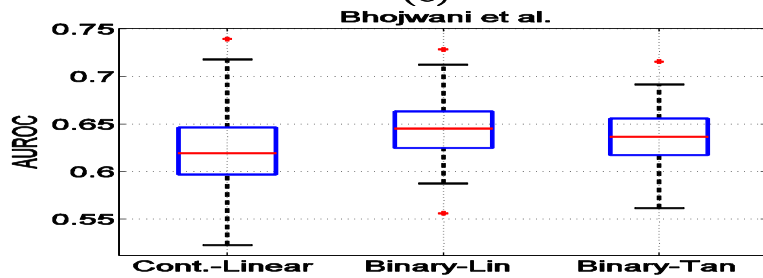
(b)



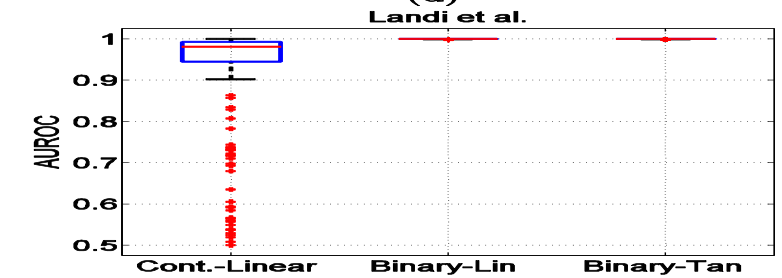
(c)



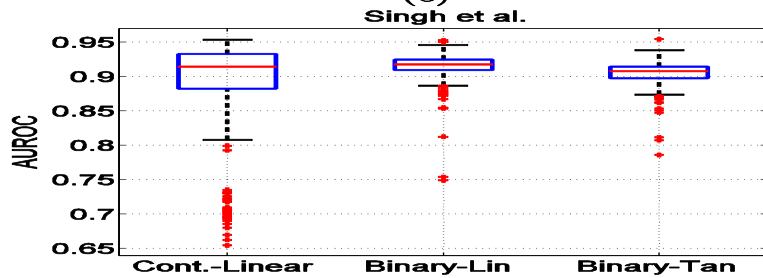
(d)



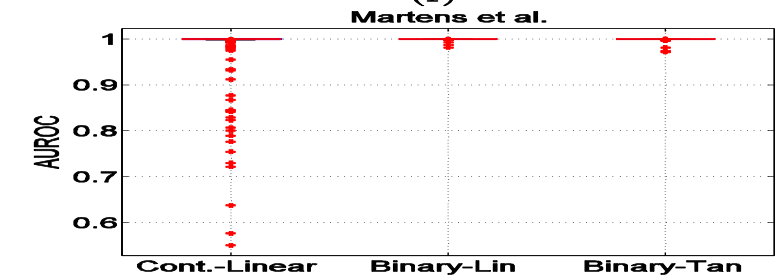
(e)



(f)

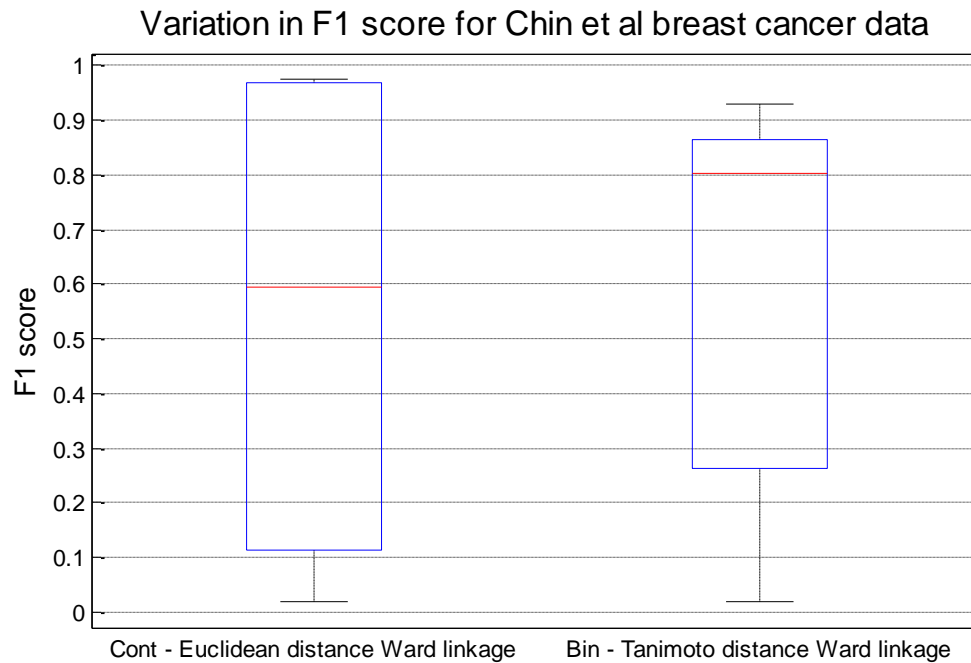


(g)



(h)

Figure : Variability in Classification due to the choice of pre-processing algorithms. Figure taken from Tuna and Niranjan (2010) *Reducing the algorithmic variability in transcriptome-based inference*, *Bioinformatics*, 26(9):1185-1191.



Conclusion

- Binary representation of microarray data with a suitable metric has the desirable effect of reducing the variability caused by the choice of the algorithmic choice and simultaneously increase the performance of inferences. This has been shown in classification and clustering using 9 different datasets.
- The variability induced from choice of pre-processing algorithms is massive. This observation should cast serious doubt on the quality of inferences drawn from microarray based studies, in this context.
- Unlike in classification, the variability reduction we are able to achieve by working in a binary space appears small.
- Binary data removes the biological variation of the microarray measurements and is more realistic.
- Tanimoto coefficient outperforming other metrics can be explained with, when there are more expressed genes in an experiment, gene expression measurements are more certain.