

Contents

Abstract	3
Contents	11
1 Introduction	17
1.1 The era of transcriptomics: from bulk to cell-level resolution . . .	19
1.1.1 Single-cell RNA-seq technologies	20
1.2 Post-transcriptional regulation of gene expression: alternative splicing and mRNA isoform expression	27
1.2.1 Computational approaches for the study of AS and iso- form transcriptomics	31
2 Motivation, aims and contributions	35
2.1 Motivation	37
2.2 Aims	38
2.3 Contributions made during the thesis	40
2.3.1 Publications	40
2.3.2 Conferences	41

2.3.3 Software development 44

2.3.4 Undergraduate and Master’s thesis supervision 45

2.3.5 Teaching 45

3 Evaluation of the limitations of single-cell RNA-seq technologies for the study of isoforms 49

3.1 Introduction 51

3.2 Ideal conditions for single-cell isoform studies 53

3.2.1 Full-length transcript capture 54

3.2.2 Sequencing depth 55

3.2.3 Cell throughput 56

3.2.4 Sequencing errors and artifacts 58

3.3 Expectations meet reality: what has been and remains to be done in single-cell isoform studies 61

3.3.1 Full-length transcript capture and isoform continuity 62

3.3.2 Capture efficiency and sequencing depth 68

3.3.3 Cell throughput 71

3.3.4 Sequencing errors and artifacts 73

3.4 Navigating the computational landscape of single-cell isoform studies 76

3.5 Assessing the theoretical limits of current technologies for single-cell isoform studies 80

3.5.1 Methods 80

3.5.2 Results	86
3.6 Discussion	94
3.6.1 Advances in short-read data generation	95
3.6.2 Advances in computational method development	97
3.6.3 Advances in long-read data generation	99
4 Integration of bulk long-reads and single-cell short-read data to enhance isoform analyses	105
4.1 Introduction	107
4.2 Methods	111
4.2.1 Data availability	111
4.2.2 Transcriptome reconstruction using long reads	113
4.2.3 Functional annotation of long read-defined isoforms	116
4.2.4 SQANTI3 machine learning-based filter	121
4.2.5 Filtering of potential artifacts using the SQANTI3 ML filter	124
4.2.6 SQANTI3 rescue strategy	127
4.2.7 Rescue of ML-filtered artifacts using SQANTI3	132
4.2.8 Single-cell data pre-processing and quality control	133
4.2.9 Single-cell multi-group Differential Expression analysis	135
4.2.10 Neural cell sampling strategy to select a consensus set of DE isoforms	136
4.3 Results	138

- 4.3.1 Defining high-quality long read transcripts: insights from pre and post curation 138
- 4.3.2 Characterizing IsoAnnotLite: performance assessment on long-read transcript annotations 144
- 4.3.3 Enabling multi-group differential expression of isoforms in single-cell data 147
- 4.3.4 Design of an enhanced, automated filtering strategy: the SQANTI3 ML filter 150
- 4.3.5 Enhancing transcriptome rescue in SQANTI3: a comprehensive approach 165
- 4.4 Discussion 171

5 A novel method to derive isoform co-usage networks from single-cell data **179**

- 5.1 Introduction 181
- 5.2 Methods 185
 - 5.2.1 Data and software availability 185
 - 5.2.2 Percentile correlation 185
 - 5.2.3 Semi-automated isoform clustering 187
 - 5.2.4 Co-expression pattern simulation 191
 - 5.2.5 Benchmarking of isoform correlation metrics for scRNA-seq data 194
 - 5.2.6 Differential Isoform Usage and co-Differential Isoform Usage across multiple groups 196

5.2.7	Functional analyses	199
5.2.8	Analysis of GABA-ergic neuron subtypes	204
5.3	Results	205
5.3.1	Detecting cell type-dependent isoform co-expression . . .	205
5.3.2	Validation of percentile correlations on simulated data .	213
5.3.3	Co-differential isoform usage analysis of single-cell iso- form expression	220
5.3.4	Functional analysis of isoforms within a coDIU network .	223
5.3.5	Functional insights on neuron-oligodendrocyte isoform co-expression patterns	231
5.3.6	Analysis of coDIU patterns in GABA-ergic neuron subtypes	236
5.4	Discussion	240
6	Conclusions	245
	References	272
I	Documentation for the acorde R package	273
I.1	Introduction	275
II	CoDIU functional analysis results	313
II.1	Introduction	315
II.2	Partially-overlapping samples z test for CoDIU vs DIU genes . .	315
II.3	Functional Diversity Analysis of DE isoform, DIU and CoDIU genes	323

II.4 Functional Diversity Analysis of oligodendrocyte-neuron clusters 327