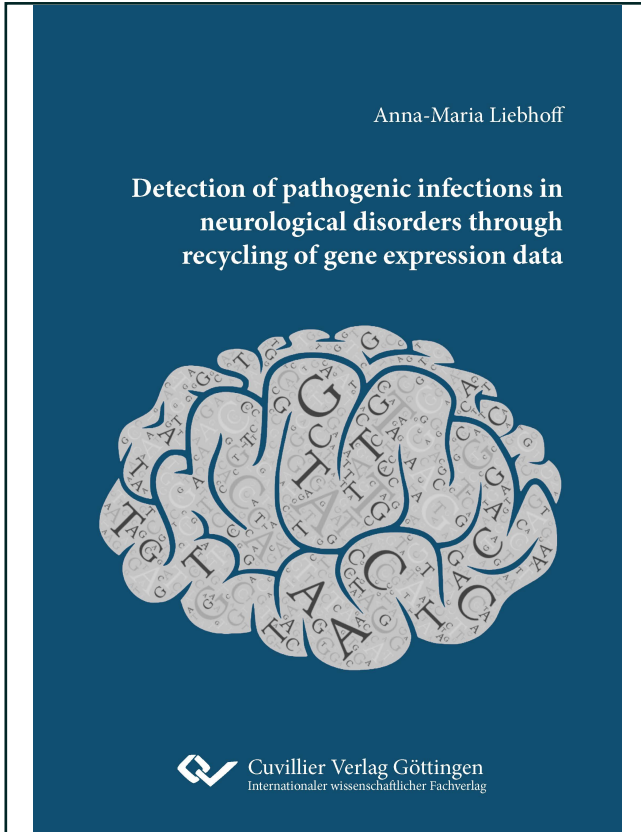




Anna-Maria Liebhoff (Autor)

**Detection of pathogenic infections in neurological disorders
through recycling of gene expression data**



<https://cuvillier.de/de/shop/publications/8428>

Copyright:

Cuvillier Verlag, Inhaberin Annette Jentsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen, Germany
Telefon: +49 (0)551 54724-0, E-Mail: info@cuvillier.de, Website: <https://cuvillier.de>

Contents

1	INTRODUCTION	1
2	BIOLOGICAL BACKGROUND	5
2.1	Ribonucleic acid (RNA)	5
2.2	Small RNA	7
2.3	Sequencing	8
2.4	Microbiology	10
2.5	Neurodegenerative diseases	12
2.5.1	Parkinson's disease	13
2.5.2	Frontotemporal dementia	14
2.5.3	Research on neurodegenerative diseases	14
2.6	Pathogens in neurological diseases	15
3	BIOINFORMATICS BACKGROUND	17
3.1	Data sources	17
3.1.1	The Gene Expression Omnibus (GEO)	18
3.1.2	The Sequence Read Archive (SRA)	18
3.2	Metadata and biological ontologies	19
3.3	Read alignment	20
3.3.1	STAR	20
3.4	Comparative analysis	21
3.4.1	Differential expression analysis	21
3.5	Metagenomic analysis	23
3.5.1	Kraken	23
3.6	Integrated analysis pipelines	24
3.6.1	Oasis2	25

4	COMPUTER SCIENCE BACKGROUND	27
4.1	Web development	27
4.1.1	Client-server model and REST APIs	28
4.1.2	Technologies and frameworks	29
4.2	Data science and machine learning	30
4.2.1	Data handling	31
4.2.2	Graphical representation of sets and intersections	32
4.2.3	PCA	32
4.2.4	t-SNE	34
4.2.5	Random forest	35
5	SMALL RNA EXPRESSION ATLAS (SEA)	39
5.1	Background	39
5.2	System architecture	41
5.3	Implementation	43
5.3.1	Data download and preprocessing	43
5.3.2	Intra-dataset analysis	45
5.3.3	Data integration and storage	48
5.3.4	Interactive analysis	50
5.3.5	External requests	54
5.4	System behavior	56
5.4.1	Type-ahead requests	56
5.4.2	Search requests for annotations	57
5.4.3	Search requests for entities	58
5.4.4	Combined search requests	60
5.4.5	Cross-comparison analysis	61
5.4.6	Dataset and comparison views	61
5.4.7	User's data	63
5.5	Results	63
6	BIOLOGICAL DISCOVERIES THROUGH SEA	67
6.1	Tissue specificity of small RNA	67
6.2	User data and Parkinson's disease	69
6.3	Pathogens in neurodegenerative diseases	72
6.4	Sex specificity of small RNA in human body	74
6.5	Discussion	77

7	PATHOGEN DETECTION IN RNA-SEQ DATA	79
7.1	Background	80
7.2	Identification algorithm	81
7.2.1	Step 1 - Alignment to NCBI nt database	81
7.2.2	Step 2 - Interpretation of Kraken alignment output	83
7.3	Performance measures	85
7.4	Evaluation	86
8	PATH(OGEN)S TO DISEASE	91
8.1	Analysis pipeline	91
8.2	<i>Burkholderia stabilis</i> in frontotemporal dementia	95
8.3	Discussion	98
9	CONCLUSION	101
	BIBLIOGRAPHY	104
A	SEA JSON COMMUNICATION OBJECTS	113
B	PATHONOA ALGORITHMS	116
C	DETAILED ANALYSIS RESULTS	118