University of Groningen

Looking through the noise

Johansson, Leonard Fredericus

DOI:
[10.33612/diss.95673752](https://doi.org/10.33612/diss.95673752)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Johansson, L. F. (2019). *Looking through the noise: novel algorithms for genetic variant detection*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.95673752>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Looking through the noise

Novel algorithms for genetic variant detection

Leonard Fredericus Johansson

Leonard Fredericus Johansson. **Looking through the noise: novel algorithms for genetic variant detection.** Thesis, University of Groningen, with summary in English and Dutch.

Printing of this thesis was financially supported by Rijksuniversiteit Groningen, University Medical Center Groningen.

Cover design and layout by L.F. Johansson. The front cover shows a variant that can only be seen when looking through the noise created by the four DNA nucleotides A, C, G and T.

Printed by Ipskamp Drukkers, Enschede.

© 2019 L.F. Johansson. All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means without permission of the author.

ISBN: 978-94-034-1858-2

ISBN (electronic version): 978-94-034-1857-5

university of
 groningen

Looking through the noise

Novel algorithms for genetic variant detection

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Wednesday 25 September 2019 at 12.45 hours

by

Leonard Fredericus Johansson

born on 29 May 1980
in Hefshuizen

Supervisors

Prof. R.H. Sijmons

Prof. M.A. Swertz

Co-supervisor

Dr. B. Sikkema-Raddatz

Assessment Committee

Prof. V.V.A.M. Knoers

Prof. M. Vihinen

Prof. J.K. Ploos van Amstel

Paranymphs

E.N. de Boer

K.K. van Dijk-Bos

Propositions

1. Depending on how samples are prepared and analyzed, next-generation sequencing is suitable for detection of both base-level variants and structural variants. (*this thesis*)
2. High coverage next-generation sequencing data is suitable for single-exon copy number variation detection. (*this thesis*)
3. Before biological variability can be detected in next-generation sequencing, first laboratory induced variability has to be minimized. (*this thesis*)
4. International screening program criteria are currently not fully met for opportunistic genetic screening. (*this thesis*)
5. In non-invasive prenatal testing, the use of multiple independent models increases the reliability of the prediction of presence of a trisomy from a single data set. (*this thesis*)
6. The same measurement outcome in non-invasive prenatal testing gives different results for women with different prior risks of carrying a child with a trisomy. (*this thesis*)
7. Noise is everything that, from a certain perspective, blocks the path between reality and measurement outcome. (*this thesis*)
8. Data can be of high and low quality at the same time (depending on what information should be retrieved from the data). (*this thesis*)
9. Understanding how or why is seldom as useful as understanding that things are. (*Robin Hobb, Fool's Assassin*)
10. It's not what you look at that matters, it's what you see. (*Henry David Thoreau*)

Contents

1	Introduction	15
1.1	A short history on chromosomes and DNA	16
1.2	Human genome variation	17
1.3	Conventional techniques for variant detection	19
1.4	Next-generation sequencing	20
1.5	Technical bias and error rates	23
1.6	DNA variant detection in genome diagnostics	24
1.6.1	Germline variants	24
1.6.2	Somatic variants	25
1.6.3	Prenatal testing	26
1.7	Aims of this thesis	26
1.7.1	Germline variant detection	27
1.7.2	Detection of somatic chromosomal translocations	28
1.7.3	Prenatal detection of trisomies	29
1.7.4	Reflection and discussion	29
	Part 1: Germline variant detection	31
2	tNGS can replace Sanger sequencing in clinical diagnostics	33
2.1	Introduction	34
2.2	Material and methods	36
2.2.1	Design of the study	36
2.2.2	Patients/samples	36
2.2.3	Targeted enrichment kit design	37
2.2.4	Sample preparation	38
2.2.5	Capturing/enrichment	38

2.2.6	Sequencing	39
2.2.7	Data analysis and variant annotation	39
2.2.8	Validation of mutations by Sanger sequencing	40
2.3	Results	40
2.3.1	Validation phase	40
2.3.2	Application phase	42
2.3.3	Reproducibility of targeted NGS	44
2.4	Discussion	44
2.5	Conclusion	48
3	CoNVaDING: single exon variation detection in NGS data	49
3.1	Introduction	50
3.2	Material and methods	51
3.2.1	General workflow CoNVaDING	51
3.2.2	Input data	51
3.2.3	Control group selection	51
3.2.4	CNV prediction score calculation	53
3.2.5	Quality control metrics	55
3.2.6	CNV calling	56
3.2.7	Implementation of CoNVaDING	57
3.2.8	Validation of CoNVaDING	57
3.2.9	Comparison to CoNIFER,XHMM, and CODEX	58
3.3	Results	59
3.3.1	Validation of CoNVaDING	59
3.3.2	Comparison to CoNIFER, XHMM and CODEX	59
3.3.3	Performance of CoNVaDING on low-coverage data	61
3.4	Discussion	62
4	Using a diagnostic gene panel for opportunistic screening	65
4.1	Introduction	66
4.2	Materials and Methods	68
4.2.1	Patient cohorts	68
4.2.2	General Dutch population cohort	69
4.2.3	Selection of genes for the NGS panel	69
4.2.4	Sequencing and alignment procedure	69
4.2.5	Data analysis and interpretation	71
4.3	Results	71
4.3.1	Sequencing quality	71
4.3.2	Patient cohort: variant analysis	72
4.3.3	Control cohorts variant analysis	75
4.3.4	Comparison patient and control cohorts	76
4.4	Discussion	76
4.4.1	Diagnostic yield	76
4.4.2	Secondary findings in families vs general population	76
4.4.3	Should we offer additional screening?	77

Part 2: Detection of chromosomal translocations	83
5 Genetic test to detect translocations in acute leukemia	85
5.1 Introduction	87
5.2 Material and Methods	88
5.2.1 Patient bone marrow cells and cell lines	88
5.2.2 TLA acute leukemia gene panel	88
5.2.3 Multiplex TLA methods	89
5.2.4 Routine genetic and cytogenetic methods	89
5.2.5 Validation of the multiplex TLA method	90
5.3 Results	90
5.3.1 Validation of the TLA multiplex panel - Training set	90
5.3.2 Validation of the TLA multiplex panel - Test set	92
5.4 Discussion	94
Part 3: Prenatal detection of trisomies	99
6 Novel algorithms for improved sensitivity in NIPT	101
6.1 Introduction	102
6.2 Material and Methods	103
6.2.1 Chi-squared-based variation reduction	104
6.2.2 Regression-based Z-score	106
6.2.3 Match QC score	107
6.2.4 Validation of algorithms	107
6.3 Results	111
6.3.1 Effect of peak correction	111
6.3.2 Effects of the two GC correction methods	112
6.3.3 Effect of chi-squared-based variation reduction	112
6.3.4 Effect of trisomy prediction algorithms	113
6.3.5 Match QC score	116
6.4 Discussion	116
6.5 χ^2 VR for chromosome 21	120
6.6 Regression model for chromosome 13	123
7 NIPTeR: an R package for NIPT analysis	127
7.1 Background	128
7.2 Implementation	129
7.3 Results	131
7.3.1 Workflow	131
7.3.2 Prediction and control group statistics	132
7.3.3 Quality control	133
7.3.4 Performance	134
7.4 Conclusion	134
7.5 Availability and requirements	135

8	NIPTRIC: a tool for clinical interpretation of NIPT results	137
8.1	Introduction	138
8.2	Results	139
8.2.1	Performance of the PPR calculator	140
8.3	Discussion	140
8.4	Material and Methods	146
8.4.1	The PPR calculator	146
8.4.2	A priori risk	146
8.4.3	Z-score	146
8.4.4	Percentage of foetal DNA	147
8.4.5	Coefficient of variation	147
8.4.6	Examples of the use of the PPR calculator	149
8.4.7	Performance of the PPR calculator	149
	Part 4: Reflection and discussion	152
9	What can I know?	153
9.1	Perspectives and measurements	154
9.2	Assumptions and biases in next-generation sequencing	158
9.3	From genotype to phenotype	162
9.4	Conclusion	164
10	What should I do?	165
10.1	Moralizing technology	166
10.2	Moral decisions in Non-Invasive Prenatal Testing	168
10.3	The potential patient	171
10.4	Revisiting existing data	173
10.5	Does your genomic information belong to your family?	174
10.6	Moralizing introduced methods and algorithms	176
10.7	Conclusion	178
11	What may I hope?	179
11.1	Germline variant testing	180
11.2	Detection of somatic chromosomal translocations	181
11.3	Prenatal detection of trisomies	182
11.4	Balancing laboratory procedures and data analysis	183
11.5	Towards a complete DNA sequencing procedure	186
11.5.1	Short-read-sequencing-based variant detection	187
11.5.2	Single cell DNA sequencing	188
11.5.3	Long-read sequencing	189
11.5.4	Chromatin organization	191
11.5.5	Prenatal variant detection	191
11.6	Point-of-care testing	193
11.7	Looking towards the future	193
11.8	Conclusion	194

Bibliography	197
List of Tables	227
List of Figures	229
Appendices	231
A Summary	233
B Samenvatting	237
C Acknowledgements	241
D About the author	245
E List of publications	247
F Other academic activities	251