References

[1] P. Bork, T. Dandekar, Y. Diaz-Lazcoz, F. Eisenhaber, M. Huynen, and Y. Yuan, "Predicting function: from genes to genomes and back", *Journal of molecular biology*, vol. 283, no. 4, pp. 707-725, 1998.

[2] D. Botstein and G. R. Fink. "Yeast: an experimental organism for 21st Century biology", *Genetics*, vol. 189, no. 3, pp. 695-704, 2011.

[3] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, ... and M.A. Harris, "Gene Ontology: tool for the unification of biology", *Nature genetics*, vol. 25, no. 1, pp. 25-29, 2000.

[4] H.W. Mewes, D. Frishman, U. Guldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Munsterkotter, S Rudd and B. Weil, "MIPS: a database for genomes and protein sequences", *Nucleic acids research*, vol. 30, no. 1, pp. 31-34, 2002.

[5] HER2 Status [Online]. Available:
http://www.breastcancer.org/symptoms/diagnosis/her2

[6] What is sickle cell disease [Online]. Available -
http://www.nhlbi.nih.gov/health/health-topics/topics/sca

[7] What are yeast [Online]. Available:
http://wiki.yeastgenome.org/index.php/What_are_yeast%3F

[8] A. Barrientos, "Yeast Models of Human Mitochondrial Diseases", *IUBMB Life*, vol. 55, no. 2, pp. 83-95, 2008.

[9] By Masur-Own work, Public Domain [Online]. Available:
https://commons.wikimedia.org/w/index.php?curid=1069017

[10] By domdomegg-Own work, CC BY 4.0 [Online]. Available:
https://commons.wikimedia.org/w/index.php?curid=46468746

[11] M.R. Duchen and G. Szabadkai, "Roles of mitochondria in human disease", *Essays in biochemistry*, no. 47, pp. 115-137, 2010.

[12] M.A. Hibbs, C.L. Myers, C. Huttenhower, D.C. Hess, K. Li, A.A. Caudy and O.G. Troyanskaya, "Directing experimental biology: a case study in mitochondrial biogenesis", *PLoS Comput Biol*, vol. 5, no. 3, e1000322, 2009.

[13] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, "Analyzing protein structure and function", *Molecular Biology of the cell* (4th ed.), New York: Garland Science, 2002.

[14] P. Radivojac, W.T. Clark, T.R. Oronn, A.M. Schnoess, T. Wittkop, A. Sokolov, K. Graim et al.,"A large-scale evaluation of computational protein function prediction",*Nature methods*, vol. 10, no. 3, pp. 221-227, 2013.

[15] G. Valentini, "Hierarchical ensemble methods for protein function prediction", *ISRN bioinformatics*, 2014.

[16] R. Nielsen, J.S. Paul, A. Albrechtsen and Y.S. Song, "Genotype and SNP calling from next-generation sequencing data", *Nature Reviews Genetics*, vol. 12, no. 6, pp. 443-451, 2011.

[17] J. Quackenbush, "Microarray data normalization and transformation", *Nature genetics*, vol. 32, pp. 496-501, 2002.

[18] A. Sanchez and M.C.R. de Villa,"A Tutorial Review of Microarray Data Analysis",Universitat de Barcelona, 2008.

[19] Bioinformatics: Introduction and Methods, Center for Bioinformatics, Peking University [Online Lecture]. Available: https://www.coursera.org/learn/bioinformatics-pku

[20] T. Can, Introduction to Bioinformatics, Middle East Technical University [Online]. Available: http://ocw.metu.edu.tr/course/view.php?id=37

[21] What is Functional Genomics [Online]. Available:
http://www.ebi.ac.uk/training/online/course/functional-genomics-introduction-embl-ebi-resource/what-functional-genomics-1

[22] L. R. Engelking, "Textbook of Veterinary Physiological Chemistry", Chapter 1 [Online]. Available:
http://booksite.elsevier.com/samplechapters/9780123848529/02~Chapter_1.pdf

[23] General structure of an amino acid [Online]. Available:
https://en.wikipedia.org/wiki/File:AminoAcidball.svg

[24] Peptide bond [Online]. Available:
http://www.ifa.hawaii.edu/UHNAI/article4.htm

[25] Amino Acid Properties [Online]. Available:
http://www.genscript.com/amino_acid_structure.html

[26] Protein Structure [Online]. Available:
http://courses.washington.edu/conj/protein/protein.htm

[27] Q. Gu, Y. S. Ding and T. L. Zhang, "An ensemble classifier based prediction of G-protein-coupled receptor classes in low homology", *Neurocomputing*, vol. 154, pp. 110-118, 2015.

[28] Protein Folds and Protein Fold Classification [Online]. Available:
http://www.proteinstructures.com/Structure/Structure/protein-fold.html

[29] Supersecondary Structures (Motifs) and Domains [Online]. Available:
https://biochemistryquestions.wordpress.com/2008/10/09/supersecondary-structures-motifs-and-domains/

[30] Protein Domains and Domain Classification [Online]. Available:
http://www.proteinstructures.com/Structure/Structure/protein-domains.html

[31] Protein Three-Dimensional Structure: Structural Levels, Motifs and Folds [Online]. Available: http://www.proteinstructures.com/Structure/protein-structure1.html

[32] G. Pandey, V. Kumar and M. Steinbach, "Computational approaches for protein function prediction: A survey", Twin Cities::Department of Computer Science and Engineering, University of Minnesota, 2006.

[33] What are protein families [Online]. Available: https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/protein-classification/what-are-protein-families

[34] J. Adams, "The proteome: discovering the structure and function of proteins", *Nature Education*, vol. 1, no. 3, pp 6, 2008.

[35] S. Halgamuge, Advanced workshop in Bioinformatics.

[36] S.Y. Rhee and M. Mutwil, "Towards revealing the functions of all genes in plants", *Trends in plant science*, vol. 19, no. 4, pp.212-221, 2014.

[37] M.M. Babu, "Introduction to microarray data analysis", *Computational genomics: Theory and application*, pp.225-249, 2004.

[38] Y.H. Yang, S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai and T.P. Speed, "Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation", *Nucleic acids research*, vol. 30, no. 4, e15-e15, 2002.

[39] B. Zhang, Gene Expression Data Analysis, Department of Biomedical Informatics, Vanderbilt University [Online]. Available: http://bioinfo.vanderbilt.edu/zhanglab/lectures/BMIF310_geneExpression-_B_ dataAnalysis_2009.pdf

[40] Z. Louxin, Gene expression data analysis, Department of Mathematics, National University of Singapore [Online]. Available: http://www.bii.a-star.edu.sg/docs/education/lsm5192/zhangLec7.pdf

[41] D. Nettleton, Normalization Methods for Two-Color Microarray Data, Iowa State University [Online]. Available: http://www.public.iastate.edu/∼dnett/microarray/05normalization2color.ppt

[42] Normalization of Microarray Data [Online]. Available: https://www.youtube.com/watch?v=FpuVrfMu-2U

[43] P. Hoen, Expression Array Normalization [Online]. Available: http://dial.liacs.nl/Courses/MicroArrayDataAnalysis/Lectures/normalization_R-course_PB.pdf

[44] GO Evidence codes [Online]. Available: http://geneontology.org/page/guide-go-evidence-codes

[45] QuickGO,A fast browser for Gene Ontology terms and annotations [Online]. Available: http://www.ebi.ac.uk/QuickGO

[46] R. Eisner, B. Poulin, D. Szafron, P. Lu and R. Greiner, "Improving protein function prediction using the hierarchical structure of the gene ontology", *In Computational Intelligence in Bioinformatics and Computational Biology*, Proceedings of the IEEE Symposium, pp. 1-10, 2005.

[47] G. Valentini, "True path rule hierarchical ensembles for genome-wide gene function prediction", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 3, pp.832-847, 2011.

[48] G. Valentini, "True path rule hierarchical ensembles", In International Workshop on Multiple Classifier Systems, pp. 232-241, Springer Berlin Heidelberg, 2009.

[49] M.N. Davies, A. Secker, A.A. Freitas, M. Mendao, J. Timmis and D.R. Flower,"On the hierarchical classification of G protein-coupled receptors", *Bioinformatics*, vol. 23, no. 23, pp. 3113-3118, 2007.

[50] J. Struyf, S. Dzeroski, H. Blockeel and A. Clare, "Hierarchical multi-classification with predictive clustering trees in functional genomics". pp. 272-283. Springer Berlin Heidelberg, 2005

[51] L. Schietgat,C. Vens, J. Struyf, H. Blockeel, D. Kocev and S. Dzeroski, "Predicting gene function using hierarchical multi-label decision tree ensembles". *BMC bioinformatics*, vol. 11, no. 1, pp. 1, 2010.

[52] Z. Barutcuoglu, R.E. Schapire and O.G. Troyanskaya, "Hierarchical multi-label prediction of gene function", *Bioinformatics*, vol. 22, no. 7, pp. 830-836, 2006.

[53] N. Alaydie, C.K. Reddy and F. Fotouhi, "Hierarchical boosting for gene function prediction", *In Proceedings of the 9th International Conference on Computational Systems Bioinformatics*, vol. 9, pp. 14-25, 2010.

[54] G. Pandey, C.L. Myers and V. Kumar, "Incorporating functional inter-relationships into protein function prediction algorithms". *BMC bioinformatics*, vol. 10, no. 1, pp. 1, 2009.

[55] Y. Guan, C.L. Myers, D.C. Hess, Z. Barutcuoglu, A. Caudy and O. Troyanskaya, "Predicting gene function in a hierarchical context with an ensemble of classifiers", *Genome biology*, vol. 9, no. 1, S3, 2008.

[56] N. Alaydie, C.K. Reddy and F. Fotouhi, "A Bayesian integration model for improved gene functional inference from heterogeneous data sources". *In Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, pp. 376-380, 2011.

[57] Q. Wu, Y. Ye, S.S. Ho and S. Zhou, "Semi-supervised multi-label collective classification ensemble for functional genomics". *BMC genomics*,vol. 15, no. 9, S17, 2014.

[58] G. Yu, C. Domeniconi, H., Rangwala, G. Zhang and Z. Yu, "Transductive multi-label ensemble classification for protein function prediction". *In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1077-1085. ACM, 2012.

[59] L. Lan, N. Djuric, Y. Guo and S. Vucetic, "MS-kNN: protein function prediction by integrating multiple data sources", *BMC bioinformatics*, vol. 14, no. 3, p.S8, 2013.

[60] S. Mostafavi, D. Ray, D. Warde-Farley., C. Grouios and Q. Morris, "GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function". *Genome Biol*, vol. 9, no. 1, S4, 2008.

[61] X.M. Zhao, Y. Wang, L. Chen and K. Aihara, "Gene function prediction using labeled and unlabeled data", *BMC bioinformatics*, vol. 9, no. 1, pp.1, 2008.

[62] N. Youngs, D. Penfold-Brown, R.Bonneau and D. Shasha, "Negative example selection for protein function prediction: the NoGO database", *PLoS Computational Biology*, vol. 10, no. 6, p.e1003644, 2014.

[63] N. Youngs, D. Penfold-Brown, K. Drew, D. Shasha and R. Bonneau, "Parametric Bayesian priors and better choice of negative examples improve protein function prediction", *Bioinformatics*, p.btt110.

[64] G. Obozinski, G. Lanckriet, C. Grant, M.I. Jordan and W.S. Noble, "Consistent probabilistic outputs for protein function prediction", *Genome Biology*, vol. 9, no. 1, pp.1, 2008.

[65] J.F. Diez-Pastor, J.J. Rodriguez, C. Garcia-Osorio and L.I. Kuncheva, "Random Balance: Ensembles of variable priors classifiers for imbalanced data", *Knowledge-Based Systems*, vol. 85, pp. 96-111, 2015.

[66] P. Yang, W. Liu, B.B. Zhou, S. Chawla and A.Y. Zomaya, "Ensemble-based wrapper methods for feature selection and class imbalance learning",

In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 544-555, Springer Berlin Heidelberg, 2013.

[67] M. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp.463-484, 2012.

[68] GO Annotations Download [Online]. Available: http://geneontology.org/page/download-annotations

[69] C. Huttenhower, M.A. Hibbs C.L. Myers, A.A. Caudy, D.C. Hess and O.G. Troyanskaya, "The impact of incomplete knowledge on evaluation: an experimental benchmark for protein function prediction", *Bioinformatics*, vol. 25, no. 18, pp. 2404-2410, 2009.

[70] D. Charif and J.R. Lobry, "SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis", *Structural approaches to sequence evolution*, pp. 207-232, Springer Berlin Heidelberg, 2007.

[71] R. Apweiler, T.K. Attwood, A. Bairoch, A. Bateman, E. Birney, M. Biswas, P. Bucher, L. Cerutti, F. Corpet, M.D.R. Croning and R. Durbin, "The InterPro database, an integrated documentation resource for protein families, domains and functional sites", *Nucleic acids research*, vol. 29, no. 1, pp. 37-40, 2001.

[72] S. Mnaimneh, A.P. Davierwala, J. Haynes, J. Moffat, W.T. Peng, W. Zhang, X. Yang, J. Pootoolal, G. Chua, A. Lopez and M. Trochesset, "Exploration of essential gene functions via titratable promoter alleles", *Cell*, vol. 118, no. 1, pp. 31-44, 2004.

[73] S. Chu, J. DeRisi, J, M. Eisen, J. Mulholland, D. Botstein, P.O. Brown and I. Herskowitz, "The transcriptional program of sporulation in budding yeast", *Science*, vol. 282, no. 5389, pp. 699-705, 1998.

[74] A.P. Gasch, M. Huang, S. Metzner, D. Botstein, S.J. Elledge and P.O. Brown, "Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p", *Molecular biology of the cell*, vol. 12, no. 10, pp. 2987-3003, 2001.

[75] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein and B. Futcher, "Comprehensive identification of cell cycle regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization", *Molecular biology of the cell*, vol. 9, no. 12, pp. 3273-3297, 1998.

[76] G.K. Smyth, N. P. Thorne and J. Wettenhall, "LIMMA: Linear Models for Microarray Data Users Guide", 2003. [Online].
Available: http://www.bioconductor.org

[77] K.V. Ballman, D.E. Grill, A.L. Oberg and T.M. Therneau, "Faster cyclic loess: normalizing RNA arrays via linear models",*Bioinformatics* vol. 20, no. 16, pp. 2778-2786, 2004.

[78] H. Bengtsson, K. Simpson, J. Bullard and K. Hansen, "aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory", Tech Report, no 745, Department of Statistics, University of California, Berkeley, February 2008

[79] T. Hastie, R, Tibshirani, B. Narasimhan and G. Chu. "impute: impute: Imputation for microarray data", 2011.

[80] K.C. Chou, "Prediction of protein cellular attributes using psuedo-amino acid composition." *Proteins: Structure, Function, and Bioinformatics*, vol. 43, no. 3, pp. 246-255, 2001.

[81] N. Xiao, Q. Xu, and D. Cao. "protr: Protein Sequence Feature Extraction with R", R package version 0.2-0, 2013 [Online]. Available: http://CRAN. R-project. org/package=protr.

[82] B. Petersen, T.N. Petersen, P. Andersen, M. Nielsen and C. Lundegaard, "A generic method for assignment of reliability scores applied to solvent accessibility predictions", *BMC structural biology*, vol. 9, no.1, pp. 1, 2009.

[83] T.N. Petersen, C. Lundegaard, M. Nielsen, H. Bohr, J. Bohr, Brunak, S., ... and O. Lund, "Prediction of protein secondary structure at 80% accuracy", *Proteins: Structure, Function, and Bioinformatics*, vol. 41, no. 1, pp. 17-20, 2000.

[84] G. Csardi and T. Nepusz, "The igraph software package for complex network research", *InterJournal, Complex Systems*, vol. 1695, no. 5, pp.1-9, 2006.

[85] L. Nanni, S. Mazzara, L. Pattini and A. Lumini, "Protein classification combining surface analysis and primary structure", *Protein Engineering Design and Selection*, vol. 22, no. 4, pp. 267-272, 2009.

[86] K.C. Chou, "Prediction of protein subcellular locations by incorporating quasi-sequence-order effect", *Biochemical and biophysical research communications*, vol. 278, no. 2, pp. 477-483, 2000.

[87] J. Shen, J. Zhang, X. Luo, W. Zhu, K. Yu, K. Chen, Y. Li and H. Jiang, "Predicting protein-protein interactions based only on sequences information", Proceedings of the National Academy of Sciences, vol. 104, no. 11, pp. 4337-4341, 2007.

[88] Y. Yang, "Identification of novel type III effectors using latent Dirichlet allocation", *Computational and mathematical methods in medicine* 2012.

[89] D.M. Blei, A. Y. Ng and M.I. Jordan, "Latent dirichlet allocation", *Journal of machine Learning research*, vol. 3, pp. 993-1022, 2003.

[90] M. La Rosa, A. Fiannaca, R. Rizzo and A. Urso, "Probabilistic topic modeling for the analysis and classification of genomic sequences", *BMC bioinformatics*, vol. 16, no. 6, pp. 1, 2015.

[91] D.M. Blei, K. Franks, M.I. Jordan and I.S. Mian, "Statistical modeling of biomedical corpora: mining the Caenorhabditis Genetic Center Bibliography for genes related to life span", *BMC Bioinformatics*, vol. 7, no. 1, pp. 250, 2006.

[92] S.G.A. Konietzny, L. Dietz, and A. C. McHardy. "Inferring functional modules of protein families with probabilistic topic models", *BMC bioinformatics*, vol. 12, no. 1, pp. 1, 2011.

[93] X.Y. Pan, Y.N. Zhang and H.B. Shen, "Large-Scale prediction of human protein protein interactions from amino acid sequence based on latent topic features", *Journal of Proteome Research*, vol. 9, no. 10, pp. 4992-5001, 2010.

[94] Y. Yang, B.L. Lu and W.Y. Yang, "Classification of Protein Sequences Based on Word Segmentation Methods." In APBC, pp. 177-186. 2008.

[95] G. Valentini and F. Masulli, "Ensembles of learning machines", *Neural Nets*,Springer Berlin Heidelberg, pp. 3-20, 2002.

[96] T.G. Dietterich, "Ensemble methods in machine learning", In International workshop on multiple classifier systems, Springer Berlin Heidelberg, pp. 1-15. 2000.

[97] J. Stefanowski, "Multiple Classifiers", Institute of Computing Sciences, PoznaÅĎ University of Technology [Online]. Available: http://www.cs.put.poznan.pl/jstefanowski/aed/DMmultipleclassifiers.pdf

[98] P. Yang, Y. Hwa Yang, B.B. Zhou and A.Y Zomaya, "A review of ensemble methods in bioinformatics", *Current Bioinformatics*, vol. 5, no. 4, pp. 296-308, 2010.

[99] N.C. Oza,"Ensemble data mining methods", NASA Ames Research Center, USA, 2004.

[100] Hal Daume III , "Ensemble methods", chapter 11, A Course in Machine Learning [Online]. Available: http://ciml.info/dl/v0_9/ciml-v0_9-ch11.pdf

[101] L. Rokach, "Ensemble methods for classifiers", *Data Mining and Knowledge Discovery Handbook*, pp. 957-980, Springer US, 2005.

[102] S. Whalen and G.K. Pandey, "A comparative analysis of ensemble classifiers: case studies in genomics", In IEEE 13th International Conference on Data Mining, pp. 807-816, 2013.

[103] T.G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes", *Journal of artificial intelligence research*, pp. 263-286, 1995.

[104] E.B. Kong and T.G. Dietterich, "Error-Correcting Output Coding Corrects Bias and Variance" In ICML, pp. 313-321, 1995.

[105] V.I. Nazarov, M.V. Pogorelyy, E.A. Komech, I.V. Zvyagin, D.A. Bolotin, M. Shugay, D.M. Chudakov, Y.B. Lebedev and I.Z. Mamedov, "tcR: an R package for T cell receptor repertoire advanced data analysis", *BMC bioinformatics*, vol. 16, no. 1, pp.1, 2015.

[106] K. Hornik, B. Grun, "topicmodels: An R package for fitting topic models", *Journal of Statistical Software,* vol. 40, no. 13, pp. 1-30, 2011.

[107] J. Holland, "Genetic algorithms", 1992.

[108] L. Kuncheva, "Genetic algorithm for feature selection for parallel classifiers", *Information Processing Letters*, vol. 46, no. 4, pp. 163-168, 1993.

[109] L. Scrucca, "GA: A Package for Genetic Algorithms in R", *Journal of Statistical Software*, vol. 53, no. 4, pp. 1-37, 2013.

[110] M. Scutari, "bnlearn: Bayesian network structure learning", R package, 2010.

[111] T. Fawcett, "ROC graphs: Notes and practical considerations for researchers", *Machine learning*, vol. 31, no. 1, pp.1-38, 2004.

[112] A.J. Viera and J.M. Garrett,"Understanding interobserver agreement: the kappa statistic", *Fam Med*, vol. 37, no. 5, pp. 360-363, 2005.

[113] J.L. Fleiss, "Measuring nominal scale agreement among many raters", *Psychological bulletin*, vol. 76, no. 5, pp. 378, 1971.

[114] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves", *In Proceedings of the 23rd international conference on Machine learning*, pp. 233-240, 2006.

[115] M Gamer, J Lemon, I Fellows and P Singh, "irr: Various Coefficients of Interrater Reliability and Agreement", R package, 2012.

Appendix A

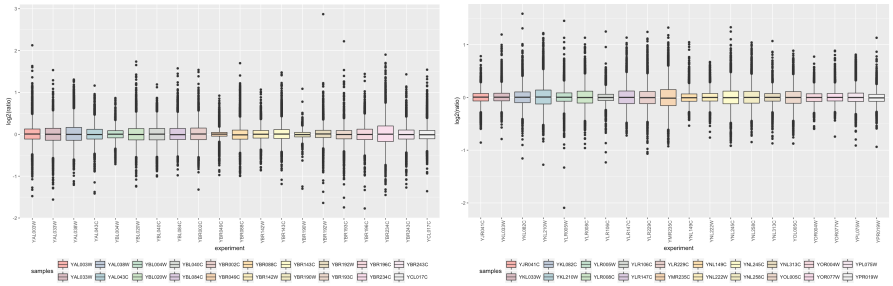
Exploratory Data Analysis

A.1 Initial Analysis of GO Annotations

There are 98,957 protein annotations with 17 variables, for 6381 distinct proteins from SGD database. 5530 distinct GO IDs have been used for annotation. 32,015 annotations were found to be duplicates (32.35% out of all). 232 annotations are explicitly noted as not being associated with the GO term. Under 18 different evidence codes: 25,412 annotations are with experimental evidence code; 5628 are with computational analysis evidence code; 481 are with author statement evidence code; 4838 are with curatorial statement evidence code; and 30,583 annotations are with IEA (Inferred from Electronic Annotation). There are 26,140 Biological Process GO term annotations; 19,129 Molecular Function GO term annotations; and 21,673 Cellular Component GO term annotations, in total. The annotations have been assigned by 8 parties: CACAO, GO Central, GOC, HGNC, InterPro, MGI, SGD and UniProt. Out of them, SGD stands for most of the annotations. All annotations belong to same database object type: 'gene', and the same taxon: 'taxon:559292', which indicates *S. cerevisiae*. Annotations have been made during the time period of 2000 - 2015. Most of the annotations have been made in 2015. All 30,583 IEA annotations have been made in 2015. However, the amount of curated annotations (36,359) surpasses the amount of electronically inferred annotations.

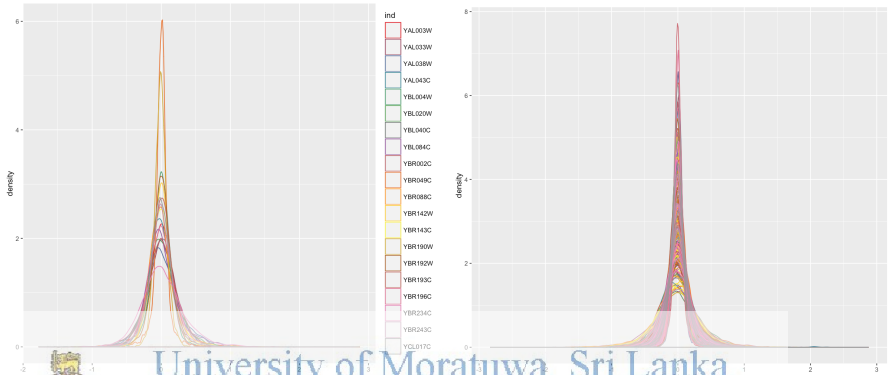
A.2 Data Visualizations

Following data visualizations were obtained using R graphic packages: *ggplot2* and *RColorBrewer*.



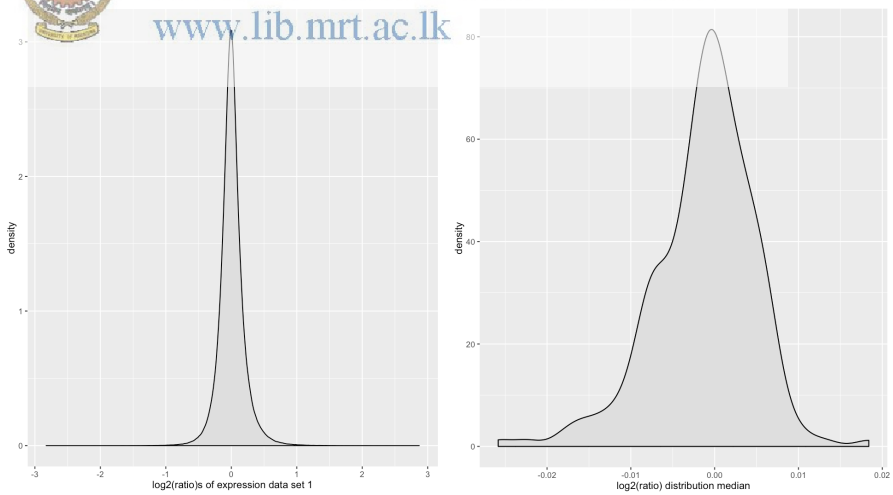
(a) First 20 experiments

(b) Last 20 experiments



(c) First 20 experiments

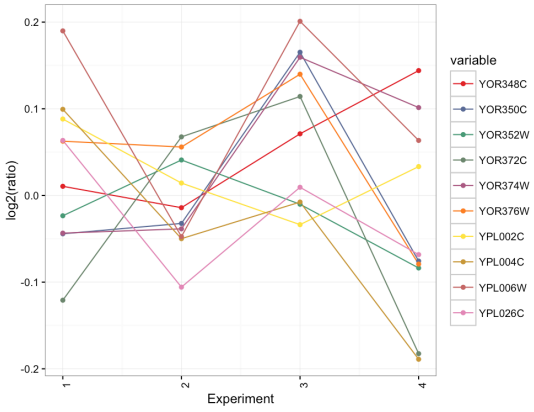
(d) All experiments



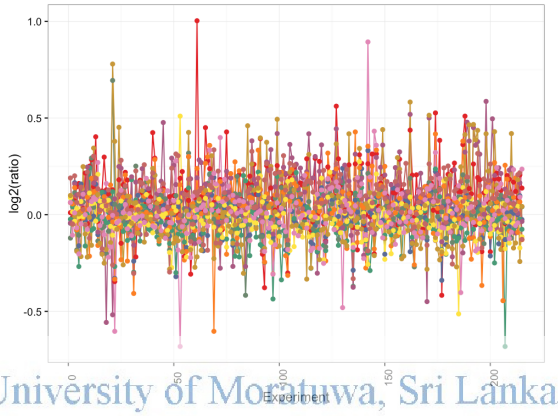
(e) All $\log_2$(ratio) values

(f) Median distribution

Figure A.2.1: Expressions 1 - Before normalization/preprocessing



(a) Expression ratio profiles of first 10 genes for the first 4 experiments



(b) Expression ratio profiles of first 10 genes for all experiments

Figure A.2.2: Expressions 1 - After normalization/preprocessing

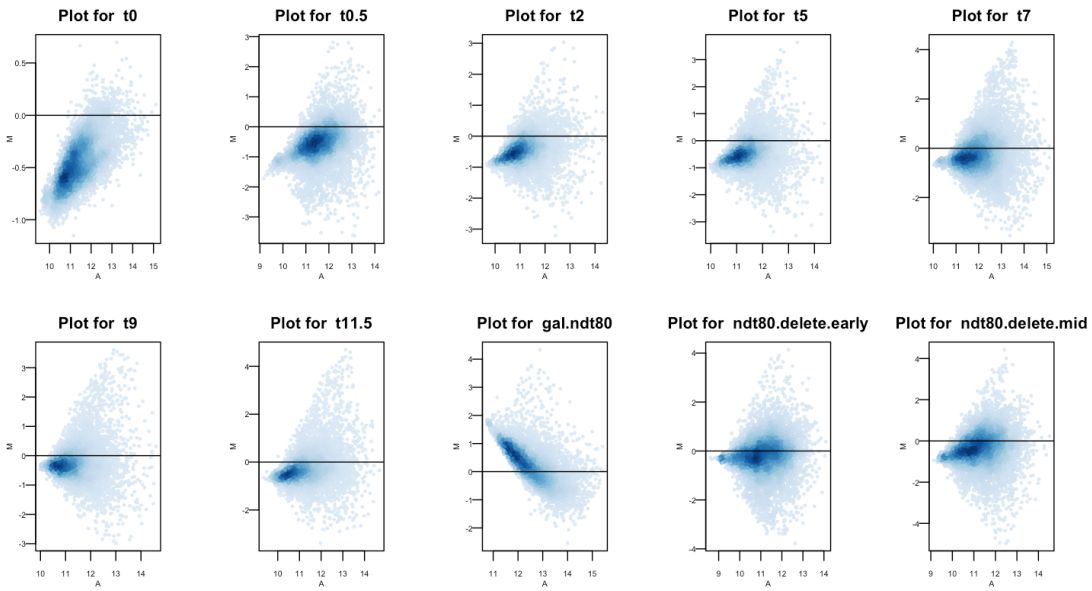


Figure A.2.3: Expressions 2 - MA plots before background correction

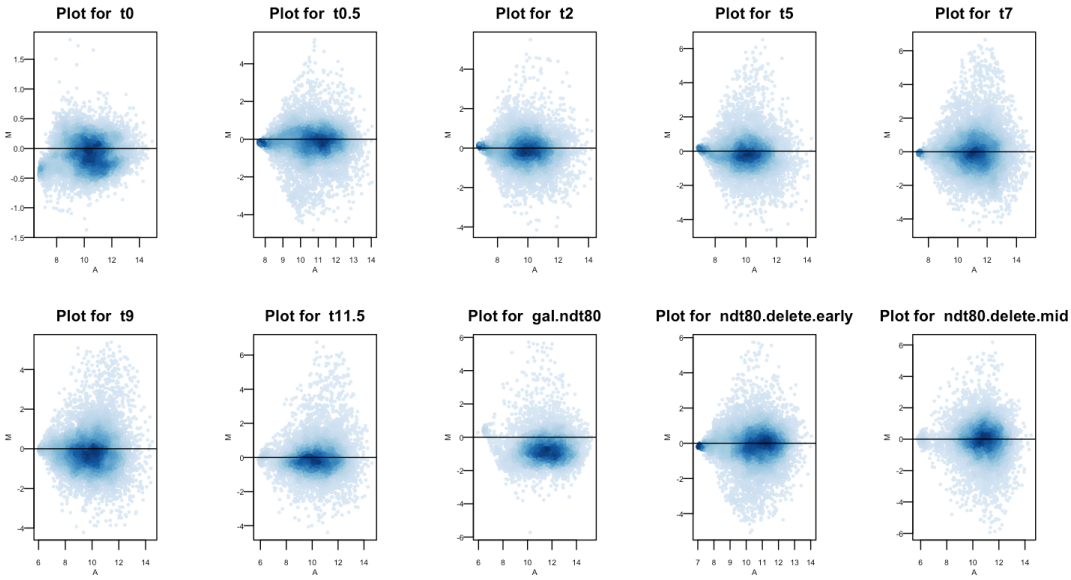


Figure A.2.4: Expressions 2 - MA plots after background correction

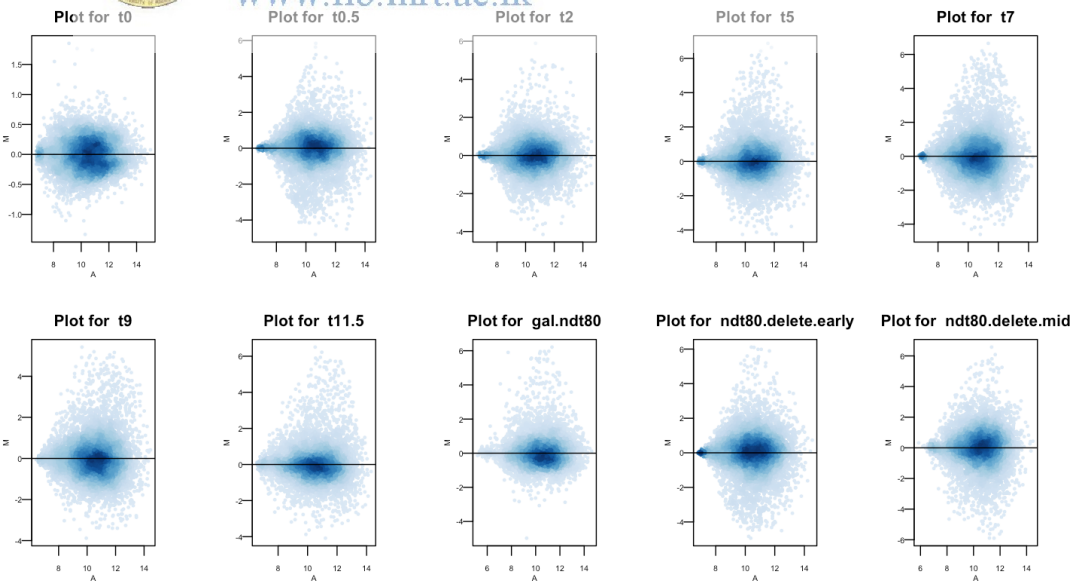
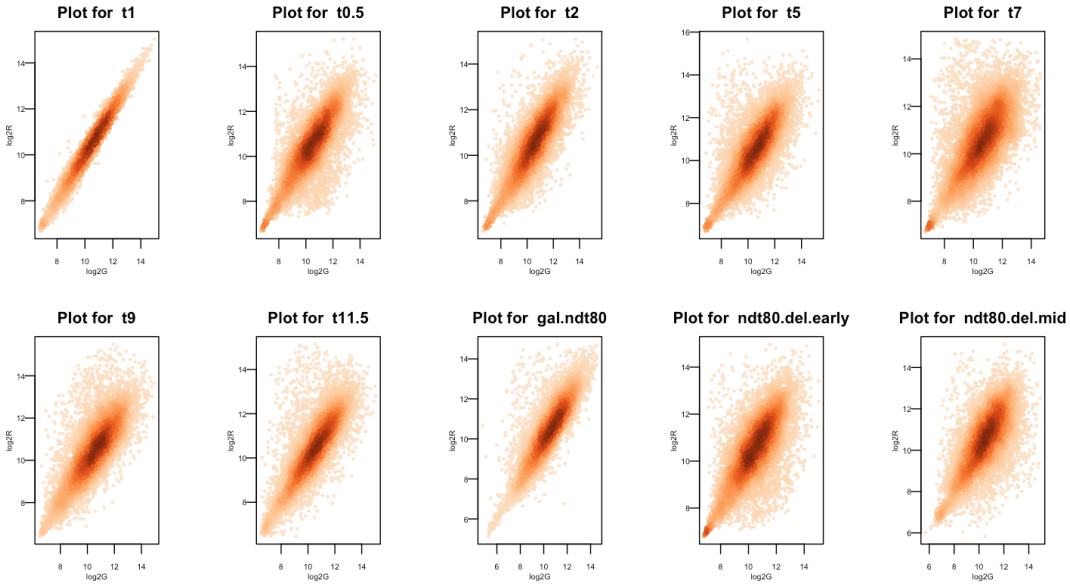
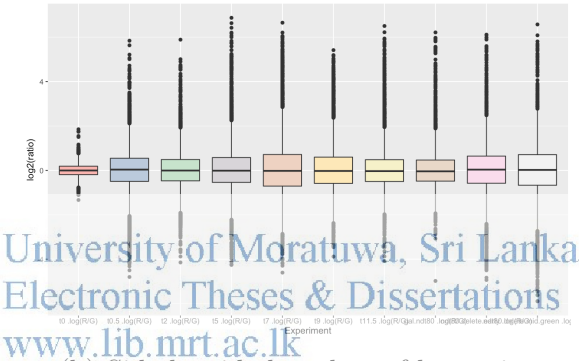


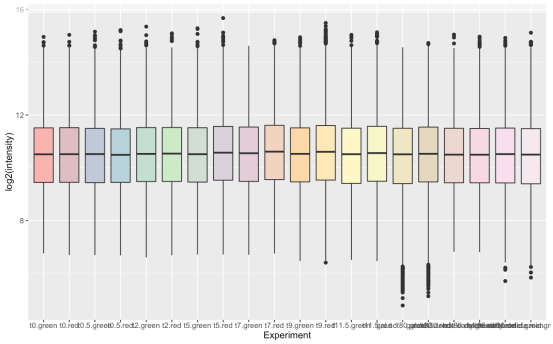
Figure A.2.5: Expressions 2 - MA plots after within/between array normalization



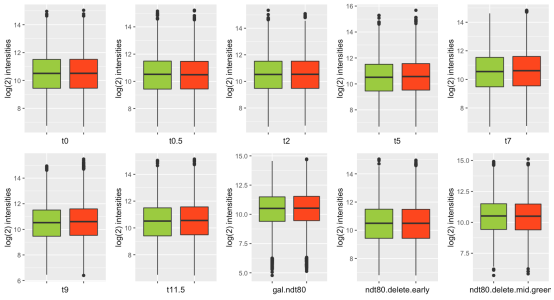
(a) Expressions 2 - log_2(G) vs log_2(R)



(b) Side-by-side boxplots of log ratios

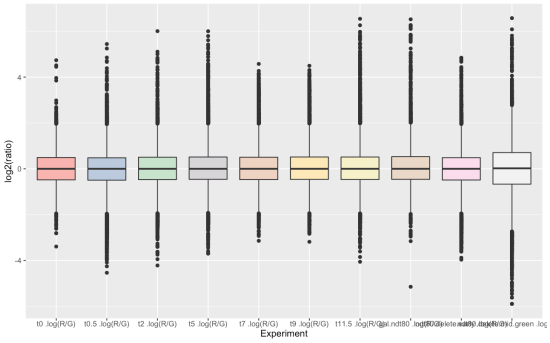


(c) Side-by-side boxplots of log_2(R) and log_2(G)

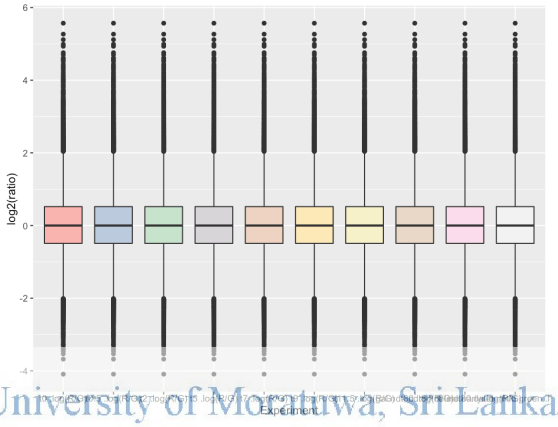


(d) Side-by-side pair boxplots of log_2(R) and log_2(G)

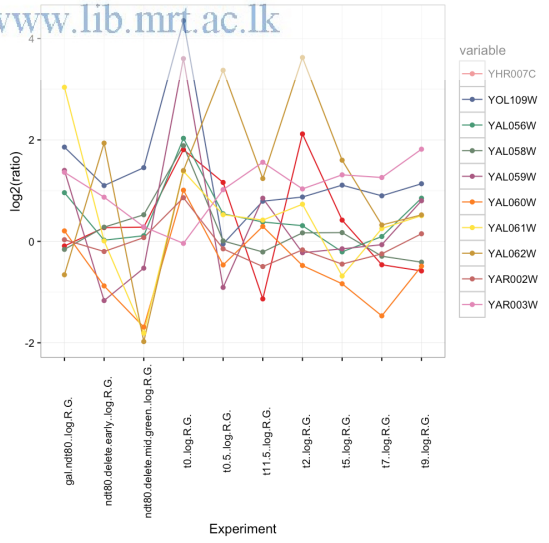
Figure A.2.6: Expressions 2 - After normalization/preprocessing



(a) After median centering and scale normalization

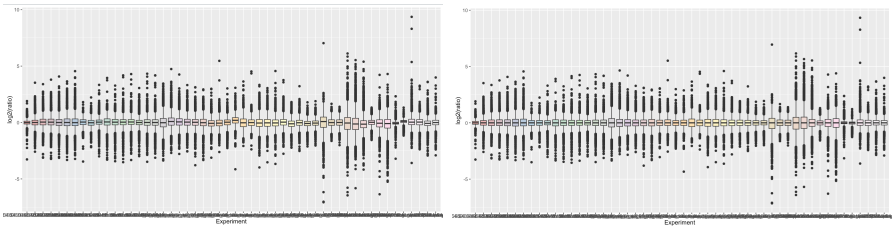


(b) After quantile normalization



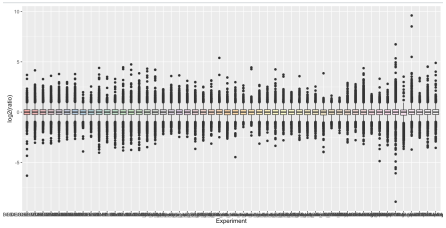
(c) Final Gene expression ratio profiles of first 10 genes

Figure A.2.7: Expressions 2 - After further normalization

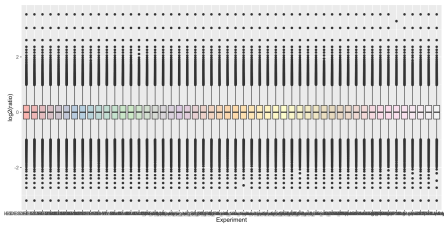


(a) Before normalization

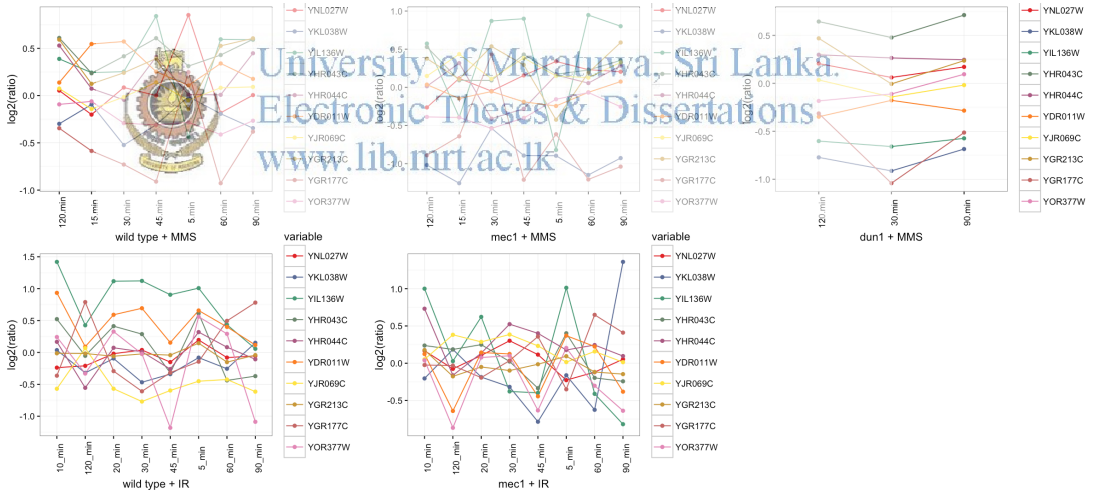
(b) After median centering



(c) After scale normalization

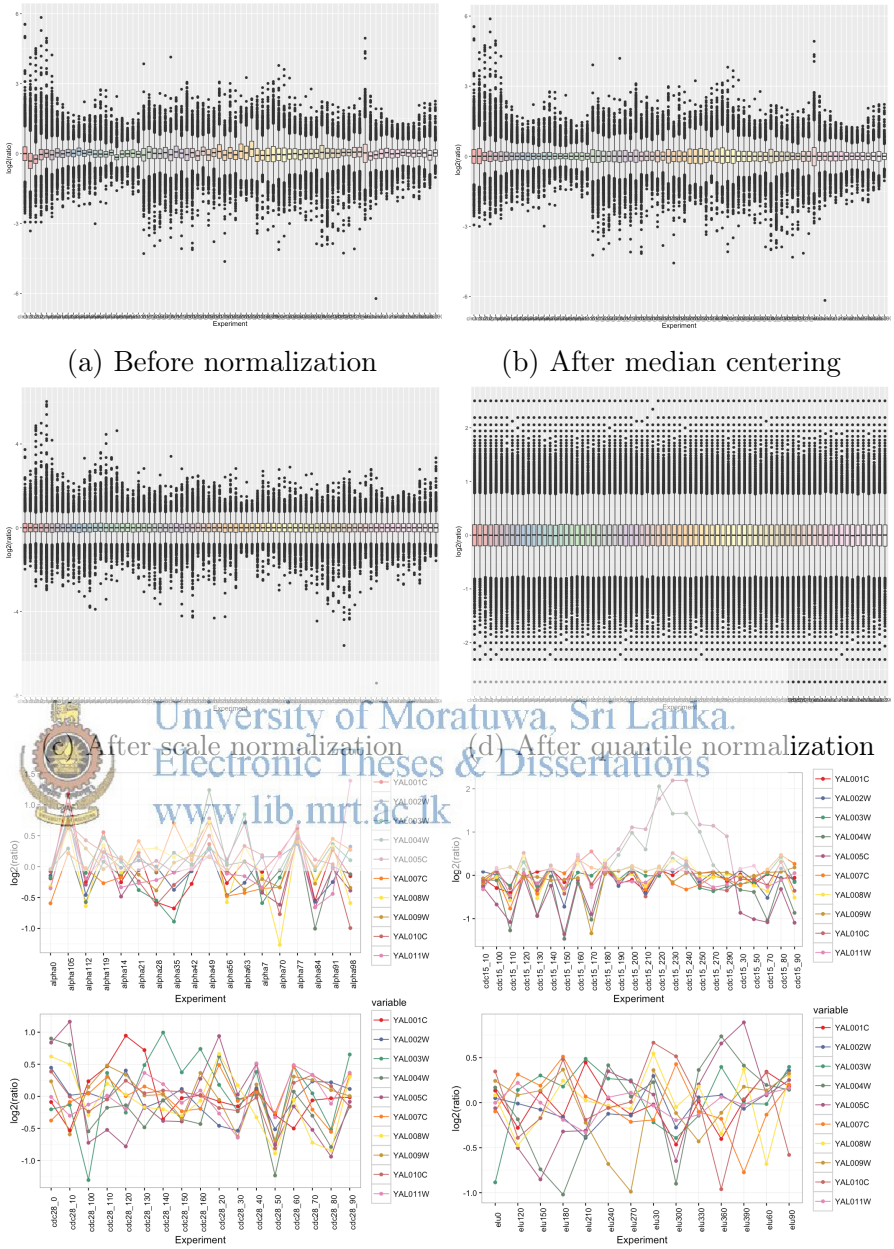


(d) After quantile normalization



(e) After normalization, expression ratio profiles of first 10 genes for the time series over each major experiment

Figure A.2.8: Expressions 3 - before & after normalization/preprocessing



(a) Before normalization

(b) After median centering

(c) After scale normalization

(d) After quantile normalization

(e) After normalization, expression ratio profiles of first 10 genes for the time series over each major experiment

Figure A.2.9: Expressions 4 - Before & after normalization/preprocessing

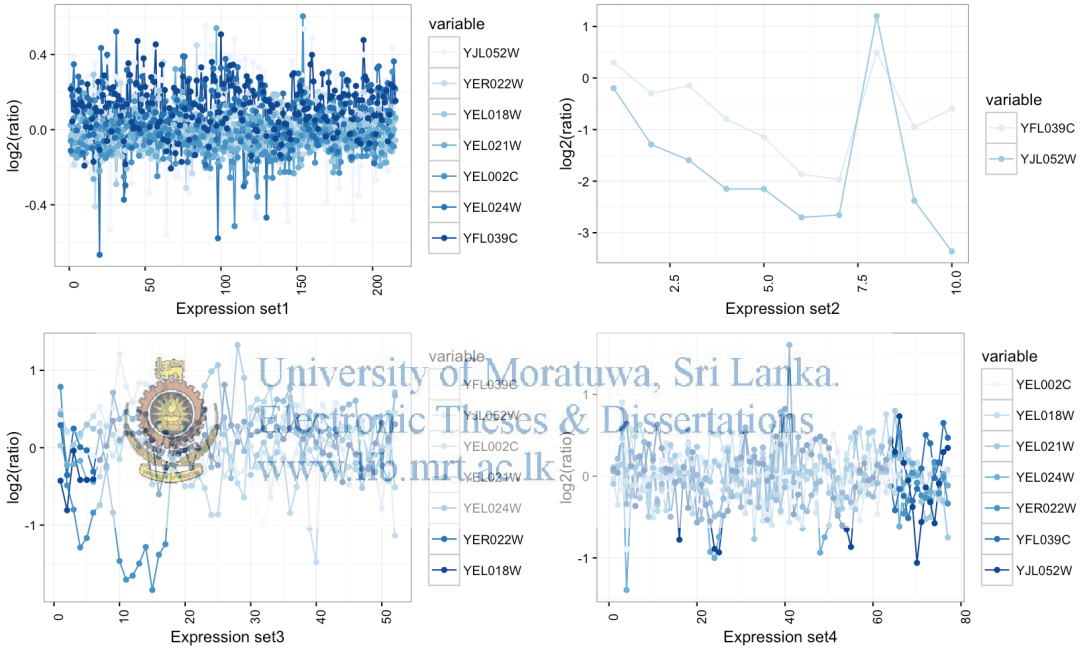


Figure A.2.10: Expression profiles of housekeeping genes

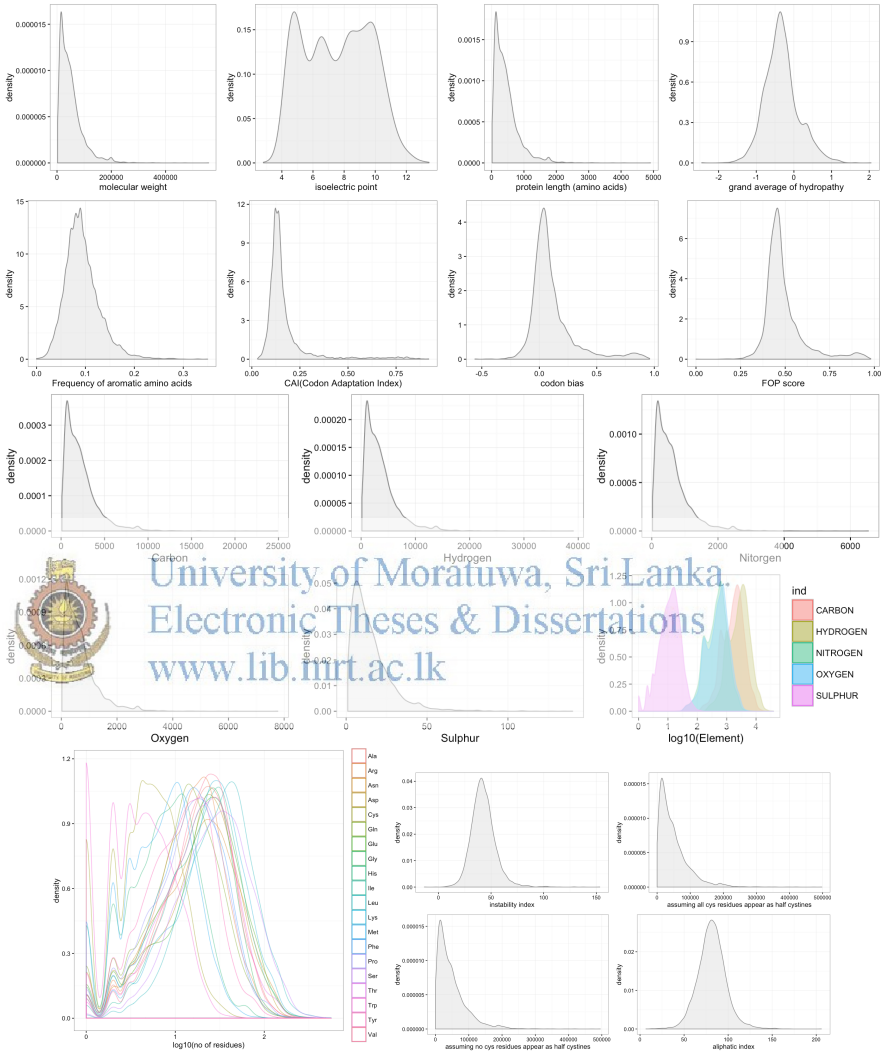


Figure A.2.11: Properties Data