Contents

1 Introduction 6
1.1 Introduction to basics . . . 6
1.1.1 Cells . . . 6
1.1.2 DNA . . . 6
1.1.3 Genes . . . 6
1.2 Gene expression . . . 7
1.2.1 Gene expression data . . . 7
1.2.2 DNA Microarray or Gene Chip . . . 7
1.2.3 Microarray gene expression data . . . 8
1.3 Co-expressed genes and coherent patterns . . . 8
1.4 Gene Regulatory Network . . . 8
1.4.1 Importance of GRN in computational biology . . . 9
1.5 Cluster Analysis . . . 9
1.5.1 Clustering approaches . . . 9
1.5.2 Existing Clustering algorithms . . . 10
1.5.3 Limitations of Clustering . . . 10
1.5.4 Biclustering fundamentals . . . 10
1.6 Motivation for bicluster ensemble . . . 10
1.7 Proposal . . . 10
1.8 Its significance . . . 11
1.9 Objective . . . 11
1.10 Contribution . . . 11
1.11 Organization of report . . . 12

2 Background 13
2.1 Microarray gene expression data . . . 13
2.1.1 Microarray Time-series gene expression Data . . . 13
2.2 Biclustering algorithms . . . 13
2.2.1 Modified Biclustering algorithms for time-series data . . . 14
2.3 Ensemble Biclusters . . . 14
2.4 Validation . . . 14
2.4.1 External Validation . . . 14
2.4.2 p-value . . . 15
2.4.3 Q-value . . . 15
2.4.4 Gene Coverage . . . 15
2.5 Existing models for inferring GRNs . . . 15
2.5.1 Relevance Networks . . . 15
2.5.2 Boolean Networks . . . 16

2.5.3 Probabilistic Boolean Networks 16
2.5.4 Bayesian Networks 16
2.5.5 Dynamic Bayesian Networks 17
2.6 **Discussion** 17

3 Related Work **18**
3.1 **Relevance Network** 18
3.2 **Bayesian and Dynamic Bayesian Network** 18
3.3 **Boolean and Probability Boolean Network** 18
3.4 **Biclustering and inferring GRNs** 19
3.5 **Discussion** 19

4 Proposed Work **20**
4.1 **Framework** 20
4.2 **Proposed Framework** 21
4.2.1 Dataset collection and its preprocessing. 21
4.2.2 Biclustering co-expressed genes with coherent expression patterns. 21
4.2.3 Identify strongly connected similar biclusters based on GO annotations. 23
4.2.4 Merge all the strongly connected biclusters and then validate the results. 23
4.2.5 Infer subnetworks using ARANCE in cytoscape. 23
4.2.6 Using an edge selection technique the edges of the three similar subnetworks are considered to construct a subnetwork. 24
4.2.7 Inferring GRNs. 25

5 Experimental Result **26**
5.1 **Environment Used** 26
5.2 **Dataset Used** 26
5.2.1 Study of Dataset and its preprocessing 26
5.3 **Biclustering Result** 27
5.3.1 CCC-biclusters 27
5.3.2 e-CCC-biclusters 27
5.4 **Ensembled Biclusters** 27
5.5 **External Validation** 27
5.5.1 p-value 28
5.5.2 Q-value 28
5.5.3 Coverage 28
5.6 **Visualization of biclusters** 28
5.7 **CCC, e-CCC and ensembled subnetwork result** 30
5.8 **Inferred subnetwork results** 31
5.9 **Inferred GRN** 31
5.9.1 Rare gene Identification 32
5.10 **Graphical User Interfaces** 32
5.10.1 Creating a MATLAB GUI 32

6 Conclusion and Future Work **37**
6.1 Conclusion 37
6.2 Future Work 37
.1 Appendix A 38
.2 Appendix B 38
.3 Appendix C 39
.4 Appendix D 40
.5 Appendix E 40

Bibliography

List of Figures

1.1 The hierarchical structure of gene.[23] 6
1.2 Central Dogma of gene expression.[24] 7
1.3 Co-expressed genes and Coherent Patterns 8
1.4 An artificial Gene Regulatory Network. 9

4.1 A pictorial overview of the proposed approach 20

5.1 Graphical representation of the mean gene expression values of CCC, e-CCC and ensembled biclusters. 29
5.2 Snapshot of the inferred subnetwork in Cytoscape 30
5.3 Snapshot of the GUI 34
5.4 Snapshot of the GUI visualizing the mean gene expression data of CCC, e-CCC and the ensembled bicluster 34

List of Tables

2.1 Microarray time-series gene expression dataset 13

4.1 Edge selection measure 25

5.1 Details of the inferred subnetworks of the CCC biclusters 30
5.2 Details of the inferred subnetworks of the e-CCC biclusters 31
5.3 Details of the inferred subnetworks of the ensembled bicluster 31
5.4 Details of the modified subnetworks result 31
5.5 p-value and Q-value of the biclusters 35
5.6 Group1 GO Annotation result 36
5.7 Group2 GO Annotation result 36
5.8 Group3 GO Annotation result 36
5.9 Group4 GO Annotation result 36
5.10 Group5 GO Annotation result 36
5.11 Group6 GO Annotation result 36