



Universiteit
Leiden
The Netherlands

Statistical methods for the analysis of complex omics data

Tissier, R.

Citation

Tissier, R. (2018, December 4). *Statistical methods for the analysis of complex omics data*. Retrieved from <https://hdl.handle.net/1887/67092>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/67092>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:

<http://hdl.handle.net/1887/67092>

Author: Tissier, R.

Title: Statistical methods for the analysis of complex omics data

Issue Date: 2018-12-04

Statistical methods for the analysis of complex omics data

Renaud Laurent Michel Tissier

Cover design: IPSKAMP Printing
Printing: IPSKAMP Printing

Copyright ©2018 Renaud Tissier, Amsterdam, the Netherlands.
All rights reserved. No part of this publication may be reproduced without prior permission of the author.

ISBN: 978-94-028-1239-8

This thesis was funded by the EU under the FP7 project MIMOmics (Methods for Integrated Analysis of Multiple Omics Datasets), (Grant/Award Number: FP7/Health/F5/2012/305280)

Statistical methods for the analysis of complex omics data

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 4 December 2018
klokke 10.00 uur

door

Renaud Laurent Michel Tissier
geboren te Louviers, Frankrijk in 1987

Promotor: Prof. Dr. J. J. Houwing-Duistermaat

Co-Promoters: Dr. M. Rodríguez-Girondo
Dr. S. Tsonaka

Leden promotiecommissie: Prof. Dr. P. E. Slagboom
Prof. Dr. J.H. Barret
· *University of Leeds, Leeds*
Dr. W.N. van Wieringen
· *Vrije Universiteit, Amsterdam*

Table of Contents

1	Introduction	1
1.1	Introduction	1
1.2	Measure of dependence: Pearson correlation coefficients	3
1.3	Family studies	5
1.3.1	Modelling between subject correlation in family studies	5
1.3.2	Modelling ascertainment in family studies	6
1.3.3	Examples of family studies	7
1.4	Secondary phenotypes	8
1.5	Correlated predictors	10
1.6	Outline of the thesis	12
2	Secondary phenotype analysis in ascertained family designs	15
2.1	Introduction	16
2.2	Methods	18
2.2.1	Retrospective likelihood approach	18
2.2.2	Mixed-effects models for the analysis of family data	19
2.2.3	Genotype probability	21
2.2.4	Estimation and statistical testing	22
2.2.5	Continuous polygenic score	22
2.2.6	Inclusion of covariates in the model	23
2.3	Simulation Study	23
2.3.1	Simulation results for a SNP	24
2.3.2	Simulation results for a polygenic score	27
2.4	Application: Analysis of the Leiden Longevity Study	29
2.4.1	Triglyceride levels analysis	29
2.4.2	Glucose levels analysis	30
2.5	Discussion	31
3	Secondary phenotype analysis in family proband designs	33
3.1	Introduction	34
3.2	Methods	37
3.2.1	Naive approach: ignoring sampling	37
3.2.2	Joint modeling under retrospective likelihood	37

3.2.3	Conditioning on probands	39
3.3	Simulation study	40
3.3.1	Simulation Setup	40
3.3.2	Simulation Results	42
3.4	The social anxiety disorder study	48
3.5	Discussion	51
4	Gene co-expression network analysis for family studies	53
4.1	Background	54
4.2	Methods	54
4.2.1	Study sample	54
4.2.2	Single probe analysis	55
4.2.3	Network constructions	55
4.2.4	Phenotype analysis	55
4.3	Results	56
4.3.1	Results obtained with all probes	56
4.3.2	Analysis of Q1	57
4.3.3	Results obtained with the 25% most heritable probes	58
4.4	Discussion	59
5	Improving stability of prediction models based on correlated omics	61
5.1	Introduction	62
5.2	Methods	64
5.2.1	Step 1: Network construction	65
5.2.2	Step 2: Hierarchical clustering	67
5.2.3	Step 3: Outcome prediction	67
5.2.4	Software implementation	69
5.3	Simulation Study	69
5.3.1	Simulation setup	69
5.3.2	Simulation results	71
5.4	Real data analysis	79
5.4.1	DILGOM: metabolites	79
5.4.2	DILGOM: Transcriptomics	80
5.4.3	Breast cancer cell lines	86
5.5	Discussion	88
6	Integration of several omic sources in prediction models	91
6.1	Introduction	92
6.2	Network-based group-penalized prediction	93
6.3	Group inference	97
6.3.1	Network construction: WGCNA	97
6.3.2	Hierarchical clustering	97
6.4	Outcome prediction	98

6.5	Simulations	99
6.5.1	Simulation setup	99
6.5.2	Simulation study	100
6.6	Real data analysis	106
6.6.1	DILGOM	106
6.6.2	Breast cancer cell lines	107
6.7	Discussion	109
	Bibliography	120
	List of Publications	121
	Summary	125
	Samenvatting	127
	Dankwoord	131
	Curriculum Vitae	133

