

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Problem Statement	4
1.3	Objectives and Research Questions	7
1.4	Research Methodology	8
1.5	Thesis Outline	9
2	Problem Investigation	11
2.1	Genomic Data Sources	11
2.2	Conceptual Modeling for Genomics	13
2.3	Data Quality for Genomics	15
2.4	Conclusions	18
3	Treatment Design	21
3.1	The SILE Method	22
3.1.1	Search	22
3.1.2	Identification	24
3.1.3	Load	27
3.1.4	Exploitation	27
3.2	Data Quality Assessment	29
3.2.1	Data Quality Methodology for Genomics	29
3.2.2	Data Quality Assessment of Genomic Repositories	30
3.2.3	Data Quality Assessment of Genomic Information	31
3.2.3.1	Types of Genetic Studies	33
3.2.3.2	Statistical Relevance of Genetic Studies	34
3.3	Conclusions	34
4	Treatment Validation	37
4.1	Building a Genomic Information System	38
4.1.1	Searching for Genomic Data Sources	39
4.1.2	Identifying Relevant Information	42
4.1.2.1	The Extraction Module	42
4.1.2.2	The Transformation Module	45
4.1.3	Information Storage	50

4.2	The Migraine Case	52
4.3	The Epilepsy Case	55
4.4	Results	59
4.4.1	Migraine Case Study Results	60
4.4.2	Epilepsy Case Study Results	60
4.4.3	Complementary Case Studies	61
4.4.3.1	Crohn Disease Case Study	61
4.4.3.2	Male Breast Cancer Case Study	62
4.5	Conclusions	65
5	Conclusions and Future Work	69
5.1	Answers to Research Questions	69
5.1.1	Results of Objective 1	70
5.1.2	Results of Objective 2	70
5.1.3	Results of Objective 3	71
5.2	Thesis Impact	72
5.2.1	Publications	72
5.2.2	Academic Works	74
5.2.3	Teaching Experience	74
5.2.4	Research Projects	74
5.3	Future Work	75
	Appendices	85
A	Conceptual Schema Description	85
B	Mapping Rules	87
C	Variant Type Mapping	90
D	Variants associated with Epilepsy	92
E	Variants associated with Crohn's Disease	93