

# LIST OF CONTENTS

	Page
Acknowledgment	i
List of Contents	ii
List of Figures	iv
List of Tables	vi
Resumen	vii
Abstract	viii
<b>1 THEORETICAL FRAMEWORK</b>	<b>1</b>
1.1 Biomarkers . . . . .	1
1.2 Omics and the variety of data . . . . .	1
1.3 Normalization . . . . .	2
1.4 Supervised and unsupervised machine learning . . . . .	3
1.4.1 Supervised methods . . . . .	3
1.4.2 Unsupervised methods . . . . .	3
1.5 Dimensionality reduction . . . . .	4
1.5.1 Principal Component Analysis . . . . .	5
1.5.2 Minimum Curvilinear Embedding . . . . .	5
1.6 Statistical evaluations . . . . .	6
1.6.1 Mann-Whitney test . . . . .	6
1.6.2 Kolmogorov-Smirnov test . . . . .	7
1.6.3 Concordance score . . . . .	7
1.6.4 Bezdek . . . . .	7
1.6.5 Kruskal-Wallis . . . . .	7
<b>2 HYPOTHESIS or DEVELOPMENT PROBLEM</b>	<b>9</b>
2.1 Hypothesis or Problem . . . . .	9
<b>3 GOALS</b>	<b>10</b>
3.1 General Goals . . . . .	10
3.1.1 Specific Goals . . . . .	10

<b>4</b>	<b>MATERIALS AND METHODS</b>	<b>11</b>
4.1	Materials . . . . .	11
4.1.1	MATLAB: . . . . .	11
4.1.2	Inkscape: . . . . .	11
4.1.3	Cytoscape: . . . . .	12
4.2	Methods . . . . .	12
4.2.1	Normalization . . . . .	12
4.2.2	Dimensionality reduction . . . . .	15
4.2.3	Statistical evaluations . . . . .	16
<b>5</b>	<b>Results</b>	<b>19</b>
5.1	First dimensions found . . . . .	19
5.2	Statistics of results . . . . .	22
5.3	Hotspot heatmaps . . . . .	28
5.4	Visualization of groups in 2D space . . . . .	36
5.5	New strategy for segregation recognition . . . . .	39
5.6	Biomarkers network . . . . .	39
<b>6</b>	<b>Discussion</b>	<b>43</b>
<b>7</b>	<b>Conclusion</b>	<b>46</b>
	<b>Bibliography</b>	<b>48</b>
	<b>Supplementary Information</b>	
<b>A:</b>	<b>Genomics Figures</b>	<b>53</b>

## LIST OF FIGURES

	Page
1.1 Dimensionality Reduction Illustration . . . . .	4
1.2 Minimum Spanning Tree used by MCE . . . . .	6
5.1 Metagenomics statistical tests first dimensions identified . . . . .	20
5.2 Metagenomics statistical evaluations first dimensions identified . . . . .	21
5.3 Normalization frequency metagenomics data . . . . .	22
5.4 singled/coupled normalization frequency metagenomics data . . . . .	25
5.5 heatmap singled/coupled normalization frequency metagenomics data . . .	26
5.6 normalization frequency metagenomics data per statistic evaluation . . . .	27
5.7 dimension frequency metagenomics data per statistic evaluation . . . . .	28
5.8 hotspot heatmap ncMCE correlation metagenomics data . . . . .	29
5.9 hotspot heatmap cMCE correlation metagenomics data . . . . .	30
5.10 hotspot heatmap ncMCE euclidean metagenomics data . . . . .	31
5.11 hotspot heatmap cMCE euclidean metagenomics data . . . . .	32
5.12 hotspot heatmap ncMCE spearman metagenomics data . . . . .	33
5.13 hotspot heatmap cMCE spearman metagenomics data . . . . .	34
5.14 hotspot heatmap ncPCA metagenomics data . . . . .	35
5.15 hotspot heatmap cPCA metagenomics data . . . . .	36
5.16 Metagenomics data segregation ncPCA . . . . .	37
5.17 Metagenomics data segregation cPCA . . . . .	38
5.18 Metagenomics data two groups segregation cPCA . . . . .	39
5.19 Newtork based in metagenomics data DR method ncMCE correlation . . .	41
A.1 Genomics statistical tests first dimensions identified . . . . .	53
A.2 Genomics statistical evaluations first dimensions identified . . . . .	54
A.3 Normalization frequency genomics data . . . . .	54
A.4 singled/coupled normalization frequency genomics data . . . . .	55
A.5 heatmap singled/coupled normalization frequency genomics data . . . . .	56
A.6 normalization frequency genomics data per statistic evaluation . . . . .	57
A.7 dimension frequency genomics data per statistic evaluation . . . . .	57
A.8 hotspot heatmap ncMCE correlation genomics data . . . . .	58
A.9 hotspot heatmap cMCE correlation genomics data . . . . .	58
A.10 hotspot heatmap ncMCE euclidean genomics data . . . . .	59
A.11 hotspot heatmap cMCE euclidean genomics data . . . . .	59

A.12 hotspot heatmap ncMCE spearman genomics data . . . . .	60
A.13 hotspot heatmap cMCE spearman genomics data . . . . .	60
A.14 hotspot heatmap ncPCA genomics data . . . . .	61
A.15 hotspot heatmap cPCA genomics data . . . . .	61
A.16 Genomics data segregation ncPCA . . . . .	62
A.17 Genomics data segregation ncMCE correlation . . . . .	63
A.18 Genomics data segregation cPCA . . . . .	64
A.19 Genomics data segregation ncMCE spearman . . . . .	65

## LIST OF TABLES

	Page
5.1 Summary of the analysis for metagenomics data . . . . .	23
5.2 Summary of the analysis for genomics data . . . . .	24
5.3 First strategy results . . . . .	40
5.4 Second strategy results . . . . .	40