

GHENT UNIVERSITY



FACULTY OF SCIENCES

Academic year 2017-2018

Data Analysis Methods for Digital PCR Quantification

Author:
Matthijs Vynck

Promoters:
Prof. dr. ir. Olivier Thas
Prof. dr. ir. Jo Vandesompele

Dissertation submitted in fulfillment of the requirements for the degree of Doctor in Statistical Data Analysis

BioStat Research Unit
Department of Data Analysis and Mathematical Modelling

Members of the examination committee

Prof. dr. Els Goetghebeur (Chair)
Faculty of Sciences, Ghent University

Prof. dr. ir. Lieven Clement (Secretary)
Faculty of Sciences, Ghent University

Prof. dr. Ward De Spiegelaere
Faculty of Veterinary Sciences, Ghent University

dr. Jim Huggett
School of Biosciences and Medicine, University of Surrey

Prof. dr. Ziv Shkedy
Faculty of Sciences, Hasselt University

Promoters

Prof. dr. ir. Olivier Thas
Faculty of Bioscience Engineering, Ghent University

Prof. dr. ir. Jo Vandesompele
Faculty of Medicine and Health Sciences, Ghent University

Dean of the Faculty of Sciences

Prof. dr. Herwig Dejonghe

Rector of Ghent University

Prof. dr. ir. Rik Van de Walle

Copyright

Several parts of this thesis have been published as articles prior to the publishing of this thesis. These parts are either published under a Creative Commons license, or the copyright has been transferred to the publisher. In the latter case, permissions for copying, altering, distributing or other use of said parts need to be obtained from the publisher. A license for reuse of previously copyrighted materials for this thesis has been granted to the author of this thesis by the publishers.

For those parts for which the author and the promoter own the copyright, they give permission to copy these parts for personal use. Any other use is subject to copyright laws, more specifically the source must be explicitly specified when using parts of this thesis.

Summary

Quantifying nucleic acids is an important tool for studying a plethora of biological processes. This quantification can be done using several molecular biological techniques. The digital polymerase chain reaction (digital PCR) is a method that was first described some 30 years ago. However, applying the technique was expensive and time-consuming. With recent technological developments in the fields of microfluidics, among others, digital PCR has become easier to implement, the result of which is several commercially available platforms. This increased availability, combined with the promise of superiority of the technique when compared to other nucleic acid quantification techniques, has resulted in an ever-increasing interest in the technique, its properties and performance, and this for a wide variety of applications.

In such a digital PCR experiment, a sample is divided into a large number of partitions, each of which may or may not contain the target nucleic acid sequence. After the PCR has finished, a digital signal, representing the absence or presence of the target nucleic acid, is obtained by measuring the fluorescence intensity of each of the partitions. Assuming a random distribution of the target nucleic acid fragments among the partitions, this digital signal, together with knowledge on the partition volumes, may be used to obtain an estimate of the contration of the target nucleic acid in the original sample.

This thesis is concerned with statistical modelling for digital PCR, with the objective of improving the resulting estimates of nucleic acid quantities. In other words, the goal is to obtain exact (unbiased) estimates of e.g. the absolute concentration of a nucleic acid sequence in a sample, along with a correct estimate of the uncertainty with which this absolute concentration was determined.

To go from sample to eventual estimate of the concentration and its associated uncertainty, several steps need to be taken. Following a general introduction to nucleic acid quantification, chapter 1 introduces several problems that may be encountered in the process of obtaining such estimates.

Chapter 2 provides an overview of several properties of digital PCR, and how these compare with different alternative nucleic acid quantification techniques. This is followed by a discussion of the problems introduced in chapter 1, and a discussion of to what extent solutions have been suggested for these issues.

Chapter 3 describes methods that allow to design several types of quantification experiments, in a such way that these allow to answer a research questions in the most optimal way, i.e. it is concerned with how different parameters affect the uncertainty with which the nucleic acid sequence is quantified. We investigate the parameters that affect this uncertainty, and to what extent, and how these may be adjusted to reduce imprecision. To allow easy exploration of several scenarios, a web application for studying the effect of these parameters is also introduced.

Chapter 4 introduces a method whose goal is to improve the determination of absence or presence of the target nucleic acid based on its fluorescence intensity signal. We tackle this problem by relying on methods from the field of extreme value theory. These methods allow to model the distribution of fluoresnce intensity maxima of those partitions that do not contain the target nucleic acid. Based on this distribution, we subsequently determine a cut-off. Partitions with a fluorescence intensity measurement below this cut-off will be considered negative (they do not contain the target nucleic acid), while those partitions with a value above the cut-off will be considered positive (they do contain the target nucleic acid). In addition, we discuss several approaches to deal with sample-to-sample shifts in the location of fluorescence intensity signals, which may eventually affect the distribution of maxima, and consequently the resulting cut-off.

Chapter 5 is concerned with the application of the generalized linear mixed model framework to the digital outcomes observed in digital PCR experiments. This framework allows to correctly model the outcomes that results from complex experimental setups, in the sense that the framework allows to model the correlations between measurements that result from the design, e.g. the repeated measurement of a single sample, or the quantification of multiple nucleic acid within a given sample. The application of the generalized linear mixed model framework to different quantification scenarios, i.e. both absolute and relative

quantification, is laid out. In addition, we introduce a method for selecting stable reference genes in relative quantification experiments that is based on estimates from these mixed models, and allows to improve the quantification, both in terms of bias and precision.

Chapter 6 is concerned with the importance of appropriate modelling of the distribution of partition volumes. In this chapter, it is shown how ignoring partition volume variability, or the use of the wrong distribution of partition volumes, leads to a bias of the resulting concentration estimates. We introduce a method that allows to use a partition volume distribution of arbitrary form, and show how it may be used to reduce the bias that arises as a consequence of partition volume variability.

Chapter 7 discusses several aspects relating to the performance of digital PCR assays and platforms. The focus of this chapter is a comparative study of methods to determine the linearity of an assay or platform. In addition, several other topics related to the determination of trueness and precision are discussed, such as the importance of the number of repeated measurements of a sample with a given concentration within a dilution series. A web application for generating a quality report is also introduced.

Chapter 8 concludes this thesis, with a short summary of the material presented in chapters 2-7, followed by a discussion of outstanding issues and (i) how solving these issues may contribute to further improving the quantification with digital PCR, and (ii) how a further distribution and increased use of methods developed in the past years by the community may be achieve. We finish with a general conclusion.

Samenvatting

Het kwantificeren van nucleïnezuren speelt een belangrijke rol in het bestuderen van een brede waaier aan moleculair biologische processen. Deze kwantificatie kan gebeuren m.b.v. verschillende moleculair biologische technieken. De digitale polymerase kettingreactie (digitale PCR), is een methode die een kleine dertig jaar geleden voor het eerst beschreven werd, maar het uitvoeren van deze methode was tot voor kort tijdsrovend en duur. Door verdere technologische ontwikkelingen op het vlak van o.a. microfluïdica, zijn van digitale PCR recent verschillende implementaties vermarkt. Deze commerciële beschikbaarheid, in combinatie met de belofte van superioriteit van deze platformen over alternatieve methoden voor kwantificatie van nucleïnezuren, heeft gezorgd voor een sterke stijging van de interesse in deze methode, de eigenschappen van de verschillende platformen en hun prestaties in het kwantificeren van nucleïnezuren in allerhande toepassingen.

In een digitale PCR experiment, wordt een staal opgedeeld in een groot aantal partities. In elk van deze partities kan een nucleïnezuursequentie van interesse aan- dan wel afwezig zijn. Na het plaatsvinden van de PCR kan deze aan- of afwezigheid a.d.h.v. de intensiteit van fluorescentie in elk van de partities bepaald worden en wordt dus het digitale signaal bekomen. Bij een willekeurige verdeling van de fragmenten van het nucleïnezuur van interesse kan a.d.h.v. dit digitale signaal, en kennis over het volume van de partities, de concentratie van het nucleïnezuur van interesse in het oorspronkelijke staal bepaald worden.

In deze thesis bestuderen we hoe we m.b.v. statistisch modelleren de kwantificatie van deze nucleïnezuren kunnen verbeteren, in termen van een correcte kwantificatie. M.a.w. het uiteindelijke doel is het exact (onvertekend) schatten van bv. de concentratie van een bepaald nucleïnezuur in een staal, met daarbij een correcte inschatting van de onzekerheid waarmee deze schatting is gedaan.

Om van een staal tot de uiteindelijke schatting van de concentratie van het nucleïnezuur, en de bijhorende onzekerheid, in dat staal te komen, moeten verschillende stappen doorlopen worden. Na een algemene introductie van de mogelijkheden voor het kwantificeren van nucleïnezuren, worden in hoofdstuk 1 de problemen die zich kunnen voordoen bij het doorlopen van deze stappen geïntroduceerd.

Hoofdstuk 2 geeft een overzicht van de eigenschappen van digitale PCR en hoe deze zich verhouden tot andere kwantificatiemethoden, gevolgd door een discussie van de problemen geïntroduceerd in hoofdstuk 1 en in welke mate er voor deze problemen reeds oplossingen werden voorgesteld.

Hoofdstuk 3 beschrijft methodes die toelaten verschillende types kwantificatie-experimenten op dergelijke wijze op te zetten zodat deze zo optimaal mogelijk een antwoord kunnen bieden op de onderzoeksvra(a)g(en), m.n. bestuderen we hoe verschillende parameters een invloed hebben op de onzekerheid waarmee de kwantificatie plaatsvindt. We bestuderen welke parameters, en in welke mate, een invloed hebben op deze onzekerheid, en hoe deze kunnen aangepast worden om een preciezere kwantificatie te bekomen. Om verschillende scenario's met gemak te kunnen bestuderen, wordt hiervoor ook een webapplicatie geïntroduceerd.

Hoofdstuk 4 introduceert een methode die als doel heeft het bepalen van de aan- of afwezigheid van het nucleïnezuur van interesse in de verschillende partities te verbeteren. We gebruiken hiervoor methoden uit de extreme waardentheorie die toelaten een uitspraak te doen over de verdeling van de maxima van fluorescentie in een partitie waarin zich geen nucleïnezuur van interesse bevindt. Op basis van deze verdeling kan vervolgens een grenswaarde bepaald worden. Beneden deze grenswaarde zullen we partities als negatief (de partitie bevat geen nucleïnezuur van interesse) of positief (de partitie bevat het nucleïnezuur wel) classificeren. We bespreken tevens hoe de intensiteit van fluorescentie kan variëren van experiment tot experiment, wat uiteindelijk de grenswaarde kan beïnvloeden, en bespreken hoe dit effect best te neutraliseren.

In hoofdstuk 5 bespreken we de toepassing van het generalized linear mixed model framework op de digitale uitkomsten van digitale PCR experimenten. Dit framework laat toe om complexere experimentele setups correct te modelleren, in die zin dat het framework toelaat rekening te houden met de correlaties tussen metingen die voortvloeien uit het uitgevoerde experiment, bv. het herhaaldelijk

kwantificeren van eenzelfde staal, of het kwantificeren van meerdere nucleïnezuren in eenzelfde staal. Dit wordt uitgewerkt voor zowel absolute als relatieve kwantificatie. Bovendien wordt ook een algoritme geïntroduceerd die, op basis van de schattingen die voortkomen uit de generalized linear mixed models, toelaat de relatieve kwantificatie te verbeteren (in termen van onvertekendheid en precisie) door het selecteren van stabiele referentiegenen.

Hoofdstuk 6 behandelt het belang van het correct modelleren van de distributie van de volumes van de partities. In dit hoofdstuk wordt aangetoond hoe het negeren van variabiliteit in de volumes van de partities, of het gebruik van een verkeerde distributie van volumes leidt tot vertekening van de geschatte concentratie. We introduceren een methode die toelaat een specifieke, arbitraire distributie van volumes te gebruiken voor het schatten van concentraties.

Hoofdstuk 7 belicht verschillende aspecten van de performantie van digitale PCR assays of platformen. De focus van dit hoofdstuk is een vergelijkende studie van methoden die toelaten de lineariteit te bestuderen. Verder worden ook enkele belangrijke aspecten in het beoordelen van juistheid en precisie besproken, zoals het aantal herhaalde metingen bij elke concentratie van een dilutiereeks. Een webapplicatie voor het genereren van een kwaliteitsrapport wordt eveneens geïntroduceerd.

Hoofdstuk 8 vormt het sluitstuk van deze thesis, met een korte samenvatting van de hoofdstukken 2-7, gevolgd door een bespreking van enkele nog resterende problemen en hoe (i) het oplossen van deze problemen kunnen bijdragen tot een verdere verbetering van kwantificatie m.b.v. digitale PCR, of (ii) een verdere verspreiding en grootschaliger gebruik van de reeds bestaande methoden bekomen zou kunnen worden. Een algemene conclusie rondt dit hoofdstuk af.

Dankwoord

Het is intussen meer dan 10 jaar geleden dat ik voor het eerst voet zette op de campus Coupure, op een van de infodagen voor de studie bio-ingenieur. Ik wist toen helemaal nog niet of ik er ooit nog zou terugkeren, en zelfs toen ik er in september van dat jaar terugkeerde, was het onduidelijk hoe lang dat avontuur zou duren. Dat ik hier intussen zo veel tijd heb doorgebracht, heb ik aan veel mensen te danken. Dat ik niet de grootste verteller ben, weten die mensen intussen ook al lang, maar het dankwoord van een doctoraatsthesis is de gelegenheid bij uitstek om diegenen die het toch zouden durven vergeten of hun rol hierin onderschatten, daar toch even aan te herinneren. Een min of meer chronologisch overzicht van hoe jullie (on)rechtstreeks bijgedragen hebben aan dit boekje.

Mama, papa, jullie hebben mij steeds aangemoedigd en de kansen gegeven de dingen te doen waar ik goed in was, toch ook af en toe eens iets te proberen waar ik soms geen zin in had, te volharden wanneer de dingen vaak ook eens minder goed gingen. Mama, dat ik als kind reeds grote fan was van de boekjes met "wiskundige" vraagstukjes, en jij mij daarbij regelmatig hielp, heeft me wellicht mee gebracht waar ik nu ben. Op dagen waarop het allemaal wat minder ging, stond je ook steeds voor ons klaar. Papa, jouw grote interesse in computers en wat je daar allemaal mee kon doen, te meer als je zelf dingetjes kon programmeren, hebben destijds die interesse ook bij mij aangewakkerd. Op schoolvakantiedagen waarop ik zo nu en dan eens wat doelloos rondliep, gaf je me een boek over programmeren en zei je (het is lang geleden, dus ik parafraseer): "Misschien moet je daar wat in lezen, maar als je iets niet begrijpt, kom het dan maar vragen." Nu ik daaraan terugdenk, toont dat ergens wel een beetje gelijkenissen met een doctoraat, en hoewel de onderwerpen van de programmaatjes toen verre van vernieuwend waren, heb ik de jaren daarna daarvan vaak de vruchten kunnen plukken, en daar plezier aan kunnen beleven.

Vrienden van *de bende*, sommigen van jullie ken ik al vele jaren, maar onze bende

is door de jaren toch flink uitgebreid. Groei is niet steeds positief, maar in het geval van onze bende is het dat zeker wel. Hoewel er strikt gezien – en voor zover ik weet – geen paginalimiet is voor een dankwoord, zou een overzicht van onze blijde avonturen en herinneringswaardige tegenslagen en hoe die bijgedragen hebben tot mij en dit doctoraat toch ietwat te veel pagina's in beslag nemen. Jullie zijn stuk voor stuk prachtige mensen, en ik hoop dat er nog vele caféavondjes, spellenavonden, etentjes, quizzen, weekendjes, … mogen volgen. Toch wil ik twee mensen met een speciale rol in dit academisch verhaal vermelden. Matthias, van alle mensen van onze bende ken ik jou het langst. We hebben vele uren samen gesleten op de middelbare schoolbanken, waar m.n. de lessen fysica in mijn herinneringen dikwijls heel bovengemiddeld plezierig waren, en ook tijdens onze unief- en doctoraatsjaren was er altijd wel tijd om zo nodig even te ontspannen, al dan niet met een paar *Westmallekes*. Alweer een reden waardoor die studies bio-ingenieur en dat doctoraat wellicht uiteindelijk gelukt zijn. Tom, door jou hoorde ik voor het eerst over het bestaan van de opleiding bio-ingenieur, en ook samen hebben we dat eerste jaar met een beperkte hoeveelheid kleerscheuren doorstaan. Uiteindelijk gingen we allen onze eigen weg, maar 's avonds even uitblazen tussen pot en pint zijn we nooit verleerd.

Dat brengt me bij de *pingponggroep*, sinds vele jaren spelen we badminton, maar om de een of andere reden zijn we toch nog de *pingponggroep*. Bedankt voor die fijne avonden, het was nu en dan eens nodig om wat frustraties eruit te meppen, op een balletje dan wel een pluimpje, en daarna het verloren vocht te compenseren met een *duifje* of iets anders van gewoonte.

Terug tussen de boeken, in de masterjaren bio-ingenieur, kregen we eindelijk wat meer keuzevrijheid. De minor computationele biologie was snel gekozen. De keuzevakken en de masterthesis bij Tim en Wim deden verder mijn interesse in bio-informatica en statistiek aanwakkeren. Midden 2013 was dan eindelijk het moment van afstuderen gekomen, en ook rond dat moment van afstuderen kwam ik via een mailtje van Tim te weten dat er kandidaten gezocht werden voor een vacature bij de onderzoeksgroep van Olivier.

Olivier, bedankt mij kansen te geven in jouw onderzoeksgroep, voor het vertrouwen en de tijd die je geïnvesteerd hebt in mij en mijn begeleiding. Dat ik als assistent startte, betekende dat er geen vastgelegd onderzoeksonderwerp was. We startten bij één van jouw schijnbaar onuitputtelijke ideeën. Wanneer we ergens

tegen een muur aanliepen, kwam je ook steeds op de proppen met één of meerdere ideetjes om die problemen op te lossen, en naarmate het onderzoek vorderde, gaf je me veel vrijheid om ook eigen ideetjes uit te werken. Je aanhoudend enthousiasme en optimisme werkten aanstekelijk, ook wanneer er eens wat minder goed nieuws kwam. Ook voor het onderwijs en het samenwerken met anderen gaf je me veel vrijheid en kansen.

Vanaf het begin van mijn doctoraat was ook Jo nauw betrokken. Jo, jouw ervaring met de meer praktische kant van het onderzoek was een grote meerwaarde. Als geen ander wist je ook de vinger te leggen op de soms nog ontbrekende bedenkingen of niet geheel duidelijke formuleringen in de neerslag van ons onderzoek.

Het voeren van wetenschappelijk onderzoek is steeds een *team effort*: zelfs die steeds zeldzamere papers met een enkele auteur zijn uiteindelijk veelal het gevolg van vele discussies met collega's, een editor van een tijdschrift en peer reviewers. Ook ik heb tijdens de voorbije jaren de kans gehad met heel wat mensen te kunnen samenwerken. Ward, Wim, Kevin, Sarah, Candy, … het was steeds vlot en aangenaam samenwerken. Door mee te kunnen werken aan jullie onderzoek, of vice versa, had ik ook de kans om eens wat breder te kijken naar de relevantie en de toepassingen van sommige componenten van ons onderzoek.

To the members of the examination committee, prof. Els Goetghebeur, prof. Lieven Clement, prof. Ward De Spiegelaere, prof. Ziv Shkedy, dr. Jim Huggett: thank you for the time and effort spent on reviewing this work (and all the administration involved). Review times were short, but still I received many interesting questions and useful comments concerning the contents of this thesis. Different specialisations resulted in comments concerning a wide range of topics, some a bit uncomforting, but especially interesting as they were somewhat outside my area of focus. Your comments and remarks made me (re)consider some areas of application where our contributions may be useful and where they may not be, and at the same time, the comments and subsequent discussions resulted in useful suggestions for potential future work. Even if evaluations are sometimes a bit unpleasant, this one was a truly useful learning experience. Thank you once more!

Lief en leed werden ook dikwijls gedeeld tussen het werken door. Yingjie, jij was in mijn beginjaren op de vakgroep bureaumaatje van het eerste uur, leerde me het reilen en zeilen op de vakgroep kennen. Jan, jij was als ex-assistent steeds bereid een woordje uitleg te geven over de onderwijs-gerelateerde zaken die voor mij,

zeker in het begin, soms wat onduidelijk waren. Alain, jij zette steeds je beste beentje voor voor het verzorgen van het onderwijs waarvoor we verantwoordelijk waren. Joris, jij stond en staat al vele jaren altijd klaar om bij te springen waar nodig, uitleg te geven over of een oplossing te vinden voor alles wat betrekking heeft tot *scientific computing* en gerelateerde *good practices*. Karel, de planten in onze bureau hebben betere tijden gekend (en dan vnl. de erfstukken en die van mij), maar gelukkig is er steeds jouw enthousiasme om de boel op te fleuren! Gedurende de jaren was er nu en dan een stoelendans, een komen en gaan van collega's, in onze bureau maar ook die van de buren. Aan alle voornoemde (ex-)BioStatters, Federico, Klaas, Stijn, Alemu, Chamberlain, Gustavo, Annelies, Thomas: bedankt voor alle hulp en leuke intermezzo's! Michiel, bedankt voor de vele leuke momenten gedurende de voorbije jaren, zowel tijdens onze studiejaren, de jaren daarna op het werk als collega-assistent, alsook daarbuiten. Timpe, Ruth, Annie, Jan: bedankt voor jullie hulp bij de administratieve en IT-gerelateerde probleempjes!

Het werken aan een doctoraat geeft veel voldoening, maar dat het soms veel aandacht en energie opeist, en soms al eens wat frustraties met zich meebrengt, zullen ze ook thuis geweten hebben. De aanhoudende steun van het thuisfront was onmisbaar in het totstandkomen van dit boekje. Irene, van alle mensen in dit dankwoord zijn onze levens het meest verweven. Elke dag weer kan ik bij jou thuiskomen. Zoals Wannes het zo mooi met een kwinkslag kan zeggen (zingen): "'t Go voe stief veel geld moeten zien, ak u ooit verkope." ;-) Maar jouw rol in dit werk en in mijn leven proberen benoemen, hoe goed geprobeerd ook, zou die rol tekortdoen. Ik koester de vele herinneringen aan de voorbije jaren en kijk uit naar de toekomst met jou. Ook de andere mensen die al lang, of wat korter, deel uitmaken van de (schoon)familie zorgen – jaar in, jaar uit – voor momenten om naar uit te kijken en daarna op terug te blikken. Mama, papa, Peter, Nui, Lieve, Lieven, Joke, Thea, Karel, … bedankt voor de vele ontspannende momenten, geruststellende en motiverende woorden, de (minder) serieuze gesprekken, de uitstapjes, …, op naar nog veel zo'n momenten!

Matthijs
Gent, mei 2018

Contents

Copyright **v**

Summary **vii**

Samenvatting **xi**

Dankwoord **xv**

1 Introduction **1**

1.1 Detecting and quantifying nucleic acids 2

1.1.1 Nucleic acids 2

1.1.2 The polymerase chain reaction 3

1.1.3 Digital PCR 6

1.1.4 Alternatives 7

1.2 Introduction to digital PCR data analysis 12

1.2.1 Estimators of the concentration 12

1.2.2 Estimators of the imprecision of the concentration estimator 14

1.3 Statistical challenges in digital PCR data analysis 15

1.4 Outline 18

2 Current state, challenges and perspectives **21**

2.1 Abstract 22

2.2 Introduction 23

2.3 Applications of dPCR in virology 24

2.4 Potential benefits of dPCR 26

2.4.1 Increased diagnostic selectivity 26

2.4.2 Increased precision 28

2.4.3 Analytical sensitivity 29

2.4.4 Quantitative trueness 29

2.4.5 Quantitative linearity and dynamic range 31
2.5 Data analysis of digital PCR 31
2.5.1 Sources of variability and bias in dPCR experiments . . . 31
2.5.2 Design of dPCR experiments: implications for precision and power calculations 32
2.5.3 Misclassification of dPCR partitions 34
2.5.4 Concentration estimation 35
2.5.5 Data analysis tools 37
2.6 General conclusion and future perspectives 38

3 On the design of experiments 39
3.1 Abstract 40
3.2 Introduction 40
3.3 Materials and methods 42
3.3.1 Data and data processing 42
3.3.2 Power equations 43
3.3.3 Simulation study 44
3.3.4 Software 45
3.4 Results and discussion 45
3.4.1 The need for designing experiments 45
3.4.2 Parameters affecting the power 45
3.4.3 Preliminary experiments 54
3.4.4 Implications for platform and assay design 54
3.4.5 Interactive web tool for power calculation 55
3.5 Conclusion 55

4 A threshold determination procedure 57
4.1 Abstract 58
4.2 Introduction 59
4.2.1 Digital PCR data analysis 59
4.2.2 Digital PCR data analysis tools 62
4.2.3 A new digital PCR data analysis tool: ddpcrquant 63
4.3 Material and methods 64
4.3.1 Digital PCR datasets 64
4.3.2 Distribution of the negative droplet population 64
4.3.3 ddpcRquant for automated digital PCR data analysis . . . 64
4.4 Results 65

4.4.1 Droplet populations are not normally distributed 65
4.4.2 ddpcRquant for automated digital PCR data analysis . . . 65
4.4.3 ddpcRquant vs. QuantaSoft on an example dataset 68
4.5 Discussion 69
4.5.1 Considerations for extreme value theory 69
4.5.2 Comparison of ddpcRquant to QuantaSoft 70
4.6 Conclusion 71

5 Improved estimates with mixed models 73
5.1 Abstract 74
5.2 Introduction 74
5.3 Materials and Methods 76
5.3.1 Absolute quantification 76
5.3.2 Copy number variation 79
5.3.3 Reference gene stability 84
5.3.4 Generic formulation 85
5.3.5 Case study data 87
5.3.6 Data analysis 88
5.4 Results and Discussion 89
5.4.1 Variance modelling 89
5.4.2 Multiple reference loci 93
5.4.3 Case study results 97
5.4.4 Comments and further research 100
5.4.5 Software implementation 101
5.5 Conclusion 101

6 Bias reduction by modelling volume variability 103
6.1 Abstract 104
6.2 Introduction 104
6.3 Materials and methods 107
6.3.1 Quantification assuming a constant volume 107
6.3.2 Huggett's approach 107
6.3.3 Verification of the partition volume distribution 108
6.3.4 A non-parametric volume modelling approach 108
6.3.5 Simulation study 110
6.3.6 Implementation 112
6.4 Results and discussion 112

6.4.1 Verification of the partition volume distribution 112
6.4.2 NPVolMod: a new approach for modelling volume variability 114
6.4.3 Absolute quantification 115
6.4.4 Relative quantification 117
6.4.5 Impact of distribution misspecification 118
6.4.6 Limitations 121
6.4.7 Importance for accurate quantification 121
6.4.8 Implications for platform development 122
6.5 Conclusion 123

7 Evaluation of assay or platform quality 125
7.1 Abstract 126
7.2 Introduction 126
7.3 Material and methods 129
7.3.1 Data and data generation 129
7.3.2 Calibration curve fitting 129
7.3.3 Scenarios 130
7.3.4 Software 132
7.4 Results and Discussion 132
7.4.1 High R^2 values do not imply good linearity 133
7.4.2 Linearity: qualitative assessments 134
7.4.3 Linearity: quantitative assessments 137
7.4.4 Direct quantification trueness 140
7.4.5 Direct quantification precision 142
7.4.6 Guidelines and recommendations for the design and analysis of a quality evaluation experiment 143
7.4.7 Case study: dynamic range of a KRAS dPCR assay 145
7.4.8 dPCalibRate: a web tool for quality assessment 147
7.5 Conclusion 148

8 Outcomes, perspectives and conclusion 151
8.1 Research outcomes 152
8.2 A look towards the future 154
8.2.1 Design of experiments and platforms 154
8.2.2 Threshold setting validation 155
8.2.3 Combining uncertainty 156

8.2.4 Open data and an open data format 158
8.2.5 Software for data analysis 160
8.3 Conclusion 161

Bibliography **163**

A Introduction **181**
A.1 Maximum likelihood estimation for digital PCR 181

B Power **183**
B.1 Derivations 183
B.1.1 Absolute quantification 183
B.1.2 Copy number quantification 187
B.1.3 Relative quantification 190
B.2 Supplementary figures 191
B.2.1 Small effect size 191
B.2.2 Relative contributions of variance sources 193
B.3 Shiny web application 194

C Thresholding **195**
C.1 Primers, probes, amplicons and cycling conditions for the inhibition assay 195
C.2 Analysis of ddPCR inhibition on fluorescence caused by primer/probe mismatches 196
C.3 QuantaSoft dataview of fluorescent output with baseline shift . . 197
C.4 Implementation of extreme value theory 198
C.4.1 Glossary 198
C.4.2 Summary 199
C.4.3 General considerations for extreme value theory 199
C.4.4 NTC (pre)processing 205
C.4.5 Sample processing 207
C.4.6 Example 207
C.5 QuantaSoft export csv file 212
C.6 ddpcRquant annotation requirements before running a ddPCR experiment 213
C.7 QuantaSoft automated analysis methods vs. ddpcRquant 214
C.8 Automated analysis by Quantasoft for the NTCs of an HIV total DNA quantification assay 215

C.9 Automated analysis by QuantaSoft for the samples of an HIV total DNA quantification assay . . . 216
C.10 Normality assessment . . . 217
C.11 ddpcRquant output for an HIV total DNA quantification assay . . . 218

D Mixed models 219
D.1 R manual . . . 219
D.1.1 Data . . . 219
D.1.2 Data analysis . . . 220
D.1.3 R code . . . 226
D.2 Overview of trueness, precision and stability estimates . . . 231
D.2.1 Stability plots . . . 231
D.3 Case study results . . . 241
D.3.1 Introduction . . . 241
D.3.2 Sample 5 . . . 242
D.3.3 Sample 6 . . . 245
D.3.4 Sample 9 . . . 248
D.3.5 Sample 13 . . . 251
D.3.6 Sample 15 . . . 254
D.3.7 Sample 16 . . . 257
D.3.8 Sample 18 . . . 260
D.3.9 Sample 19 . . . 262
D.3.10 Sample 20 . . . 264
D.4 Statistical derivations and considerations . . . 266
D.4.1 Absolute quantification . . . 266
D.4.2 Copy number variation . . . 266
D.4.3 Confidence intervals . . . 268
D.4.4 Exchangeability . . . 269
D.5 Shiny manual . . . 270
D.5.1 Data management . . . 270
D.5.2 Analysis . . . 272
D.5.3 Figures . . . 276
D.5.4 Report . . . 276
D.5.5 Help . . . 277
D.6 Supplementary tables . . . 277
D.6.1 PCR assays . . . 277
D.6.2 Final copy number estimates and confidence intervals . . . 279

E Volume **281**
E.1 Algorithm for obtaining an unbiased estimate for absolute quantification . . . 281
E.2 A numerical example of obtaining an unbiased estimate . . . 282
E.3 Approach for relative quantification . . . 283
E.3.1 Algorithm for obtaining unbiased estimates . . . 283
E.3.2 Obtaining conditional densities . . . 283
E.3.3 A model for relative quantification estimates and associated variance . . . 283
E.4 Location - shape - scale . . . 285
E.5 Simulation results . . . 286
E.5.1 First distribution . . . 286
E.5.2 Second distribution . . . 289
E.5.3 Third distribution . . . 292
E.5.4 Fourth distribution . . . 295

F Quality **299**
F.1 Data simulation . . . 299
F.2 Ordinary versus (robust) weighted least squares . . . 299
F.3 Linearity . . . 303
F.4 Trueness . . . 305
F.5 R code . . . 305

G Curriculum vitae **307**
G.1 Education . . . 307
G.2 Employment . . . 307
G.3 International peer-reviewed journal publications . . . 307
G.3.1 In preparation, under review or revision . . . 308
G.4 Conference contributions . . . 309
G.5 Teaching . . . 311
G.6 Awards . . . 312
G.7 Other . . . 312

1 Introduction

1.1 Detecting and quantifying nucleic acids

This thesis is concerned with the detection and quantification of nucleic acids. More specifically, it is concerned with statistical challenges encountered in the detection and quantification of these nucleic acids using a specific method, termed the digital polymerase chain reaction, which will be introduced later (Section 1.1.3). We first focus on these nucleic acids: what are they? Perhaps more interesting to the applied scientist than asking what they are, is asking: why should we care, why are they so important?

1.1.1 Nucleic acids

Nucleic acids are strands or polymers of nucleotides. These nucleotides consist of a base (i.e. adenine, cytosine, guanine, thymine in the case of deoxyribonucleic acid or DNA; thymine is replaced with uracil in the case of ribonucleic acid or RNA) and a sugar (ribose or deoxyribose) phosphate, with the sugar phosphates forming the backbone of the strand of nucleic acid. A DNA molecule consists of two DNA strands, held together by hydrogen bonds between the complementary bases (forming base pairs): adenine pairs with thymine, cytosine with guanine (Figure 1.1). RNA typically occurs as a single strand, though it may occur as a double-stranded molecule as well.

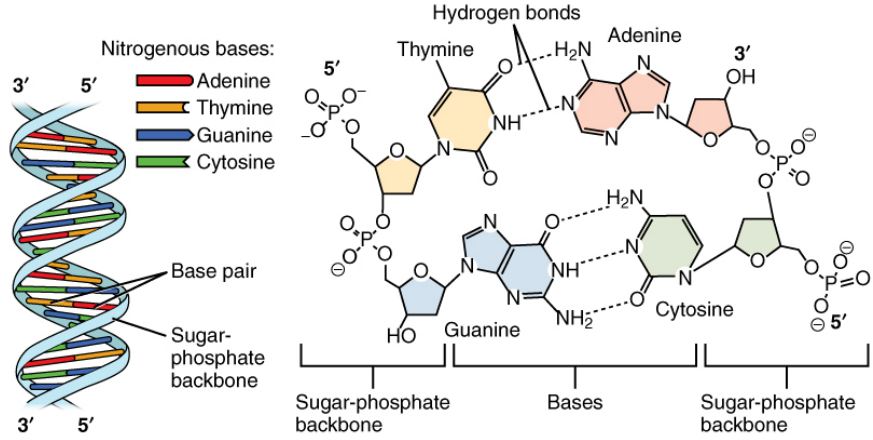


FIGURE 1.1: Structure of a DNA molecule. (Figure retrieved from OpenStax, Rice University.)

It is no surprise that many academic biochemistry and cell biology books commence the reader's journey of the cell's life with an introduction to these nucleic acids: they are essential to all life as we know it. Without nucleic acids, life as we know it would not exist, and I would probably not be writing this introduction, nor would you be reading it. Alberts et al. (2008) commence their cell biology classic with the following statements:

- all cells store their hereditary information in the same linear chemical code (DNA - deoxyribo*nucleic acid*),
- all cells replicate their hereditary information by templated *polymerization*,
- all cells transcribe portions of their hereditary information into the same intermediary form (RNA - ribo*nucleic acid*),
- ...
- all cells function as biochemical factories dealing with the same basic molecular building blocks.

It should be clear from these statements that nucleic acids form the basis of all of life, without exception. They are essential in storing and transmitting the organism's hereditary information.

A change in the nucleic acid i.e. the sequence of nucleotides or in the amount of nucleic acid present may have a tremendous effect on many aspects of the cell in which it is contained. A nucleic acid may give rise to proteins that will go on to catalyze reactions within the cell. A change in the cell's functioning may subsequently influence the functioning of the tissue and organism in which it is contained. Depriving the cell of a particular nucleic acid sequence, or the superfluous presence of a particular nucleic acid sequence within the cell can lead to a malfunctioning cell, and consequently to a malfunctioning tissue, organism, or disease. Studying particular nucleic acid sequences in the healthy and aberrant cell, tissue and organism can thus yield much insight into how this healthy or aberrant cell came to be the way it is.

1.1.2 The polymerase chain reaction

Suppose we are interested in studying a specific nucleic acid sequence: is it present, and how much of it is present? To study whether a DNA sequence is present, a

technique named the polymerase chain reaction (PCR) may be used[1]. To study how much of it is present, alternatives such as the digital PCR (among others) may be used (Section 1.1.3). Let's first focus on the conventional PCR.

Consider the following case: we are trying to find whether a piece of DNA (e.g. a sequence that is associated with increased chance of developing a disease) is present in the blood sample of a patient. This DNA sequence is often termed the *target* DNA. The human blood contains many fragments of circulating DNA. To find this particular piece of target DNA (of which just a few down to a single copy may be present) can be considered as a gigantic "Where's Wally" (Waldo) puzzle, with the complication that Wally (the target DNA) may not be present at all in the picture being assessed: among many disturbing factors, such as characters resembling Wally (DNA fragments resembling the target DNA), it is often quite hard to find Wally (the target DNA). The idea of the PCR is then to *amplify* the target DNA, that is to make more copies of the target DNA if it is present. If instead of a single Wally, there would be millions to billions of copies of Wally (while all other characters remain equal in numbers), it would be much easier to see whether Wally is present in the picture or not. As Wally's relative abundance is increased it would become hard to see anything else but Wally.

Generating more of the target DNA to make it better detectable was a puzzle that got a breakthrough solution in the early-mid 1980s when Kary Mullis invented the PCR, a feat for which he was later awarded the Nobel prize (Mullis et al., 1986; Rabinow, 2011). As was noted in Section 1.1, *polymerization* is a way of replicating nucleic acids using some template, a process occurring in every living cell. The contribution of Mullis et al. (1986) was making this process possible in vitro in such a way that a specific fragment of DNA can be polymerized in a reasonably short span of time (minutes)[2]. Without going into too much detail, some essential components for the polymerization to take place are the building blocks of the nucleic acid (nucleotides), a stable polymerization enzyme and two primers (oligonucleotides) that serve to demarcate the part of the DNA to be amplified (one at the start, one at the end). Additional components are added so that the polymerization takes place in an optimal, controlled (*buffered*) environment. A rough overview is presented in Figure 1.2. For the interested reader, a gentle but

[1] RNA may be detected, after the RNA has been converted to complementary DNA by reverse transcriptase (RT), i.e. RT-PCR.

[2] The typical length of human cell division (of which the DNA replication is part) in cell culture is approximately 24 hours (Cooper, 2000).

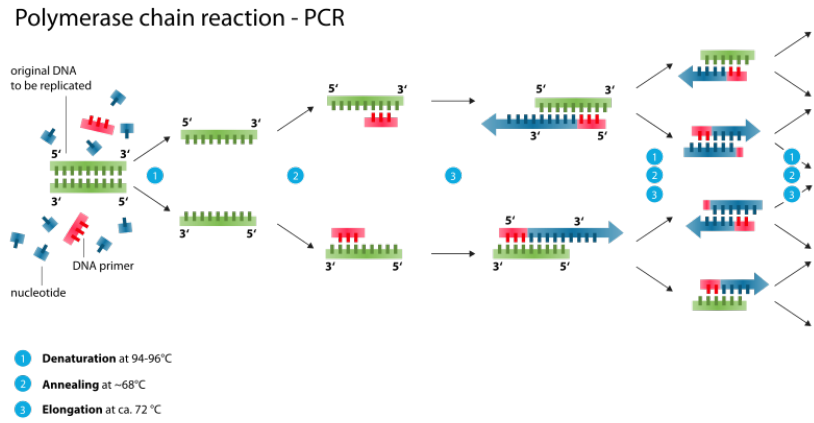


FIGURE 1.2: The amplification process can be considered a three-step procedure. It is initiated by increasing the temperature so that the double-stranded DNA breaks up into two single-stranded molecules (*denaturation*). In a second step, the temperature is decreased and primers bind to their complementary sequences on the single-stranded molecules (*annealing*). In a third and last step, the temperature is again increased and a polymerization enzyme (DNA polymerase, such as the Taq polymerase) then elongates the primer, constructing double-stranded DNA from a single strand of DNA to which it annealed (*elongation* or *extension*). This procedure is repeated many times, the result of which is an enormous number of DNA fragments. (Figure by Enzoklop, retrieved from Wikimedia Commons.)

more profound introduction is given by e.g. Lo and Chan (2006)[3].

The *chain reaction* arises from the repetition of the polymerization step. When a single template DNA (target DNA) is polymerized, then, after polymerization has completed, there is not one, but two copies of the target DNA. In a second polymerization step, these two target DNA fragments are again polymerized, so that we obtain four of them. This process is repeated many times (*cycled*), and in ideal circumstances the process results in the doubling of the target fragment with each cycle. After a typical 35-40 cycles, a single copy of the target fragment will then have turned into billions of copies, making its detection much easier. Detection of the target fragment is done by (i) measuring fluorescence intensity using dyes

[3]For the more audibly / visually inclined, there is also a PCR song and video: `http://bit.ly/2zPhlGz`. While you're at it, there is also a(n, arguably, even better) GTCA song / video, if you tend to forget your nucleotides: `http://bit.ly/1BTbPb1`.

intercalating double-stranded DNA, or (ii) the use of reporter probes, i.e. probes that emit fluorescence after hybridizing to a complementary DNA sequence.

1.1.2.1 Duplex and multiplex PCR

Multiple targets (two (duplex) or more (multiplex)) may be amplified concurrently. This is achieved quite easily by adding multiple sets of primers to the reaction mix. Multiplexing is advantageous in the sense that it reduces the amount of sample material needed and that certain sources of variability, e.g. inter-reaction variability, may be eliminated to some extent.

1.1.3 Digital PCR

In a digital PCR experiment, traditional PCR is performed, but instead of performing a single PCR on the complete reaction mix, this mix is partitioned into a large number (up to 10 million on modern platforms) of subreactions (*partitions*) that are typically assumed to be of equal size. Unlike traditional PCR, which allows for detection but not quantification of the target nucleic acid, this approach enables the quantification of the number of target fragments in the original sample.

Quantification is achieved by counting the number of subreactions in which the target nucleic acid is detected after amplification by PCR. For example, when 100 target nucleic acid fragments are contained within the reaction mix, and these 100 molecules are divided among 10 million partitions, it is expected that in about 100 of the 10 million partitions the target fragment will be detected (positive subreactions), while there will be no detection in the remainder (negative subreactions)[4]. The ratio of positive subreactions to the total number of subreactions can then be divided by the volume of the subreactions to obtain an estimate of the concentration of the target fragment in the reaction mix.

In practice, several complications arise in the detection and quantification of nucleic acid fragments using digital PCR. Many of these complications may be addressed using suitable statistical approaches, and they are the topic of this thesis. An introduction to the basic statistical analysis of digital PCR experiments

[4]While this statement is correct for a very low abundant target, it will no longer hold that x fragments are likely to result in x positive partitions when the abundance of the target is relatively large: the probability that a single partition contains more than a single fragment increases for an increasing abundance of fragments. A suitable estimator is presented in Section 1.2.

is given in Section 1.2, while (statistical) challenges are introduced in Section 1.3 and further discussed in Chapter 2. Some of these challenges are addressed in subsequent chapters of this thesis.

While digital PCR has been around since the late 1980s, commercial platforms only became available about two decades later [5]. Development of these commercial platforms has led to a sharp and still ongoing increase in the use of digital PCR (Section 1.1.4.5), supported by claims of superiority over quantitative PCR (see Section 1.1.4.1; a comparative assessment is presented in Chapter 2). A growing interest in digital PCR and the challenges the technique is faced with, motivate the development of suitable data analysis approaches.

1.1.4 Alternatives

Several alternatives for the detection and / or quantification of nucleic acids exist and are in common use. They will be introduced briefly. A comprehensive introduction is beyond the scope of this thesis, yet it is important to be aware of the existence of these alternatives, their strengths and weaknesses. It should be noted that in many cases these techniques are complementary rather than competing (Section 1.1.4.5).

1.1.4.1 Quantitative PCR

In a traditional PCR experiment a target fragment of nucleic acid is amplified, after which presence of the target fragment may be determined, for example with a fluorescent dye or a fluorescent probe which becomes fluorescent upon binding to the fragment sequence. It is therefore also named "end point PCR": whether the target fragment is present is determined at the end of the reaction[6].

Quantitative PCR differs from this approach in that the fluorescence intensity is measured continuously. The idea is that the fluorescence intensity represents the number of target fragments present: as more PCR cycles are performed, the number of target nucleic acids will increase exponentially. When a sufficient number

[5] Digital PCR was described several years before quantitative PCR, but quantitative PCR (Section 1.1.4.1) was commercialized before digital PCR was (Simmonds et al., 1990; Higuchi, 1992; Morley, 2014). Digital PCR is thus often called the third generation of PCR, the conventional PCR (Section 1.1.2) being the first, and quantitative PCR being the second (Bizouarn, 2014).

[6] Digital PCR is in most, but not all, commercial platforms also an end point PCR: fluorescence intensity in each of the partitions is measured only after the amplification cycles have finished.

of target fragments has been generated, the fluorescence intensity will also start to increase exponentially. The cycle at which a predetermined fluorescence intensity level is reached, is then representative of (and inversely proportional to) the original number of target fragments. To allow for absolute quantification of the number of target fragments in a given sample, a calibration curve generated from a dilution series with a known concentration of target fragments, is needed.

1.1.4.2 Blotting

In a Southern blot (for detection of DNA) or northern blot (for detection of RNA) the presence of specific sequences is interrogated using a hybridization probe that is complementary to the target sequence. The hybridization probe is typically radioactively or fluorescently labeled, and imaging is done by radiography or fluorescence measurement. Intensity of the resulting band is representative for the quantity of the fragment.

Before detecting absence or presence, or quantifying the number of target sequences, the nucleic acid fragments are separated by gel electrophoresis: due to the nucleic acid's negative charge, fragments will migrate through the gel away from the negative electrode (cathode) and towards the positive electrode (anode). The speed with which they move through the gel depends on their size: smaller fragments migrate more quickly as they experience less resistance from the gel. By including labeled fragments of known size (known as a *ladder* or *size marker*), the size of the fragments may be determined.

1.1.4.3 Sequencing

Sequencing approaches exist in many different forms, and a detailed discussion is beyond the scope of this thesis. A recent introduction to the history and current state of sequencing is given by e.g. Shendure et al. (2017).

So-called second generation, massively parallel or high-throughput sequencing approaches are now commonly used to detect and quantify RNA and DNA fragments. Their focus is on determining the particular sequence of these nucleic acid fragments, and do so in a parallelized fashion, i.e. to produce typically millions of sequences. The focus is on detecting and quantifying many different sequences, often across the genome or transcriptome, whereas PCR techniques

typically focus on a single or few sequences of interest. Depending on the particular application, this higher throughput comes at the cost of a lower precision and sensitivity (Section 1.1.4.5).

1.1.4.4 Microarrays

Just like sequencing approaches, the focus of microarrays is on detecting and quantifying many sequences, rather than just one or a few. As with sequencing, the target fragments may be either DNA (e.g. array comparative genomic hybridization) or RNA (e.g. expression profiling arrays). The general principle is the hybridization of fluorescently labeled sample nucleic acid to nucleic acid sequences present on the array. The latter are referred to as probes and are fixed on the array at known positions. Non-binding nucleic acids are washed away, after which fluorescence intensity is measured, indicating if (presence of a signal above background) and how much (comparative strength of the signal) of the specific nucleic acid is present. Many variations exist, and their applications are diverse.

1.1.4.5 Competing or complementing?

Notwithstanding that several of these methods may be used to answer the same question, there is often a single method that is preferable over the others due to its specific characteristics. Examples of factors to consider include sample throughput (how many samples do I need to analyze?), nucleic acid throughput (how many targets within a sample do I need to study?), cost (how much will the total experiment cost and what results do I get in return?), precision (how precise do my measurements need to be to be able to answer my question?), …

Perhaps the most obvious "competition" for digital PCR is the quantitative PCR (Section 1.1.4.1): both are able to detect and quantify nucleic acids in a reasonably precise way when compred to other alternatives. Yet, both have their place in the current-day lab: while quantitative PCR beneftis from e.g. lower per target cost and higher throughput, digital PCR may be the preferred technique when highly precise measurements are needed, or when sample quality is substandard (parts of Chapter 2 provide comparative analyses of both techniques). For example, when the biologically relevant quantity change is about two-fold or higher, qPCR will

typically suffice. When, on the other hand, the relevant change is (much) less than 100%, the often increased precision of dPCR may be required.

The northern blot (for RNA) and Southern blot (for DNA) provide some advantages over digital (reverse transcriptase (RT)) PCR in that they readily supply information on e.g. fragment size and alternative splicing (different fragment sizes) (Streit et al., 2009). However, one of its major disadvantages is its limited quantitative nature, relative large amount of nucleic acid requirement, and low sensitivity as compared to digital (RT) PCR (Streit et al., 2009).

Both microarrays and sequencing technologies allow for detection and quantification of nucleic acids, yet their strengths are very different from that of the PCR-based approaches: microarrays and sequencing allow for quantifying thousands of nucleic acid sequences simultaneously, albeit at a lower precision and sensitivity than PCR-based approaches. Moreover, microarrays and especially sequencing approaches are more costly, and typically require more time and bioinformatics before final results are obtained. Because of their characteristics, microarrays and sequencing approaches are typically considered tools for discovery settings, e.g. in research environments, while the modern PCR-based approaches are considered more appropriate for confirmation and / or precise and sensitive detection and quantification, e.g. in diagnostic settings. [7]

1.1.4.6 Changing popularity

Each of these approaches have their specific strengths and weaknesses and consequently may co-exist, yet some methods seem to be going out the door and are being replaced by alternatives: the older blotting methods have been consistently declining in popularity over the last decade, and also microarrays have been declining in recent years. Quantitative PCR has more or less maintained its popularity. There has been a steady increase in the use of digital PCR and sequencing approaches (Figure 1.3)[8].

[7] Note that is not generally true. For certain applications, sequencing is the method of choice in a diagnostic setting.

[8] Note that the data given in Figure 1.3 represent the number of publications that mention these specific techniques. In how many publications each of the techniques was effectively applied was not assessed. Moreover, the number of publications per year is estimated to be growing by 8-9% annually (Bornmann and Mutz, 2015), so that the absolute number of mentions should be interpreted with care.

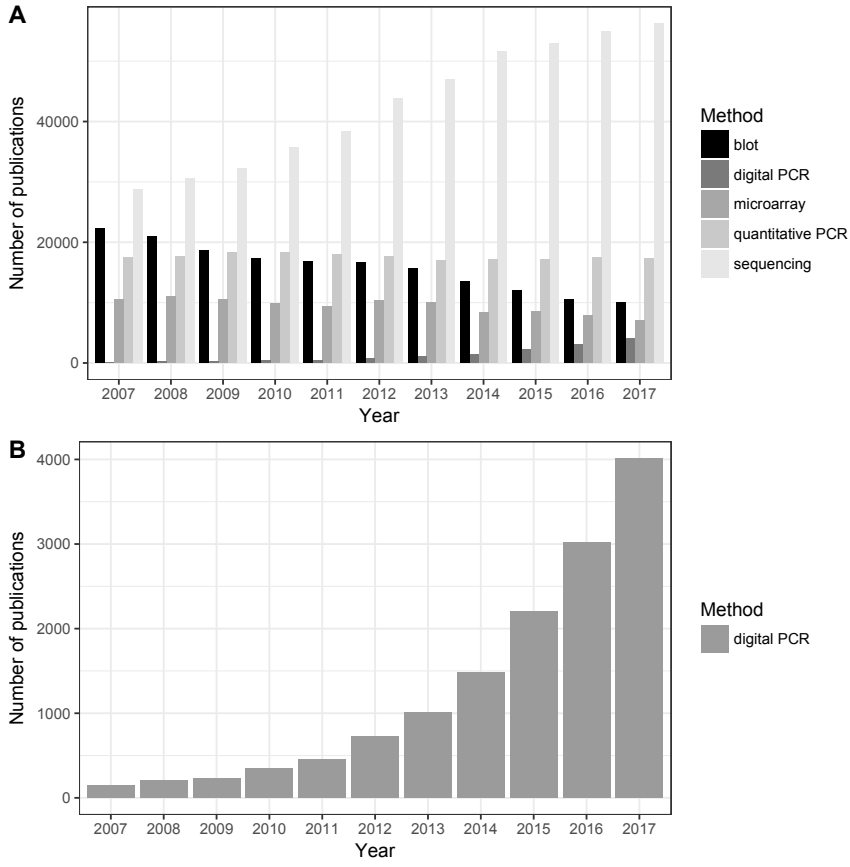


FIGURE 1.3: Number of publications by year; retrieved from Google Scholar using the search query "digital PCR", "quantitative PCR", "northern blot" or "southern blot", "DNA microarray", and "DNA sequencing" or "RNA sequencing". Data was retrieved on March 6, 2018. (A) Comparison of different nucleic acid quantification methods. (B) Increasing popularity of digital PCR during the last decade.

1.2 Introduction to digital PCR data analysis

1.2.1 Estimators of the concentration

In a digital PCR experiment, a total number of target fragments l is present in the PCR reaction mix. Before performing the PCR, this mix is randomly partitioned into n partitions. Denote k the number of positive partitions, i.e. the partitions that contain at least one copy of the target fragment. Given a relatively small ratio of target fragments to total reactions, it is likely that the l target fragments will result in an equal number of positive partitions, i.e. $k = l$, and a simple estimator of the average number of target copies per partition, λ, is then

$$\hat{\lambda}_{\text{simple}} = \frac{k}{n}, \tag{1.1}$$

while an estimator of the concentration c is given by

$$\hat{c}_{\text{simple}} = \frac{k}{n \times V_{\text{partition}}}, \tag{1.2}$$

where $V_{\text{partition}}$ is the volume of a partition, here considered a fixed constant.

However, when the fraction of occupied partitions increases, the estimator above is no longer suited as with an increasing fraction of occupied partitions there is an increasing probability for a partition to contain more than a single fragment. Assuming that the fragments are partitioned randomly among the subreactions, the counts of target fragments in a partition can be described by a Poisson distribution with mean λ. An estimator for the mean may be obtained using the fraction of negative partitions. When Y_i denotes the number of target copies in partition i, then

$$P(Y_i = 0) = \exp(-\lambda)\frac{\lambda^0}{0!} = \exp(-\lambda), \tag{1.3}$$

from which we find the estimator

$$\hat{\lambda} = -\log\left(1 - \frac{k}{n}\right). \tag{1.4}$$

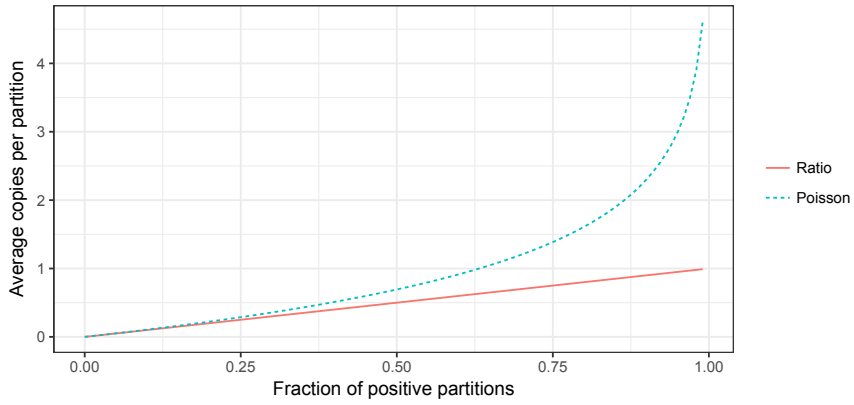


FIGURE 1.4: Comparison of two estimators for a total of 10 000 partitions. The ratio estimator corresponds to Equation 1.1. The Poisson estimator corresponds to Equation 1.4. The latter is unbiased, while the former shows little to no bias for (very) small fractions of positive partitions.

Hence, an estimator of the concentration is given by

$$\hat{c} = -\frac{1}{V_{\text{partition}}} \log\left(1 - \frac{k}{n}\right). \tag{1.5}$$

The difference between Equations 1.1 and 1.4 is displayed in Figure 1.4.

Equation 1.4 is thus preferable as it gives unbiased estimates for λ for a large range of fractions of positive partitions.

Note that a third and most efficient (Section 1.2.2) estimator of the average number of copies per partition is just the sample mean of the number of copies per partition,

$$\hat{\lambda}_{\text{count}} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \tag{1.6}$$

yet given that only the digital signal is observed, it is not applicable to the digital PCR.

The estimators presented in this section provide estimates of the average number of copies per partition, or the concentration of the sample that was analysed. Only a limited number of partitions was analyzed, causing some uncertainty about

these estimates: if we were to repeat the experiment, the random partitioning of the nucleic acid fragments may result in a higher or lower fraction of positive partitions than was observed in the first experiment. Thus, even when the number of nucleic acid fragments in the original sample remains the same, the resulting estimate of e.g. the average number of copies per partition may differ from experiment to experiment. This necessitates the characterization of the imprecision of these estimators to allow a meaningful conclusion about the sample as a whole. Estimators of the imprecision are introduced next.

1.2.2 Estimators of the imprecision of the concentration estimator

An estimator of the variance of $\hat{\lambda}$ may be derived (Appendix A.1) as

$$\widehat{\mathrm{var}}(\hat{\lambda}) = \frac{1-\exp(-\hat{\lambda})}{n\exp(-\hat{\lambda})}. \tag{1.7}$$

The estimator using the digital outcome (Equation 1.4) is less efficient than the estimator which uses the unobservable counts (Equation 1.6), as is illustrated in Figure 1.5.

An approximate 95% confidence interval of λ is easily constructed from its estimated value and variance,

$$\left[\hat{\lambda} - 1.96\sqrt{\frac{1-\exp(-\hat{\lambda})}{n\exp(-\hat{\lambda})}}, \hat{\lambda} + 1.96\sqrt{\frac{1-\exp(-\hat{\lambda})}{n\exp(-\hat{\lambda})}}\right]. \tag{1.8}$$

Likewise, an approximate 95% confidence interval of c is given by

$$\left[\hat{c} - 1.96\sqrt{\frac{1-\exp(-\hat{\lambda})}{nV_{\text{partition}}^2\exp(-\hat{\lambda})}}, \hat{c} + 1.96\sqrt{\frac{1-\exp(-\hat{\lambda})}{nV_{\text{partition}}^2\exp(-\hat{\lambda})}}\right]. \tag{1.9}$$

The formulae derived within this Section (Section 1.2) are only applicable for very simple experimental setups, i.e. analysis of a single reaction, and under a strict set of assumptions, e.g. that the partition volume $V_{\text{partition}}$ is constant, that turn out to be violated quite often in practice. These challenges are introduced next.

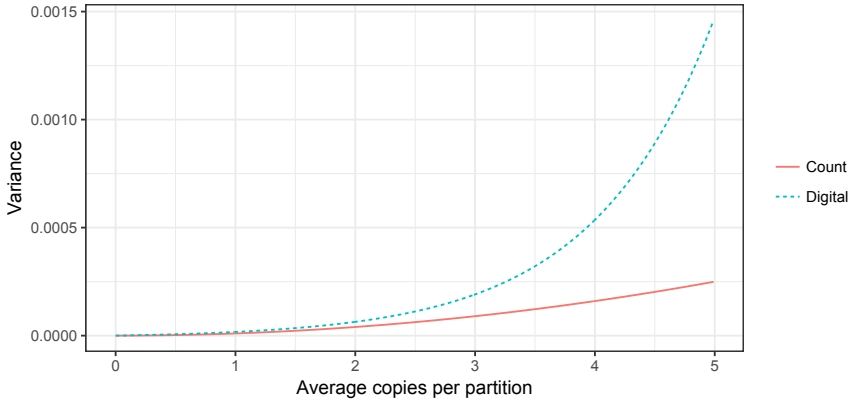


FIGURE 1.5: Comparison of variance of the two estimators for a total of 10 000 partitions. The variance of the estimator using the digital outcome is larger than when the exact counts would have been available, i.e. it is a less efficient estimator. The difference is especially large for larger values of the average number of copies per partition.

1.3 Statistical challenges in digital PCR data analysis

While the statistical analysis presented in Section 1.2 is quite straightforward, complications arise in practice. A good and comprehensive overview is given in Jacobs et al. (2014) and Majumdar et al. (2015). We summarize the main issues.

- It is not always easy to determine whether target nucleic acid is present or absent in a given partition. The absence or presence of the target nucleic acid in a given partition is determined from a continuous fluorescence intensity measurement of that partition. In well-designed and optimized assays and for relatively pure samples (e.g. no inhibitors, no contaminants, ...) there is often a clear distinction between the fluorescence intensity measurements. A fixed fluorescence intensity threshold resulting in a good separation between the fluorescence intensity measurements from positive and negative partitions may, in such cases, quite easily be determined. The distinction may be more difficult for samples of lower quality. Indeed, in such cases the fluorescence intensities may not be well separated and determination of absence or presence of the target nucleic acid is no longer straightforward, resulting in the misclassification of some partitions.

Figure 1.6 represents fluorescence intensity measurements from some typical samples. Misclassification is of little concern in Figure 1.6A: there is clear separation between both clusters, moreover, there are many partitions clearly belonging to one of both clusters so that misclassification of some partitions will have a minor effect on the resulting estimates. Figure 1.6B represents a sample similar in target concentration to Figure 1.6A, yet the separation between both clusters is less clear. However, as noted before, misclassification of a limited number of partitions will have a minor influence on the final estimates as both the positive and negative cluster are quite large.

Figure 1.6C and Figure 1.6D show samples with a low amount of target DNA. As long as separation is quite clear (Figure 1.6C), misclassification will be limited and the resulting estimates useful. Figure 1.6D represents a difficult sample: there are little positive partitions (if any), and misclassification of even a handful of partitions will have a major effect on the resulting estimates. If, for example, only 10 partitions are truly positive, but an additional five are erroneously classified as positive, the bias will be $\sim$50%. This as opposed to the sample in Figure 1.6B: even when 100 partitions are misclassified, the bias will remain limited to $\sim$10%.

In summary, preventing misclassification can be of high importance if the method is to result in unbiased (limited bias) estimates.

- Researchers across research disciplines in the life sciences will often perform experiments that consist of (i) multiple samples, (ii) multiple replicates of the same sample, (iii) multiple targets within the same sample, (iv) assessment of samples from the same individuals at different time points, … Depending on the exact experimental design dependencies among measurements may arise. They need to be taken into account when calculating e.g. concentration estimates, and especially imprecision estimates for e.g. the concentration estimates. Indeed, analyzing the data with a model different than that suggested by the experimental design will often lead to unbiased first moments, but (often downwardly) biased second central moments. Optimistic (too small) imprecision estimates and, when hypothesis testing is involved, too low p-values and an inflation of the false positive rate may result.

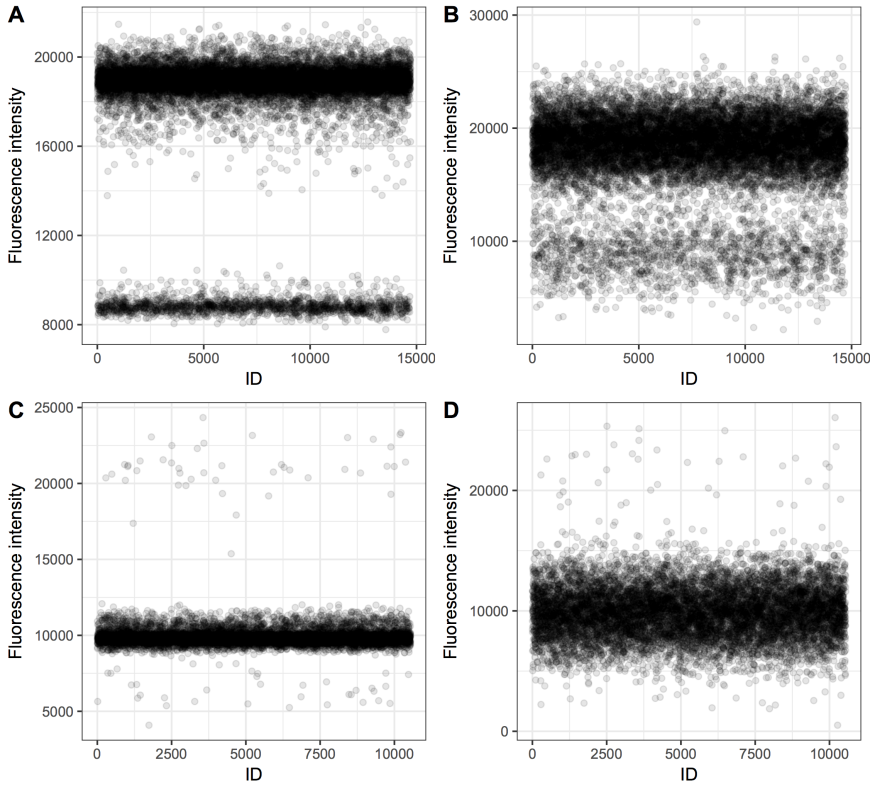


FIGURE 1.6: Examples of typical fluorescence intensities obtained for singleplex digital PCR experiments. Panels A and B represent samples with high levels of the target DNA (many positive partitions). Panels C and D represent samples with low levels of the target DNA (few positive partitions). In panels A and C positive and negative partitions are nicely separated, while the opposite is true for panels B and D. The latter two samples are likely the suffer from misclassification of partitions.

In addition, underestimation of the imprecision will invalidate power calculations: required number of technical replicates, required number of partitions, … will typically be underestimated, and the resulting experiments will consequently be underpowered. Ultimately, when the power calculation ignores these dependencies but the analysis accounts for them, the result may be false negatives, i.e. an unacceptable false positive risk arises (Colquhoun, 2014).

- Estimation of the final concentration in a given sample requires knowledge of the partition volume (Equation 1.5). This, obviously, requires knowledge of the exact partition volume. Yet, (i) the partition volume for at least one system has been revised following scientific evidence suggesting an overestimation of that partition volume (Pinheiro et al., 2011) and (ii) there is evidence that suggests that even for a given digital PCR system the partition volume changes between experiments (Kosir et al., 2017). In addition, the estimators of Section 1.2 implicitly assume that the partition volume is a fixed constant for all partitions. When the partition volume varies over partitions, ignoring the partition volume variability will lead to a downward bias of the concentration estimate (Jacobs et al., 2014; Huggett et al., 2015; Majumdar et al., 2015).

1.4 Outline

The remainder of this thesis consists of published or submitted manuscripts (Chapters 2-7) and is concluded with a discussion (Chapter 8). These manuscripts share a focus on data analysis procedures for digital PCR experiments. These procedures are presented in the order in which they are typically applied by the practitioner analyzing her/his data.

Chapter 2 is a review with a focus on applications of digital PCR within the field of virology, and a review of data analysis challenges and procedures for digital PCR.

Chapter 3 provides power equations for the design, or post-hoc power verification, of digital PCR experiments, for both absolute and relative quantification. A discussion of parameters affecting the power is provided, as well as a web application for power calculation.

Chapter 4 describes a thresholding procedure for the conversion of continuous fluoresence intensity measurements into a digital outcome (absence or presence of target nucleic acid). This research is motivated by the lack of automated thresholding procedures taking the characteristics of the fluorescence intensity measurements into account. Indeed, all of the approaches available at the time of publication assumed a normal distribution of fluorescence intensity measurements. Our contribution is a thresholding procedure that is non-parametric in the sense that it does not assume a particular distribution of the fluorescence intensity measurements.

Chapter 5 describes the application of the generalized linear mixed model framework for the analysis of experiments that go beyond the "single target, single measurement" design (introduced in Section 1.2). It was (is) common to ignore the design of experiments in the subsequent data analysis of digital outcomes. We describe how different generalized linear mixed models may be used to estimate absolute quantities of target nucleic acid, as well as relative quantities and copy numbers, while accounting for interreplicate variability, variability between samples, … In addition, we describe a measure based on estimates resulting from these generalized linear mixed models, to select stable reference genes in relative quantification and copy number experiments.

Chapter 6 provides an approach that models the typical variable size of digital PCR partitions. We describe how an arbitrary distribution of partition sizes can be incorporated in the generalized linear (mixed) model framework introduced in Chapter 5, so that bias due to the partition size variability may be reduced. A comparison is provided with a competing approach and the common practice of simply ignoring partition size variability.

Chapter 7 is concerned with the study of digital PCR assay quality / performance parameters. This chapter is motivated by the adoption of the R^2 measure and seemingly relevant graphical displays for linearity assessment, as commonly used in quantitative PCR and other (bio)analytical chemistry methods. We show why they are inadequate and suggest some alternatives. We provide some guidelines on (designing experiments for) assessing linearity, trueness and precision in digital PCR experiments, along with a web application for easy assessment of these assay quality parameters.

Chapter 8 provides a discussion of the presented work. A look towards the future with still outstanding statistical issues and important work remaining concludes

this thesis.

The manuscripts making up Chapters 2 through 7 are the result of collaborations with many different people. The contributions of each of the collaborators is mentioned at the start of each chapter.

2 Current state, challenges and perspectives

This chapter is an adaptation from the paper: Vynck, M., Trypsteen, W., Thas, O., Vandekerckhove, L. and De Spiegelaere, W. (2016). The future of digital polymerase chain reaction in virology. Molecular Diagnosis and Therapy, 20(5), 437-447.

Contributions: Matthijs Vynck and Ward De Spiegelaere conceived the contents and structure of, and wrote, the manuscript. Wim Trypsteen, Olivier Thas, and Linos Vandekerckhove reviewed and commented on a first draft of the manuscript. All authors agreed on the final version.

2.1 Abstract

Driven by its potential benefits over currently available methods and the recent development of commercial platforms, digital PCR has received increasing attention in virology research and diagnostics as a tool for the quantification of nucleic acids. The current technologies are more precise and have high trueness, but may not be much more sensitive compared to qPCR applications. The most promising applications with the current technology are the analysis of mutated sequences, such as emerging drug resistant mutations. Guided by the recent literature, this review focuses on three aspects that demonstrate the potential of digital PCR for virology researchers and clinicians: the applications of digital PCR within both virology research and clinical virology, the benefits of the technique over the currently used real-time quantitative PCR and the importance and availability of specific data analysis approaches for digital PCR. Comments are provided on current drawbacks and often overlooked pitfalls that need further attention to allow a widespread implementation of digital PCR as an unbiased and precise tool within the field of virology.

Key points:

- Digital PCR may provide an accurate direct quantification for nucleic acids of viral origin.
- Current digital PCR platforms are in several cases more unbiased and more precise, but not more sensitive compared to qPCR.
- New data analysis tools are required to enable accurate digital PCR quantification.

2.2 Introduction

The invention of the polymerase chain reaction (PCR) in 1985 by Kary Mullis to detect and amplify minute amounts of target DNA is marked as one of the most revolutionary techniques that allowed the development of current day biomolecular research (Mullis et al., 1986; Bartlett and Stirling, 2003). Shortly after its invention, researchers explored techniques to make PCR a quantitative tool for the determination of nucleic acid concentrations. To date, the real-time quantitative PCR (qPCR), developed in 1992, has become the most widely used method for PCR-based analysis (Higuchi, 1992). However, an alternative technique, nowadays called digital PCR (dPCR), was already explored in the late 1980s, well before the invention of qPCR (Simmonds et al., 1990; Morley, 2014).

dPCR is a method that involves the partitioning of a sample in multiple partitions, forming isolated replicate PCRs. This partitioning, performed on sufficiently diluted samples, results in few template DNA molecules per partition. In this limiting dilution setting, the number of target molecules in the different partitions follows a Poisson distribution. Hence, an absolute quantification can be obtained when the number of positive and negative partitions is measured (Figure 2.1) (Vogelstein et al., 1999).

Even though this is straightforward in theory, the generation of multiple partitions for a single sample suffered from practical and cost constraints, hampering the broad application of dPCR for quantitative purposes. However, with the recent development of microfluidic techniques, multiple partitions can now be generated at reasonable costs. Indeed, since 2010 the number of publications on dPCR has grown drastically, indicating an increasing application of this technology. Furthermore, dPCR offers a method of direct quantification that does not rely on a predesigned standard curve, as is the case for qPCR. This has two important benefits. First, standard curves are usually calibrated by spectrophotometry which may introduce extra bias in the quantification results (Hayden et al., 2015). Second, qPCR assumes that the PCR efficiency in the unknown samples is exactly equal to that in the standard curve. In reality, this assumption is never entirely met, resulting in additional variation, especially at low-level target concentrations (Pfaffl, 2001; Trypsteen et al., 2015a).

To this day, qPCR remains the method of choice for a wide range of applications in diagnostic virology but dPCR has received increasing attention in this field.

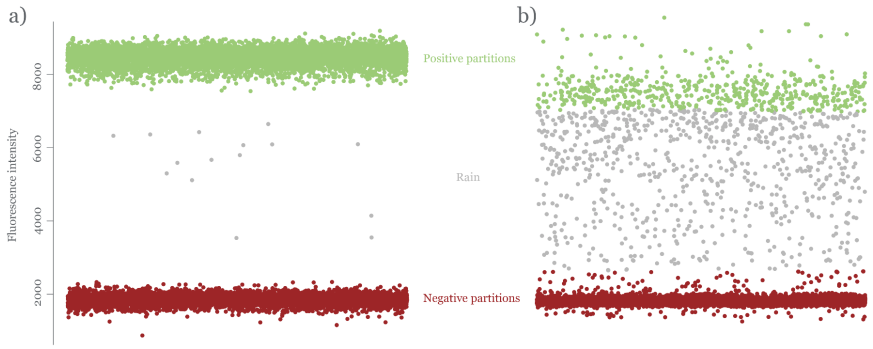


FIGURE 2.1: Fluorescence readouts from a typical dPCR experiment, showing the endpoint fluorescent values of the partitions (droplets). **a** The negative partitions (red) are clearly distinguishable from the positive partitions (green), but a small amount of partitions with intermediate fluorescence (grey) can be observed and are called rain. **b** Due to the presence of a large amount of rain (grey), the distinction between negative (red) and positive partitions (green) is not clear, and setting an inappropriate threshold may result in a large number of false positives or false negatives.

Nowadays, dPCR is increasingly being used in virological research settings and several reports have provided initial data for its possible use in a clinical setting. In accordance with quantitative PCR, digital PCR applications will require an extensive validation (Bustin et al., 2009; Huggett et al., 2013). In addition, dPCR requires its own data analysis workflow. To this date, only few dPCR data analysis strategies have been described, likely because of the very recent rise in popularity of this technology. The present review will explore and focus on three important aspects of dPCR for its possible use in virology. We provide a brief overview of possible dPCR applications in virology, both in a research setting and in a diagnostic setting. We describe the benefits of digital PCR platforms compared to qPCR platforms, and lastly, we discuss recent developments in dPCR data analysis.

2.3 Applications of dPCR in virology

In current virology research, absolute quantification by dPCR has mostly been explored and compared with qPCR for standard viral load testing (Hayden et al., 2013) and low-level detection of viruses (Brunetto et al., 2014; Henrich et al., 2012;

Kiselinova et al., 2014a; Strain et al., 2013; White et al., 2012). Examples of the latter are the quantification of persisting viruses, such as reservoirs of the human immunodeficiency virus (HIV) in HIV infected individuals on antiretroviral therapy (Henrich et al., 2012; Strain et al., 2013; Kiselinova et al., 2014a, 2015, 2016) or low-level detection of GB Virus Type-C (GBV-C) in cell lines (White et al., 2012), human T-lymphotropic virus (HTLV) in cerebrospinal fluid (Brunetto et al., 2014) and covalently closed circular DNA (cccDNA) in HBV infected individuals (Mu et al., 2015). dPCR was recently described for direct quantification of viral nucleic acids without initial DNA or RNA isolation. Hence, dPCR provides a quantitative strategy that minimizes pre-PCR processing variation and prevents an excessive loss of target sequences due to the isolation procedure (Pavsic et al., 2016). Alternatively, dPCR has been frequently explored for testing low-levels of viral contaminants in environmental samples (Kishida et al., 2014; Racki et al., 2014a; Coudray-Meunier et al., 2015). These studies indicate that dPCR may indeed form a promising tool in standard diagnostic virology. However, an extensive validation of this platform will still be required to move to a clinical setting.

Apart from absolute quantification, dPCR has also been developed to quantify ratios of specific template molecules to assess emerging mutations or copy number variations (CNV) of particular genes (Hindson et al., 2011). The higher quantitative trueness and precision of the direct quantification of multiplex dPCR experiments make this technology especially interesting to monitor small changes in ratios between a target gene and a reference; for example gene copy number variations or rare mutant detection (Hindson et al., 2011). In oncology, the percentage of emerging mutations can be identified and monitored more accurately with dPCR compared to qPCR (Pekin et al., 2011; Wang et al., 2010). This dPCR approach may apply well in the setting of virology because emerging drug resistance mutations also need to be screened/monitored. In this context, a small number of papers have already explored this possibility, indicating that dPCR is more sensitive compared to qPCR at detecting low levels of emerging drug resistant mutations (Mukaide et al., 2014; Taylor et al., 2015; Whale et al., 2016). Alternatively, dPCR can also be used for comparing the ratio of viral sequences to the number of haploid cellular genomes. In a method reminiscent of CNV analysis, dPCR was employed for assessing the ratio of Human Herpes Virus 6 (HHV-6) molecules to cellular genomes. dPCR can be used to reliably asses the ratio of HHV-6 genomes to human genomes in order to avoid unnecessary treatment of persons with chromosomally integrated HHV-6 (Sedlak et al., 2014a).

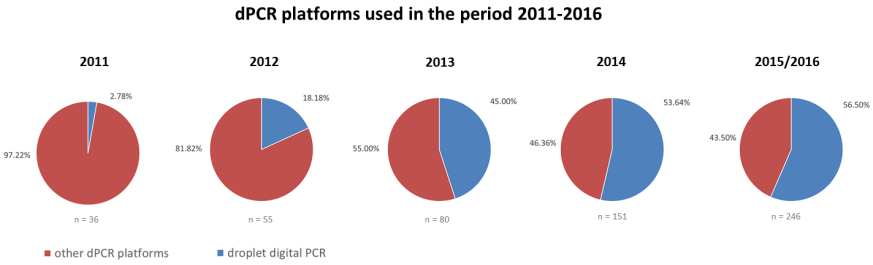


FIGURE 2.2: Indication of the use of the droplet dPCR platform (Bio-Rad) compared with other dPCR platforms in the recent literature.

Similarly, Ota et al. (2012) used dPCR to show different copy numbers of Merkel cell polyomavirus (MCPyV) compared to the number of haploid genomes in skin tumours.

The possible applications of dPCR in virology seem numerous, but a careful analysis of the current dPCR platforms is needed to assess if the theoretical advantages over qPCR are also attained with the current state-of-the-art dPCR technology. Next, we will review these technological aspects/benefits of current dPCR platforms in a virology setting and we will comment on new data analysis approaches for dPCR experiments. It should be noted that most of the present research has been performed on the Bio-Rad droplet digital PCR (ddPCR) platform (Figure 2.2).

2.4 Potential benefits of dPCR

2.4.1 Increased diagnostic selectivity

Diagnostic selectivity is the ability of a method to correctly identify or quantify analytes in the presence of interfering substances. In PCR these substances are generally called PCR inhibitors, which interfere with the efficiency of the PCR to amplify all template DNA molecules with each round of amplification (De Spiegelaere et al., 2008). This PCR inhibition is generally caused by the presence of PCR inhibitory substances, too high concentrations of DNA in the reaction mixture or mismatches in primer/probe sequences. PCR inhibition is a major issue in qPCR experiments, because it impacts PCR reaction efficiency and causes quantification bias when samples are compared to a standard curve with a different efficiency

(Pfaffl, 2001; De Spiegelaere et al., 2008; Suslov and Steindler, 2005). One of the claimed advantages of dPCR over qPCR is its higher tolerance to PCR inhibition. Since dPCR relies on an end-point read-out of the PCR, the effect of moderate changes in efficiency will not bias quantification to the same extent as for qPCR, as long as the positive signals are well discernible from the negative signals.

2.4.1.1 PCR inhibitory substances

Sedlak et al. (2014b) found a better resistance to inhibitory substances of ddPCR compared to qPCR for detection of CMV and human adenovirus species F (AdV) in stool samples. dPCR could be used on undiluted samples, whereas qPCR could only be performed on diluted stool samples because of the high concentrations of PCR interfering substances that affect the sensitivity of the qPCR reaction. Similarly, a better tolerability of RT-dPCR to PCR inhibition was shown to detect viral Hepatitis A and Norovirus RNA in lettuce and bottled water (Coudray-Meunier et al., 2015). The higher tolerance to PCR inhibition may also allow direct quantification of viruses from samples without prior extraction, as shown by Pavsic et al. (2016) on CMV viruses, yielding a higher concentration of detectable DNA copies in comparison to extracted DNA.

Interestingly, the notion that dPCR is less refractory to PCR inhibition may not be generalizable to all inhibitory substances. Dingle et al. (2013) showed that measuring cytomegalovirus (CMV) by ddPCR is less susceptible to inhibition by sodium dodecyl sulfate (SDS) and heparin, but not by Ethylenediaminetetraacetic acid (EDTA), when directly added to the reaction mix. A later study by Nixon et al. (2014), supports this finding, as they indicated that dPCR devices using chip-based systems (cdPCR) are less susceptible to ethanol and plasma inhibition, but not to inhibition by K2 EDTA when compared with qPCR.

ddPCR is also known to be robust at high levels of input DNA concentration; up to 1μg per 20μl reaction can be added in ddPCR without impacting quantitative variation (Strain et al., 2013; Malatinkova et al., 2015). Additional experiments from other scientific fields than diagnostic virology are in line with these findings, mostly showing a decreased susceptibility to inhibition in dPCR (Hoshino and Inagaki, 2012), and reverse transcription (RT)-dPCR experiments (Racki et al., 2014a,b; Yang et al., 2014). Furthermore, dPCR has been reported to be less refractory to PCR inhibition caused by mismatches between the target sequence

and the primers (Strain et al., 2013) or the probes (Trypsteen et al., 2015b) used in the assay. This topic is especially important in a virology setting, as viral sequences may vary substantially between, but also within samples. High mutation rates of viral sequences will have a smaller impact on dPCR platforms, compared to qPCR platforms as long as the end-point PCR signal can be well discriminated from the background signal of the negative partitions.

It is furthermore worthwhile to note that due to the differences in chip-based dPCR and droplet dPCR, the resistance to inhibition may differ. Sedlak and Jerome (2014), for example, suggest that the oil used in ddPCR (but not cdPCR) may sequester some of the inhibitory components, but this has so far not been empirically demonstrated. Hence, different mechanisms of PCR inhibition may differentially affect dPCR platforms and a careful validation of the effect of inhibitory substances on dPCR devices will be needed. In this context, an improved trueness in inhibition-prone samples can be obtained with the use of a spiked internal control, e.g. in a reference plasmid which is quantified in a multiplex reaction together with the target sequence (Sedlak et al., 2014b).

2.4.2 Increased precision

Precision is defined as the closeness of agreement between independent test results. There is a growing consensus that dPCR is more precise than qPCR; both the within-run or intra-assay variability (repeatability) and between-run or inter-assay variability (reproducibility) have been shown to be smaller in an increasing number of studies in the field of virology and beyond (Sedlak and Jerome, 2014; Pinheiro et al., 2011). A higher precision may allow for more consistent measurements and may result in e.g. an improved monitoring of latent HIV reservoirs. In a diagnostic virology setting, increased precision may allow for a better diagnosis, follow-up and treatment of patients.

Differences in precision between qPCR and dPCR have been assessed for most widely studied human viruses. For HIV, Strain et al. (2013) showed a 4-fold to 20-fold decrease in coefficient of variation (CV) for HIV Pol and 2-LTR quantification. When ddPCR was compared to a semi-nested qPCR for HIV quantification, ddPCR was more precise for measuring HIV DNA (Bosman et al., 2015) but showed an equal precision for the quantification of HIV RNA forms (Kiselinova et al., 2014a). For CMV, Hayden et al. (2013) and Sedlak et al. (2014c) found a higher

precision at high CMV concentrations, although equal precision was found for lower levels of CMV. Similar findings have been reported for HCV (Mukaide et al., 2014), HBV (Mu et al., 2015) and influenza (Taylor et al., 2015).

2.4.3 Analytical sensitivity

In contrast to an increased precision of dPCR over qPCR, mixed findings are reported on the analytical sensitivity, also referred to as limit of detection of dPCR. Sensitivity in a dPCR setup is hampered by technical limitations of the currently available platforms; a limited reaction volume results in a decrease of sensitivity (Hayden et al., 2013; Nixon et al., 2014). If reaction volumes for qPCR and dPCR are equal, sensitivity of both platforms have been shown to be comparable (Sedlak et al., 2014c). Notably, false positive partitions have been encountered on several dPCR platforms and can cause a decrease in sensitivity (Kiselinova et al., 2014a; Strain et al., 2013; Pinheiro et al., 2011; Bosman et al., 2015). These false positives seem assay dependent and may also differ between the dPCR platform used (Kiselinova et al., 2014a; Bosman et al., 2015). For diagnostic purposes it will be necessary for each new application to compare the sensitivity of the dPCR to that of the current gold standard assay and to assess the clinical significance of these results.

Importantly, some recent studies have shown an increased sensitivity of dPCR over qPCR for detection of sequence variations to detect emerging drug resistant mutations. Both Taylor et al. (2015) and Whale et al. (2016) studied the sensitivity to detect influenza A mutants and found an increase in sensitivity of more than 50- and 30-fold, respectively. Such an increased sensitivity, combined with the precision of dPCR, may allow a diagnostic application of dPCR, especially when specific mutations can be targeted by allele specific PCR.

2.4.4 Quantitative trueness

The quantitative trueness, i.e. the closeness of the measurement to the actual quantity of the analyte, of dPCR should theoretically be superior to qPCR. qPCR is an indirect quantification method, whereby the output (the kinetics of the fluorescent signal) is compared to that of a predefined standard curve. Hence, the

quantitative trueness of a qPCR relies on an equal PCR efficiency in the standard curve and in the samples, but also on the trueness of the standard itself. In contrast to qPCR, dPCR is a direct quantitative tool, which does not rely on a predefined standard curve. Although this is true for absolute DNA quantification, RT-dPCR to quantify RNA will still necessitate a calibrator. Since dPCR amplifies cDNA and not RNA, an incomplete reverse transcription biases absolute quantification (Kiselinova et al., 2014a; Coudray-Meunier et al., 2016). In this regard, a calibrator to assess RT efficiency will remain crucial for RNA quantification (Kiselinova et al., 2014b; Sanders et al., 2013).

In viral diagnostics a high agreement of diagnostic assays to quantify viruses is crucial to enable comparability between laboratories and/or different time points. For this purpose, reference materials are produced, however, the method by which these standards are developed and validated may vary. Indeed, using dPCR, Hayden et al. (2015) showed that commonly used standard curves for CMV quantification by qPCR produces variable results. Nowadays, dPCR has been suggested as a reference measurement procedure to certify new standards (Pavsic et al., 2015). In addition to this, digital PCR can be used as a reference method to assess the commutability of reference standards for qPCR, i.e. the extent to which reference standards behave like patient samples and are consistent across different assays (Tang et al., 2016). Recently, dPCR was used to validated standard reference materials for several viruses, such as CMV (Haynes et al., 2013) or Ebola Virus (Mattiuzo et al., 2015). In this context, it will be imperative to carefully investigate the impact of different dPCR platforms and the effect of the conformation of the standard DNA, i.e. circular plasmids or linear DNA (Dong et al., 2015).

Despite the theoretical advantages, dPCR may also suffer from some degree of quantification bias. The concept of dPCR relies on certain assumptions, these being a fully random distribution in the sample, a clear discrimination between negative and positive partitions, no variation in partition size and no molecular dropout (i.e. template molecules that do not become amplified in certain partitions). These assumptions are not fully met in current dPCR platforms (Huggett et al., 2013; Jacobs et al., 2014; Corbisier et al., 2015). Also assay design (e.g. primer pairs), master mix choice, extraction method etc. may cause bias (Devonshire et al., 2015). Apart from an improvement of current dPCR platforms, improvements in design and analysis methods will help us better cope with these issues.

2.4.5 Quantitative linearity and dynamic range

A quantitative linearity over a wide dynamic ranges may be required for diagnostic virological tools to be applicable in a clinical setting. It is widely agreed that dPCR has a good quantitative linearity over a wide dynamic range (Hayden et al., 2013; Henrich et al., 2012; Bosman et al., 2015; Hindson et al., 2013; Sedlak and Jerome, 2013). However, it should be noted that the dynamic range of any dPCR device depends on the amount of partitions per reaction. Quantification by dPCR can only be performed when a number of partitions remain negative. Hence, applications that acquire more partitions per μL reaction will have a higher dynamic range (Pinheiro et al., 2011). This can be obtained by either increasing total reaction volumes, by minimizing partition volumes or by dilution of the sample (Huggett et al., 2013, 2015).

2.5 Data analysis of digital PCR

Various factors influence the precision (variability) and trueness (bias) of dPCR experiments. As most of these factors are inherent to the dPCR instruments and the associated sample handling, these can only be accounted for in the downstream analysis of the results. Only in the past few years data analysis procedures have been developed to account, and possibly correct for these sources of bias and variability. Studying these sources of variability and bias is of profound importance when the goal is to obtain a clearer picture of the reliability and the scope of applicability of dPCR experiments. Especially for certain applications (e.g. absolute quantification), specific sources of bias and variability may hinder the use of dPCR as a reliable tool.

2.5.1 Sources of variability and bias in dPCR experiments

The standard dPCR formula to calculate a concentration c is given by

$$\hat{c} = -\frac{\log\left(1 - \frac{k}{n}\right)}{V_p}$$

which involves three parameters: the partition volume V_p, the number of positive partitions k and the total number of partitions n (Hindson et al., 2011). These parameters are all potential sources of bias and/or variability and they should be studied in detail.

The problem is, however, not limited to these three parameters constituting the basic dPCR formula: other sources of variability such as sample handling and pipetting errors may alter the analyte concentration and cause additional variability of the concentration estimate. These sources cannot be measured using a single dPCR experiment, and technical replicates are needed to allow taking these extra sources of variability into account.

Below we review some of the most important aspects of designing and analyzing dPCR experiments that have recently been proposed.

2.5.2 Design of dPCR experiments: implications for precision and power calculations

As in any good scientific study, the goal of the dPCR analysis should be clearly formulated so that an appropriate experimental setup can be designed. Statements about the desired power or desired precision should be part of the formulation of the study objectives. For example, if the goal of the experiment is to detect an absolute concentration deviating from a clinically relevant threshold by a certain amount, the experiment should be designed so that the resulting concentration estimate will be precise enough to allow making such a statement with confidence. In other words, the statistical power of the experiment should be high enough so that correct conclusions can be made with a high probability. Another example is controlling the coefficient of variation of a concentration estimate to a sufficiently low level: the precision of the experiment should be controlled. Failing to properly design the experiment may lead to results that are of no practical use. For dPCR this typically implies the control of a number of variables and will require some preliminary experiments.

One of the main factors influencing the precision of the estimated concentration is the number of partitions: a larger number of partitions will result in a more precise estimate. The number of partitions depends on the platform used for conducting the dPCR experiments. If the number of partitions is too low and

sufficient sample material is present, replicating experiments may increase the number of partitions and consequently the precision.

A second parameter that affects the precision is the fraction of positive partitions: an optimal experiment will have a fraction of positive partitions that is neither too high nor too low. A sample that contains many target molecules will result in a too large fraction of positive partitions, which decreases the precision of the estimate. The same holds true for a sample containing very few target molecules in which a very low fraction of positive partitions will result in a decreased precision. For absolute quantification the optimal number of copies per partition is around 1.5-1.6, corresponding to about 80% positive partitions (Jacobs et al., 2014; Bhat et al., 2009; Huggett et al., 2016). Huggett et al. (2016) (Figure 2) and (Jacobs et al., 2014) (Figure 4) demonstrate the influence of the number of positive partitions on the precision of the estimate. If feasible in practice (e.g. the sample material is not dilute already), a proper dilution may increase the precision.

Also, a clinically relevant difference needs to be determined to achieve the desired power associated with the hypothesis test. For instance, a lower precision will be required to detect a 5% concentration threshold exceedance than to detect a 1% exceedance.

Several authors have proposed methods to determine the power associated with certain levels of these parameters for different types of experiments such as absolute quantification (Hindson et al., 2013; Dorazio and Hunter, 2015), copy number variation (Dorazio and Hunter, 2015; Evans et al., 2012; Whale et al., 2012) and dilution series (Kreutz et al., 2011), but, with the exception of Hindson et al. (2013) who proposed a method that accounts for technical replicates, none of the above methods account for additional sources of variability that may lower the precision of the final concentration estimate. Such sources may originate from sample preparation, pipetting errors, run-to-run variability, cartridge effects, etcetera. Using simulated and real data, Jacobs et al. (2014) and Vynck et al. (2016a) respectively, show that those additional sources of variability may substantially lower the precision and that ignoring this additional variability may result in overly optimistic results. Thus, when ignoring these sources of variability in the design of experiments, the obtained results may also turn out to be of limited practical use.

2.5.3 Misclassification of dPCR partitions

Classification of dPCR partitions is based on the fluorescence readout of each of the partitions. The number of positive partitions may be biased when partitions are misclassified as positive or negative. Some partitions contain genetic material of interest that fails to amplify (molecular dropout), resulting in false negatives, with an underestimation of the analyte concentration as a consequence. On the other hand, partitions containing no target material may display a higher than expected fluorescence intensity, resulting in false positives and an overestimation of the analyte concentration.

Several explanations for the origin of misclassification have been suggested: coagulation of droplets in ddPCR, variation in PCR efficiency due to sequence variation, inhibition or non-homogenous distribution of template or PCR mix components (Dingle et al., 2013; Trypsteen et al., 2015b). It must be noted that these rainy droplets cannot be neglected as mere artifacts, as we and others have shown that droplets with intermediate fluorescence are likely to contain target that is amplified less efficiently (Trypsteen et al., 2015b; Lievens et al., 2016). In certain settings such as low-level quantification (i.e. very few positive partitions), false positives may significantly alter the resulting concentration estimate. Also in experiments where a lot of rain (partitions with an intermediate fluorescence intensity not clearly belonging to the negative or positive fraction (Figure 2.1)) is present, a large number of false positives may occur as a consequence of a threshold being placed too low. Similarly, placing the threshold too high may result in a large number of false negatives. Thus, setting an appropriate threshold to discriminate positive from negative partitions is of importance.

In software provided by dPCR manufacturers, this threshold can be set manually or it is determined by the software. Both may result in problems: allowing the researcher to set the threshold may introduce investigator specific bias, while relying on manufacturers' software may result in a threshold being too close to e.g. the population of negative partitions, leading to false positives (Trypsteen et al., 2015b).

To counteract these drawbacks several authors have proposed algorithms for the calculation of optimal thresholds. Strain et al. (2013) were the first to suggest a clustering-based approach. The method relies on normal distributions to assign partitions to either the positive or negative fraction. Jones et al. (2014) suggested

a similar approach based on the k-nearest neighbours algorithm while assuming a normal distribution of the partitions' fluorescence intensities. A third approach uses the mean of the negative intensities of control samples and adds a certain amount of standard deviations (Dreo et al., 2014). Trypsteen et al. (2015b) recently showed that relying on such distributional assumptions is not optimal and proposed an approach that does not depend on the distribution of the fluorescence intensities.

2.5.4 Concentration estimation

Concentration estimation is the final step in the data analysis procedure and depends critically on both a correct partition volume and a correct combination of the fraction of positive/negative partitions from replicate experiments. In obtaining a correct estimate it is important to take the precision and trueness of both these factors into account.

2.5.4.1 Partition volume

The partition volume may be under- or overestimated. If a suitable reference gene for normalization is not used within the same reaction volume, this results in a biased concentration estimate: an underestimation of partition volume will lead to an overestimation of analyte concentration and vice versa. The partition volume as provided for the Bio-Rad platform was 0.91 nL, which was used in a majority in publications without further verification (Corbisier et al., 2015), even though Pinheiro et al. (2011) analyzed droplet volume on the Bio-Rad QX100 platform and found an average volume lower than was supplied by the manufacturer. Furthermore, Pinheiro et al. (2011) discussed the importance of an unbiased volume estimation in more detail and Corbisier et al. (2015) found that concentrations of reference materials were underestimated due to real droplet volumes 8% lower than the provided 0.91 nL. This finding was confirmed by Dong et al. (2015) who furthermore found that the droplet volume provided for the Raindrop ddPCR platform was also overestimated (4.39pL instead of 5pL or a 14% overestimation). They also found a larger partition volume for the Biomark platform than earlier reported by Bhat et al. (2009). The data as collected by Dong et al. (2015) and Pinheiro et al. (2011) furthermore indicate that the partition volume differs

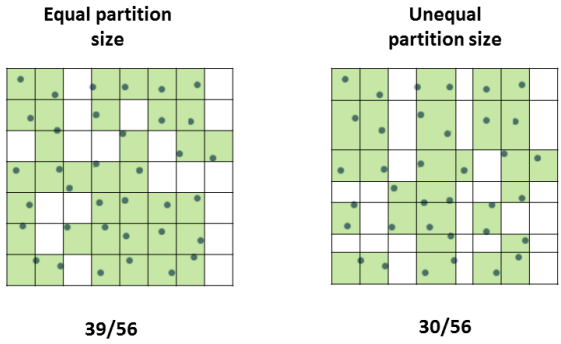


FIGURE 2.3: Illustration of the influence of unequal partition volumes. When a variation in volume is present, smaller partitions will have a higher chance of having no template compared with partitions with a normal or bigger volume. Consequently, when variation in volume is present, a higher number of negative partitions will be recorded than would be expected by the Poisson distribution. Currently used analysis procedures assume an equal concentration of template molecules in each partition; hence, variation in partition size always leads to an underestimation of the actual concentration

between samples and/or cartridges, further providing evidence of the need of partition volume verification. The partition volume used by the QX100 and QX200 ddPCR systems (Bio-Rad) has since been updated to 0.85nL.

The partition volume is often considered to be a fixed constant. However, it has been shown that the partition volume variability may be substantial, and that not accounting for this may lead to both a bias of the estimated concentration and an underestimation of the uncertainty of the estimated concentration (Pinheiro et al., 2011; Jacobs et al., 2014; Huggett et al., 2015; Bhat et al., 2009; Dong et al., 2015) (Figure 2.3). The partition volume variability has minor effects (reduction in trueness of about 1%) if the average number of copies per partition is quite small (less than a copy per partition). However, the effects become substantial when the volume variability is 10% or higher and for higher number of copies (several copies per partition) (Jacobs et al., 2014; Huggett et al., 2015). To account for this partition volume variability, Huggett et al. (2015) recently proposed a model by allowing the copies per partition to vary according to a gamma distribution, i.e. to account for partition volume variability. The parameters of the gamma distribution determine the amount of partition volume variability, and may be set to reflect values as obtained from partition volume analysis.

2.5.4.2 Concentration estimation

For simple experiments, such as calculating absolute concentrations, the estimation of the concentration is fairly straightforward and can be performed by using the software supplied by the platform manufacturers. However, correct analyses, and particularly correct error propagation, can become quite complicated when dealing with technical replicates or more complicated analyses such as calculating copy number variations or when dealing with dilution series. In earlier years, specific methods were developed to deal with some of these more complicated setups such as copy number variation (Whale et al., 2012; Dube et al., 2008) or dilution series/multivolume partitions (Kreutz et al., 2011; Debski et al., 2015). Recently a general statistical framework based on generalized linear models has been described, allowing the analysis of both simple and more complicated setups (Dorazio and Hunter, 2015; Vynck et al., 2016a)

2.5.4.3 Sample-to-sample variation

Another common practice in the application of dPCR is the pooling of technical replicates (e.g. Whale et al. (2012); Morisset et al. (2013); Yu et al. (2015)). As discussed in the design subsection of this review, various factors such as sample preparation and pipetting errors may violate the assumptions necessary to allow a pooling of technical replicates. As a consequence, pooling may lead to overly optimistic results. Although it does not impact the trueness of the estimate, it may lead to an underestimation of the uncertainty of the concentration estimate (Jacobs et al., 2014; Vynck et al., 2016a).This can however be dealt with by extending the framework proposed by Dorazio and Hunter (2015) to allow for additional sources of variability such as that observed or expected between technical replicates (Vynck et al., 2016a).

2.5.5 Data analysis tools

If new data analytical approaches are to be adopted by the dPCR community, practical tools for using these state-of-the-art methods easily are needed. Several authors have made available easy-to-use standalone web tools that allow for applying these data analysis methods to the researchers' own data (Jones et al.,

Table 2.1: Web tools for dPCR data analysis

Tool	Use	URL
definetherain	Threshold setting	http://definetherain.org.uk
ddpcRquant	Threshold setting	http://statapps.ugent.be/dPCR/ddpcrquant
GLM/GLMM	Quantification	http://statapps.ugent.be/dPCR/dPCR

2014; Trypsteen et al., 2015b; Vynck et al., 2016a). An overview of these web tools is given in Table 2.1.

2.6 General conclusion and future perspectives

In the past few years dPCR applications, instruments and their properties have received increasing attention and this has led to a substantial increase in knowledge about the advantages and drawbacks of the technology. Further research concerning the drawbacks is necessary to allow a widespread and routine use of dPCR as a diagnostic tool. These include the elimination of false positive results (especially important for low-level quantification), user-friendly instrumentation that allows to monitor partition volume variability, awareness of pitfalls when designing experiments and analyzing data and increased usability of data analysis procedures. When these issues are addressed, the clear advantages of dPCR over qPCR and the increased knowledge about the technology makes the use of dPCR as a diagnostic technology increasingly likely.

Many assays on qPCR can be directly transferred to a dPCR platform. However, in the case of diagnostic testing an extensive validation should precede any clinical application. Assay specific limits of detection and different influences on PCR inhibitory substances may result in a better or worse performance of dPCR compared to qPCR. In addition, increases in precision and trueness may not be clinically significant in some contexts and may thus not be cost-effective with current dPCR platforms. Future platforms that enable high throughput dPCR at low cost with a sufficiently high dynamic range are likely to replace many qPCR applications in the future.

3 On the design of experiments

This chapter is an adaptation from the paper: Vynck, M., Jo Vandesompele and Thas, O. On the design of digital PCR experiments. *Under revision.*

Contributions: Matthijs Vynck, Jo Vandesompele and Olivier Thas conceived of the contents of this paper. Matthijs Vynck and Olivier Thas wrote the R code. Matthijs Vynck, Jo Vandesompele and Olivier Thas analyzed the data. Matthijs Vynck wrote an intial draft of the manuscript. Matthijs Vynck, Jo Vandesompele and Olivier Thas revised the manuscript. All authors agreed with the contents of the final paper.

3.1 Abstract

The experimental design that will be carried out to evaluate a nucleic acid quantification hypothesis determines the cost and feasibility of digital PCR studies. Experiment design involves the calculation of the number of technical measurement replicates and the determination of the characteristics of those replicates, and this in accordance with the capabilities of the available digital PCR platform. Available digital PCR power analyses suffer from one or more of the following limitations: narrow scope, unrealistic assumptions, no sufficient detail for replication, lack of source code and user-friendly software. Here, we discuss the nature of six parameters that affect the statistical power, i.e. desired effect size, total number of partitions, fraction of positive partitions, number of replicate measurements, between-replicate variance, and significance level. We also show to what extent these parameters affect power, and argue that careful design of experiments is needed to achieve the desired power. A web tool, dPowerCalcR, that allows interactive calculation of statistical power and optimization of the experimental design is available.

3.2 Introduction

The digital polymerase chain reaction (digital PCR) is a molecular technique that allows direct absolute quantification and relative quantification of nucleic acids. While the relatively limited throughput and high cost of nucleic acid quantification with digital PCR hamper its adoption for routine nucleic acid quantification, an area where real-time quantitative PCR (qPCR) is still considered the gold

standard, the technique enjoys increasing attention within various life science disciplines (Baker, 2012; Sedlak and Jerome, 2014; Morisset et al., 2013): several applications have been identified for which the high and scalable precision of digital PCR is required. These applications include the detection and quantification of low-level analytes such as viral titers (Henrich et al., 2012; Kiselinova et al., 2014a), rare mutations in e.g. influenza (Whale et al., 2016), small copy number alterations in e.g. non-invasive prenatal testing (El Khattabi et al., 2016), among many others. The research questions of interest for these applications often involve the comparison of absolute or relative quantities with a predetermined value, e.g. is the viral titer above a clinically important threshold?

As with any quantitative scientific investigation, thoughtful experimental design before starting the research project is of primordial importance. It provides the researcher with information on the expected reagent cost, on the expected hands-on time and on the feasibility of the envisioned study and it allows efficient execution of experiments needed to answer the research question.

Specification of the experimental design, and sample size or power calculations are common in many fields, notably clinical trials, but they are generally lacking in quantification experiments using digital PCR. However, the use of e.g. inadequate platforms (e.g. platforms with too little partitions for the desired precision), or the analysis of too few replicates, may result in inadequate power to detect biologically or clinically meaningful differences in nucleic acid quantities. This may ultimately result in false negatives and a large false discovery rate, and consequently affects the reproducibility of published findings (Colquhoun, 2014). On the other hand, experiments may be overpowered which unnecessarily increases the cost, the use of precious sample material and the time required to execute.

Several papers have been published related to power calculations for digital PCR experiments, yet (i) they often rely on unreasonable assumptions, e.g. they do not account for between-replicate variance, which we will show is a major effecter of the power (Whale et al., 2012), (ii) they do not provide sufficient detail to repeat the analyses, and consequently not enough details to calculate power for the practitioner's own experiments (Evans et al., 2012), (iii) source code is lacking and user-friendly software for power calculation in digital PCR experiments is non-existent. As such, more appropriate power calculations, and their availability in a user-friendly way may be of benefit to the digital PCR user community. In addition, awareness among digital PCR practitioners of the parameters affecting the

power may improve their experiments.

In this paper, we first derive equations for power calculation for a given experiment. This is done for both absolute and relative quantification (including gene copy number) scenarios. Secondly, we provide simulation results to demonstrate the influence of all relevant parameters on the power. Finally, we provide an intuitive and freely available web tool for the calculation of power with parameters as specified by the user.

3.3 Materials and methods

3.3.1 Data and data processing

We collected publicly available data on a total of 1058 digital PCR reactions, of which 500 related to absolute quantification(Jones et al., 2014; Lievens et al., 2016) and 558 to relative quantification (Vynck et al., 2016b).

In case partition fluorescence intensities were available, thresholding was performed. In particular, when no-template controls were available, the data were thresholded using ddpcRquant (Trypsteen et al., 2015a), whereas Cloudy was used if no-template controls were absent (Lievens et al., 2016). After thresholding, the numbers of positive, negative and total partitions were used for further processing.

Absolute concentrations were estimated assuming a Poisson distribution of copies per partition:

$$\hat{C} = -\frac{\ln\left(1-\frac{k}{n}\right)}{V_p},$$

with $\hat{C}$ the estimated concentration, ln the natural logarithm, V_p the average partition volume, k the number of positive partitions and n the total number of partitions.

Relative quantities were estimated as the ratio of two absolute quantities, i.e.

$$\widehat{RQ} = \frac{\ln\left(1-\frac{k_1}{n_1}\right)}{\ln\left(1-\frac{k_2}{n_2}\right)} N_b,$$

where k_1 and k_2 denote the numbers of positive partitions, and n_1 and n_2 denote the total numbers of partitions of the target and reference nucleic acid, respectively, and N_b denotes the ploidy level of the organism. The constant N_b is set to one for relative quantification of RNA levels, and is set to e.g. two for calculating copy numbers in diploid organisms.

3.3.2 Power equations

3.3.2.1 Absolute quantification

Suppose that from a single replicate measurement, the estimate of the absolute concentration is denoted by $\widehat{C}_i$ and suppose that from r replicate experiments the absolute concentration is estimated by $\widehat{C} = \frac{1}{r}\sum_{i=1}^{r}\widehat{C}_i$, which is approximately normally distributed. Hence, for a given concentration C_0, an appropriate test statistic for testing $H_0 : C = C_0$ against $H_1 : C > C_0$ is given by

$$T = \frac{\widehat{C} - C_0}{s},$$

where s is a consistent estimate of the standard deviation of $\widehat{C}$. The corresponding variance estimator is given by

$$s^2 = \frac{v_{P,C} + v_{B,C}}{r}$$

where $v_{B,C}$ is the between-replicate variance and $v_{P,C}$ is the variance caused by the Poisson process (for a single replicate). Derivations and final equations for the calculation of $v_{B,C}$ and $v_{P,C}$ are provided in Appendix B.1.1.

The power of the test, when performed at the α level of significance, for detecting a concentration $C_\delta > C_0$, i.e. an effect size $C_\delta - C_0$, is given by

$$1 - \Phi(\Phi^{-1}(1-\alpha) - (C_\delta - C_0)/\sqrt{s^2})$$

where s^2 is the variance of the absolute concentration estimate obtained from replicate reactions and Φ and Φ^{-1} the cumulative distribution function and quantile function of the standard normal distribution, respectively.

For a two-sided alternative $H_1 : C \neq C_0$, the power of the test is given by

$$1 - \Phi(\Phi^{-1}(1-\alpha/2) - (C_\delta - C_0)/\sqrt{s^2}) + \Phi(\Phi^{-1}(\alpha/2) - (C_\delta - C_0)/\sqrt{s^2}).$$

3.3.2.2 Relative quantification

Suppose that from a single replicate measurement, the estimate of the relative quantity is denoted by $\widehat{\mathrm{RQ}}_i$ and suppose that from r replicate reactions the RQ is estimated by $\widehat{\mathrm{RQ}} = \frac{1}{r}\sum_{i=1}^{r} \widehat{\mathrm{RQ}}_i$, which is again approximately normally distributed and that an appropriate test statistic for testing $H_0 : \mathrm{RQ} = \mathrm{RQ}_0$ against $H_1 : \mathrm{RQ} > \mathrm{RQ}_0$ is given by

$$T = \frac{\widehat{\mathrm{RQ}} - \mathrm{RQ}_0}{s},$$

where s is a consistent estimate of the standard deviation of $\widehat{\mathrm{RQ}}$. The corresponding variance estimator is given by

$$s^2 = \frac{v_{P,\mathrm{RQ}} + v_{B,\mathrm{RQ}}}{r}$$

where $v_{B,\mathrm{RQ}}$ is the between-replicate variance and $v_{P,\mathrm{RQ}}$ is the variance caused by the Poisson process (for a single replicate). Derivations and final equations for the calculation of $v_{B,\mathrm{RQ}}$ and $v_{P,\mathrm{RQ}}$ are provided in Appendix B.1.2. Note that the calculation of $v_{P,\mathrm{RQ}}$ differs from the calculation of $v_{P,C}$.

The power of the test is obtained as for absolute quantification.

3.3.3 Simulation study

We used the empirical data to obtain realistic estimates of parameters affecting the power. Using these estimates, we evaluated the effect of each of the parameters on the resulting power by letting a single parameter take (biologically) relevant values, keeping all remaining parameters fixed or letting a second parameter vary to assess the joint effect of both parameters on the power. The relevant values of the parameters were determined as the 95% percentile of the empirically observed values for that parameter that results in the least power (Tables 3.1 and 3.2. The absolute or relative concentration is also kept fixed throughout the simulations.

3.3.4 Software

All analyses were implemented in R 3.4.1 (R Core Team, 2018). Source code for the data analyses, the simulations as well as the R/Shiny application dPowerCalcR is freely available from `http://github.com/CenterForStatistics-UGent/dPowerCalcR` under the GPL3 license.

3.4 Results and discussion

3.4.1 The need for designing experiments

The minimal relevant change (effect size), the ability to dilute or concentrate reaction mixes, the optimal fraction of negative partitions, the quantity of sample material needed, the number of samples needed, the number of required partitions, ... are all factors that may be specified and should consequently be considered in the design phase of the experiment. Failing to acknowledge the influence of these parameters on the ability to detect predetermined biologically or clinically relevant changes may result in erroneous conclusions, irreproducible findings and a waste of resources.

3.4.2 Parameters affecting the power

Six parameters are part of the power equations, and consequently affect the power for a given design: the effect size, the total number of partitions, the fraction of negative (or positive) partitions, the number of replicate measurements, the between-replicate variance and the significance level. We will discuss the effect of each of the parameters on the power while keeping the other parameters fixed at realistic values. These values were determined from empirical data. The observed ranges of parameter values are given in Tables 3.1 and 3.2. It should be noted that these parameters may vary considerably from one experiment to the next, depending on e,g, the setup of the experiment and type of sample material. The empirical data serve to demonstrate the effects of the different parameters in a realistic setting.

TABLE 3.1: Parameters in empirical absolute quantification data affecting the power. For each of the parameters the range, the median value and a pessimistic 95th percentile are presented. The pessimistic 95th percentile corresponds to that end of the parameter range resulting in the least power.

Parameter	Range (min, max)	Median	Pessimistic 95th percentile
Number of partitions	(1814, 16985)	12545	8331
Fraction of negatives	(0.0119, 0.9999)	0.6066	0.9806
Number of replicates	(1, 4)	2	1
Between-replicate variance	(0, 0.03057)	0.00001	0.0046

TABLE 3.2: Parameters in empirical gene copy number quantification data affecting the power. For each of the parameters the range, the median value and a pessimistic 95th percentile are presented. The pessimistic 95th percentile corresponds to that end of the parameter range resulting in the least power.

Parameter	Range (min, max)	Median	Pessimistic 95th percentile
Number of partitions	(9437, 16030)	14065	9437
Fraction of negatives	(0.8151, 0.9999)	0.8956	0.9780
Number of replicates	(2, 3)	3	2
Between-replicate variance	(0,0.0577)	0	0.0119

The significance level is typically set to 5% and the effect size to 10%. The presented results are for relative quantification, but similar results are obtained for absolute quantification.

3.4.2.1 Effect size

Figure 3.1 displays the influence of the effect size on power for effect sizes in the range 1%-100%, i.e. effect sizes for which the superior precision of digital PCR, when compared to qPCR, is needed. With the given parameters, effect sizes of 1% will almost never be detected, but the power increases rapidly for larger effect sizes and reaches desirable levels for effect sizes of 10% and higher. Detecting effect sizes lower than 10% is hard but not impossible: increasing the number of replicates substantially, for example, may even make detection of effect sizes of 2-3% achievable (Figure 3.2). Detection of these very small effects is considered to

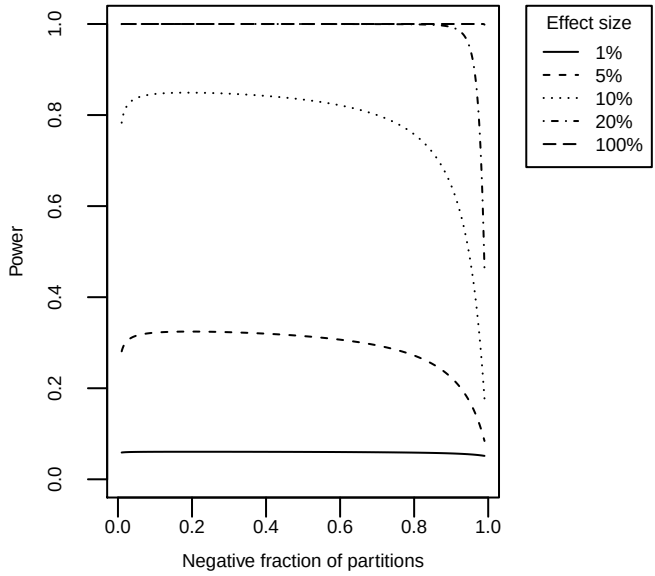


FIGURE 3.1: Power to detect effect sizes in the range 1%-100% for given fractions of negative partitions, with three replicates and with pessimistic values of the remaining parameters observed in the empirical data (Table 3.2).

be one of the key strengths of digital PCR, and effect sizes of a few percent occur in practice, e.g. in non-invasive prenatal testing (El Khattabi et al., 2016).

3.4.2.2 Fraction of negatives

Figure 3.1 also demonstrates the influence of the fraction of negatives (positives): power often decreases drastically when the fraction of negative partitions is either very low or very high. This is a known effect and is related to the increase of stochastic variability for such values of the fraction of negative partitions (Jacobs et al., 2014; Majumdar et al., 2015; Huggett et al., 2016). These very low or high fractions should, if possible, be avoided. An ideal fraction of negatives is typically around $\sim 35\%$ for relative quantification(Huggett et al., 2016) ($\sim 20\%$ for absolute quantification (Jacobs et al., 2014)), but this depends on parameters such as the ratio of quantities in a relative quantification scenario.

This has important implications for e.g. low-level quantification where often few

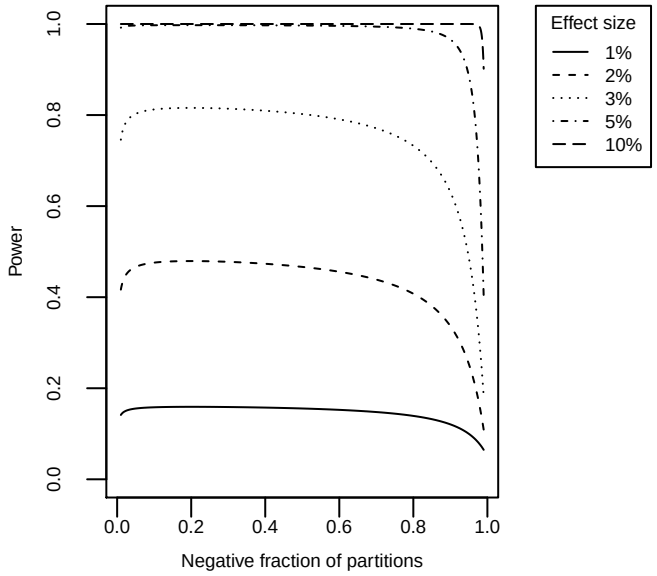


FIGURE 3.2: Power to detect effect sizes in the range 1%-10% for given fractions of negative partitions, with thirty replicates and with pessimistic values of the remaining parameters observed in the empirical data (Table 3.2).

positive partitions are observed. Examples include the quantification of viral fragments and mutation quantification (Vynck et al., 2016a; Whale et al., 2016). A large number of replicates may be needed to achieve the desired power in such low-level quantification experiments.

In addition, the smaller the effect size, the stronger the relative effect of the fraction of negative partitions on the resulting power: small changes in the negative fraction may result in a large relative drop in power (Figure B.3). This effect manifests most strongly for absolute quantification. This may be explained by the increasingly smaller difference in quantities, that is often not offset by the decrease in variance, especially when the between-replicate variance contributes substantially to the total variance. Decreasing between-replicate variance lessens this effect (Figure B.3), but increasing the number of replicates may lessen the effect more strongly as it reduces both the contribution to the total variance of the stochastic variance and the between-replicate variance (Figure B.3).

3.4.2.3 Number of partitions

The number of partitions may be thought of as one of the major contributors to power. While it is true that the power increases for increasing numbers of partitions, the power may level off relatively quickly (Figure 3.3). Increasing the number of partitions typically has a limited effect on the power beyond the range of several tens of thousands of partitions. However, this strongly depends on the between-replicate variance. If the between-replicate variance is low, increasing the number of partitions may still provide benefits in terms of power gain to detect small effect sizes (Figure 3.4).

The reason for this is easily seen in the calculation of the variance of the absolute or relative quantity. The variance involves both the stochastic (Poisson) variance, as well as the between-replicate variance. The contribution of the stochastic error to the total variance will decrease with increasing numbers of partitions, but the between-replicate variance remains unaffected. Thus, for large numbers of partitions, the total variance is dominated by the between-replicate variance and a further increase of the number of partitions will have little effect on the reduction of the total variance. Conversely, if the between-replicate variability can be minimized, the total variance will largely be determined by the number of partitions. In such cases, an increase in the number of partitions may help to achieve the desired power. Moreover, the stochastic variance is inverse proportional to the number of partitions, so that the decrease in stochastic variance is initially very steep, but addition of more partitions contributes increasingly less to this reduction (Figures 3.3, 3.4).

3.4.2.4 Number of replicates

As was argued for the required number of partitions, the between-replicate variance contributes to the total variance and its contribution can only be reduced by running additional replicates. Increasing the number of replicate measurements may increase the power of the experiment substantially (Figures 3.2, 3.5). Keeping other parameters constant, the effect size - replicate relationship to achieve a given power is non-linear and the detection of increasingly smaller effect sizes will result in drastic increases of the required number of replicates. For example, whereas triplicate measurements allow detection of effect sizes of 10% with a

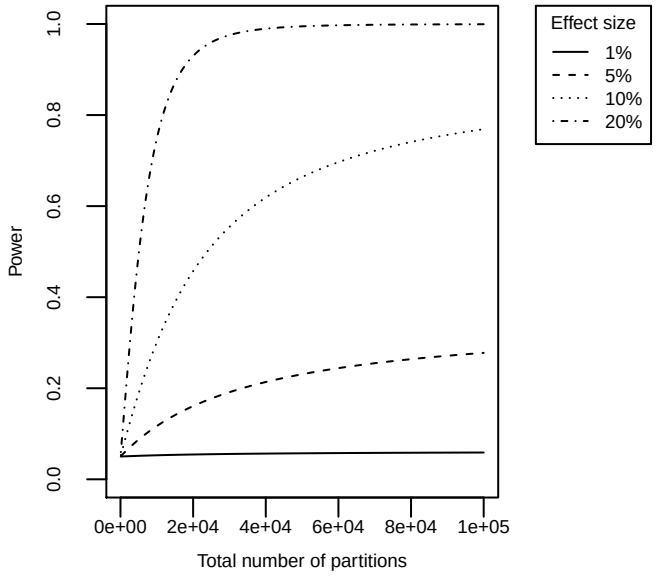


FIGURE 3.3: Power to detect effect sizes in the range 1%-20% for given total numbers of partitions, with pessimistic values of the remaining parameters in the empirical data (Table 3.2).

power of 80% in the empirical data, detection of effects of 5% requires $\sim$ 10 replicates and detection of effects of 3% requires $\sim$ 30 replicates. Detection of an effect of 1% becomes practically infeasible, requiring hundreds of replicates to achieve acceptable power (i.e. a power of 80%).

Determination of the required number of replicates is likely one of the major tasks of power analysis. The number of replicates is the factor that is most easily changed, this as opposed to e.g. the number of partitions which is more or less determined by the platform, or the fraction of negatives which requires preliminary knowledge on the concentration of the sample and whose alteration may not be feasible when the analyte is dilute already. Careful selection of the number of replicates is furthermore needed as it is one of the major factors affecting the cost of the experiment.

3.4.2.5 Between-replicate variance

Between-replicate variance may vary among experiments and may have a substantial effect on the power (Figure 3.6). A realistic estimate of this variance is

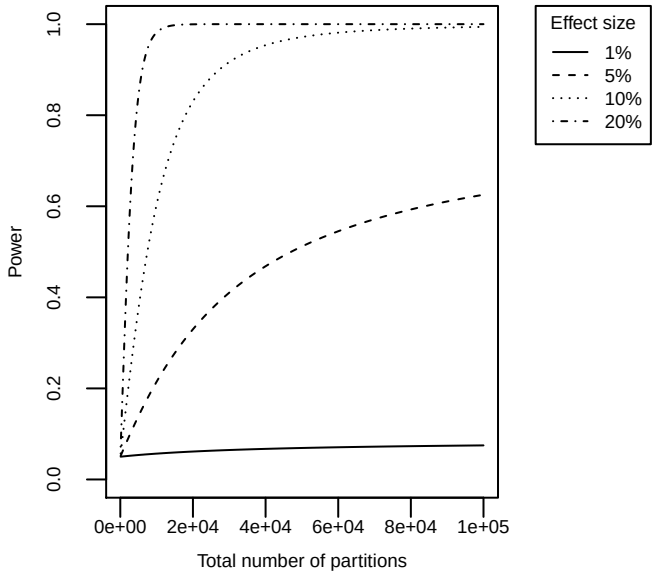


FIGURE 3.4: Power to detect effect sizes in the range 1%-20% for given total numbers of partitions, with pessimistic values of the remaining parameters in the empirical data (Table 3.2) but with between-replicate variance being three times smaller than the pessimistic value.

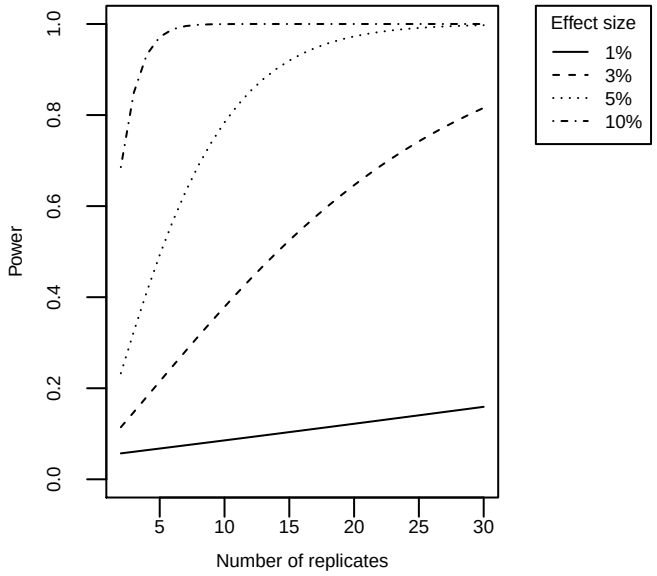


FIGURE 3.5: Power to detect effect sizes in the range 0%-10% for given numbers of replicates, a negative fraction of 20%; and with pessimistic values of between-replicate variance and number of partitions in the empirical data (Table 3.2).

consequently highly important. In our data we observed between-replicate variances ranging from 0 to 0.0577, or a contribution to the total variance of 0% to 84% in relative terms (Appendix B.2.2). In the example in Figure 3.6, this corresponds to a power of nearly 100% for the lowest variance, but results in only around 20% power for the largest between-repliacte variance. The true between-replicate variance is likely somewhere between these two extremes as these empirical estimates are typically based on just three replicates in the empirical data, resulting in a reasonably imprecise estimate of the true between-replicate variance. Yet it should be clear that underestimation of the between-replicate variance (or ignoring it altogether) may result in severe overestimation of the power.

When studying between-replicate variance, it is of importance to consider what the between-replicate variance reflects. This will depend on the level at which the replicates were generated in the preanalytical phase. Generating replicates at an early point in the preanalytical phase (e.g. collecting multiple replicate samples from the same individual rather than generating replicates once the PCR mix has been made) is likely to result in a larger-between replicate variance. The choice of the level at which replicates are generated will influence the interpretation of

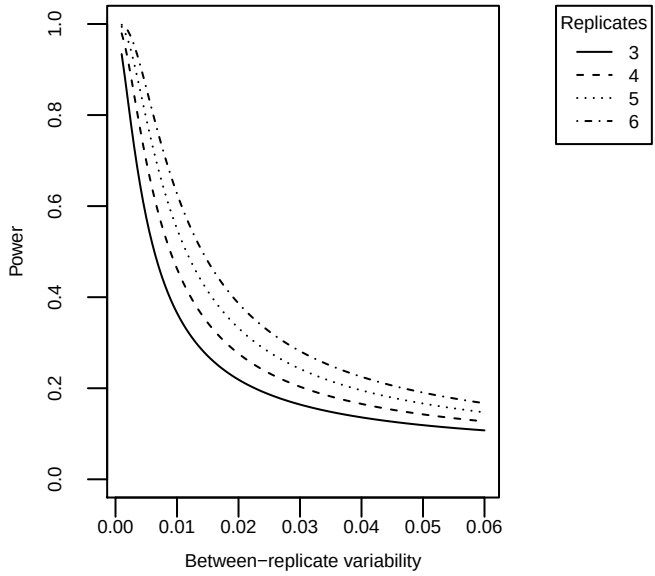


FIGURE 3.6: Power to detect effect sizes of 10% with three to six replicates and for given between-replicate variances, 20% of negative partitions, and a total of 10000 partitions.

the resulting power calculations, and is likely to impact on the results.

3.4.2.6 Significance level

The power increases along with the significance level: increasing the significance level will allow for easier rejection of the null and will hence lead more quickly to the conclusion that a significant effect has been detected. Yet, increasing the significance level also increases the probability of false positives, and should thus be kept to a sufficiently low level, i.e. the commonly accepted 5% or 1%. In case multiple targets are assessed, the significance level may be adjusted to retain a family-wise error rate (Noble, 2009) or false discovery rate (Benjamini and Hochberg, 1995) of e.g. 5%. This adjusted may be done using typical error rate control procedures such as the Holm-Bonferroni procedure (Holm, 1979), though alternative strategies may be used.

3.4.3 Preliminary experiments

Several parameters have a strong effect on the power calculation. These parameters are dependent on the sample material. Preliminary experiments need to be performed to get an idea about typical values for these parameters. Such experiments will typically be limited in scope and parameters derived from these experiments will often be determined with a rather large imprecision.

The parameters used in the power calculation are ideally chosen towards the pessimistic end of the spectrum as overestimation of the true power is to be avoided. For example, if the range of accepted partitions in a preliminary experiment is 12000 - 17000, it is better to base power calculations on 12000 partitions as this will result in the lowest power deemed appropriate. The resulting designed experiment will consequently achieve at least the desired power. A post-hoc evaluation of achieved power using the full results and predetermined relevant effect size is recommended to verify whether the experiment was indeed sufficiently powered, particularly when no significance of the statistical test was obtained, as this may indicate whether the obtained result is likely to be a false negative, in which case the experiment may be redone with e.g. an increased number of replicates.

3.4.4 Implications for platform and assay design

Studying experimental designs and their associated power to detect an effect of interest may also suggest improvements for the design of the dPCR platforms themselves. Increased numbers of partitions or higher throughput of samples so that higher numbers of replicates are more easily obtained may extend the application domain of the technique. Reduction of technical error may additionally reduce variance among replicates, leading to an increase of power.

Our analyses are applicable to commercially available platforms. Other platforms have been described in literature, such as multi-volume designs, but our power calculations are only applicable to those platforms employing a fixed-volume partition size.

Optimization of assays may help to further reduce between-replicate variance. The presence of so-called rain (partitions with intermediate fluorescence intensity) that may arise from e.g. non-specific amplification or suboptimal reaction conditions (Jones et al., 2014; Trypsteen et al., 2015a; Lievens et al., 2016) may

result in increased variance of absolute or relative quantities between samples. Optimizing the assay to prevent e.g. non-specific amplification may increase the power achievable without need for e.g. increasing the number of replicates.

3.4.5 Interactive web tool for power calculation

An R/Shiny web tool for power calculation, dPowerCalcR, is available at `http://statapps.ugent.be/dPCR/dPowerCalcR`. Relevant parameters as discussed in this manuscript can be specified by the user and the resulting power will be calculated. Graphical displays are provided to help optimize the experiment to achieve the desired power.

3.5 Conclusion

We have derived power equations for various digital PCR applications and have demonstrated how various parameters affect the power. Careful consideration of these parameters is needed. Of note, the power is sensitive to the between-replicate variance, a factor previously ignored in power calculations, and its careful determination is warranted. Ultimately, power analysis may aid researchers in designing effective experiments and may provide an idea of the feasibility of specific experiments.

4 A threshold determination procedure

This chapter is an adaptation from the paper: Trypsteen, W., Vynck, M., De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., ... and De Spiegelaere, W. (2015). ddpcRquant: threshold determination for single channel droplet digital PCR experiments. Analytical and Bioanalytical Chemistry, 407(19), 5827-5834.

Contributions: Maja Kiselinova, Eva Malatinkova and Karen Vervisch performed the wet-lab experiments. Wim Trypsteen, Pawel Bonczkowski and Ward De Spiegelaere analyzed the ddPCR data. Matthijs Vynck, Jan De Neve, Olivier Thas, Wim Trypsteen, Ward De Spiegelaere developed the statistical methodology. Wim Trypsteen and Matthijs Vynck wrote the R code. Wim Trypsteen, Ward De Spiegelaere, Matthijs Vynck and Jan De Neve wrote the paper. All authors critically read, reviewed, and approved the final version of the manuscript.

4.1 Abstract

Digital PCR is rapidly gaining interest in the field of molecular biology for absolute quantification of nucleic acids. However, the first generation of platforms still needs careful validation and requires a specific methodology for data analysis to distinguish negative from positive signals by setting a threshold.

The currently described methods to assess droplet digital PCR (ddPCR) are based on an underlying assumption that the fluorescent signal of droplets is normally distributed. We show that this normality assumption likely does not hold true for most ddPCR runs, resulting in an erroneous threshold. We suggest a methodology that does not make any assumptions about the distribution of the fluorescence readouts: by modelling the extreme values in the negative droplet population using extreme value theory, a threshold is estimated. Furthermore, the method takes shifts in baseline fluorescence between samples into account.

An R implementation of our method is available, allowing automated threshold determination for absolute ddPCR quantification using a single fluorescent reporter.

4.2 Introduction

With recent technological developments, digital PCR (dPCR) is becoming widely available and has the potential to become the gold standard technique/reference measurement procedure for absolute quantification of nucleic acids in molecular biology. Digital PCR enables direct quantification without the need for a standard dilution series and, in theory, offers an increased accuracy compared to quantitative PCR (Hindson et al., 2013; Sedlak and Jerome, 2013). However, the first generation of digital PCR platforms still needs careful validation before being used broadly in clinical diagnostics.

To date, the droplet digital PCR system (ddPCR, Bio-Rad) is the most widely used dPCR platform. By using microfluidics, it theoretically generates 20.000 droplets from an initial 20 μl PCR mix with each droplet representing an isolated endpoint PCR reaction. Subsequently, fluorescently labeled probes enable a simple readout of the endpoint fluorescence in each droplet and allow the classification of droplets as positive or negative (Pinheiro et al., 2011; Hindson et al., 2011). As a result, absolute quantification can be performed by assuming a Poisson distribution and the concentration of nucleic acid templates in the total reaction can be calculated based on the frequency of positive to negative droplets, the total number of droplets and the droplet volumes (Hindson et al., 2011; Dube et al., 2008).

4.2.1 Digital PCR data analysis

Most dPCR platforms use fluorescent probes or intercalating dyes which provide a fluorescent signal for the droplets that contain amplified PCR product. Accordingly, positive droplets will have a higher fluorescence amplitude compared to the background fluorescence of the negative droplets. Despite the simplicity of this principle, there is no well-defined indication for the exact position of the threshold value to distinguish positive from negative droplets. This particularly holds true for assays that produce droplets where no clear differentiation is observed between negative and positive droplets, as is often the case in a heterogeneous pool of DNA templates, i.e. HIV and/or other micro-organisms.

Another important aspect to consider when positioning the threshold are the droplets that fall between the negative and positive population, often referred

to as 'rain'. The origin of this phenomenon still remains a topic of discussion (Figure 4.1A, B). This rain may arise from coagulation of multiple droplets that result in a higher background fluorescence. Another reason can be a suboptimal PCR amplification due to variation in PCR efficiency between droplets caused by a non-uniform distribution of the PCR mix components or by small differences in sequence composition between individual HIV DNA molecules. The latter condition will certainly hold true for absolute quantification of viral or bacterial genomes which may be characterized by a high frequency of sequence variation (e.g. HIV). Hence, both coagulating droplets as well as target sequence variation may play a role in the formation of rain.

A final consideration regarding the current generation of the ddPCR platform arises from false positive droplets that sporadically occur. These false positive droplets are sometimes indistinguishable from the true positive droplets and have been observed in several studies (Pinheiro et al., 2011; Kiselinova et al., 2014a; Strain et al., 2013; Sedlak et al., 2014a).

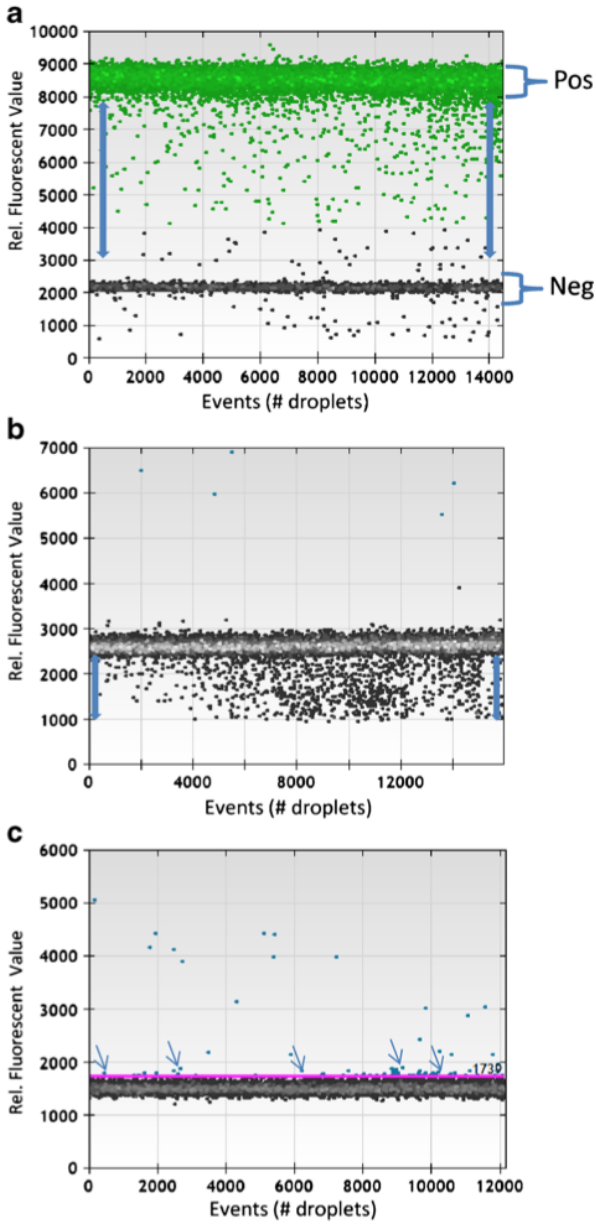


FIGURE 4.1: Examples of QuantaSofts' ddPCR fluorescent readout. (A) Positive sample showing positive droplet population, negative droplet population, and positive rain (droplets between the arrows). (B) Negative template control showing negative rain (droplets between the arrows). (C) QuantaSofts' automated threshold determination: The threshold (straight line) falls in the negative droplet population resulting in false- positive droplets (arrows)

4.2.2 Digital PCR data analysis tools

The ddPCR setup invokes a different methodology of analyzing data compared to existing quantitative real time PCR (qPCR). Currently, only a few methods have been described for threshold setting in ddPCR reactions (Strain et al., 2013; Jones et al., 2014). According to the instructions provided by the manufacturer of the ddPCR QX100 platform (Bio-Rad), it is possible to set a manual threshold, allowing the investigator to make a visual interpretation of the background, rain and false positive droplets but this may also introduce investigator specific bias. In addition, automated systems for threshold settings are also available in the QuantaSoft software that comes with the ddPCR platform. However, this threshold setting is not always accurate as the threshold often falls within the population of negative droplets or is not possible to calculate (Figure 4.1C). Moreover, the exact method of threshold calculation is not disclosed by the manufacturer, making interpretation impossible.

Apart from the QuantaSoft software, first attempts to enable a more robust automated threshold setting were made by Strain et al. (2013) and more recently by Jones et al. (2014). These methods use clustering algorithms to identify the positive and negative droplet populations and do not use one threshold but define the 95/99% confidence intervals for the negative and positive population. Droplets that fall within the intervals are allocated to the respective population and rain is discarded and excluded from further calculations.

Although these clustering approaches appear promising, the exclusion of rain may cause an underestimation of the true concentration if this rain is resulting from suboptimal PCR reactions of target template molecules. This could be the case if the PCR reaction is partly inhibited or when minor mismatches occur between the primer/probe sequences and the target molecules. To investigate this hypothesis, a series of ddPCR reactions were performed with HIV DNA amplicons containing a mismatch in the probe binding region (Appendix C.1 and Figure C.1). This resulted in intermediate fluorescence when a mismatch was present, proving that rain may contain true positive droplets and is not always an artifact. Accordingly, Dreo et al. (2014) recently showed that the clustering method can result in discarding true positive droplets when testing low target concentrations of Erwinia amylovora in plant material. Therefore, a single threshold may still be preferred, since the rain arising from coalescence can be assessed by running multiple (>3) negative template controls (NTC) containing template free genomic

DNA. This was also proposed by Dreo et al. and in this context a new method was described where a global manual threshold (MTg) is defined as the averaged fluorescence signal in the NTCs plus six times the standard deviation (Dreo et al., 2014).

A final consideration concerning available data analysis methods is that the calculation of confidence intervals in both clustering and MTg methods is performed under the assumption of a normal distribution of the fluorescent amplitude of the negative droplet fluorescence and without correcting for possible shifts in baseline fluorescence between NTCs (Strain et al., 2013; Jones et al., 2014). This may have a major influence on the correct allocation of droplets in case the distribution of fluorescent amplitudes does not follow a normal distribution.

4.2.3 A new digital PCR data analysis tool: ddpcrquant

Here, a new method of threshold setting is proposed for the quantification of DNA templates by ddPCR. The method is based on extreme value theory and resolves the described problems. The method does not make assumptions about the distribution of the droplet fluorescence, and consequently deviation from normality of the negative droplet fluorescent amplitudes will not influence the correctness of the method. Moreover, it results in a single threshold and does not discard any of the droplets.

In addition, small shifts in baseline fluorescence of the negative droplet population can be often observed (Figure C.2). QuantaSoft and all other algorithms don't account for these small shifts. This can possibly lead to an inaccurate quantification since a threshold value is calculated based on the NTC and transferred to all samples. Consequently, discrepancies in baseline fluorescence can have an influence on the number of positive droplets (Figure C.2). Therefore a baseline correction to account for these small shifts has to be included.

4.3 Material and methods

4.3.1 Digital PCR datasets

A set of 29 ddPCR runs was used as raw data for all experiments performed in this article. These runs mainly included HIV related absolute quantification assays on patient samples: HIV 2LTR (Buzon et al., 2010(@), total HIV DNA (van der Sluis et al., 2013; Liszewski et al., 2009), HIV unspliced cell associated RNA (Palmer et al., 2003), HIV multiple spliced cell associated RNA (Fischer et al., 2004), WPRE (Lizee et al., 2003). PCR cycling was performed with the QX100™ system (Bio-Rad, Pleasanton, CA) as previously described (Kiselinova et al., 2014a; De Spiegelaere et al., 2013). QuantaSoft absolute quantification single and multiple well data analysis methods were used.

4.3.2 Distribution of the negative droplet population

Normality of the negative droplets of 154 NTCs from 29 assays was assessed statistically by the Anderson-Darling test and visually by constructing QQ plots.

4.3.3 ddpcRquant for automated digital PCR data analysis

The ddpcRquant method provides an automated analysis of one dimensional/single colour ddPCR data by calculating a single threshold value based on the negative template controls for each assay in a ddPCR run. The script/package is written in the R language (R version 3.0.3) and is made available at `http://www.ddpcrquant.ugent.be` (R Core Team, 2018).

Briefly, the ddpcRquant workflow is designed to use the combined data of multiple replicate NTCs of an assay to model the extreme values by extreme value theory (see results and Appendix C.4 for a detailed description). Subsequently, the threshold is set as a predefined percentile of the fitted extreme value distribution. Next, this threshold is used to calculate the negative and total droplets for each sample. Finally, absolute quantification of DNA template in the ddPCR mix is performed assuming a Poisson distribution, resulting in following formula:

$$C = -\log\frac{N_{neg}}{N} \times \frac{1000 \times D}{V_d}$$

with C: concentration (copies/μL stock), N_{neg}: negative droplets, N: total droplets, V_d: 0.91 nL, D: dilution factor (volume mix (μL)/volume sample (μL)), as described earlier (Hindson et al., 2013; Pinheiro et al., 2011).

4.4 Results

4.4.1 Droplet populations are not normally distributed

QQ plots suggest a substantial deviation from normality in most NTCs. Although to be interpreted with care due to the large sample size, p-values from the Anderson-Darling test confirmed that the fluorescence readouts from a large number of NTCs did likely not follow the normal distribution: 128 p-values were smaller than 10^{-6}, five were larger than 0.001 and only two p-values were larger than 0.05. Falsely assuming normal distribution can lead to a threshold too close to the negative droplet population and inaccurate quantification results (Figure 4.1C).

4.4.2 ddpcRquant for automated digital PCR data analysis

With the knowledge gained from the previous experiments, a new automated method for one-dimensional ddPCR data analysis is proposed. The ddpcRquant workflow sets a threshold based on the negative template controls which are performed with each ddPCR assay (Figure 4.2). The script was designed to use the combined data of replicate NTCs in case more than one is performed per assay. Consequently, the selected threshold is applied to the samples and the concentration is calculated.

4.4.2.1 Step 1: Input and automating process

The QuantaSoft software that comes with the ddPCR automatically generates a head file with .csv (comma separated value) extension which contains the metadata of each well in a ddPCR run (assay, sample, etc.). In addition to the head file, a second file is required per well which includes the fluorescent amplitude data for each separate droplet (Figure 4.2 Step1). These files need to be manually exported from the QuantaSoft software (Figure C.13). These amplitude files together with the head file are used as input for the script.

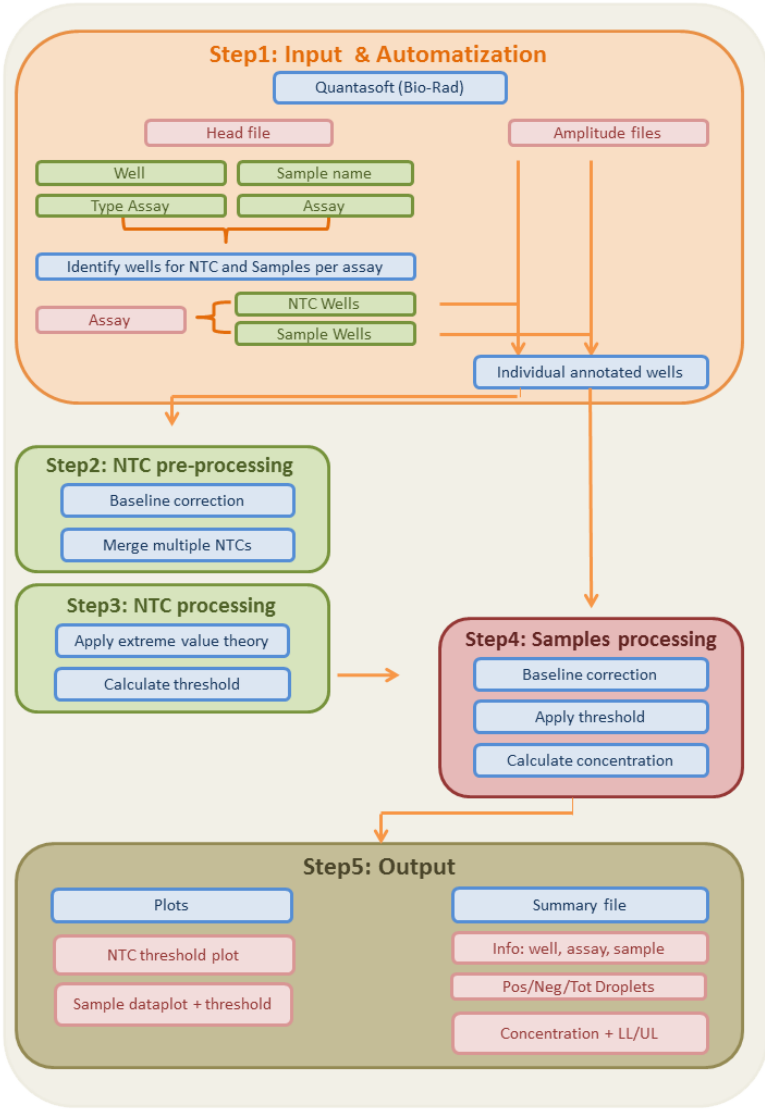


FIGURE 4.2: Overview of the ddpcRquant workflow. (Step1) Input and automatization process showing the required input information from the head file to automatically allocate the correct amplitude file to the individual wells. (Step2-3) The different steps in NTC preprocessing resulting in a calculated threshold based on extreme value theory. (Step4) Sample processing that applies the calculated threshold and Poisson statistics to return the concentration in each sample. (Step5) Output generation: Different plots and a summary file are generated and returned to the user together with an HTML output file containing all information

For automated data analyses purposes, the script uses four information parameters (well, assay, sample, type assay) stored in the headfile.csv to automatically locate NTCs and samples per assay and fetch the corresponding raw fluorescence data/amplitude files per well. Therefore, a consistent annotation is required (Figure C.14).

4.4.2.2 Step 2: NTC preprocessing

Because small shifts in baseline fluorescence can be observed, the NTC pre-processing comprises a baseline correction before merging fluorescence data of multiple NTCs per assay (Figure 4.2 Step2 and Figure C.2). For each sample the Robertson-Cryer mode estimator (also known as the half sample mode) (Robertson and Cryer, 1974; Bickel and Fruhwirth, 2006) is calculated and subtracted from all fluorescence intensities. After this baseline correction the data points of the individual NTCs of an assay are merged.

4.4.2.3 Step 3: NTC processing and threshold setting

The merged NTC dataset is used for the estimation of the threshold using methods relying on extreme value theory. The droplet fluorescence intensities are randomly assigned to k groups of equal size. The maximum fluorescence intensity within each of these groups is calculated. The k group maxima are then used to estimate the parameters of a generalized extreme value distribution using maximum likelihood (R evd package (Stephenson, 2002)). The 0.995 percentile (or other, user-determined) is then considered to be the threshold. This process of random grouping and parameter estimation is repeated 100 times after which the average of the 100 thresholds is taken to be the final threshold.

4.4.2.4 Step 4: Samples preprocessing and processing

The fluorescence intensities are baseline corrected by calculating the Robertson-Cryer mode estimator for the fluorescence intensities below a certain cutoff c and subsequently subtracting the estimated mode from the fluorescence intensities (Robertson and Cryer, 1974). This cutoff c is taken to be the average of the Robertson-Cryer estimates of the NTCs plus the threshold determined in step 3. Alternatively, c can be specified by the user.

By comparing the baseline-corrected intensities to the threshold of step 3, the droplets are classified into positive or negative droplets. The number of negative and positive droplets is used to calculate the concentration of DNA template in the ddPCR mix as previously described (Pinheiro et al., 2011; Hindson et al., 2011) (Figure 4.2 Step4).

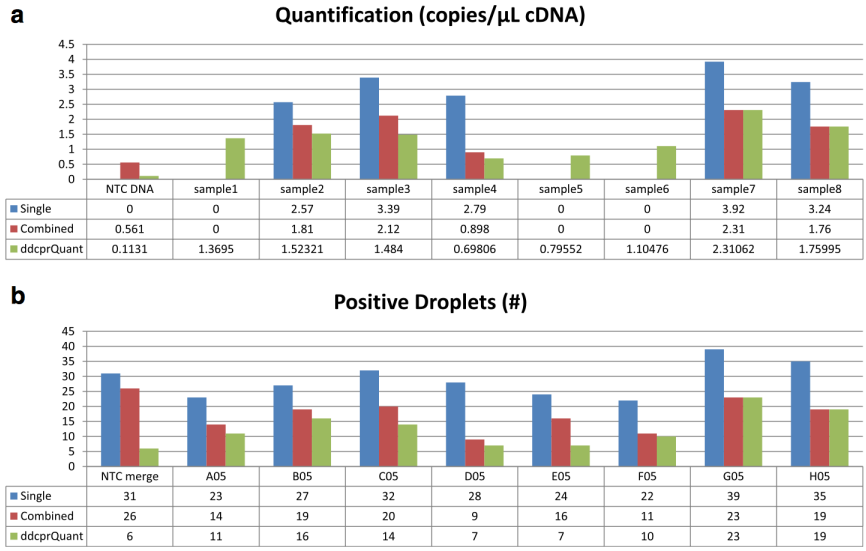
4.4.2.5 Step 5: Output

As a summary, a .csv file per assay is generated containing information of the well, assay, name, type, concentration (positive templates per μl ddPCR mix), upper and lower CI, number of positive, negative and total droplets. Next, plots are returned for distribution fitting of the NTC, the data plot of the NTC with calculated threshold and the sample data plots with the threshold (Figure 4.2 Step5). As a total overview, all data output is saved in an HTML (HyperText Markup Language) file that is opened automatically upon ending of the analysis.

4.4.3 ddpcRquant vs. QuantaSoft on an example dataset

In order to compare the method of ddpcRquant to the QuantaSoft software provided by the manufacturer (BioRAD), 1 of the 29 HIV quantification datasets was used as an example (Liszewski et al., 2009). Here, both the single and combined well automated analysis of QuantaSoft was compared to the ddpcRquant algorithm (Figure 4.3 and Figure C.15).

Processing of the combined NTCs shows that in the automated analysis methods of QuantaSoft (Single and Combined Well auto analysis) the number of positive droplets was respectively 31 and 26 droplets, whereas the ddpcRquant analysis results in only 6 positive droplets (Figure 4.3B and Figure C.15). Consequently, sample processing/quantification by the QuantaSoft analysis methods shows an overall higher concentration level of template (Figure 4.3A and Figure C.15). In addition, in some samples no concentration was calculated by the QuantaSoft methods in comparison with ddpcRquant.



	NTC DNA	sample1	sample2	sample3	sample4	sample5	sample6	sample7	sample8
Single	0	0	2.57	3.39	2.79	0	0	3.92	3.24
Combined	0.561	0	1.81	2.12	0.898	0	0	2.31	1.76
ddcprQuant	0.1131	1.3695	1.52321	1.484	0.69806	0.79552	1.10476	2.31062	1.75995

	NTC merge	A05	B05	C05	D05	E05	F05	G05	H05
Single	31	23	27	32	28	24	22	39	35
Combined	26	14	19	20	9	16	11	23	19
ddcprQuant	6	11	16	14	7	7	10	23	19

FIGURE 4.3: Comparison of QuantaSoft auto-analysis methods (single and combined well analysis) and ddpcrQuant on an example dataset of an HIV total DNA quantification assay. Overview of the calculated concentration (A) and number of positive droplets (B) in the NTC and samples

4.5 Discussion

4.5.1 Considerations for extreme value theory

Extreme value theory studies the distribution of a series of maxima of large samples. In the current context the droplets from an NTC can be distributed over a fixed number of groups so that each group still contains a large number of droplets. Further calculations will only require the largest intensity from each group. Provided that the number of observations in each group is large, extreme value theory implies that the maxima are distributed as a generalized extreme value (GEV) distribution, whatever distribution of the original observations (known as the parent distribution) (Fisher–Tippett theorem). This is asymptotically a nonparametric theory because the GEV holds whatever the parent distribution of the droplet intensities that maxima were sampled from, as long as the number of observations in each group is sufficiently large. Hence, our method does not impose any distributional assumptions on the droplet intensities.

A second argument in favor of extreme value statistics is that setting a threshold as a percentile of the intensity parent distribution is difficult because the percentile is typically in the far end of the tail. In particular, with a parametric fit of a hypothesized parent distribution (e.g. a normal distribution) the correctness of the percentile is very sensitive to the distributional assumption, and with a nonparametric density estimate the far end percentile cannot be estimated accurately (Malec and Schienle, 2014).

In applying the extreme value methodology, the user should be aware of setting certain parameters and how they influence the interpretation of the obtained results. This is discussed extensively in Appendix C.4. In summary the user should consider block size, number of NTCs and the tradeoff between specificity and sensitivity.

Our methodology furthermore corrects for NTC-to-NTC (or sample-to-sample) variability in baseline fluorescence levels by calculating a robust estimate of the mode which is to be used as a normalization constant. This robust mode estimator results in stable estimates of the extreme value distribution parameters and threshold when pooling data from several NTCs (Figure C.8).

4.5.2 Comparison of ddpcRquant to QuantaSoft

The major reason for the observed discrepancy in concentration levels is that the calculated threshold by QuantaSoft is placed too close to negative droplet population (Figure C.16 and Figure C.17). This can be caused by: 1) no correction for the observed shift in baseline/background fluorescence between NTCs/samples (Figure C.16), 2) assuming a normal distribution of the negative droplet population without performing normality testing/distribution fitting experiments. Indeed, distribution fitting showed that the droplet population is not normally distributed (Figure C.18).

In the output generated by ddpcRquant these problems are resolved and the threshold is placed correctly (Figure C.19).

4.6 Conclusion

The ddpcRquant algorithm is developed for single color analysis and is applicable to high and low level detection. Furthermore, the ddpcRquant method accounts for two observed problems with current data analysis methods: 1) baseline shift between samples, 2) the requirement of the normal distribution. By implementing a baseline correction and extreme value theory, these issues are resolved.

In addition, the type of probe and the composition of the sample (e.g. DNA, cDNA, eluted DNA, direct lysate) can affect both the distribution as well as the amount of background fluorescence. Our method is able to cope with these situations, provided that the NTCs are run in the same background (e.g. DNA, lacking the specific template) as the samples.

Finally, we emphasize that ddpcRquant is a data driven method to assist the end-user determine a threshold and perform absolute quantification. Nonetheless, the end-user is still encouraged to evaluate the threshold setting and responsible for the interpretation/quality control of the ddPCR experiment.

5 Improved estimates with mixed models

This chapter is an adaptation from the paper: Vynck, M., Vandesompele, J., Nijs, N., Menten, B., De Ganck, A. and Thas, O. (2016). Flexible analysis of digital PCR experiments using generalized linear mixed models. Biomolecular Detection and Quantification, 9, 1-13.

Contributions: Matthijs Vynck, Jo Vandesompele and Olivier Thas conceived the analysis strategy. Björn Menten collected the samples. Jo Vandesompele, Nele Nijs, Björn Menten, Ariane De Ganck designed and performed the wet-lab experiments. Matthijs Vynck, Jo Vandesompele, Ariane De Ganck and Olivier Thas analyzed and interpreted the data. Matthijs Vynck, Jo Vandesompele and Olivier Thas developed the statistical methodology. Matthijs Vynck, Jo Vandesompele and Olivier Thas wrote the manuscript. Matthijs Vynck and Olivier Thas wrote the R code. All authors critically reviewed and agreed with the contents of the paper.

5.1 Abstract

The use of digital PCR for quantification of nucleic acids is rapidly growing. A major drawback remains the lack of flexible data analysis tools. Published analysis approaches are either tailored to specific problem settings or fail to take into account sources of variability. We propose the generalized linear mixed models framework as a flexible tool for analyzing a wide range of experiments. We also introduce a method for estimating reference gene stability to improve trueness and precision of copy number and relative expression estimates. We demonstrate the usefulness of the methodology on a complex experimental setup.

5.2 Introduction

The number of publications on digital PCR (dPCR) have markedly increased during the last decade, with a rapid growth of publications in the field of biomedical sciences in recent years. This adoption has in part been possible due to an increase of commercially available, user-friendly instruments (Morley, 2014; Baker, 2012) and is further stimulated by positive reports on dPCR demonstrating the advantages over quantitative PCR (qPCR) (Huggett et al., 2015), particularly for applications such as low-level quantification (Sanders et al., 2011; Strain et al.,

2013), absolute quantification (Sanders et al., 2011; Strain et al., 2013) and copy number variation (CNV) determination (Whale et al., 2012).

Despite the advantages and increasing popularity of dPCR and as a consequence of the technique still being in its infancy, one major drawback of dPCR remains the lack of dedicated data analysis tools taking full advantage of the specific digital nature of the data. Most published papers rely on data-analysis software provided by hardware manufacturers. These software suites are typically black-box tools providing the user with a limited amount of information on the algorithms. They furthermore do not allow the user to analyze more complicated experimental setups such as the correct use of technical replicates or the use of multiple reference loci for determining CNVs, even though such approaches may be advisable (Vandesompele et al., 2002; Versluis et al., 2015; Zmienko et al., 2015).

Although several papers have been published that propose data analysis methods, these methods have been developed to analyze very specific experimental setups. For example, Whale et al. (2012) and Dube et al. (2008) developed ad hoc methods for calculating CNVs, but these methods can only be used to calculate CNVs using a single reference locus and do not take into account interreplicate variability. Extending these methods to cope with other experimental setups would require significant work, tailored to each of these specific designs. A major difficulty is the correct estimation of standard errors and confidence intervals.

In this paper, we detail how the established generalized linear mixed model (GLMM) framework (Stroup, 2012) can be used to analyze dPCR data from a wide range of experimental setups, ranging from simple experiments such as absolute quantification to complicated studies such as CNV estimation with multiple reference loci normalization and handling of variable numbers of technical replicates, while correctly accounting for various sources of variability. The basis of this GLMM framework has recently also been described by Dorazio and Hunter (2015). We argue that known sources of variability should be accounted for and that the approach of pooling counts of technical replicates used for analysis by Dorazio and Hunter (2015) (among others, e.g. Yu et al. (2015)) may lead to incorrect estimation of standard errors and confidence intervals.

Further, a novel approach for selecting stable reference loci for CNV studies from a pool of candidate reference loci is developed and successfully applied. An approach for reference gene selection in relative expression experiments is also suggested.

To demonstrate the flexibility of the approach, our methodology is used to analyze a dataset consisting of droplet digital PCR (ddPCR) data for 14 individuals who have been screened for chromosomal abnormalities using 14 genes on 6 chromosomes. The performance in terms of trueness and precision is evaluated for calculating CNVs using both a single reference locus and multiple reference loci.

5.3 Materials and Methods

5.3.1 Absolute quantification

dPCR splits a sample mixture into partitions. Each of these partitions is subsequently called as containing target nucleic acid, or having no target nucleic acid. A positive signal thus indicates that one or more target copies may be present. As a consequence of the random partitioning of copies, the number of copies in a partition is assumed to follow a Poisson distribution with parameter λ which has the interpretation of the average number of copies per partition. If Y_j^* denotes the unobserved number of copies in partition j ($j = 1,\ldots,J$, with J the number of partitions), then we can write the observed digital outcome as the binary variable Y_j:

$$Y_j = \min(Y_j^*, 1) = \begin{cases} 0 & \text{if } Y_j^* = 0 \\ 1 & \text{otherwise.} \end{cases} \tag{5.1}$$

Having observed the digital outcomes, the λ parameter of the Poisson distribution can be estimated from the probability of zero copies, relying on the probability mass function of the Poisson distribution (Equations 5.2-5.3):

$$\mathrm{P}\left\{Y_j^* = 0\right\} = \frac{\lambda^0}{0!}\exp(-\lambda) = \exp(-\lambda) \tag{5.2}$$

$$\lambda = -\log \mathrm{P}\left\{Y_j^* = 0\right\} = -\log \mathrm{P}\left\{Y_j = 0\right\} \tag{5.3}$$

The final equality in Equation 5.3 follows from the construction of the binary outcomes (Equation 5.2). Since a probability of a binary event can be estimated from simple counts, an estimate of λ is given by

$$\hat{\lambda} = -\log\left(\frac{\text{number of negative partitions}}{\text{total number of partitions}}\right). \tag{5.4}$$

$\hat{\lambda}$ can also be obtained using a Generalized Linear Model (GLM). The GLM for the unobserved counts Y_j^* is specified by a Poisson distribution with mean λ related to a parameter β_0 through a log-link function,

$$\log \lambda = \beta_0. \tag{5.5}$$

Using Equation 5.3, the observed binary outcomes Y_j can be described by a binomial distribution with probabilities

$$\begin{aligned} \mathrm{P}\left\{Y_j = 0\right\} &= \mathrm{P}\left\{Y_j^* = 0\right\} = \exp(-\lambda) \\ &= \exp(-\exp(\beta_0)) \\ \mathrm{P}\left\{Y_j = 1\right\} &= \mathrm{P}\left\{Y_j^* > 0\right\} = 1 - \exp(-\lambda) \\ &= 1 - \exp(-\exp(\beta_0)). \end{aligned} \tag{5.6}$$

Equations 5.6 state a GLM for a binomial distribution with a complementary log-log link. The more conventional model formulation is:

$$\log(-\log(\mathrm{P}\left\{Y_j = 0\right\})) = \beta_0, \tag{5.7}$$

where β_0 is the same as in Equation (5.5). Since the digital outcomes Y_j are observed, GLM software can be used for estimating β_0. If $\hat{\beta}_0$ denotes the estimate, an estimate of λ is then given by

$$\hat{\lambda} = \exp(\hat{\beta}_0). \tag{5.8}$$

Using Equation 5.4 or Equation 5.8 will result in the same estimate for λ.

Assuming a constant volume of the partitions, say $V_{\text{partition}}$, the concentration can be estimated from the average number of copies per partition (Equation 5.9):

$$\hat{c} = \frac{\hat{\lambda}}{V_{\text{partition}}}. \tag{5.9}$$

To obtain a reliable estimate of the concentration, an experiment is typically replicated. We now define Y_{ij}^* as the number of copies in partition j of replicate i ($j = 1, \ldots, J_i$, with J_i the number of partitions in replicate i, $i = 1, \ldots, I$, with I the number of replicates). As before, the counts are not observable, but upon

applying Equation 5.1, binary outcomes Y_{ij} can be calculated. To take the replicate variability into account, we introduce a random effect for the replicate in the Poisson model. Within a replicate, the counts are still Poisson distributed. The statistical model is formulated hierarchically. In particular, within a replicate:

$$Y_{ij}^* \mid R_i \sim \text{Poisson}(\lambda_i) \tag{5.10}$$

where

$$\log \lambda_i = \beta_0 + R_i, \tag{5.11}$$

with R_i the effect of replicate i on the Poisson mean. These replicate effects R_i are described by a normal distribution,

$$R_i \sim N(0, \sigma^2). \tag{5.12}$$

The model results again in a binomial regression model with a complementary log-log link for the observed digital outcomes. In particular, within a replicate

$$\log(-\log(\text{P}\left\{Y_{ij} = 0 \mid R_i\right\})) = \beta_0 + R_i, \tag{5.13}$$

with β_0 and R_i as before. The model is a special case of a GLMM (Stroup, 2012). Statistical software is available for estimating the model parameters (e.g. R(R Core Team, 2018), an environment often used for analysis of PCR experiments (Rodiger et al., 2015)), including random effect variances (Bates et al., 2015).

The objective is to estimate the mean number of copies, averaged over all replicates, i.e. $\text{E}\left\{Y_{ij}^*\right\}$ is the quantity of interest for absolute quantification. Statistical theory (Appendix D.4.1) gives

$$\text{E}\left\{Y_{ij}^*\right\} = \exp(\beta_0 + 0.5\sigma^2). \tag{5.14}$$

From the estimate of β_0 (say $\hat{\beta}_0$), the estimate of the variance σ^2 of the random effect (say $\hat{\sigma}^2$) and from Equation 5.9 a concentration estimate can subsequently be calculated as

$$\hat{c} = \exp(\hat{\beta}_0 + 0.5\hat{\sigma}^2)/V_{\text{partition}}. \tag{5.15}$$

The statistical software also gives the estimated standard errors of the estimates $\hat{\beta}_0$ which can be used for the calculation of an approximate confidence interval of the concentration (Appendix D.4.3). Example analyses are given in Appendix

D.1.2.1.

5.3.2 Copy number variation

For the estimation of CNV, data on both a target and at least one reference must be available. Several experimental designs are appropriate for obtaining target and reference measurements. Figure 5.1 shows six examples, ranging from single reference settings with single channel experiments (panel A) or duplex experiments (panel B) to multiple reference studies with single channel (panel C) or duplex (panel D, E) or multiplex (panel F) experiments. In this section, a GLMM methodology is outlined that is applicable to all of these designs, also in the presence of replicates.

The general guideline for obtaining valid statistical estimation, error propagation and hypothesis testing, is that the data analysis method should account for dependencies and sources of variability implied by the experimental setup. For example, as in Section 5.3.1, random replicate effects should be included in the model to take care of the dependence between droplets within the same replicate.

5.3.2.1 Single reference designs

The same notation (Y_{ij} and Y^*_{ij}) as before is used, but the partition index j may now refer to a measurement which can be from a target or a reference. The distinction between target and reference is made by a dummy regressor X_{ij} which is defined as zero when partition (i, j) comes from the target and one when it comes from the reference. For the designs A and B (Figure 5.1), the model for the unobservable number of copies is written as

$$Y^*_{ij} \mid R_i \sim \text{Poisson}(\lambda_{ij}) \tag{5.16}$$

where

$$\log \lambda_{ij} = \beta_0 + X_{ij}\beta_1 + R_i \tag{5.17}$$

and

$$R_i \sim N(0, \sigma^2). \tag{5.18}$$

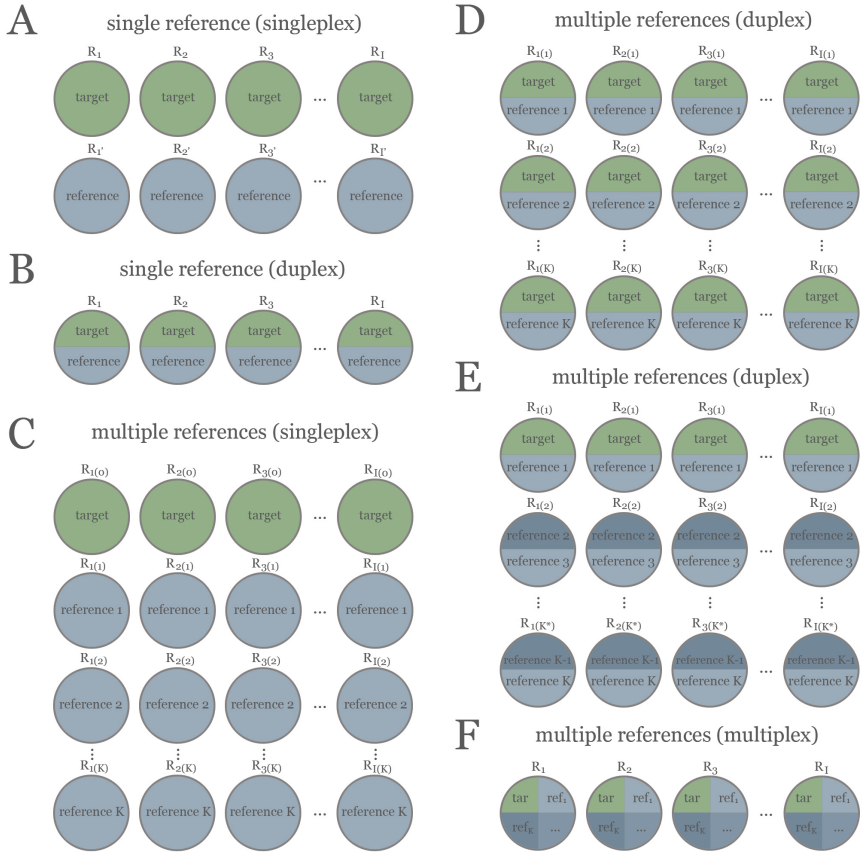


FIGURE 5.1: Experimental designs for calculating copy numbers using digital PCR. R denotes a replicate. (A) Singleplex experiments with a single reference gene. (B) Duplex experiments with a single reference gene. (C) Singleplex experiments with multiple reference genes. (D) Duplex experiments with multiple reference genes and repeated analysis of the target gene. (E) Efficient duplex experiments with multiple reference genes. (F) Multiplex experiments with multiple reference genes. For each design, a single droplet is shown for each replicate. The circles represent partitions and they show whether a target or a reference is included in either singleplex, duplex or multiplex. The replicates are indicated with the letter *R*, the total number of replicates is given by I. Setup-specific subscripts (prime, asterix) as used in the model specifications of Section 5.3.

Thus within replicate i, the mean number of target copies per partition again equals $\exp(\beta_0 + 0 \times \beta_1 + R_i)$, and the mean number of reference copies per partition equals $\exp(\beta_0 + 1 \times \beta_1 + R_i)$.

Let $c_{\text{target},i}$ and $c_{\text{ref},i}$ denote the concentrations of target and reference in replicate i, respectively, and N_b the ploidy of the organism. For design A (Figure 5.1), the CNV based on replicate i for the target and replicate i' for the reference, is given by

$$\text{CNV}_{i,i'} = \frac{c_{\text{target},i}}{c_{\text{ref},i'}} N_b = \frac{\exp(\beta_0 + 0 \times \beta_1 + R_i)/V_{\text{partition}}}{\exp(\beta_0 + 1 \times \beta_1 + R_{i'})/V_{\text{partition}}} N_b = \exp(-\beta_1 + R_i - R_{i'}) N_b. \tag{5.19}$$

The overall CNV is then given be the average of $\text{CNV}_{i,i'}$ over all replicates (see Appendix D.4.2), resulting in

$$\text{CNV} = \text{E}\left\{\text{CNV}_{i,i'}\right\} = \exp(-\beta_1 + \sigma^2) N_b. \tag{5.20}$$

As before, the model parameters may be estimated by reformulating the model for the digital outcomes Y_{ij}. In particular, a GLMM with a complementary log-log link is obtained:

$$\log(-\log(\text{P}\left\{Y_{ij} = 0 \mid R_i\right\})) = \beta_0 + X_{ij}\beta_1 + R_i, \tag{5.21}$$

with β_0, β_1 and R_i as in model 5.17.

For design B (Figure 5.1), the CNV based on replicate i, which now contains droplets with both target and reference (duplex), is given by

$$\text{CNV}_i = \frac{c_{\text{target},i}}{c_{\text{ref},i}} N_b = \frac{\exp(\beta_0 + 0 \times \beta_1 + R_i)/V_{\text{partition}}}{\exp(\beta_0 + 1 \times \beta_1 + R_i)/V_{\text{partition}}} N_b = \exp(-\beta_1) N_b. \tag{5.22}$$

Note that the random effect cancels out and that the CNV does not depend on i. Hence, an overall CNV estimate is given by $\exp(-\hat{\beta}_1) N_b$, with the estimates again calculated from the GLMM with a complementary log-log link. The random effect can however not be omitted altogether, as it influences the variance on the fixed effect parameters, and thus the inclusion of the random effect is essential for a correct error propagation.

5.3.2.2 Multiple reference designs

The model can be further extended to contain multiple reference loci. The number of copies and the deduced binary outcome for partition (i, j) are denoted by Y^*_{ijk} and Y_{ijk}, respectively, in which the index k refers to the reference $k = 1, \ldots, K$, with K the number of reference loci and with $k = 0$ referring to the target. Consider the dummy X_{ijk}, which is defined as one when the signal belongs to the kth reference and zero when the signal comes from the target. Reference-to-reference differences are allowed by making use of nested random effects.

For designs C and D, for a given replicate i and for a given target or reference k, the Poisson model for the unobserved counts Y^*_{ijk} has log-mean

$$\begin{aligned} \log \mathrm{E}\left\{Y^*_{ijk} \mid S_k, R_{i(k)}\right\} &= \log \lambda_{ijk} \\ &= \beta_0 + \beta_1 X_{ijk} + S_k X_{ijk} + R_{i(k)} \end{aligned} \tag{5.23}$$

with S_k the effect of reference k on the log-mean, and $R_{i(k)}$ the effect of the ith replicate of the experiment with the PCR mix containing reference k (or $k = 0$ for target in design C). The variability of these two random effects are described by independent normal distributions:

$$S_k \sim N(0, \sigma_1^2) \text{ and } R_{i(k)} \sim N(0, \sigma_2^2). \tag{5.24}$$

Hence, S_k is a random effect for the between reference locus variation and $R_{i(k)}$ is a random effect for the interreplicate variation nested within a given target or reference. Note that the model formulation assumes that the random effects are exchangeable (see Appendix D.4.4 for more information).

The same model applies to design E, except that the index k in $R_{i(k)}$ should be replaced by an index k^* which is an indicator of the unique PCR mix (each row in panel E of Figure 5.1 represents a unique PCR mix). The model for design F is also similar, except that the replicate effect $R_{i(k)}$ does not depend on reference k, because in this multiplex experiment all references are potentially included in all partitions, i.e. in each replicate all references are included in the PCR mix. Hence, the nested random effect $R_{i(k)}$ in model 5.23 has to be replaced by R_i.

As before, the model parameters can be estimated from the corresponding GLMM for the binary outcome:

$$\log(-\log(\mathrm{P}\left\{Y_{ijk}=0 \mid S_k, R_{i(k)}\right\})) = \beta_0 + \beta_1 X_{ijk} + S_k + R_{i(k)}. \tag{5.25}$$

For design C the CNV is first given for target versus a single reference k, based on replicates i and i':

$$\mathrm{CNV}_{i,i';k} = \frac{\exp(\beta_0 + R_{i(0)})}{\exp(\beta_0 + \beta_1 + S_k + R_{i'(k)})} N_b = \exp(-\beta_1 - S_k + R_{i(0)} - R_{i'(k)}) N_b. \tag{5.26}$$

The overal CNV is obtained by averaging over all replicates and all references (see Appendix D.4.2 for details):

$$\mathrm{CNV} = \mathrm{E}\left\{\mathrm{CNV}_{i,i';k}\right\} = \exp(-\beta_1 + \frac{1}{2}\sigma_1^2 + \sigma_2^2) N_b. \tag{5.27}$$

For design D, which allows for CNV calculation w.r.t. a single reference k within a replicate i (duplex), we first give

$$\mathrm{CNV}_{ik} = \frac{\exp(\beta_0 + R_{i(k)})}{\exp(\beta_0 + \beta_1 + S_k + R_{i(k)})} N_b = \exp(-\beta_1 - S_k) N_b. \tag{5.28}$$

Note that the replicate effect has been eliminated (due to combining target and reference k in the duplex). After averaging over references, the overall CNV becomes

$$\mathrm{CNV} = \mathrm{E}\left\{\mathrm{CNV}_{ik}\right\} = \exp(-\beta_1 + \frac{1}{2}\sigma_1^2) N_b. \tag{5.29}$$

For design E, the CNV w.r.t. a single reference measured in duplex with the target is given by Equation 5.28. CNVs for the other $k-1$ references measured in duplex with respect to one another are given by Equation 5.26. An overall CNV is obtained as

$$\mathrm{CNV} = \exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2\right) N_b \frac{(1 + \exp(\sigma_2^2)(k-1))}{k}. \tag{5.30}$$

For design F (multiplex) we also start with the CNV calculation w.r.t. a single reference k within a replicate i :

$$\mathrm{CNV}_{ik} = \frac{\exp(\beta_0 + R_i)}{\exp(\beta_0 + \beta_1 + S_k + R_i)} N_b = \exp(-\beta_1 - S_k) N_b. \tag{5.31}$$

Also here the replicate effect is eliminated. After averaging over references, the overall CNV becomes

$$\mathrm{CNV} = \mathrm{E}\{\mathrm{CNV}_{ik}\} = \exp(-\beta_1 + \frac{1}{2}\sigma_1^2) N_b. \tag{5.32}$$

Detailed derivations are available in Appendix D.4.2. Example analyses in R are given in Appendix D.1.2.2.

5.3.3 Reference gene stability

Reference locus stability for CNV estimation is calculated for each reference locus k ($k = 1, \ldots, K$) as

$$\mathrm{Stability}_k = \frac{\sum_{l \in S_k} (\hat{\beta}_{1kl}^2 + \mathrm{Var}(\hat{\beta}_{1kl}))}{K - 1} \tag{5.33}$$

where S_k is the set $\{1, \ldots, K\} \setminus k$. If only data from one sample are available, the $\hat{\beta}_{1kl}$ are parameter estimates that result from fitting the model

$$\log \lambda_{ij} = \beta_{0kl} + \beta_{1kl} X_{ij} + R_i \tag{5.34}$$

with $R_i \sim N(0, \sigma^2)$ and X_{ij} defined as zero when partition (i, j) comes from reference k and one when it comes from reference l. For fitting model 5.34 only the data from references k and l are used. An example analysis is given in Appendix D.1.3.3.

For across-sample stability in the CNV case, the model needs to account for variability between samples: similarly to the introduction of the between-reference gene random effect, we introduce a between-sample random effect S_s that accounts for the sample-specific effects. Consequently, model 5.34 is replaced by

$$\log \lambda_{ij} = \beta_{0kl} + \beta_{1kl} X_{ij} + R_{i(s)} + S_s \tag{5.35}$$

with $R_{i(s)} \sim N(0, \sigma_1^2)$, $S_s \sim N(0, \sigma_2^2)$ and X_{ij} defined as before. Again only the data from references k and l are used.

The rationale for this stability measure is based on two arguments. First, when given a set of candidate reference genes in a CNV setup, we expect these reference genes to have similar copy numbers. If two candidate reference genes, say k and l, are in agreement (e.g. they both show a close to diploid copy number in e.g. a human sample), the estimate $\hat{\beta}_{1kl}$ is expected to be close to zero, or, equivalently, the ratio of their estimated concentrations will be close to one (Equations 5.20, 5.22). Thus, a reference k that gives a large $\sum_{l \in S_k} \hat{\beta}_{1kl}^2$ is said to be a biased reference.

Second, a good reference gene will be stable across replicates or samples, warranting the inclusion of the $\mathrm{Var}(\hat{\beta}_{1kl})$ estimate: if copy numbers across replicates or samples are highly variable, this will be reflected in a large variability of $\hat{\beta}_{1kl}$ (i.e. a large $\mathrm{Var}(\hat{\beta}_{1kl})$). Thus, references that are highly variable are less ideal and will be penalized. A detailed discussion is presented in Section 5.4.2.

To allow for different quantities of reference genes in a relative expression scenario (see detailed discussion in Section 5.4.2), model 5.35 is used, but the relative expression gene stability is simplified to

$$\mathrm{Stability}_k = \frac{\sum_{l \in S_k} \mathrm{Var}(\hat{\beta}_{1kl})}{K - 1} \tag{5.36}$$

i.e. only taking variability into account.

5.3.4 Generic formulation

Generally, the model can be written as a combination of fixed effect parameters – such as those of target and reference genes but also confounders or baseline variables, for example, age and gender – and random effect parameters such as interreplicate variation, interreference gene variation but also to account for e.g. variation between multiple laboratories. The general model can be written as

$$\log \lambda_{\cdot} = \beta_0 + \beta_1 X_{1\cdot} + \cdots + \beta_p X_{p\cdot} + Z_{1\cdot} + \cdots + Z_{q\cdot} \tag{5.37}$$

for p fixed effects and q random effects, where the subscripts which are here denoted by $\cdot$, refer to the variables used for constructing the X and Z variables.

The model can further be extended to allow for interactions, and random effects may be nested if implied by the study design.

One example of such an extension could be relative concentration estimation with multiple patients and with an effect of gender on the target concentration:

$$\log \lambda_{ij} = \beta_0 + X_{1ij}\beta_1 + X_{2ij}X_{3ij}\beta_2 + Z_{1i} \tag{5.38}$$

for partition j obtained from patient i, X_{1ij} is one if partition j of patient i is from the reference gene and zero if it is from the target gene, X_{2ij} is zero if patient i is male and one if the patient is female and X_{3ij} is zero for reference genes and one for target genes. The CNV for a male remains as in Equation 5.22, but for females it becomes:

$$\widehat{\text{CNV}} = \frac{\hat{c}_{\text{target}}}{\hat{c}_{\text{ref}}} N_b = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_2)}{\exp(\hat{\beta}_0 + \hat{\beta}_1)} N_b = \exp(\hat{\beta}_2 - \hat{\beta}_1) N_b. \tag{5.39}$$

Similarly, random effects can be added to account for e.g. stratification in multicenter trials. Absolute concentration estimation with an age effect, a patient effect and a center effect, could be modelled as:

$$\log \lambda_{ijc} = \beta_0 + X_{1ijc}\beta_1 + Z_{1c} + Z_{2i(c)} \tag{5.40}$$

where for center c, patient i and partition j, β_1 is the age effect, Z_{1c} is the random center effect (with variance σ_1^2) and $Z_{2i(c)}$ is the random patient effect (nested within the center effect, with variance σ_2^2). Using results similar as in Equation 5.14, the mean concentration for a patient of age *age* can then be estimated as

$$\hat{c} = \hat{\lambda}/V_{\text{partition}} = \exp(\hat{\beta}_0 + age \times \hat{\beta}_1 + 0.5\hat{\sigma}_1^2 + 0.5\hat{\sigma}_2^2)/V_{\text{partition}}. \tag{5.41}$$

For all models, the corresponding binary GLMM with complementary log-log link function is available in statistical software (e.g. R (R Core Team, 2018; Bates et al., 2015)).

5.3.5 Case study data

14 genes of interest (13 target loci located on chromosomes 13, 18, 21, X and Y along with a single reference locus for normalization, *RPP30*, located on chromosome arm 10q for normalization) from 10 samples with chromosomal abnormalities and 4 control samples were analyzed. DNA was extracted from blood samples using the QIAamp DNA Blood Mini Kit (Qiagen) according to the manufacturer's instructions, after which DNA concentration was measured using UV spectrophotometry (Nanodrop). All patient samples and healthy control samples were diluted to 5 ng/μl using nuclease-free water and 10X carrier solution (Roche's tRNA from brewer's yeast, cat-n°10109517001, 50 ng/μl) making the final tRNA carrier concentration 5 ng/μl. The no template control (NTC) sample also contained 5 ng/μl carrier. Two μl of the diluted DNA sample was added into the final ddPCR reaction resulting in a 10 ng sample input. Primers and probes were diluted to a work solution of 5 μM and 2 μM, respectively, using 1X IDTE buffer, pH8 (cat-n° 11-05-01-09). One ddPCR master mix was created per assay/probe pair (FAM/HEX or VIC) to perform 3 reactions per sample, control and NTC. One 20 μl reaction consisted out of following reagents: 10 μl 2X ddPCR Supermix for probes (Bio-Rad; cat-n° 1863010), 250 nM of each forward and reverse primer (final concentration), 100 nM of each probe (final concentration), and 2 μl of sample DNA (5 ng/μl, 10 ng final input) or nuclease-free water as NTC. The 20 μl ddPCR reaction mix was added to the droplet generator cartridge together with 70 μl droplet generation oil (Bio-Rad, cat-n° 1863005). Droplets were generated using a Bio-Rad QX100 droplet generator, followed by gentle transfer to a Twin.tec semi-skirted 96-well PCR plate. Using a Bio-Rad T100 thermal cycler, the following temperature cycling program was used for target amplification: 10 min 95 °C activation, 40 cycles of 30 sec 95 °C and 1 min 59 °C, followed by 10 min 98 °C. After PCR amplification, the plate was analyzed using a QX100 droplet reader using Quantasoft software. Data was exported as a CSV file for further processing.

Primers and hydrolysis probes were designed using an in house developed primerXL assay design engine (Lefever *et al.*, in preparation; http://www.primerxl.org), avoiding SNPs under the primer and probe annealing sites, avoiding secondary structures, and assessing primer specificity using genome wide Bowtie (Langmead et al., 2009) primer alignment, avoiding primers that have fewer than 3 mismatches

to a possible off-target homologous sequence. All primers and probes were ordered at Integrated DNA Technolgies, except for the *RPP30* probe which was ordered from Life Technologies. Sequences are available in Table D.5.

To allow comparison with digital PCR results, samples were also analyzed using arrayCGH, karyotyping and/or FISH in an ISO 15189 accredited lab at the Center for Medical Genetics, Ghent University Hospital as described before (Menten et al., 2006; Buysse et al., 2009; Menten et al., 2009).

5.3.6 Data analysis

To assess the influence of ignoring interreplicate variation in the case of absolute quantification, GLM and GLMM models were fitted. The GLM model (Equation 5.5) did not account for interreplicate variation, and thus the analysis was based on pooling the negative and positive droplet counts. The GLMM (Equation 5.11), on the other hand, was fitted using a random effect to account for the interreplicate variation. Estimates of both approaches and their variances were compared.

Using the proposed framework, the copy numbers for each of the target loci was calculated by using the *RPP30* locus as a reference (model Equation 5.20 or 5.22, Figure 5.1 panel A or B, depending on the target locus). Copy numbers were also calculated, again by constructing GLMM models, by using all loci located on autosomal chromosomes with close to normal diploid copy number as references (model Equation 5.30, Figure 5.1 panel E), except for sample 10 where no loci with normal diploid copy number were distinguishable. Model evaluation was done by calculating the mean absolute deviation from the closest integer copy number over all non-reference locus copy numbers in a given sample.

To assess trueness and precision with an increasing number of reference loci, target locus copy numbers were estimated by sequentially adding reference loci. To limit computational burden this was done for both a best case scenario, where the most stable reference loci were added first, and a worst case scenario, where the least stable reference loci were added first. After each step of adding a reference locus, trueness was assessed by calculating the mean absolute deviation from integer copy numbers of the targets. Precision was assessed by calculating the mean width of all target 95% confidence intervals.

For each of the patient samples in the case study, reference locus stability was determined using the stability measure described in Section 5.3.3 (model 5.34).

Retaining the most stable reference loci only, final copy numbers were determined for each of the target loci (Equation 5.30, setup as in Figure 5.1, panel E).

5.4 Results and Discussion

There is a clear need for more flexible data analysis tools for dPCR experiments. The GLMM framework is ideally suited to accommodate a wide range of dPCR experimental setups. As outlined in the Materials and Methods section, the framework can be used for absolute quantification, CNV calculation and gene expression analysis (with or without replicates), but it can also accommodate other applications such as mutation quantification both in singleplex, duplex or higher multiplexing mode, making it compatible with all existing dPCR instruments. As further demonstrated in the Materials and Methods section, the framework also allows adjusting for e.g. clinical baseline covariates such as age or gender, including treatments effects, or analysing multicenter trials.

The type of the data we have collected, allows us to assess differences between classical approaches and those described in this paper. The initial setup of the experiment was to study 13 target loci along with a single reference locus for normalization. Because the patients in this study suffer from only one type of chromosomal abnormality, loci located on the non-affected chromosomes will still retain their normal copy number status. In principle, this allows the use of loci located on unaffected chromosomes as reference loci, in addition to the *RPP30* reference locus. Studying chromosomal loss or gain furthermore has the advantage that under ideal reaction conditions and assuming no mosaicism, copy numbers are expected to be integers. This is in contrast to e.g. expression levels that may take any (non-integer) value. This property of the study allows to measure trueness as a deviance from an integer copy number.

5.4.1 Variance modelling

The need for modelling the different sources of variability becomes apparent from the estimates as obtained from the GLM models. As evidenced by Figure 5.3 there is often substantial variability between technical replicates as well as between reference loci. Not accounting for these sources of variability may lead to overly

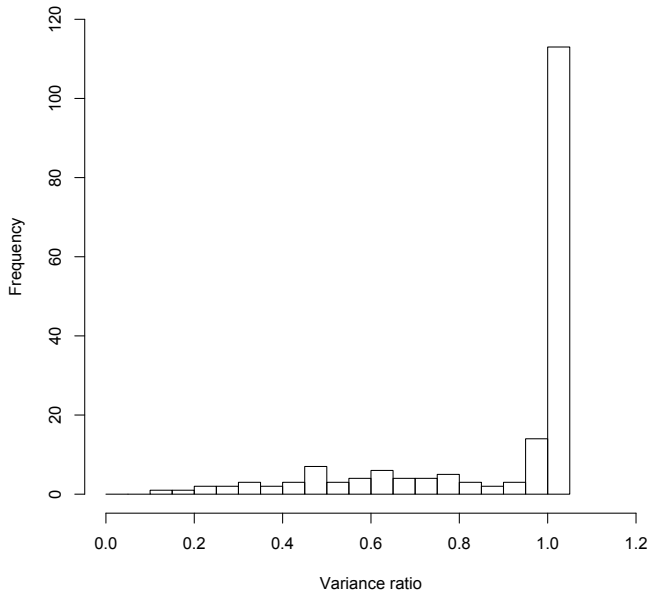


FIGURE 5.2: **Importance of modelling variability** Histogram of the ratio of variances of the parameter estimates of the naive pooling method relative to the GLMM. A ratio smaller than one indicates that the pooling method is too optimistic: the variance estimate of the pooling method is smaller than the variance estimate accounting for interreplicate variation. Ignoring this variation results in a variance estimate that is too small and consequently confidence intervals that are too narrow. It is recommended to always account for the possibility of variation between replicates to get an accurate variance estimate.

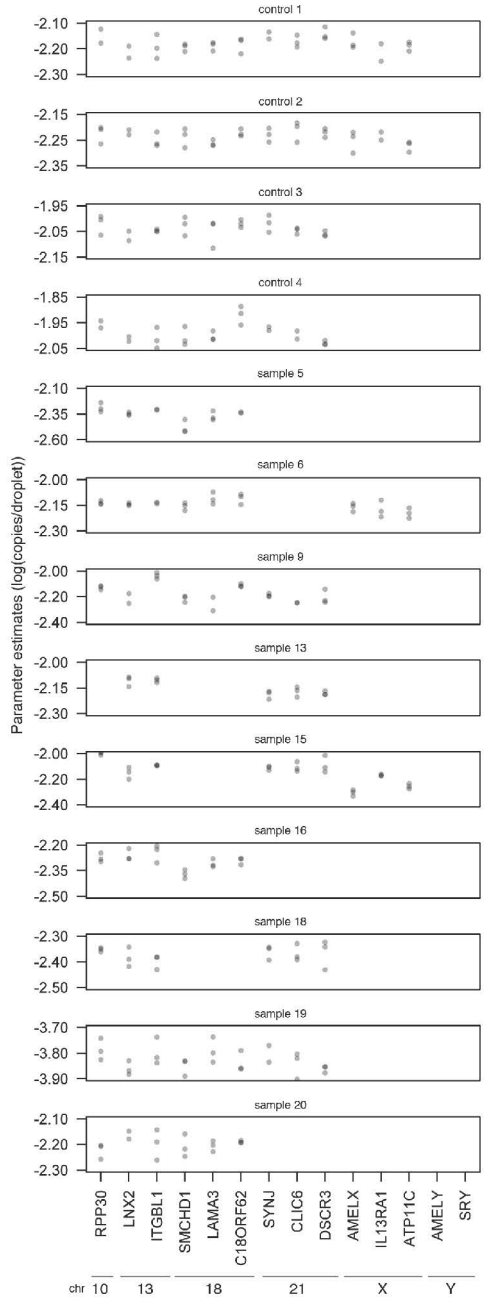


FIGURE 5.3: **Variability across replicates and reference loci** Parameter estimates of absolute quantities of the 14 candidate reference loci in 13 samples. Loci displayed are those with close to diploid copy number. Loci with a single or three copies are not displayed. For each reference locus and each sample, technical replicates are shown. The graphs demonstrate the presence of substantial variability between technical replicates, between samples and between reference loci.

optimistic uncertainty estimates. This is observed when comparing the uncertainty estimates of a naive pooling strategy (ignoring interreplicate variability) with those of a GLMM model accounting for interreplicate variability in an absolute quantification scenario. Figure 5.2 shows a histogram of the ratio of the variance of the absolute quantification obtained with the pooling strategy relative to the variance calculated from the GLMM method. Even though the variability estimates of both methods are often close in our experimental data (in a majority of samples the fold difference was approximately one), the variability was underestimated up to a factor ten when using a naive pooling strategy (Figure 5.2). It is thus recommended to always incorporate the interreplicate variability to avoid overoptimistic variance estimates. This has also been demonstrated by Jacobs et al. (2014) through simulation: pooling replicate data results in faulty confidence intervals, while accounting for interreplicate variation results in correct variance estimates and, consequently, correct confidence intervals.

The same holds true for accounting for variability between reference loci: there is often a discrepancy between different candidate reference loci (Figure 5.3) and ignoring this may result in an underestimation of the uncertainty on relative quantity or CNV estimates.

Dorazio and Hunter (2015) argue that accounting for additional sources of variability should only be considered if concentrations are expected or can be shown to be different. One can argue that due to the imperfectness of e.g. sample processing in a typical experiment (e.g. using replicates) such sources of variation will always be present and should always be considered (Jacobs et al., 2014). Dorazio and Hunter (2015) furthermore propose the use of deviance statistics to assess model goodness of fit. They argue that their model is not worse than the saturated model based on this deviance statistic, but they do not make a comparison with models accounting for additional sources of variability such as interreplicate variability. Even if such a comparison would demonstrate that a model not accounting for interreplicate variability is not significantly worse than one accounting for interreplicate variability, this may be a weak conclusion as goodness of fit tests may suffer from a lack of power (Hosmer and Hjort, 2002). For example, using our data we found that a lack of fit was detected only when the uncertainty using the GLMM method was approximately at least three times that of the uncertainty estimated using the GLM method (results not shown). When unsure about the presence or power of detecting additional sources of variation, it is thus better to be cautious and to allow for the possibility of these additional

sources of variability, rather than to take the risk to obtain too optimistic results.

5.4.2 Multiple reference loci

It is recommended to use multiple reference loci, for relying on a single reference locus for calculating a CNV may result in reduced trueness (Vandesompele et al., 2002; Dhaene et al., 2010). Especially in the case of e.g. erroneous amplification (e.g. due to inhibition) or copy number alteration of the candidate reference locus, estimated copy numbers may be far from unbiased. This is clearly observed in the results from the case study, particularly in sample 15: when relying solely on the locus located on the *RPP30* gene, which is a popular reference gene in CNV studies (Hindson et al., 2011), an underestimate of copy numbers of target loci is obtained. This has also been observed in other studies. For example, Versluis et al. (2015) observed a gain of the *TERT* reference gene in two samples, resulting in erroneous results. They argue that relying on 3 to 4 reference loci would circumvent this problem.

These problems can indeed be remediated by relying on multiple reference loci, in which case e.g. erroneous amplification of one of the reference loci will only partially influence the estimated copy number. This is illustrated in Figure 5.4 where relying on the predetermined reference locus on the *RPP30* gene often results in much larger mean absolute deviations from integer copy numbers when compared to relying on multiple reference loci. In those samples where the single locus normalization performs much worse, the *RPP30* locus is unsuitable for normalization (Appendix D.2.1).

The trueness of the estimate can be further improved by using a selection of multiple stable reference loci. This can also be done using the GLMM model (possibly accounting for interreplicate variability). Our measure of stability takes both bias and variance in account. Because stable reference loci are expected to have the same copy number, the ratio of their concentrations should be close to one, or, in terms of model 5.34 or 5.35, the parameter estimates $\hat{\beta}_{1kl}$ should be close to zero. Hence, a ratio of one, or a parameter estimate of zero corresponds to no bias. By making all pairwise comparisons between a reference locus and all other reference loci, and calculating the sum of the squared deviations from zero, an estimate of the total squared bias of the reference locus is obtained: optimal reference loci

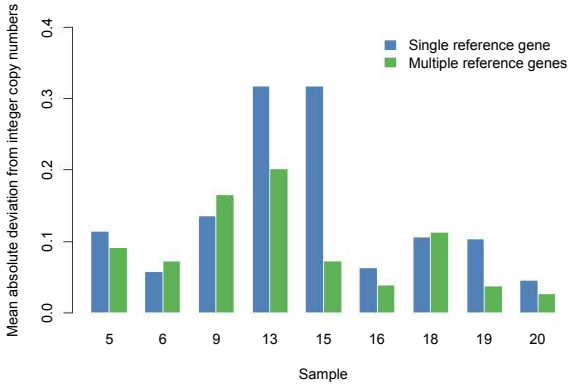


FIGURE 5.4: **Effect of normalization strategy** Effect of normalization strategy on the mean absolute deviation from integer copy numbers for nine samples with chromosomal abnormalities. A low mean absolute deviation indicates that the obtained estimates are close to the expected copy numbers. Using multiple stable reference loci can result in a substantial increase in trueness if the single reference locus was biased. If it is not beneficial, the difference with the single reference locus is small.

will have a low bias while loci with large bias indicate that they are not in agreement with the other candidate reference loci. Variance is taken into account by using the estimate of the variance of the bias parameter estimate: loci that have a more variable concentration across replicates or samples (as propagated in the variance of the estimate) are less ideal reference loci.

The stability measure is thus penalising aberrant reference loci in two ways: loci that have a concentration deviating from other reference loci will be penalised as well as those with higher variance. Optimal reference loci are both in accordance with other reference loci and show little variation when replicated. Reference loci having similar bias will be ranked from lower to higher variance, and likewise loci with similar variance will be ranked from lower to higher bias. Figure 5.5 indicates that for example for patient sample 5, the genes *LNX2*, *LAMA3* and *C18ORF62* are most stable and can be used as reference loci.

When sequentially adding reference loci from stable to less stable (best case scenario), the uncertainty of the target estimate first decreases, but increases again when more unstable reference loci are added (U shape of the curve, Figure 5.5). Excluding unstable reference loci may thus lead to a more unbiased estimate while inclusion of multiple stable reference loci typically also enhances the precision of the estimate. When all reference loci are stable, more reference loci means less uncertainty and exclusion of reference loci is disadvantageous (Appendix D.2.1).

Sequentially adding reference loci in reverse order (worst case scenario), i.e. using one or more less stable reference loci, may result in severe bias (Figure 5.5): only when the more stable reference loci are added to the pool of reference loci the bias gets close to that of the best case scenario. Even though the bias may converge to that of the best case scenario when using just a few reference loci, the uncertainty of the estimate may still remain higher than what is observed under a best case scenario.

Relative expression stability determination necessitates a different approach as the ratio of referene gene concentrations is generally not close to one, but should be stable across all samples (Vandesompele et al., 2002). Taking a bias parameter (deviation from a ratio of one) into account would wrongly discount good reference genes for relative expression estimation. We suggest a modified stability criterion for this scenario, taking only variability of the ratio into account. Thus, genes that have a variable concentration ratio across samples and within samples

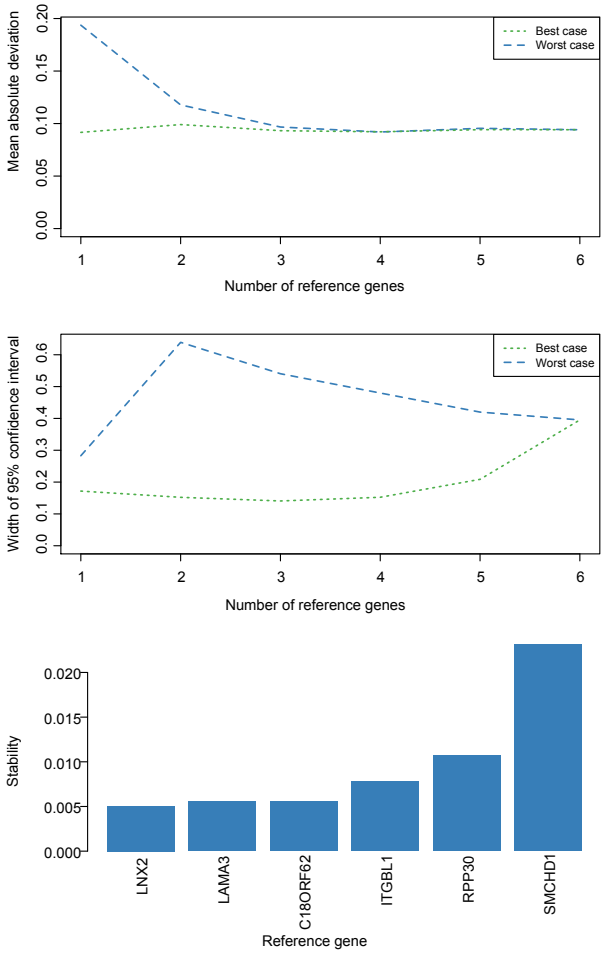


FIGURE 5.5: **Assessing reference stability Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference loci in sample 5. The reference loci are added in the order of decreasing stability (best case) or increasing stability (worst case). Using an increasing number of reference loci results in a copy number estimate that is less biased. Especially when only a few reference loci are used and those reference loci are not stable, the copy number bias may be considerably larger. **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference loci in sample 5. The reference loci are added in the order of decreasing stability (best case) or increasing stability (worst case). In terms of obtaining precise estimates, it is better to select the most stable reference genes and omitting the less stable reference genes. In this case, the optimal number of reference genes is 3 as the uncertainty is lowest in this case. **Bottom:** Stability of candidate reference loci in sample 5. Loci on *LNX2*, *LAMA3* and *C18ORF62* are more stable than the others.

(technical replication) will be considered less ideal than genes displaying a stable ratio across and within samples. This approach is similar to the widely adopted stability measurement described by Vandesompele et al. (2002).

5.4.3 Case study results

Copy numbers were measured in duplex (Figure 5.1E) for 10 patient samples and analyzed using the appropriate models given in the Materials and Methods section. A summary of all obtained aberrations is listed in Table 5.1. A full list of estimated copy numbers and confidence intervals is available as Appendix D.6. For 9 out of the 10 samples at least one target locus (located on the abberant chromosome) was confirmed, i.e. the 95% confidence interval contained the integer copy number. In those 9 samples there was also a clear aberration of the non-confirmed loci located on the chromosome with the aberration, supporting the evidence for the confirmed loci (Appendix D.6). Sample 10 displayed an unusual profile which we were unable to identify (no loci with (close to) normal diploid copy number). The obtained aberrations using ddPCR correspond to those obtained using state-of-the-art methods (karyotyping, FISH, arrayCGH), except for the previously mentioned sample 10 that was not identifiable using our ddPCR results (identified to be be a 45,X (Turner syndrome) sample using karyotyping, FISH and arrayCGH).

We discuss the case of patient sample 15, but the approach for the other samples is similar (Appendix D.3.10). The setup of the study was to assess chromosomal

Table 5.1: Overview of detected chromosomal abnormalities with confirmed loci.

Sample	Detected aberration	Confirmed loci
5	Chromosome 21 trisomy	*CLIC6*
6	Chromosome 21 trisomy	*DSCR3*, *SYNJ1*
9	Isochromosome Xq	*AMELX*
10	-	-
13	Chromosome 18 trisomy	*C18ORF62*, *LAMA3*
15	Chromosome 18 trisomy	*C18ORF62*, *LAMA3* , *SMCHD1*
16	Chromosome 21 trisomy	*CLIC6*, *DSCR3*, *SYNJ1*
18	Chromosome 18 trisomy	*LAMA3*
19	Isochromosome Xq	*AMELX*, *IL13RA1*, *ATP11C*
20	Chromosome 21 trisomy	*CLIC6*, *DSCR3*, *SYNJ1*

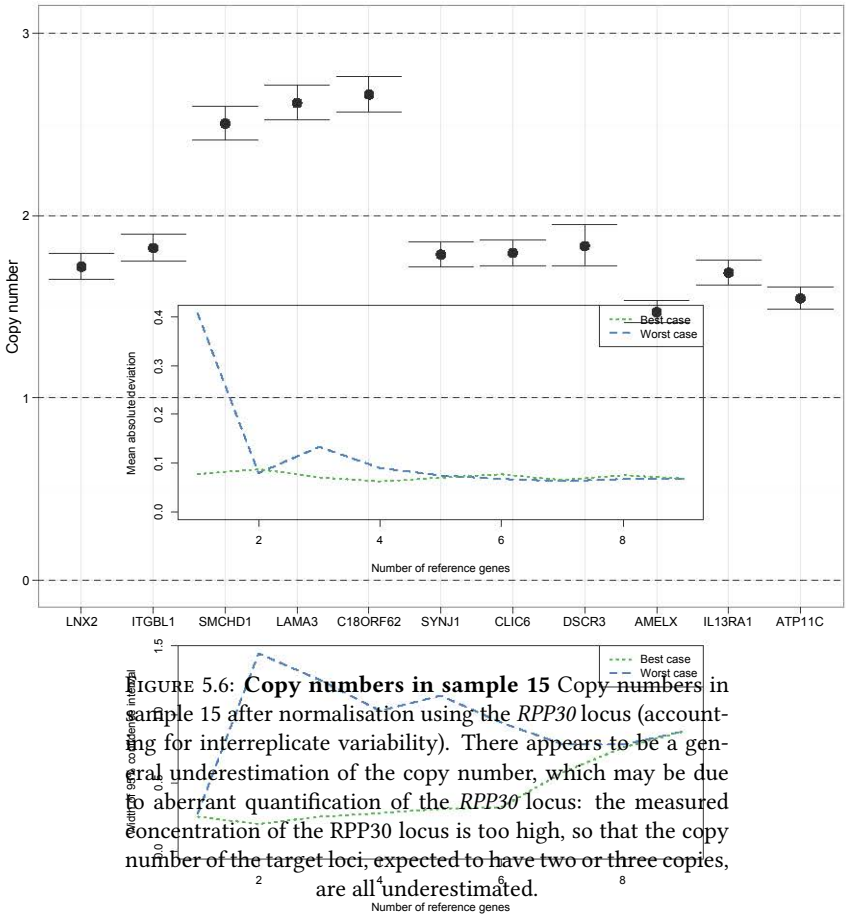


FIGURE 5.6: **Copy numbers in sample 15** Copy numbers in sample 15 after normalisation using the *RPP30* locus (accounting for interreplicate variability). There appears to be a general underestimation of the copy number, which may be due to aberrant quantification of the *RPP30* locus: the measured concentration of the RPP30 locus is too high, so that the copy number of the target loci, expected to have two or three copies, are all underestimated.

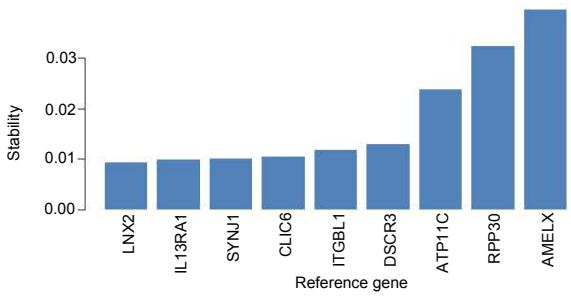


FIGURE 5.7: **Stability of candidate reference loci in sample 15** Loci on *LNX2*, *IL13RA1*, *SYNJ1*, *CLIC6*, *ITGBL1* and *DSCR3* are of similar stability, with rising instability thereafter for the loci on *ATP11C*, *RPP30* and *AMELX*.

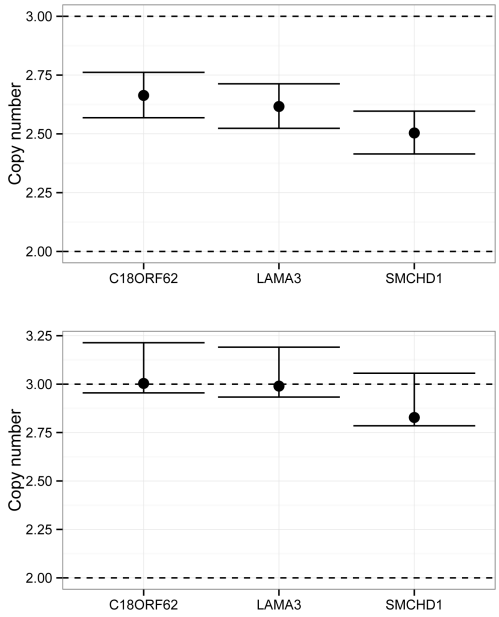


FIGURE 5.8: **Effect of normalization strategy** Effect of normalization strategy on the estimated copy numbers of 3 loci on chromosome 18 for a patient with chromosome 18 trisomy. **Top:** a single reference locus is used (*RPP30*), **bottom:** three reference loci are used simultaneously. When using multiple reference loci the copy numbers are on average closer to the expected value of three (i.e. less bias). Error bars denote 95% confidence intervals, whereby the multiple reference locus normalized results encompass the expected integer copy number for all three loci. Using an unstable (biased) single reference locus may result in biased copy number estimates. Relying on multiple reference loci typically has the advantage that the bias is reduced: if one of the reference loci is biased, this effect is partially cancelled by averaging together with two stable reference loci. See Appendix D.3.10 for more examples.

aberrations, i.e. deviations from diploid copy numbers on 13 target loci. The reference locus was located on the *RPP30* gene. The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure 5.6. It is clear we are dealing with a sample belonging to a female (absence of positive droplets of the Y chromosome loci, Appendix D.6), but all copy numbers seem to be deviating: there is a general underestimation of the copy number, which may be due to aberrant quantification of the *RPP30* locus. It can furthermore be seen that the *C18ORF62*3, *LAMA3* and *SMCHD1* loci (all located on chromosome 18) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 18 trisomy is likely. We thus use the loci on chromosomes X, 13, 21 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Materials and Methods section. The stability plot (Figure 5.7) indicates that the loci on *LNX2*, *IL13RA1*, *SYNJ1*, *CLIC6*, *ITGBL1* and *DSCR3* are of similar stability, with rising instability thereafter for the loci on *ATP11C*, *RPP30* and *AMELX*. The latter three are thus excluded as reference loci and the copy numbers of the three chromosome 18 loci recalculated using the six stable reference loci. The results are shown in the bottom panel of Figure 5.8: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for all three loci for a trisomy 18 case.

5.4.4 Comments and further research

Even though the assumption of only one chromosomal aberration turned out to be correct in our case study, it is generally to be recommended to determine candidate reference loci upfront. For CNV determination it will not always be clear whether e.g. a non-integer increase in a chromsome 18 loci combined with a non-integer decrease in chromosome 13 loci corresponds to a chromosome 18 trisomy or a chromosome 13 monosomy. This can furter be complicated due to contamination of sample material by e.g. maternal DNA. When using the framework for detection of non-integer copy numbers, the reference locus selection approach used in our case study can no longer be used and determination of stable reference loci upfront is needed.

We evaluated our method by calculating a deviation measure from integer copy

numbers, assuming that the samples did not display any mosaicism. This approach is only valid under the ideal circumstances of no mosaicism, no contamination, ... and deviation from this ideal scenario may have affected our case study findings.

We demonstrated that accounting for variation between replicates and/or reference genes is necessary to obtain correct standard errors of estimated absolute or relative quantities. These are sources of variations that can be accounted for in a majority of experiments as the use of replicate experiments and multiple reference genes are nowadays widely implemented. We want to stress that there are additional sources of variation, and depending on the experimental setup these can also be taken into account by e.g. adding additional random effects for plate effects, run effects, cartridge effects, ...

5.4.5 Software implementation

R code and a tutorial demonstrating the implementation of these models is available in Appendix D.1.3.3: we give examples of how to analyze different types of data using the models proposed in this manuscript. By providing easy to use functions, the practitioner is able to determine reference gene stability and perform absolute quantification (with or without replicates) and CNV/relative expression analysis (with one or more reference genes and with or without replicates) with a few lines of code. Additionally, a Shiny web application is available at http://statapps.ugent.be/dPCR/dPCR/ that does not require any type of programming, but relies on a spreadsheet-like data input and point and click interface. A manual for using this web interface is also available (Appendix D.5.5).

5.5 Conclusion

Currently available methodology and software could be used to analyze e.g. experiments with multiple reference genes, but these methods would not properly account for sources of variability introduced by these specific designs. The data analyses presented in this paper suggest that not accounting for various sources of variability can result in extremely unreliable estimates of variability of nucleic acid concentrations or copy numbers.

We introduce a general framework that covers the analysis of a wide range of experimental setups, correctly accounting for various sources of variability and allowing researchers to analyze data from experiments for which previously no appropriate methods were proposed.

A method for reference gene selection relying on this framework is suggested, allowing for an improved estimation of copy numbers or relative quantities: findings indicate that selecting stable reference genes may be beneficial in two ways: (i) bias may be reduced, (ii) uncertainty can be decreased by selecting a suitable subset. User-friendly R scripts, a Shiny web interface and tutorials are available to facilitate the use of the methodology.

6 Bias reduction by modelling volume variability

This chapter is an adaptation from the paper: Vynck, M. and Thas, O. Reducing bias in digital PCR quantification experiments: the importance of appropriate modelling of volumetric variability. *Submitted*.

Contributions: Matthijs Vynck conceived of the initial analysis strategy. Matthijs Vynck and Olivier Thas revised the analysis strategy. Matthijs Vynck wrote the R code, analyzed the data and wrote an intial draft of the manuscript. Matthijs Vynck and Olivier Thas revised the manuscript. Both authors agreed with the contents of the final paper.

6.1 Abstract

Multiple simulation studies have shown that volume variability of partition sizes in digital PCR (dPCR) causes bias of the resulting concentration estimates and their associated standard errors. These biases are especially apparent when volume variability is large and the targeted nucleic acid concentration high. Currently only a single method for the elimination or reduction of these biases is available, and it assumes a fixed class of models for the volume variability. We show that the form in which volumetric variability occurs in empirical data is variable and cannot be modelled by a single distribution. We propose a new volume modelling method, NPVolMod, that takes volume variability of arbitrary form into account, and that is applicable to both absolute and relative quantification. The method is non-parametric in the sense that no distributional assumption is needed. Moreover, the volumes of each of the individual partitions are not needed. We empirically demonstrate by simulation that NPVolMod nearly eliminates the biases caused by volumetric variability and that it often outperforms the existing method. The possibility of proper modelling of volume variability may have implications for platform design and may increase the performance of existing dPCR platforms in terms of e.g. their trueness and linear dynamic range.

6.2 Introduction

In a digital polymerase chain reaction (digital PCR, dPCR) experiment, a nucleic acid sample mixed with PCR reagents is divided into a large number of partitions. Based on the partitions' end-point fluorescence intensity signals, each of

these partitions is subsequently classified as a partition containing the specific target nucleic acid (positive partitions) or not containing this nucleic acid (negative partitions). The number of target copies in a partition is assumed to follow a Poisson distribution, and the mean parameter of the Poisson distribution can be estimated from the number of negative and positive partitions by relying on the probability mass function of the Poisson distribution (Baker, 2012; Morley, 2014; Vynck et al., 2016a). Subsequently a fixed partition volume is used to convert the estimated number of copies per partition to e.g. an absolute concentration estimate. Hence, an unbiased estimate of the partition volume is needed for unbiased concentration estimation (Vynck et al., 2016b; Kosir et al., 2017).

A complication arises when the partition volume is not constant. It has been shown that in the presence of non-negligible partition volume variation both the concentration and its associated precision estimates may show substantial bias (Bhat et al., 2009; Dong et al., 2015; Huggett et al., 2015; Jacobs et al., 2014; Pinheiro et al., 2011). Under certain circumstances, such as low partition occupancy (i.e. less than one copy per partition) and low partition volume variability (i.e. not exceeding about 5%) the consequences of partition volume variability are minor (about 1-2% bias) (Jacobs et al., 2014; Vynck et al., 2016b). In contrast, severe consequences arise when partition occupancy reaches higher levels or for platforms exhibiting non-negligible partition volume variability (Jacobs et al., 2014; Vynck et al., 2016b). This problem reduces the practical dynamic range of dPCR platforms, which is to this day still inferior to that of quantitative PCR (qPCR), and thus reducing the competitiveness of dPCR (Jones et al., 2016). Moreover, for a technique that aims to compete with qPCR, which is a well established technique in routine use for accurate nucleic acid quantification, any loss of accuracy is to be avoided.

Multiple efforts have been made to obtain a correct partition volume for different dPCR platforms (Kosir et al., 2017; Bhat et al., 2009; Dong et al., 2015; Pinheiro et al., 2011; Corbisier et al., 2015), but there has been limited research into dealing with partition volume variability and its consequences. Huggett et al. (2015) recently proposed a method, further referred to as Huggett's method, to account for partition volume variability by allowing the mean number of copies per partition to vary according to a gamma distribution. Although this approach is the first to account for volume variability that is directly applicable to currently available commercial systems, this method has some drawbacks.

Firstly, the average number of copies per partition is assumed to follow a gamma distribution. This has some mathematically convenient consequences, most importantly the fact that if the mean number of copies per partition is gamma distributed, the number of copies per partition follows a Poisson-Gamma mixture distribution, which is more generally known as the Negative Binomial distribution, and which is a well-characterised distribution. Although this is theoretically convenient, it is also the weakest link of the methodology: the number of copies per partition is assumed to follow a fixed distribution, and this may not be the case in practice (Pinheiro et al., 2011). Wrongfully relying on this assumption may thus result in inaccurate concentration estimates.

Secondly, it is applicable to absolute quantification only and incorporation of other sources of variability and correct error propagation for more complicated settings may be cumbersome (Vynck et al., 2016a).

In this paper we first study the distribution of digital PCR volumes using data published earlier by Dong et al. (2015) and verify whether the assumption underlying Huggett's method holds. We then propose NPVolMod, which is a method that accounts for volume variability and that does not require a distributional assumption, and we show that it often outperforms alternative strategies, especially when volume variability is large. We demonstrate this superior performance by means of a simulation study which includes NPVolMod, Huggett's method and a naive method that ignores the information available. We study the performance in terms of bias and variance of the concentration estimates, under a variety of empirical volume distributions and both for absolute and relative quantification experiments. In addition, we study the impact of misspecification of the volume distribution, in terms of shape and scale. Finally, we discuss the results and benefits and limitations of the studied methods.

6.3 Materials and methods

6.3.1 Quantification assuming a constant volume

The mean paramater λ of the Poisson distribution in a dPCR experiment may be determined from the probability of observing a negative partition ($Y = 0$):

$$\mathrm{P}\{Y = 0|\lambda\} = \frac{\lambda^0}{0!}\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda}, \tag{6.1}$$

from which

$$\lambda = -\log(\mathrm{P}\{Y = 0\}). \tag{6.2}$$

An estimate of the average number of copies per partition is obtained from an intercept-only generalized linear regression model with complementary log-log link function (Vynck et al., 2016a):

$$\hat{\lambda} = \exp(\hat{\beta}_0), \tag{6.3}$$

with $\hat{\beta}_0$ an estimate of the intercept.

Assuming a constant volume of the partitions, say V_p, the absolute concentration can be estimated from the average number of copies per partition:

$$\hat{c} = \frac{\hat{\lambda}}{V_p}. \tag{6.4}$$

The method is readily extended to relative quantification (Vynck et al., 2016a).

6.3.2 Huggett's approach

Huggett et al. (2015) suggest to let the λ parameter of the Poisson distribution vary between partitions according to a gamma distribution. Conceptually, this corresponds to a change in the number of copies within a partition depending on the partition volume. The density of the number of copies per partition is then

given by a Poisson-Gamma or, equivalently, a Negative Binomial distribution. With this distribution the probability of a negative partition becomes

$$P(Y = 0|\alpha, \beta) = \left(\frac{\alpha}{1+\alpha}\right)^{\beta}, \tag{6.5}$$

with Y the number of copies per partition, α the scale and β the shape of the gamma distribution. It can equivalently be stated that V_p is described by a gamma distribution: when λ has a gamma distribution then so does V_p, as the average number of copies per partition is directly and linearly related to the partition volume under typical assumptions made for dPCR quantification. In what follows, we will verify whether the empirically observed partition volumes comply with this assumption of a gamma distribution.

6.3.3 Verification of the partition volume distribution

Dong et al. (2015) studied droplet sizes of the Biorad droplet digital platform and provided the droplet sizes as Supplementary material. We use these droplet sizes to study the volume variability for this system. This is done qualitatively by inspection of density curves and QQ plots, and quantitatively by performing pairwise comparisons of the observed distributions using Kolmogorov-Smirnov tests (Smirnov, 1939). For the Kolmogorov-Smirnov tests, p-values below the significance level of 5% after Holm-Bonferroni multiple-testing correction (Holm, 1979) were considered evidence for a difference in partition volume distribution.

6.3.4 A non-parametric volume modelling approach

If the exact volumes of all partitions are known, they can be used as offsets in regression models for the analysis of digital PCR data (Vynck et al., 2016a). However, currently available commercial platforms do not measure partition volumes along with their positive/negative status. Yet, when the marginal partition volume distribution is known, it is possible to correct for volume variability. An outline of our approach follows.

We let V denote the volume, N the unobservable count of target copies in a partition, and Y the corresponding digitalised count, and we introduce an index i to refer to a partition. We have observed n digital outcomes y_i, and we have a set of

volumes, but we do not know which volume corresponds with which partition. Let $\bar{v}$ denote the average volume.

For partitions with volume v, we assume that the counts follow a Poisson distribution with mean

$$\mathrm{E}\{N \mid V = v\} = \lambda \frac{v}{\bar{v}}, \tag{6.6}$$

such that the λ parameter has the same interpretation as in Equation 6.1, i.e. λ is the average copy number for a partition with volume $\bar{v}$.

Let $g(v)$ denote the marginal density of the volume. We estimate this from the data nonparametrically and let $\hat{g}$ denote this estimate. Samples from $\hat{g}$ are obtained using rejection sampling (Gilks et al., 1995).

From the density of the volume and the observed numbers of negative and positive partitions, we subsequently derive a concentration estimate taking into account partition volume variability:

1. Estimate λ from Equation 6.3 (i.e. without using the volumes): $\hat{\lambda}_0$. This will be an underestimate of λ (Jacobs et al., 2014; Huggett et al., 2015). By starting with $\hat{\lambda}_0$ as an initial estimate, an unbiased estimate is obtained by means of an iterative algorithm (see Appendix E.1 for details and Appendix E.2 for a numerical example). Denote the unbiased estimate as $\hat{\lambda}_{\text{final}}$.

2. Based on $\hat{\lambda}_{\text{final}}$, estimate $f(v \mid Y = 0)$ and $f(v \mid Y = 1)$, i.e. the two conditional densities of the volume: generate counts N_i from Poisson distributions with mean $\hat{\lambda}_{\text{final}} \frac{v_i}{\bar{v}}$. Those v_i resulting in $N_i = 0$ are used for estimating $f(v \mid Y = 0)$, while v_i resulting in $N_i > 0$ are used for estimating $f(v \mid Y = 1)$. $\hat{f}(v \mid Y = 0)$ and $\hat{f}(v \mid Y = 1)$ denote these estimates.

3. For all observations $y_i = 0$, sample volumes v_i from $\hat{f}(v \mid Y = 0)$, and for all observations $y_i = 1$, sample volumes v_i from $\hat{f}(v \mid Y = 1)$.

4. With the partition status y_i and their associated partition volume v_i, the concentration of the target in the original sample may be obtained from a Poisson regression model with log-mean

$$\log(\lambda) = \beta_0 + \log(v). \tag{6.7}$$

The parameter estimate β_0 is provided by statistical software. Following Vynck et al. (2016a), an estimate of the concentration is

$$\hat{c} = \exp(\hat{\beta}_0), \tag{6.8}$$

and an estimate of the variance of the concentration estimate, by application of the delta method, is

$$\mathrm{Var}(\hat{c}) \approx c^2 \mathrm{Var}(\hat{\beta}_0), \tag{6.9}$$

which can be estimated by $\hat{c}^2 \widehat{\mathrm{Var}}(\hat{\beta}_0)$, with $\widehat{\mathrm{Var}}(\hat{\beta}_0)$ an estimate of $\mathrm{Var}(\hat{\beta}_0)$. The required estimates are again provided by the statistical software used for fitting the Poisson regression model (Vynck et al., 2016a).

A similar approach may be taken to account for volume variability for relative quantification (see Appendix E.3 for details).

6.3.5 Simulation study

To assess the performance of NPVolMod and compare it with ignoring the volume variability and with Huggett's method, a simulation study was performed. We simulated volume data by sampling from the densities estimated from the rescaled Dong et al. (2015) data. The scale of the densities was adjusted so that the robust relative deviations of the volumes corresponded to predetermined levels of 5%, 10% or 20%. The robust relative deviation is defined as the ratio of a robust estimate of the scale of the volumes S_n (as defined by Rousseeuw and Croux (1993)), divided by the median volume. The robust relative deviation is a more appropriate measure than the relative standard deviation for non-normal and asymmetric distributions (Rousseeuw and Croux, 1993). Data were generated for different scenarios:

1. Absolute quantification,
   (a) volumes with a robust relative deviation of 5%, 10% and 20%,
   (b) expected copies per partition of average volume ($\lambda / \bar{v}$) of 0.5, 1 and 5.
2. Relative quantification,

(a) volumes with a robust relative deviation of 10% and 20%,

(b) expected copies per partition of average volume ($\lambda/\bar{v}$) of 0.5, 1 and 5 for the target,

(c) ratios of 1, 2 and 5.

These partition occupancies correspond to levels for which the impact of partition volume variability is small to considerable (Jacobs et al., 2014; Huggett et al., 2015), and which are of practical relevance (Pinheiro et al., 2011; Henrich et al., 2012; Dong et al., 2014; Te et al., 2015).

To study the impact of partition volume distribution misspecification, counts per partition were generated using the median-centered and scaled partition volume distribution of one of Dong et al. (2015)'s samples, but analyzed using the partition volume distribution of one of Dong et al. (2015)'s other samples. For example, counts per partition were generated according to the volume distribution of sample 1 in Dong et al. (2015)'s data, but analyzed using the volume distribution as given by sample 2 in Dong et al. (2015)'s data. In this context we refer to the former as the data-generating volume distribution and to the latter as the working distribution. Hence, in the first misspecification study, the location and scale of the partition volume working distribution will still be correct, but the shape of the working distribution will be incorrect (Figure E.1).

In a second study to assess the impact of misspecification, counts per partition were generated using the median-centered and scaled partition volume distribution of one of Dong et al. (2015)'s samples and analyzed using that same partition fvolume distribution, but scaled differently. Robust relative deviations for the data-generating process were 5%, 10% and 15%, while the assumed robust relative deviations were 50% lower or higher than the robust relative deviation used for data-generation. In this case, the scale of the partition volume working distribution will be incorrect, but the location and the shape of the working distribution will be correct.

Each of the scenarios involved 500 simulations, e.g. for absolute quantification with an occupancy of 0.5 expected copies per average partition volume and 10% robust relative deviation, the sampling and the various concentration estimation procedures were performed 500 times.

Assessment criteria were the bias of the concentration or relative quantity estimate and the bias of the associated standard error.

6.3.6 Implementation

All analyses were implemented in R. Additional implementation details are provided in Appendix E and our R code, including examples, is available at `http://github.com/CenterForStatistics-UGent/dPCRVolMod`.

6.4 Results and discussion

6.4.1 Verification of the partition volume distribution

Figure 6.1 shows the non-parametric density estimates of the four droplet volume samples from Dong et al. (2015). Visual assessment learns that these do not closely follow a known parametric distribution and that the densities seem to differ between the samples. Figure 6.2 shows QQ plots comparing the empirical quantiles of the droplet volumes with the quantiles of a gamma distribution fitted to the droplet volume sizes using maximum likelihood. The figures indicate deviation from this gamma distribution for at least some of the data sets.

We furthermore assessed the difference in distribution between the four samples by performing pairwise Kolmogorov-Smirnov tests (Table 6.1). The resulting p-values indicate that there are indeed differences in distribution between the samples that have been analyzed.

Table 6.1: Kolmogorov-Smirnov test p-values (Holm-Bonferroni corrected) for pairwise equality of distribution for the four volume samples by Dong et al. (2015).

	Sample 2	Sample 3	Sample 4
Sample 1	0.0250	0.0012	0.9579
Sample 2		<0.0001	0.0002
Sample 3			0.0002

From these analyses it may be concluded that (a) the Huggett's method may inadequately account for volume variability as the assumption of the gamma distribution may be violated, and (b) not a single parametric distribution correctly models the volumes in all samples of the Dong et al. (2015) data.

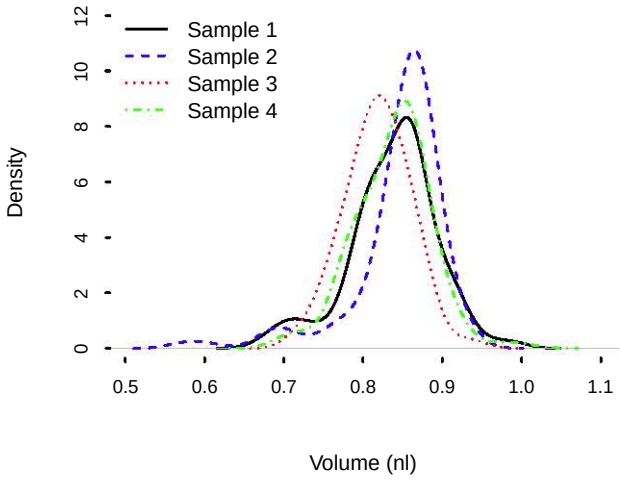


FIGURE 6.1: Non-parametric volume density estimates of the four volume samples of Dong et al. (2015)

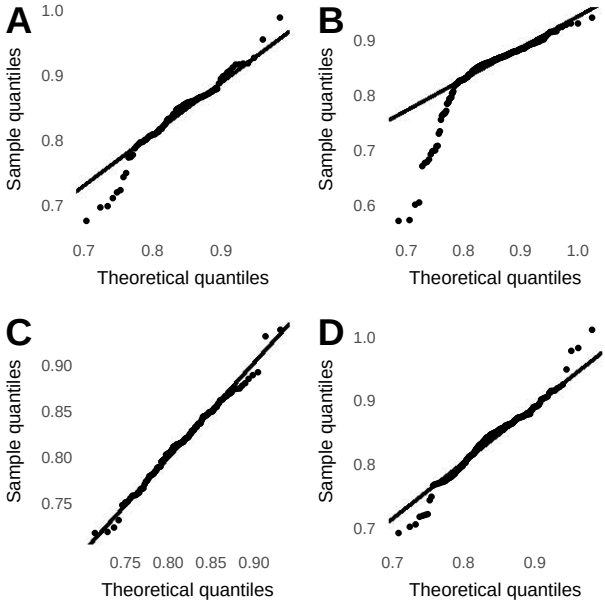


FIGURE 6.2: QQ plots comparing the observed volumes with a gamma distribution, for all four samples of Dong et al. (2015). (A) Sample 1. (B) Sample 2. (C) Sample 3. (D) Sample 4.

This is in accordance with evidence provided by others. Figure 4 in Tellinghuisen (2016), for example, provides evidence that the distribution of volumes is not normal. Their Figure 4 shows that the volume distribution is left skewed, indicating that a gamma distribution will not adequately describe the partition volume distribution either. Figure 4 in Dagata et al. (2016), on the other hand, suggests that a lognormal distribution describes the partition diameters well. A lognormal distribution for the diameters translates into a lognormal distribution for the volumes as well. As a lognormal distribution is right-skewed, as opposed to the left-skewed distribution in Tellinghuisen (2016), this provides additional evidence that a single parametric distribution may not describe the partition volumes adequately for all studies, but that this distribution may differ from experiment to experiment, or from platform to platform.

In addition, there may be a difference between average partition volumes between cartridges, droplet generators and reagents used, as has been demonstrated by several laboratories (Pinheiro et al., 2011; Dagata et al., 2016; Kosir et al., 2017). This evidences that also the location of the distribution may differ from experiment to experiment. This difference in average partition volume is sometimes substantial, and may cause a major bias when used for calculating the concentration of the nucleic acid in the sample.

Taken together, there is evidence that the location, the shape and the scale of the partition volume distribution may vary and that these distributional parameters are affected by several experimental parameters. If the goal is unbiased quantification, then there is a need to acquire both an accurate estimate of the average partition volume, and of the partition volume distribution.

6.4.2 NPVolMod: a new approach for modelling volume variability

Previous research has demonstrated that partition volume variability in digital PCR experiments may hamper accuracte quantification (Jacobs et al., 2014). Although the phenomenon of volume variability has been studied in detail by many researchers, solutions to deal with this volume variability have been given limited attention.

The inability to quantify the individual partition volumes along with its fluorescence intensity (i.e. whether it is positive or negative) complicates proper data

analysis. If the volume were known alongside the status of a partition, regression models for the analysis of digital PCR experiments can be used by using the volume as an offset (Vynck et al., 2016a). This approach is unfortunately not possible with current dPCR platforms, and hence a different path is needed. We proposed a method which requires knowledge of the overall data-generating volume distribution only. This data-generating distribution may be of any arbitrary form. This form may be supplied by the user, or it can be estimated from volume data when such data are available.

NPVolMod depends crucially on the assumption that the number of target molecules being present within a given partition is linearly related to the volume of that partition. This is a reasonable assumption, as that same assumption is needed to allow correct direct quantification by relying on the Poisson distribution, which is an assumption that has been widely accepted and is consistently used in the dPCR literature.

6.4.3 Absolute quantification

Figure 6.3 shows the bias of the concentration estimates for the different simulation scenarios of a single sample (see Figures E.2, E.6, E.10 and E.14 for all samples). As expected (Jacobs et al., 2014; Huggett et al., 2015), there is no bias for low partition occupancy, but increasing bias with occupancy increasing. Ignoring the volume variability may lead to biases of up to $\sim 22\%$ for the highest occupancy at a volume variability of 20%, and this bias decreases to $\sim 16\%$ when the volume variability is set to 10%. For 5% volume variability, the bias remains reasonably limited, not exceeding $\sim 3\%$. Huggett's method performs better than the naive method, with a maximum bias close to the 4% level in the case of the most severe deviation from the gamma distribution. NPVolMod performs best overall, with a maximum bias of $\sim 0.1\%$.

Figures E.3, E.7, E.11 and E.15 display the bias of the estimates of the variance of the concentration estimate of a single sample. Again, ignoring volume variability may lead to substantial bias. In the most severe case, the variance is underestimated with $\sim 60\%$. This happens with 20% volume variability. For volume variabilities of 5% and 10%, the bias is $\sim 10\%$ and $\sim 50\%$, respectively. For Huggett's method, the bias in the most severe case reaches levels of $\sim 14\%$, $\sim 14\%$ and $\sim 1.6\%$ for levels of volume variability of 20%, 10% and 5%, respectively. In all

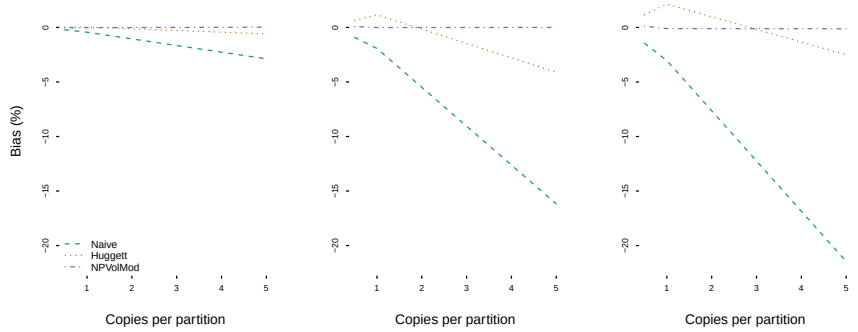


FIGURE 6.3: Bias of concentration estimates for volume variabilities of (A) 5%, (B) 10% and (C) 20%.

scenarios, NPVolMod keeps the bias below $\sim 0.5\%$.

Overall, NPVolMod outperforms Huggett's method for absolute quantification, even under conditions for which Huggett's method performs quite well: when the distribution is misspecified (i.e. it is not gamma), bias of the estimates obtained with Huggett's method remains reasonably limited in many cases and typically outperforms ignoring volume variability altogether. This suggests that modelling the scale of volume variability is an important contributor to reducing bias. Thus, even if only an approximate distribution for the partition volumes is known, significant improvement in estimates may be obtained. This may be important for platforms suffering from non-negligible volume variability, as has been reported for example for the RainDance and Stilla platforms (Dangla et al., 2013; Dong et al., 2015). The concentration estimates resulting from these platforms may be improved if an approximate form of volume variability is known. Nonetheless, knowledge of the data-generating distribution and not just the scale of the distribution further reduces bias and imprecision. The benefits of exact specification become more apparent as the partition volume variability increases (Figure 6.3). In addition, even when the volume variability is reasonably small, as has been observed for e.g. the Biorad platform (Dong et al., 2015), accounting for volume variability is still beneficial for a quantification technique that aims to be as unbiased and precise as possible. It can still reduce the bias in estimates by 2-3%, which may be a relevant bias for some applications (Lun et al., 2008; El Khattabi et al., 2016).

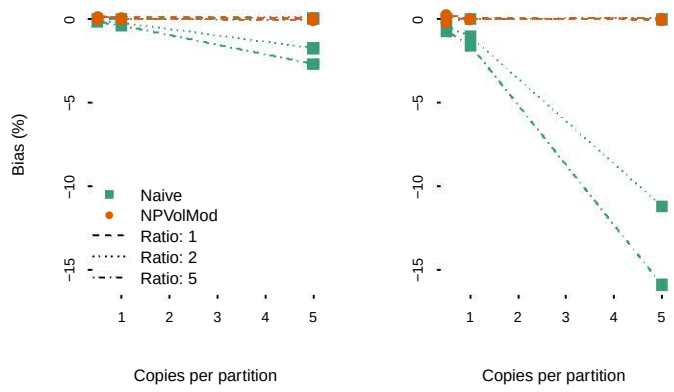


FIGURE 6.4: Bias of relative quantity estimates for volume variabilities of (A) 10% and (B) 20%.

6.4.4 Relative quantification

We have for the first time studied the effects of volume variability in a relative quantification scenario, and propose the first approach for dealing with volume variability for relative quantification.

A similar pattern as for absolute quantification is observed for relative quantification in the simulation study (Figure 6.4, Figures **??**, Figures E.4, E.8, E.12 and E.16 for all samples). Ignoring volume variability may lead to substantial biases of the estimates (up to $\sim 16\%$), whereas accounting for volume variability limits the bias ($< 1\%$). Likewise, the bias of the variance estimates remains limited when accounting for volume variability, but may be underestimated by up to $\sim 45\%$ when ignoring volume variability (Figures Figures E.5, E.9, E.13 and E.17).

The observed biases can be explained as follows. The effect of volume variability is eliminated if the ratio of the two nucleic acid concentrations is close to one: even when partition occupancy is high, both estimates are downwardly biased but to the same extent, and hence not influencing the eventual ratio of quantities (Figure 6.4). Moreover, the uncertainty of the ratio estimate will be lower, so that large volume variability is actually advantageous (Figure 6.4 and Figures Figures E.5, E.9, E.13 and E.17). The practical advantage of this decreased uncertainty is however limited, as interest typically lies in detecting non-identical quantities in relative quantification/copy number alteration applications.

When the ratio is much larger or smaller than one, and partition occupancy is high for one of the nucleic acids, a substantial bias may still result. Indeed, the nucleic acid with high partition occupancy will have a larger bias than the nucleic acid with low partition occupancy, resulting in a bias for the estimated ratio. This has important consequences for relative quantification for which the target nucleic acid may have a much higher concentration than the reference nucleic acid, e.g. in case of significant amplification of the target gene in copy number studies (Versluis et al., 2015).

Huggett's method is currently applicable to absolute quantification only, hampering its use for more complicated experimental setups, and especially for correct error propagation in such situations (Vynck et al., 2016a). NPVolMod is compatible with a regression framework for digital PCR data analysis, allowing to account for volume variability in a wide range of applications and with correct standard error estimation (Vynck et al., 2016a).

6.4.5 Impact of distribution misspecification

Accurate knowledge of the volume distribution is required. In practice this requires the measurement of a sufficient number of partition volumes, so that these measurements allow for accurate estimation of the partition volume distribution. If an insufficient number of partition volumes is measured, or if these volume measurements do not reflect the true partition volume distribution, then the partition volume distribution may be misspecified and this may result in biased concentration estimates. We will refer to the true volume distribution as the data-generating distribution and to the distribution used in the data analysis as the working distribution.

The impact of a misspecified distribution location has been the topic of previous investigations, and it naturally results in a wrong average partition volume, therefore affecting the bias of the estimates (Pinheiro et al., 2011; Corbisier et al., 2015; Kosir et al., 2017). In order to further investigate the impact of distribution misspecification, the effect of a discrepancy between the data-generating distribution and the working distribution in terms of (i) a different shape, and (ii) a different scale was investigated.

Figure 6.5 displays the biases that result from using one of Dong et al. (2015)'s

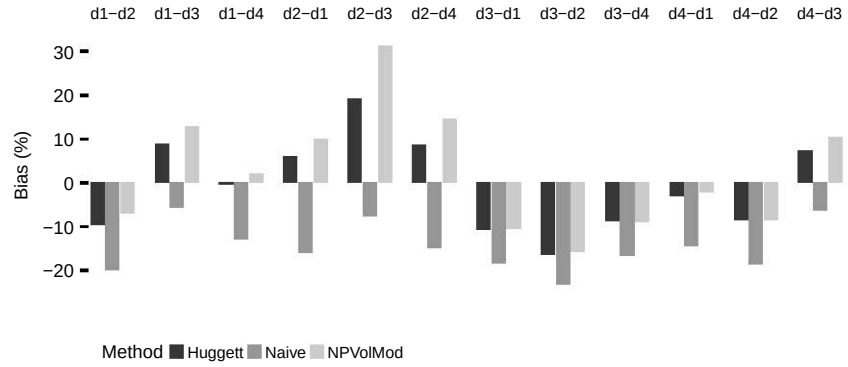


FIGURE 6.5: Bias (%) of the concentration estimate when the working distribution has the wrong shape but correct location and scale. The column titles refer to the distributions: the first distribution is the data-generating distribution, while the second is the working distribution. In particular, di-dj means that data set i was used for data-generation and data set j for estimating the working distribution. Results as obtained for a partition volume relative standard deviation of 20% and an average of five copies per partition.

data sets for estimating the working distribution, while the data-generating distribution was one of the other distributions of Dong et al. (2015), after centering and scaling the distributions. Hence, the working distribution has a wrong shape (see also Figure E.1). The naive approach of ignoring partition volume variability performs worst overall in terms of bias. Only when the data-generating distribution is very different from the working distribution, as is the case for the sample 2-sample 3 (d2-d3) comparison (see also Figure 6.2), ignoring volume variability results in a substantially lower bias. Depending on how much the data-generating distribution agrees with a gamma distribution, as assumed by Huggett's method, either Huggett's method or NPVolMod performs best. NPVolMod performs better when the data-generating distribution is not close to a gamma distribution. This does not detract from the findings presented in Figures 6.3-6.4: NPVolMod performs best when the working distribution is correctly specified, which should naturally be aimed for. Only in case of a rather severe difference between the data-generating and working partition volume distribution, Huggett's method may perform better than NPVolMod.

Likewise, Figure 6.6 displays the biases that result from assuming a wrong scale

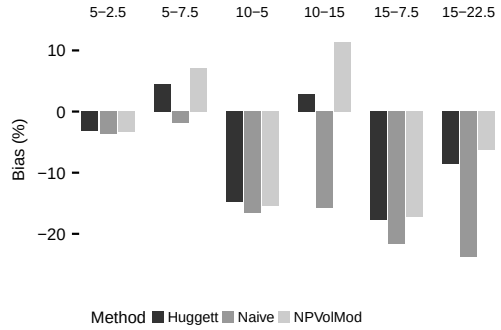


FIGURE 6.6: Bias (%) of the concentration estimate when the working distribution has the wrong scale but correct location and shape. The column titles refer to the scale of the data-generating and the working distributions: where the first number is the scale (relative robust deviation) of the data-generating distribution, and the second number is the scale of the working distribution. Sample 2 of the Dong et al. (2015) data was used as the data-generating distribution with an average of five copies per partition.

for the working distribution. As expected, assuming a scale that is smaller than the scale of the data-generating distribution, results in a negative bias or concentration underestimation: larger partition volume variability leads to larger downwards biases, and while there is some correction for partition volume variability, the correction is not strong enough due to underestimation of the scale. Likewise, assuming a scale that is larger than the scale of the data-generating distribution, results in a correction that is too strong: the downward bias is overcompensated, resulting in a positive bias. Note that this observation seems to be reversed in the simulation with the largest partition volume variability (22.5%), but this stems from truncation at volumes of zero after rescaling, leading to some change in the shape of the partition volume distribution. In addition, in those cases where the bias is overcompensated, Huggett's method seems to outperform NPVolMod, but this stems from the negative bias that arises from shape misspecification.

Except for the smallest partition volume variabilities, slight misspecification of the scale will thus be less detrimental in terms of bias than ignoring partition volume variability altogether. It should be noted that a 50% under- or overestimation of the scale is quite severe, and unlikely to occur often in practice. The

consequences of misspecification are thus expected to be lower than those obtained in the simulation study.

6.4.6 Limitations

NPVolMod also suffers from some drawbacks. Most importantly, the method is computationally more intensive than Huggett's method. Although it takes only seconds for a single iteration, the imprecision of the resulting estimates is reasonably large as the approach is sampling-based. Multiple iterations are ideally performed to reduce the imprecision of the estimates, increasing computational time to the order of minutes for a single sample on a modern-day computer. Given the limited bias for low partition occupancy it is consequently worthwhile to only perform bias correction for those samples for which a reasonably large and relevant bias is expected. Secondly, a large number of partition volume measurements may be required to capture the granularities of the partition volume distribution. As previously argued, insufficient partition volume measurements may result in an inaccurate estimate of the partition volume distribution, for which the consequences are expected to be a smaller yet often significant reduction in bias, similar to what is observed for wrongfully assuming a gamma distribution of partition volumes.

6.4.7 Importance for accurate quantification

Focus in the literature has been largely directed towards acquiring correct average partition volumes (Pinheiro et al., 2011; Dong et al., 2015; Kosir et al., 2017), a task that has recently been shown to be complex due to the influence of many experimental parameters (Kosir et al., 2017). Yet, bias not only arises from using an incorrect average partition volume, but, as has been shown, also from using an incorrect partition volume distribution, or from entirely ignoring partition volume variability.

The bias that arises as a consequence of specifying a wrong average partition volume may be substantial (Dong et al., 2015; Kosir et al., 2017), and occurs at all levels of partition occupancy. Yet, for those reactions with high partition occupancy, together with a reasonably large partition volume variability, the contribution to bias from ignoring partition volume variability may exceed the bias that

arises from assuming a wrong average partition volume. If the goal is accurate quantification across a wide dynamic range, then more attention to modelling the partition volume distribution, and not just the average partition volume, will be needed.

6.4.8 Implications for platform development

These results may be important for further development of digital PCR platforms, where volume variability has typically been considered something to be avoided (Huggett et al., 2015). If focus is shifted away from minimizing partition volume variability, as this is a factor that may be corrected for in downstream analysis, this may create opportunities for e.g. reducing platform or consumables cost or increasing the dynamic range by reduction of partition size and increasing the number of partitions, ultimately increasing the competitiveness of dPCR with other nucleic acid quantification technologies. For example, the RainDance platform has a theoretically large dynamic range due to its very large number of partitions, but its manufacturers suggest to limit partition occupancy to less than 10%, likely due to the effect of its reasonably large volume variability (Dong et al., 2015; Jones et al., 2016). Knowledge of the (approximate) partition volume distribution may thus increase the practical dynamic range of this platform.

While the focus of this manuscript is on modelling the entire distribution of partition volumes, obtaining an accurate estimate of the average partition volume has been shown to be of primordial importance for obtaining accurate concentration estimates (Pinheiro et al., 2011; Kosir et al., 2017). That this average volume depends on several experimental parameters such as the reagents used, or the exact installation of the platform (Kosir et al., 2017) implies that it may be necessary to characterize this average partition volume routinely. In doing so, information on the partition distribution is acquired as well, thus making the method introduced in this manuscript directly practicable.

The need to characterize the partition volume distribution and/or average partition volume for acquiring accurate concentration estimates, may prevent the routine use of digital PCR as a very accurate quantification approach with large dynamic range. It is thus not unlikely that platforms will need to provide the user with information on these parameters if digital PCR is to be used routinely. Yet,

as the partition volume distribution may depend on many experimental parameters (Pinheiro et al., 2011; Dagata et al., 2016; Kosir et al., 2017), there may be a need to measure these volumes routinely. Platforms that are able to measure not only partition fluorescence intensities but also partition volumes, or a proxy, may thus be able to yield more accurate estimates. Such volume measurements may be obtained from imaging the partitions, as is done, although currently not very precisely, in e.g. the Stilla Naica platform (Madic et al., 2016), or through measuring other quantities that may relate to partition size, such as fluorescence intensity or light scatter (Tzur et al., 2011).

6.5 Conclusion

We have proposed NPVolMod, which is a new method that does not assume a particular distribution of partition volumes, but that can use partition volume measurements to non-parametrically estimate the volume distribution. We have performed an empirical evaluation of methods using empirically determined partition volumes, and have shown that accounting for volume variability improves the accuracy of the concentration estimation in both absolute and relative quantification scenarios as compared to ignoring the available volume information. We have also shown that NPVolMod often outperforms Huggett's method, which relies an a gamma volume distribution and which is only applicable for absolute quantification. At the same time, we argue that a major contribution to bias reduction is modelling the scale of volume variability. Our results indicate that even approximate knowledge of the partition volume distribution may reduce bias and imprecision in quantification when this volume variability is taken into account. Nevertheless, an accurate specification of the distribution location and shape further reduces the bias.

7 Evaluation of assay or platform quality

This chapter is an adaptation from the paper: Vynck, M., Vandesompele, J. and Thas, O. (2017). Quality control of digital PCR assays and platforms. Analytical and Bioanalytical Chemistry, 409(25), 5919-5931.

Contributions: Matthijs Vynck conceived the initial analysis strategies. Matthijs Vynck, Jo Vandesompele and Olivier Thas revised and refined the analysis strategies. Matthijs Vynck wrote the R code, analyzed the data and wrote an intial draft of the manuscript. Matthijs Vynck, Jo Vandesompele and Olivier Thas revised the manuscript. All authors critically reviewed and agreed with the contents of the paper.

7.1 Abstract

Digital polymerase chain reaction (digital PCR, dPCR) is a direct nucleic acid quantification method, thus requiring no standard curves unlike quantitative real-time PCR (qPCR). Nevertheless, evaluation of the linear dynamic range, trueness and precision of an assay or platform is recommended, as there are several potential causes of important non-linearity, bias and imprecision. Ignoring these quality issues may lead to erroneous quantification. This necessitates an approach akin to the construction of standard curves. We study the pitfalls associated with the evaluation of such an experiment, and provide guidelines for the assessment of linearity, trueness and precision in dPCR experiments. We present simulation results and a case study supporting the importance of a thorough evaluation. Further, typically presented plots and statistics may not reveal problems with linearity, trueness or precision. We find that a robust weighted least squares approach is highly advisable, yet may also suffer from an inflated false positive rate. The proposed assessments are also applicable to other analyses, such as the comparison of results obtained from qPCR and dPCR. A web tool for quality evaluation, dPCalibRate, is available.

7.2 Introduction

Quantitative real-time polymerase chain reaction (qPCR) for the quantification of nucleic acids has become a mainstream technique since its conception over two

decades ago (Higuchi, 1992). Currently, a lot of interest is directed at an alternative nucleic acid quantification method termed digital PCR (dPCR) (Baker, 2012). Its many potential advantages such as its simplicity, precision and sensitivity is steadily turning dPCR into an appealing method for many researchers across a multitude of medical and life sciences disciplines (Corbisier et al., 2010; Doi et al., 2015; Li et al., 2014; Morisset et al., 2013; Vynck et al., 2016a).

A common application is that of quantifying the amount of nucleic acid in a sample with unknown concentration. For qPCR, this problem is typically tackled by constructing a standard or calibration curve as obtained from a series of dilutions of a source with a known concentration of nucleic acid. From the calibration curve, the concentration of the unknown samples can then be estimated. Further, key quality control parameters such as the efficiency of the PCR or the R^2 of the linear fit can be determined. Reporting of these parameters is considered essential by the MIQE guidelines to allow a correct interpretation of PCR results (Bustin et al., 2009). The calibration curve can also be used to determine the linear dynamic range of the assay, i.e. the range in which the results can be suitably described by a linear model.

Because dPCR is a direct quantification method, a calibration curve is not strictly required to estimate the concentration of an unknown sample. Nevertheless, a calibration curve can (and should) still be used to determine, for example, the linear dynamic range, trueness and precision of a dPCR assay or platform (Jones et al., 2016; Huggett et al., 2013). Nonlinearities, lack of trueness and imprecision of an assay or platform may invalidate the concentration estimates. In any case, it provides the researcher with more information on the assay or platform under consideration (Huggett et al., 2013).

A common misconception in the evaluation of such experiments is that high R^2 values are taken to be proof for good linearity, or as a proof of a dynamic range spanning the concentrations used to construct the calibration curve (e.g. Morisset et al. (2013); Verhaegen et al. (2016); Pavsic et al. (2016); Arvia et al. (2017), among many others). As we will show, this is a logical fallacy: it is perfectly possible to obtain high R^2 values in the presence of a serious deviation from linearity. This fallacy may lead to a false sense of security and/or wrongful conclusions. Consequently, it is of importance that linearity is thoroughly investigated and the results clearly reported. To address this problem, we will describe additional quality control procedures, some of which may be widely known but are very

rarely reported. We will argue that it is essential that these are detailed to allow suitable scientific scrutiny.

Moreover, trueness of the results is often not thoroughly studied nor extensively discussed (e.g. Jones et al. (2016); Verhaegen et al. (2016), among many others). If quantification by dPCR has high trueness, it is expected that the observed concentration (i.e. the output concentration, as measured by the dPCR instrument) is equal to the expected concentration (i.e. the input concentration with known concentration such as a certified reference material or internatational standard). There are several reasons why dPCR may not have high trueness, especially at high concentrations, including, but not limited to, sample or assay dependent parameters such as amplification inhibition and platform dependent parameters such as volume variability (Vynck et al., 2016a).

Finally, the measurement uncertainty is often not studied (e.g. Jones et al. (2016); Verhaegen et al. (2016), among many others), while it is of paramount importance when determining the practical use of these measurements, especially at very high or low concentrations as this is where problematically large variability tends to occur.

It is important that researchers and reviewers are aware of the pitfalls accompanying assessment of linearity, trueness and precision. Ignoring these may invalidate research findings (Bustin, 2014, 2017).

In this paper, we show that substantial bias may lead to what are typically considered sufficiently high R^2 values. We show that, perhaps surprisingly, a bias at more than one dilution level may lead to a higher R^2 value than a bias at a single dilution level. We demonstrate that commonly used plots for visual inspection of the linear curve can be misleading, and suggest some alternative visualisations. Furthermore, we highlight the importance of inspecting the residuals from the model fit and what to look out for. We study several statistical tests and procedures that may allow to detect deviation from linearity. We detail a qualitative and quantitative method to assess trueness and comment on the importance of precision assessment. Finally, by reanalysing publicly available dPCR data, we show that relying solely on R^2 values and visual inspection may indeed obscure potential issues.

Throughout these sections, we will compare the use of an ordinary least squares and a robust weighted least squares approach for calculating the linear fit, and

show the superiority of the latter. At the same time, we highlight the lesser-known and thus frequently disregarded inflated type I error rate of the (robust) weighted least squares approach.

7.3 Material and methods

7.3.1 Data and data generation

We obtained concentrations from an HIV genome fragment dilution series from a publicly available source (Jones et al., 2016). The aim of the experiment was to determine the dynamic range of an HIV assay on the RainDance RainDrop dPCR platform. The dilution series consists of seven ten-fold dilutions. At each of these dilution levels, estimated concentrations of five replicates are available. These measurements were used to set up a simulation experiment.

For each of the seven dilution levels we calculated the sample variance of the five observed concentrations at a given dilution point. The expected concentrations and the sample variances observed at those concentrations may then serve as a model from which dilution series data can be generated. We assume a normal distribution at each of the dilution levels. Additional information is given in Appendix F.1.

We simulated dilution series data as follows: there are seven dilution points, each dilution being a tenfold dilution from the previous dilution concentration. For each of these dilution levels, we generated a given number of replicates, depending on the scenario (see further). As the true data-generating mechanism is known, we can assess the performance of various approaches using Monte Carlo simulations. Each of the Monte Carlo simulation studies involved 10 000 simulation runs.

7.3.2 Calibration curve fitting

Throughout all assessments we fitted calibration curves in two different ways. We obtained a first fit using the (arguably) most commonly used method: ordinary least squares. The second fit is obtained using a robust weighted least squares approach. This latter approach is a multi-step procedure, motivated by research on

heteroscedastic linear models, i.e. linear models for which the variability of the dependent variable depends on the value(s) of the independent variable(s) (White, 1980; MacKinnon and White, 1985; Carroll and Cline, 1988; Long and Ervin, 2000). First, we fitted a weighted least squares model with weights estimated as the inverse of the sample variance at the given dilution level. Subsequently, we performed inference using HC3-type robust covariance estimators (MacKinnon and White, 1985; Long and Ervin, 2000). Our simulation results show that this latter method is best at retaining a correct type I error rate. Additional information is given in Appendix F.2.

7.3.3 Scenarios

7.3.3.1 Impact of bias on R^2

In a first experiment, we aimed to assess the impact of different downward biases on the value of the R^2 statistic, and to compare the R^2 values as obtained from a robust weighted least squares fit versus an ordinary least squares fit. We generated five replicates at each of the seven dilution levels. To mimic scenarios where biases occur at the highest concentration levels, we consider five scenarios: (i) no bias at any concentration level, (ii) a single bias of 10% at the highest concentration level, (iii) a single bias of 20% at the highest concentration level, (iv) a double bias where the highest concentration level has a bias of 20% and the second highest level a bias of 10% and (v) a double bias where both the highest and second highest concentration level have a bias of 20%. We subsequently calculated R^2 values from linear models fitted with both ordinary least squares and robust weighted least squares. We then compared R^2 values from both approaches and under the different simulation scenarios w.r.t. discriminatory power: can we use R^2 values effectively to establish a deviation from linearity and if so, under what circumstances?

7.3.3.2 Linearity: qualitative assessment

We generated a single calibration curve (five replicates, seven dilution levels) with no deviation from linearity (no bias at any concentration level). Next, we plotted both the original calibration curve and that same calibration curve but with an introduction of a 10% bias at the highest concentration level in multiple ways.

We constructed three types of plots: (i) a plot displaying the unbiased data, with the log-transformed observed concentration as a function of the log-transformed expected concentration, along with the fit of an ordinary least squares simple linear regression model, (ii) a plot displaying those same quantities, but with a 10% bias at the highest concentration level and (iii) a plot displaying the biased data, but with a robust weighted least squares fit.

We furthermore constructed plots displaying the observed and fitted concentrations for each of the concentration levels separately, for both the biased and unbiased scenario. We generated a table displaying the signs of the residuals from the robust weighted least squares models fitted to both the unbiased and the biased data.

7.3.3.3 Linearity: quantitative assessment

We generated data with the same biases as outlined for the assessment of R^2 values, but with numbers of replicates ranging from 2 until 8. We fitted models with both the expected concentration as well as the squared expected concentration as predictors (i.e. a quadratic model) and using both ordinary least squares and robust weighted least squares. Finally, we compare the p-value for the quadratic term to a 5% level of significance. This allows us to assess the type I error rate (false positive rate): in the absence of any bias, we expect the type I error rate to be controlled at the 5% significance level. We may also calculate power to detect a deviation from linearity (i.e. a quadratic coefficient that is significantly different from zero) for the scenarios where a bias was introduced.

We generated data in the same way to assess the type I error rate and power for the Wald-Wolfowitz runs test (Wald and Wolfowitz, 1940) (as implemented in the randtests R package version 1.0) and the frequency within a block test (Bassham et al., 2010), relying on the residuals of a simple linear regression model to assess the randomness of the residuals. We performed these assessments using both ordinary least squares and robust weighted least squares.

In a fourth and last assessment, we performed an F-test for lack of fit (Kutner et al., 2005), and for both ordinary and robust weighted least squares fits.

7.3.3.4 Direct quantification trueness

We generated data for all seven dilutions levels, with the number of replicates ranging from 2 until 8. We generated data in three ways: (i) without bias, (ii) with a bias of 5% for all dilution levels and (iii) with a bias of 10% for all dilution levels, i.e. with an expected slope of 1, 0.95 and 0.9, respectively. We subsequently calculated the type I error rate and power to detect a slope different from one for all scenarios.

To allow a qualitative assessment of trueness, we constructed plots displaying the observed concentrations and the expected concentration for each of the concentration levels. To construct the calibration curves, we generated data as before, with five replicates at seven dilution levels and with a 10% bias at the highest concentration level.

7.3.3.5 Direct quantification precision

We evaluated precision in terms of the coefficient of variation (CV) by generating data for a given dilution level, with the number of replicates ranging from 2 until 8. For each Monte Carlo simulation run, we calculated the CV (estimated as the sample standard deviation divided by the sample mean). Eventually, we normalized the CVs by dividing by the true data-generating CV. The result is indicatory of the (relative) bias and the variability of the precision estimate.

7.3.4 Software

We performed all simulations using R version 3.3.2, including the packages sandwich version 2.3-4, lmtest version 0.9-35 and randtest version 1.0. Our R code is available at https://github.com/CenterForStatistics-UGent/dPCalibRate.

7.4 Results and Discussion

We considered multiple scenarios to study the performance of several approaches aiming to assess linearity, trueness and precision. A scenario without any deviation from linearity or trueness was included, as well as several scenarios where

we introduced biases. More specifically, we introduced a single bias at the highest concentration level, or a double bias, at the two highest concentration levels. These bias-scenarios correspond to non-linearities and biases that are likely to occur in a typical dPCR experiment. Too high DNA concentrations may hinder the PCR reaction, for example due to high concentrations of inhibitors (Opel et al., 2010). High DNA concentrations may also lead to underestimation as a consequence of e.g. volume variability (Huggett et al., 2015; Jacobs et al., 2014).

While these scenarios represent only a subset of cases that do not show good linearity and/or trueness, they serve to demonstrate that the currently reported assessments in the dPCR literature are unsatisfactory.

7.4.1 High R^2 values do not imply good linearity

Figure 7.1 displays the distributions of R^2 values obtained from a set of simulations of dPCR dilution series data without bias and with varying degrees of bias. The R^2 values for ordinary least squares under different scenarios are all very high, with a large majority of R^2 values above 0.995. This immediately indicates that ordinary least squares R^2 may be inadequate to detect deviations from linearity and that alternative measures may be required: as in qPCR, R^2 values of 0.985-0.99 and higher are generally considered sufficiently high to indiciate linearity in dPCR experiments (Harwood et al., 2011; Racki et al., 2014b; Hayden et al., 2013). Yet, even in the severely biased scenarios, high R^2 values around and above 0.985-0.99 may be obtained. This finding is partially alleviated when using a robust weighted least squares approach, but still a large majority of R^2 values in the bias-scenarios are larger than 0.985 (Figure 7.1b).

Figure 7.1a furthermore indicates that higher R^2 values may be obtained when a bias is present at more than one dilution level. Indeed, R^2 values are lowest when there is a single bias of 20% at the highest concentration, but R^2 values shift closer to 1 when biases are present at the second highest concentration level. Perhaps most strikingly, the best R^2 among the considered biased scenarios are obtained when the two highest concentrations suffer from the highest biases considered (both 20%), thus confirming the inadequacy of the R^2 measure for the assessment of linearity. In contrast, when using a robust weighted least squares method, lower R^2 values typically correspond to more biased measurements (Figure 7.1b).

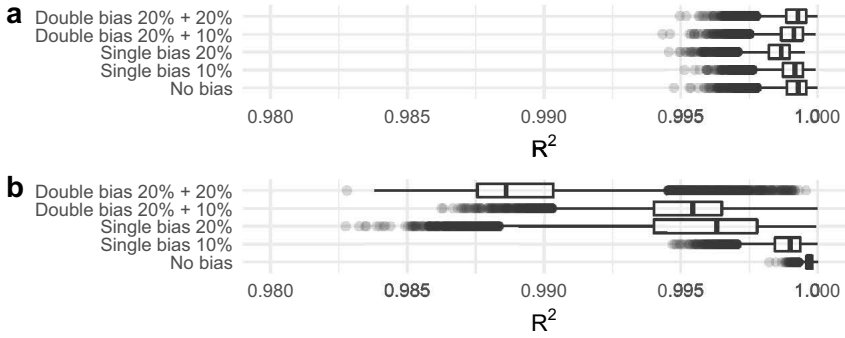


FIGURE 7.1: R^2 values as obtained from data generated under different scenarios for (a) ordinary least squares, (b) robust weighted least squares. Single bias scenarios introduce a 10% or 20% downward bias at the highest concentration level. Double bias scenarios introduce a 20% downward bias at the highest concentration level and a 10% or 20% downward bias at the second highest concentration level

Still, there is significant overlap in the distribution of R^2 values for different scenarios. Consequently, determination of linearity based on the R^2 value of a given experiment is not to be recommended. It is not unlikely that high R^2 values may still be obtained even in the presence of large biases. Reporting of R^2 values alone is thus inadequate to determine linearity. Especially if the R^2 value has been obtained from an ordinary least squares fit, which is currently common practice in the dPCR literature, its informativeness is typically limited.

7.4.2 Linearity: qualitative assessments

Qualitative assessments of linearity may be performed in at least two ways: a visual assessment of the linearity using suitable plots and an inspection of the residuals obtained from a linear fit.

Visual assessment may show obvious deviations from linearity, but the clarity of such deviations depends on the type of plot constructed. Often, log-log plots are constructed, as they seem to allow an assessment at each of the dilution levels. Such plots are shown in Figure 7.2, panels a-c, and it seems that there is reasonable linearity along the whole range of dilutions studied, even though the highest concentration in both panel b and c has a bias of 10%. From such plots it is hard

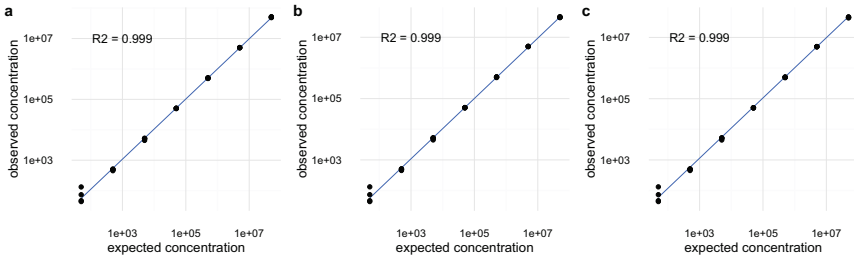


FIGURE 7.2: Log-log plots of observed and expected concentrations. Panel a: no bias is introduced. Panels b and c: a bias of 10% is introduced at the highest concentration. Panels a and b show an ordinary least squares fit. Panel c shows a robust weighted least squares fit

to observe a deviation from linearity as deviations become less pronounced on the log scale. An assessment of linearity using this type of plot is unsatisfactory.

A clearer visualisation of deviation from linearity is obtained by zooming in on the individual dilution levels of the plots, as exemplified in Figure 7.3: in case of no bias, points are randomly scattered around the estimate obtained from the regression line (panel a), while there are clear patterns in case of deviation from linearity (panel b).

This plot is similar to an assessment of the residuals, but a typical residuals plot is not useful in the case of PCR experiments: due to the large dynamic range and heteroscedasticity, the residuals at low concentrations are dwarfed by those at high concentrations, making inspection of the residuals at low concentrations cumbersome. However, it remains possible to inspect the signs of the residuals, which should not display obvious patterns such as long sequences of positive or negative residuals. An example of sign inspection is given in Table 7.1 for an unbiased dilution curve and a biased dilution curve (10% bias at the highest concentration level). For the latter scenario it is apparent that the signs display a clear pattern, i.e. long stretches of positive or negative residuals, in the biased scenario. Note that there is no natural ordering of signs within a given concentration level, which may affect the sequences of positive or negative residuals.

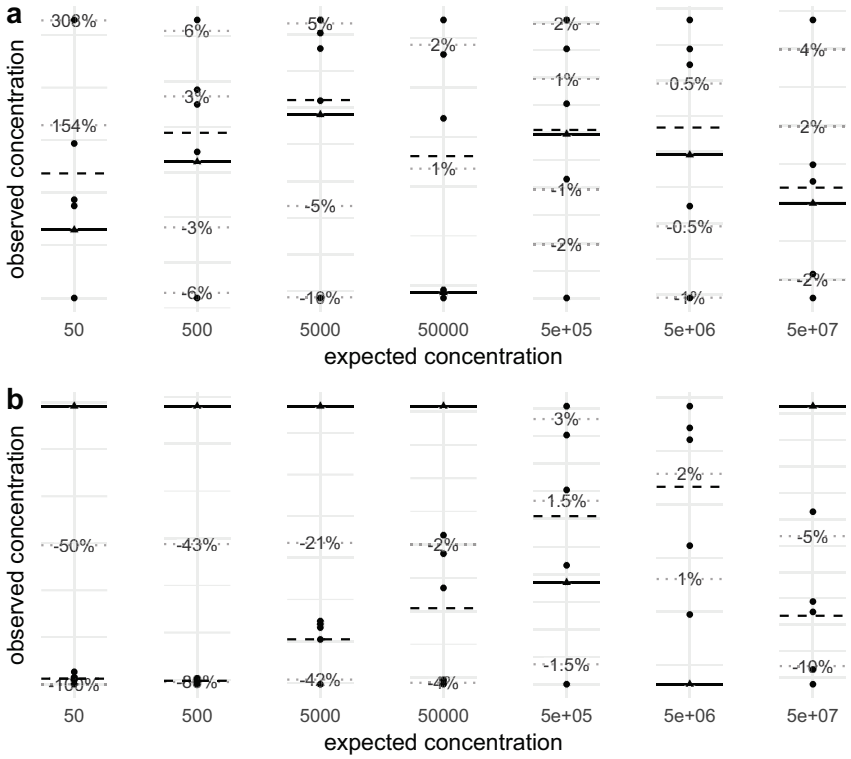


FIGURE 7.3: Observed (circles), average of observed (dashed line) and fitted (triangles, full line) values at different concentrations. Percentages (gray) indicate the percentage deviation from the fitted values, serving to reflect the variability of individual estimates and discrepancies between fitted and average of observed values. Fitted values were obtained from a robust weighted least squares fit. Note that the scale of the y-axis is variable and has been omitted for compactness. Panel a: no bias. Panel b: a bias of 10% is introduced at the highest concentration

TABLE 7.1: Signs of the residuals in an unbiased (top row) and biased (bottom row) scenario

5×10^7	5×10^6	5×10^5	5×10^4	5×10^3	5×10^2	5×10^1
−+−++	−+++−	++−−+	−++++	+++−+	++−++	++++−
−−−−−	+++++	+++−+	−−−−−	−−−−−	−−−−−	−−−−−

7.4.3 Linearity: quantitative assessments

There are several quantitative ways to assess deviations from linearity. Firstly, one can model the observed concentration by means of a regression model with a linear and a quadratic effect term. A coefficient representing the quadratic term that is significantly different from zero indicates a deviation from linearity.

Secondly, quantitative assessment may also be done using a Wald-Wolfowitz runs test on the residuals of the model (Wald and Wolfowitz, 1940). The runs test aims to detect unusual patterns (stretches of postive or negative residuals, so-called runs) in the residuals. When linearity is satisfied, it is expected that there will be an equal number of positive and negative residuals and that their ordering is completely random, i.e. long runs are unlikely to occur. In case of deviation from linearity, equal numbers of positive and negative residuals are not necessarily expected, and long runs are more likely to occur. The runs test allows to statistically determine the significance of the runs as qualitatively observed in Table 7.1.

A third quantitative approach is the frequency within a block test (Bassham et al., 2010). This test can be used to determine whether the signs of the residuals within a block are random, i.e. the frequency of positive residuals is equal to the frequency of negative residuals, or whether either negative or positive residuals are more prevalent than can be explained by chance, indicating a deviation from linearity within that block. A block can here be considered as a dilution level.

A fourth quantitative assessment of linearity is an F-test for lack of fit (Kutner et al., 2005). This lack of fit assessment consists of comparing the pure error of the *full model* (the within-dilution level variability) with the total error from a *reduced model* (the residual error as obtained from the linear model, which is a combination of the pure error and error as a consequence of deviation from the linear fit). If the total error is significantly larger than the pure error, there is evidence of a deviation from the linear model.

Of note, even if deviations from linearity are present, these tests may fail to reach significance if there is not enough statistical power to detect such deviations (e.g. too few replicates or dilution levels). Table 7.2 shows the type I error rate and power for the four different quantitative methods and for data simulated under scenarios with different biases, for even numbers of replicates ranging from 2 until 8 (full results are given in Tables F.1 and F.2).

Table 7.2: Type I error rate (no bias) and power to observe deviation from linearity using an OLS and robust WLS approach and for different biases and replicate sizes (Rep). Methods studied are a quadratic regression (Quad), a runs test (Runs), a frequency within a block test (Freq) and an F test for lack of fit (LoF)

		Robust WLS					OLS				
		Type I	Power				Type I	Power			
Quad	Rep	No bias	10%	20%	20% + 10%	20% + 20%	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.18	0.82	0.93	0.81	0.40	0.08	0.52	0.89	0.57	0.05
	4	0.10	0.99	1.00	1.00	0.82	0.00	0.55	0.99	0.66	0.00
	6	0.08	1.00	1.00	1.00	0.99	0.00	0.72	1.00	0.83	0.00
	8	0.07	1.00	1.00	1.00	1.00	0.00	0.87	1.00	0.95	0.00
Runs	Rep	No bias	10%	20%	20% + 10%	20% + 20%	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.03	0.07	0.14	0.30	0.38	0.05	0.00	0.00	0.34	0.15
	4	0.03	0.14	0.34	0.86	0.96	0.55	1.00	1.00	0.95	0.85
	6	0.04	0.21	0.55	0.98	1.00	0.65	1.00	1.00	0.99	0.98
	8	0.04	0.30	0.70	1.00	1.00	0.76	1.00	1.00	1.00	1.00
Freq	Rep	No bias	10%	20%	20% + 10%	20% + 20%	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.02	0.22	0.52	0.96	0.99	0.95	1.00	1.00	1.00	1.00
	6	0.01	0.44	0.86	1.00	1.00	0.99	1.00	1.00	1.00	1.00
	8	0.01	0.72	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LoF	Rep	No bias	10%	20%	20% + 10%	20% + 20%	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.43	0.35	0.35	1.00	1.00	0.03	0.02	0.02	0.47	0.84
	4	0.20	0.16	0.17	1.00	1.00	0.00	0.00	0.00	0.71	1.00
	6	0.14	0.12	0.12	1.00	1.00	0.00	0.00	0.00	0.93	1.00
	8	0.11	0.10	0.10	1.00	1.00	0.00	0.00	0.00	0.99	1.00

The most apparent conclusion from this simulation is that this exploration has little value when not using robust weighted least squares. When using an ordinary least squares approach, the power is extremely low, even for some of the scenarios with major deviations from linearity.

As a consequence of heteroscedasticity, the type I error rate (false positive rate) is furthermore not controlled when using ordinary least squares: depending on the type of assessment, the type I error rate is either deflated (quadratic regression approach, F test for lack of fit) or severely inflated (runs test, frequency within a block test), with increasingly undesirable behaviour for increasing numbers of replicates. Thus, even if a runs test and frequency within a block test have high power to detect deviations from linearity when a bias is present, the high type I error rate renders these approaches useless as it is not possible to distinguish between false positives and true positives. Indeed, the runs test and frequency within a block test will often indicate that there is deviation from linearity when in fact there is none.

With respect to type I error rate control, the robust weighted least squares performs much better, although power is still unsatisfactory for some scenarios (e.g. the 20% + 20% bias scenario in the quadratic regression approach) when the number of replicates at each dilution level is limited. Moreover, the robust weighted least squares approach also suffers from an inflated false positive rate for the quadratic regression approach and the F test for lack of fit approach. An explanation can be found in the estimation of the weights: with relatively few replicates there is insufficient data for precise weight estimates, resulting in variable standard error estimates and consequently in unstable p-values and hence poor type I error rate control, even when using a best-in-class robust approach (Appendix F.2). Similarly as with ordinary least squares, this complicates the interpretation of p-values obtained from such an analysis, albeit to a lesser extent. Using larger numbers of replicates (partially) alleviates this complication, providing another argument for performing these experiments with sufficiently high numbers of replicates.

It is interesting to note that some methods perform uniformly better than others. When using the superior robust weighted least squares method, a frequency within a block test achieves power higher than the runs test (except for the two replicate scenario). This may result from the fact that it is a more appropriate test than the runs test: the ordering of the residuals for the runs test is not clear as

there are multiple measurements at the same dilution level, while the frequency within a block test naturally deals with multiple measurements at the same dilution level as it considers the signs of the residuals at each dilution level separately.

In many cases, however, the best method depends on the scenario under consideration. For single biases at the highest concentration level, the quadratic regression approach outperforms all other methods. This may stem from the fact that it suffers from an inflated type I error rate for low numbers of replicates, but its superiority is maintained for higher numbers of replicates where the inflation of the type I error rate is limited.

For the most severely biased scenario under consideration (20% + 20% bias scenario), the quadratic regression approach is outperformed by the frequency within a block test and the F test for lack of fit, although this latter approach suffers from a severe inflation of the type I error rate, hampering its practical use.

In conclusion, the quadratic regression approach is to be preferred, but one should be wary of its result when the number of replicates is low (i.e. less than five). In such a case, a frequency within a block test may be preferred, as it does not suffer from type I error rate inflation. Its power is however lower in most scenarios under consideration, in part because the frequency within a block test tends to be conservative.

7.4.4 Direct quantification trueness

Additionally, for direct quantification methods such as dPCR, it is of interest to investigate the trueness of the quantification. If the direct quantification has high trueness, observed concentrations are expected to be equal to expected concentrations, in other words the slope of the linear fit should be one.

If trueness assessment is to be performed, two practices are advisable: (1) to check whether the slope is indeed one or deviates from one, and (2) to force the slope to be one and the intercept zero, after which a residual inspection may indicate concentrations for which direct quantification is either biased or unbiased. The reporting of the regression equation alone is not sufficient, as it does not give any information about the uncertainty on the estimated slope and intercept, and thus whether they may differ from the ideal scenario.

The type I error rate and power to detect deviations from a slope equal to one

TABLE 7.3: Type I error rate (no bias) and power to detect deviation from a slope of one using a robust weighted least squares and an ordinary least squares approach for different replicate sizes (Rep)

	Robust WLS			OLS		
	Type I	Power		Type I	Power	
Rep	No bias	5%	10%	No bias	5%	10%
2	0.41	0.99	1.00	0.61	0.96	1.00
4	0.19	1.00	1.00	0.50	1.00	1.00
6	0.14	1.00	1.00	0.47	1.00	1.00
8	0.11	1.00	1.00	0.46	1.00	1.00

for different biases and number of replicates is listed in Table 7.3. The power to detect deviations from a slope of one are very high, even for small numbers of replicates (e.g. two), but this may stem from an underestimation of the standard error, as noted earlier for quantitative assessment of linearity and as outlined in Appendix F.2. Indeed, the type I error rate is highly inflated. This observation is even worse for ordinary least squares, for which the type I error remains high even for larger numbers of replicates. This again provides evidence that robust weighted least squares poses an advantage over ordinary least squares.

Secondly, a visual assessment similar to Figure 7.3 may be produced to inspect deviations from a slope of one by plotting the observed in function of the expected concentrations at each of the dilution levels. Figure 7.4 shows a case where a bias has been introduced at the highest concentration, which can be readily observed. Note that the difference with Figure 7.3 is that the triangles now represent the expected concentration (i.e. those corresponding to a slope of one and an intercept of zero), rather than those obtained from the (robust weighted) least squares fit. Moreover, this plot allows an inspection of trueness at each of the dilution levels.

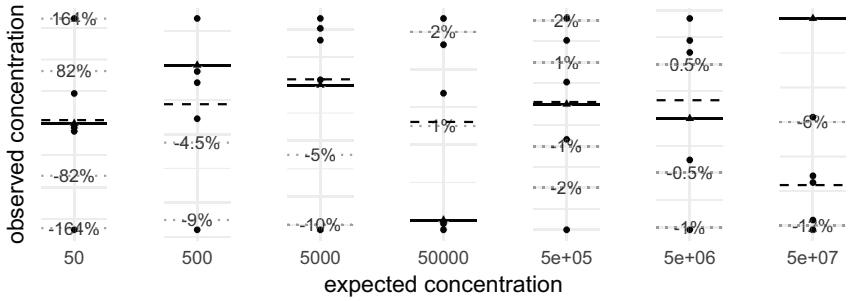


FIGURE 7.4: Trueness assessment plot: A bias of 10% is introduced at the highest concentration. Triangles indicate the expected concentration for perfect trueness, i.e. corresponding to a slope of one and an intercept at zero. Dots indicate the observed concentrations. Percentages indicate deviation from perfect trueness

7.4.5 Direct quantification precision

Besides the importance of the quantification having high trueness, i.e. being correct on average, it is furthermore of importance that the quantification is not too imprecise, i.e. individual estimates should not deviate too much from the average. If the assay or platform is imprecise, more technical replicates will be needed to obtain a reasonably correct estimate of the quantity of interest.

This warrants an investigation of the variability at each of the dilution levels in order to determine the practically useful dynamic range of an assay or platform. The variability at each of the dilution levels should be relatively small compared to the concentration at that dilution level. In other words, the coefficient of variation (CV) should be sufficiently low.

A fact that is often neglected, is that the CV is variable itself. Thus, the uncertainty with which it is estimated from the data also depends on the number of replicates at a given dilution level. Figure 7.5 shows the bias and variability of the CV for varying numbers of replicates. Figure 7.5 indicates that the uncertainty of the CV is very high for the typical duplicate or triplicate analyses and only diminishes slowly with increasing numbers of replicates. For a typical experiment performed in duplicate or triplicate, the true CV may be more than twice the observed CV with reasonable probability. Moreover, for low numbers of replicates, the CV is typically an underestimate of the true CV. When reporting a CV, it is

thus highly advised to report it along with its confidence interval (MacKay, 1932). Comparisons of CVs, for example those obtained from a qPCR and dPCR experiment, should furthermore take into account that the CVs have been determined with uncertainty.

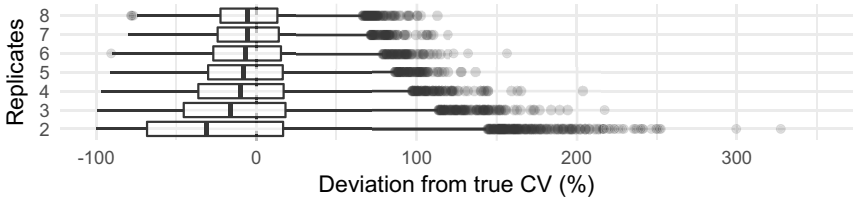


FIGURE 7.5: Precision and bias of the CV for varying numbers of replicates

7.4.6 Guidelines and recommendations for the design and analysis of a quality evaluation experiment

Unlike for other quantitative assays, for example the bioanalytical method validation guidelines for drug substances of the FDA (Food and Drug Administration, 2001), there are currently limited guidelines as to what constitutes acceptable trueness, precision or linearity for nucleic acid quantification procedures. What is acceptable should be determined by the field expert (Clinical and Laboratory Standards Institute, 2014). Moreover, how these quality parameters need to be assessed is often not specified or the suggested assessment procedures are inadequate as shown by our analyses.

The EPA, for example, lists quality parameters that need to be assessed for nucleic acid quantification procedures, including linearity, trueness and precision, but does not detail what is considered acceptable (Environmental Protection Agency, 2009). They mention that linearity in PCR experiments can be assessed using the R^2 statistic (Environmental Protection Agency, 2009), which we have shown is inadequate in dPCR experiments.

The FAO does provide some guidelines for nucleic acid quantification in food samples, stipulating an imprecision (more specifically, repeatability) of not more than 25% (CV) over the entire dynamic range, a deviation from trueness (bias) not exceeding 25% and an "acceptable" linearity (Food and Agricultural Organization,

2010). How to assess linearity is not specified, though it is mentioned that "a sufficient number of standards" should be used.

CLSI guideline EP06 details that linearity may be assessed using polynomial regression, and they suggest the use of two replicates at each of five dilution levels (Clinical and Laboratory Standards Institute, 2003). Our analyses suggest that this may result in poor power to detect deviation from linearity in some cases and that higher numbers of replicates and/or dilution levels are recommended. CLSI guideline EP06 furthermore mentions that weighted regression with weights as the inverse of the variances of the replicates may be used (Clinical and Laboratory Standards Institute, 2003). We have shown that this indeed improves linearity assessment, but that using a robustified version performs even better (Appendix F.2). Guideline EP05 discusses the impact of the number of replicates on the uncertainty of the precision estimate (Clinical and Laboratory Standards Institute, 2014). Likewise, we have demonstrated that the typical duplicate or triplicate experiment may result in large uncertainty of the precision estimate. This may hamper a conclusive interpretation of repeatability and/or reproducibility.

We summarize our findings as six recommendations. Note that these recommendations are based on the data analysed within this paper, and that these recommendations depend on the amount of variability in the replicated measurements, the number of dilution levels and the dynamic range studied. It is not possible to give exact numbers of dilution points or replicates as this will depend on the dPCR instrument, assay characteristics, … In general, the following statements can be made: increased variability warrants more replicates, increased numbers of dilution levels trade off with the number of replicates and an increased dynamic range requires more replicates as the uncertainty at very low and very high concentrations is typically increased.

1. A robust weighted least squares approach should be used as the ordinary least squares approach suffers from a loss of control of the type I error rate.

2. Even if a best-in-class robust weighted least squares approach is adopted, the type I error rate is often inflated and a careful interpretation is needed. The inflation of the type I error rate decreases with increasing number of replicates.

3. The quadratic regression approach gives a clear indication of deviation from linearity and is to be preferred above a runs test, frequency within

a block test or F test for lack of fit. Both the runs test and the frequency within a block test have lower power to detect deviations. The highly inflated type I error rate of the F test for lack of fit limits its practical use. However, care also needs to be taken when interpreting the results from a quadratic regression approach, as it suffers from an inflation of the false positive rate for small numbers of replicates. In those cases, the result from the frequency within a block test may proof useful, as it also displays power to detect deviations but does not suffer from a type I error rate inflation.

4. The number of replicates at each dilution level is ideally five or more. This will ensure a high power to detect deviations and allow to determine the CV at each of the dilution levels with reasonable precision. The inflation of the type I error rate of the assessments is furthermore limited for five or more replicates.

5. If calculated from a robust weighted least squares fit and linearity has been confirmed, R^2 values may still be useful for comparative assessments e.g. to compare the relative performance in terms of precision of two assays as in such a situation, the R^2 values only reflect the imprecision of the assay.

6. Qualitative assessments such as graphical displays for linearity and trueness assessment often yield useful insight, even when the quantitative approaches fail to detect significant deviations.

7.4.7 Case study: dynamic range of a KRAS dPCR assay

The data analysed for this case study were obtained from an experiment aiming to assess the dynamic range of a KRAS mutation dPCR assay (Jones et al., 2016). Note that solution concentrations reported for the assay were limited to one significant digit (Jones et al., 2016), preventing a conclusive statement on trueness or bias for this particular assay. The trueness and bias findings presented should consequently be considered educational.

The authors reported log-log plots similar to those displayed in Figure 7.2 and R^2 values for the fitted linear regression lines using an ordinary least squares approach. Based on those findings, they concluded that there was good linearity across the concentrations studied. As outlined before, these assessments are insufficient for a thorough evaluation of linearity. Furthermore, there was no

mention of trueness or precision of the assay. We therefore applied the methods described in this paper to assess linearity, trueness and precision.

From the preferable tests, neither the quadratic regression approach ($p = 0.42$) nor the frequency within a block test ($p = 0.08$) indicated a deviation from linearity. It should be noted that the power to detect deviations for intermediate concentrations (as observed qualitatively in this case) is very low for the quadratic effect test. Typically, only biases at the end-points of the dilution curve (very high and/or low concentrations) have a major effect on the estimated quadratic effect. This highlights the utility of the qualitative assessments.

Figure 7.6 shows the graphical procedures applied to the data. Linearity seems to be good, except for the replicates containing 5000 copies, where an underestimation of copies is observed. Trueness assessments confirm the underestimation for the replicates containing 5000 copies, and additionally shows an overestimation for the replicates containing five million copies and an underestimation for those containing 500 and 50 copies, with estimated biases of respectively 5.2%, 2.4%, 14% and 4.8%. Furthermore, the outlying value with the highest concentration at the highest concentration level (five million copies) may be worrisome, given the overestimation at the second highest concentration level. A larger number of replicates could have resolved whether there are indeed potential issues.

Additionally, precision turns out to be potentially worrisome for the two lowest concentration levels, with coefficients of variation of 37% (95% CI [21% to 142%]) and 21% (95% CI [12% to 64%]).

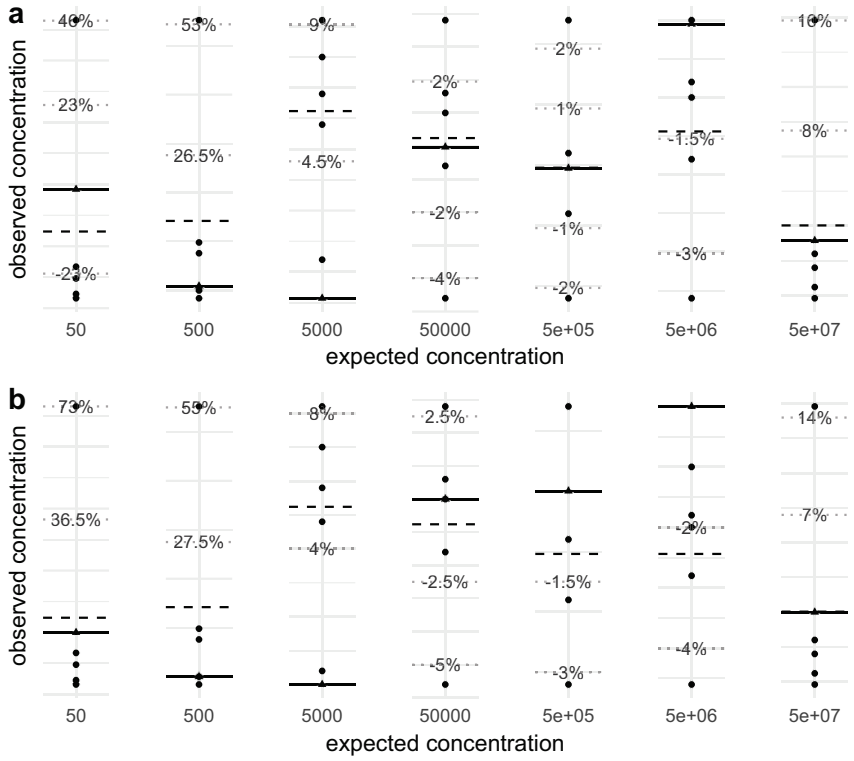


FIGURE 7.6: (a) Linearity assessment plot for a dPCR KRAS assay. Triangles indicate the fitted values of the robust weighted least squares fit. Dots indicate the observed concentrations. Percentages indicate deviation from the fitted value. (b) Trueness assessment plot for a dPCR KRAS mutation assay. Triangles indicate the expected concentration for perfect trueness, i.e. corresponding to a slope of one and an intercept at zero. Dots indicate the observed concentrations. Percentages indicate deviation from perfect trueness

7.4.8 dPCalibRate: a web tool for quality assessment

The techniques described in this paper have been implemented in the web tool dPCalibRate. dPCalibRate requires a series of expected and observed concentrations only, that can be entered in a spreadsheet-like interface or read from a comma separated value (csv) file. Using the data supplied by the user, a report is subsequently generated. This report contains all qualitative and quantitative

assessments as studied and discussed in this paper, allowing easy and fast quality control of linearity, trueness and precision. dPCalibRate is freely available at http://statapps.ugent.be/dPCR/dPCalibRate. The application code is available at https://github.com/CenterForStatistics-UGent/dPCalibRate.

7.5 Conclusion

Using real and simulated data, we have demonstrated that a thorough assessment of linearity, trueness and precision for dPCR experiments is required. The current common practice of reporting of the R^2 value and regression equation only is far from sufficient as it may not reveal potential issues. We have demonstrated the use of several qualitative and quantitative analyses that are able to properly demonstrate deviations from linearity and trueness and have commented on the importance of precision assessment. We have shown that the use of ordinary least squares is inappropriate and that the advantages of a robust weighted least squares approach mandate its adoption, notwithstanding the fact that a robust weighted least squares approach also has its limitations as it often suffers from an inflated type I error rate.

Our assessments were performed for data generated on the RainDance dPCR platform. The RainDance platform is the platform that currently has the highest theoretical dynamic range, due to its high number of partitions. A higher range of expected concentration corresponds with higher leverage for the observations at the highest concentrations, while the variability is also larger for higher concentrations, increasing the effect of heteroscedasticity. Consequently, the effect of heteroscedasticity and influential observations is most likely more extreme for the RainDance platform than for any of the other commercially available dPCR platforms. All of the observed problems such as inflation of the type I error rate are also present for these other platforms, but may not be as manifest, as a consequence of a lower dynamic range. The proposed assessments remain equally applicable and will enable improved scrutiny of linearity, trueness and precision compared to currently reported approaches.

It is noteworthy that the use of the proposed assessments is not limited to the applications presented, but can also be used to study the agreement between qPCR and dPCR results. Typically, only R^2 values are reported, but for these agreement analyses it is also interesting to study the linearity, i.e. to study whether

there is an agreement between both methods along the whole range of concentrations studied or whether there is an agreement for certain concentrations but a non-linear relation at higher concentrations. Similar to the trueness assessment presented, deviation from a slope of one can be assessed to determine whether the methods agree completely over the entire or part of the range studied. When comparing the precision of a qPCR and dPCR assay, the uncertainty on the precision estimate (CV) should also be taken into account, for example by reporting a confidence interval for the CV.

Ultimately, our findings demonstrate the need for a careful and thorough assessment and interpretation of key quality parameters in any dPCR experiment. It is finally up to the subject expert to determine whether these comply with subject-specific guidelines and to determine potential impact on the application of interest. Moreover, in the absence of guidelines, increased awareness and reporting of these important quality parameters and how to properly assess them may eventually lead to the development of application-specific guidelines.

8 Outcomes, perspectives and conclusion

8.1 Research outcomes

Quantifying the amount of one or multiple fragments of nucleic acid in a sample using digital PCR in an unbiased and precise way is not always straightforward. Notwithstanding that once the digital data has been obtained, an estimated quantity may be obtained by applying a simple formula (Equation 1.5), all sorts of difficulties may arise:

1. How do we obtain the digital signal, given that the readout from a digital PCR machine is a continuous fluorescence measurement for each of the partitions?
2. Given the digital signals of several replicates, how do we estimate the number of targeted nucleic acid fragments and its associated imprecision?
3. How many replicates do we run to begin with?
4. Given an estimate of the average number of targeted fragments per partition, how do we obtain an absolute or relative quantity of the target nucleic acid if the volume cannot be considered a constant?
5. How do we verify that the results are reliable, i.e. that they are unbiased and precise?

In this thesis we have proposed approaches to tackle these problems.

In Chapter 3 we have introduced power equations that can be used in the design stage of digital PCR experiments. In particular, we have provided an answer as to how an experiment should be set up so that it allows to test the research hypothesis of interest. Perhaps most importantly, the approach allows to determine the number of replicates needed to be able to make a conclusion reliably, e.g. to be able to conclude that a person suffers a disease when he/she does, or to conclude the opposite when he/she does not. Moreover, we have discussed how other parameters, typically depending on the composition of the sample or the digital PCR platform, influence this decision process.

In Chapter 4 we have introduced a new method to determine the partition status in single channel experiments, i.e. experiments targeting a single nucleic acid sequence per reaction. Our approach solves two problems that existed at the time. Firstly, other metjods assume a normal distribution of partition fluorescence intensities, ultimately resulting in a theoretically incorrect threshold of which the

influence on partition misclassification in practice should still be further investigated (Section 8.2.2). Secondly, we introduced a normalisation step that removes shifts in the location of the cluster of fluorescence intensities of negative partitions. Ignoring these shifts results in a theoretically incorrect threshold, a consequence of which may be the misclassification of some of the partitions in the samples to which the threshold is subsequently applied.

Later, others have published other approaches to tackle the same problem. Lievens et al. (2016) introduced a method that models the skewness of the distribution of negative fluorescence intensities, thus taking into account that these typically do not follow a normal distribution. Jacobs et al. (2017) introduced a method that is essentially very different from the other solutions in that they work with a soft thresholding procedure (as opposed to hard thresholding): the outcome of their procedure is a probability that the partition is positive, rather than a binary classification of the partitions as with the other methods. Their approach is also flexible in the sense that it models the densities non-parametrically, and thus does not assume a particular distribution of the fluorescence intensities.

In Chapter 5 we have outlined how the generalized linear mixed model framework may be used to obtain absolute or relative quantity estimates and their associated imprecision in a wide variety of experimental setups. This was motivated by the common malpractice of pooling the numbers of negative and positive partitions across replicates, thus ignoring the experimental design. The pooling of replicates ultimately results in an underestimation of the imprecision of the estimated quantities. In addition, we introduced a reference gene stability measure built upon the estimates resulting from the generalized linear mixed models and we demonstrated that relying on multiple reference genes typically improves the estimates, in terms of both trueness and precision.

In Chapter 6 we have tackled the question of how to obtain unbiased estimates of absolute and relative quantities when the partition volume cannot be considered a given constant. It has been long known that partition volume variability leads to an underestimation of absolute quantity estimates, yet flexible methods to take into account the partition volume variability were lacking. We have introduced a method to model partition volume variability of arbitrary form to eliminate this bias. In addition, we showed that such a correction is in some cases also needed for relative quantification, and we provided a correction procedure for the relative quantification scenario.

Finally, in Chapter 7, we studied several qualitative and quantitative methods to assess linearity, trueness and imprecision of digital PCR assays. This was motivated by our belief that the conventional assessments of these quality parameters were not sufficient to detect potentially worrisome non-linearity, bias or imprecision.

Together, these methods form a series of statistically motivated building blocks that allow to build a digital PCR data analysis pipeline, starting from a sample and, ultimately, leading to a more reliable, scrutinized estimate of the concentration of one or several target nucleic acids within that sample.

8.2 A look towards the future

We have addressed several issues in the digital PCR data analysis pipeline in the preceding chapters of this thesis. Yet, in the process of addressing questions, new questions inevitably arise. Several outstanding issues remain, and these are discussed in the subsequent sections.

8.2.1 Design of experiments and platforms

We have presented power equations that are helpful for designing digital PCR experiments. These equations apply to simple experiments, i.e. those experiments that aim to detect deviations from a predetermined reference value. While our method can be used for many of the experiments for which digital PCR is currently used, the derivation of power equations for more complex scenarios, such as two-group comparisons (e.g. differential expression) may be worthwhile when the cost and throughput of digital PCR reaches more competitive levels. Appropriate power calculations for such more complex setups may require a simulation-based approach, e.g. relying on a mixed model approach (Johnson et al., 2015; Kain et al., 2015). The evaluation of many different designs using such simulation-based approaches may however not be feasible, especially when relying on computationally intensive models such as generalized linear mixed models.

Studying experimental designs and their associated power to detect an effect of interest may also suggest improvements of the dPCR platforms themselves,

such as the development of a platform with partitions that have variable but predetermined volumes. Kreutz et al. (2011), for example, use a maximum likelihood approach to demonstrate how a multi-volume platform may provide a large dynamic range with relatively few partitions when these partitions are chosen thoughtfully, e.g. some with large volumes and some with small valumes to allow both quantification of high and low concentrations, without the need for dilution. A similar approach is discussed by Debski et al. (2015). Importantly, the design of such platforms may also be optimized using the generalized linear (mixed) model approach outlined in Chapter 5 by using an offset for the volume, similar to our approach in Chapter 6 for volume variability. This would allow flexible calculation of power for different designs, be it single or multi-volume platforms, with or without technical replicates, for absolute or relative quantification, … (Johnson et al., 2015; Kain et al., 2015).

8.2.2 Threshold setting validation

Several methods have been proposed for the determination of hard thresholds for the classification of partitions into positives and negatives (Strain et al., 2013; Dreo et al., 2014; Jones et al., 2014; Trypsteen et al., 2015b; Lievens et al., 2016), or, in the case of soft thresholding, to assign a probability that a partition contains the target nucleic acid (Jacobs et al., 2017). In addition, several methods have been introduced for the determination of partition status in the bivariate case, i.e. for two-channel experiments (Attali et al., 2016; Dobnik et al., 2016; Chiu et al., 2017; Lau et al., 2017).

While these methods have been shown to behave favorably on example datasets, there has been no rigorous investigation of their performance under various circumstances. Such an investigation is warranted. For example, some methods rely on assumptions that are violated in practice, such as a normal distribution of partition fluorescence intensities in the negative or both the negative and positive cluster (Trypsteen et al., 2015b). Other methods discard rain partitions which are likely to contain the target nucleic acid (Lievens et al., 2016). It is likely that (i) the extent of non-normality, (ii) the separation between clusters (or resolution, as termed by Lievens et al. (2016)), (iii) the amount and form of rain present and (iv) the amount of target nucleic acid in the sample, will affect the performance of each of these methods.

In addition, there is some evidence that shifts occur in the location of the negative (and positive) cluster from one sample to the next, and that not accounting for these shifts may negatively affect the performance of the thresholding procedures (Trypsteen et al., 2015b; Jacobs et al., 2017). Others have mentioned that such a correction is not necessary as the extent of these shifts is negligible, though without providing argument (Lau et al., 2017). Further investigation of the extent of shifts and their impact on the performance of the thresholding procedures and downstream data analysis is warranted.

8.2.3 Combining uncertainty

It is common to set hard thresholds to classify the partitions as negative or positive: a fluorescence threshold is determined and each partition having a level below this fluorescence intensity is said to not contain nucleic acid (negative), while partitions having a fluorescenve intensity above this threshold are said to be positive. These 0/1 measurements can subsequently be used within e.g. our GLMM framework (Chapter 5).

In the light of correct error propagation, this approach is not ideal: the determination of partition status is uncertain. It is easy to imagine that a partition which has a fluoresence intensity far below the fluorescence threshold is much more likely to be truly negative than a partition whose fluorescence intensity is only just below the fluorescence threshold.

To deal with this uncertainty, Jacobs et al. (2017) have recently developed a method that assigns a probability of being positive (negative) to each partition, along with an imprecision measure on that probability. For example, a partition is no longer said to be negative with 100% certainty, but it is said to be negative with a probability of 84% with 95% confidence interval [70% to 98%].

This change in outcome requires a different data analysis procedure than those for a 0/1 binary outcome. We provide here an outline of a possible solution for this problem.

Let $\hat{p}_j$ denote an unbiased estimator of the probability of partition j being positive, i.e. $\mathrm{E}(\hat{p}_j) = P(\text{partition } j \text{ is positive})$. We want to construct a model to estimate a parameter vector β based on the probability of observing a positive partition:

$$\text{logit}(P(p_j \text{ is positive}|\boldsymbol{X})) = \boldsymbol{X}\boldsymbol{\beta}, \tag{8.1}$$

where the logit is used to constrain the outcome to the range $[0,1]$ and $\boldsymbol{X}$ is the regressor vector. The structure of $\boldsymbol{X}$ and $\boldsymbol{\beta}$ will depend on the scenario under consideration (Vynck et al., 2016a). We can rewrite Equation 8.1 as

$$\text{E}(\hat{p}_j|\boldsymbol{X}) = \text{expit}(\boldsymbol{X}\boldsymbol{\beta}). \tag{8.2}$$

Without further imposing distributional assumptions on $\hat{p}_j$, the β parameter vector can be estimated by solving the estimated equation

$$\sum_j \boldsymbol{A}_j \left(\hat{p}_j - \text{expit}(\boldsymbol{X}_j\boldsymbol{\beta})\right) = 0, \tag{8.3}$$

where $\boldsymbol{X}_j$ is the regressor vector associated with partition j and where $\boldsymbol{A}_j = \boldsymbol{D}\boldsymbol{V}^{-1}$ is the product of a partial derivatives matrix

$$\boldsymbol{D} = \partial\left(\text{expit}(\boldsymbol{X}_j\boldsymbol{\beta})\right)/\partial\boldsymbol{\beta}, \tag{8.4}$$

and, assuming independence among the $\hat{p}_j$ for now, the inverse of a covariance matrix

$$\boldsymbol{V} = \text{diag}\left(V(\text{expit}(\boldsymbol{X}_1\boldsymbol{\beta})), \ldots, V(\text{expit}(\boldsymbol{X}_n\boldsymbol{\beta}))\right), \tag{8.5}$$

with V a known variance function (Dunlop, 1994; Stefanski and Boos, 2001). The variance function may be derived from the variances of the $\hat{p}_j$, using the method described by Jacobs et al. (2017).

The formulation above can be altered to accomodate specific scenarios. For example, if we have technical replicates, the variance matrix V may be organized as block diagonal and with compound symmetry within blocks as we do not expect the replicates' correlations to differ from one replicate to the next.

8.2.4 Open data and an open data format

There are currently many digital PCR platforms available on the market. An issue with many of these platforms is the lack of data accessibility: the data that are generated by the platform are not available to the user, or they are only available to the user after processing by software provided by the manufacturer. This may be problematic in several ways. We will focus on some issues that we have encountered with the Bio-Rad platforms.

Filtering of partitions may affect the partition classification tools and consequently also the results from the downstream analysis of partition statuses. In some of the data analysed, there obviously has been a removal of some partitions from the experiment. Moreover, an increase in the numbers of retained partitions has been reported with introduction of new software versions (Gerdes et al., 2016), indicating a change in the partition filtering algorithm. It is not clear how and why partitions are removed, but heterogeneity in partition intensity measurements may introduce additional technical variability that may otherwise be removed if all data was made available.

Partition intensities are sorted by intensity before being made available to the user. This prevents an analysis of e.g. signal drift in the intensity measurements. When intensities would be available to the user in the sequence as measured, detection of drift and correcting for drift may be possible and this may reduce technical variability (Figure 8.1), leading to improved absent/present calling of the target nucleic acid. Details on the strength of within-sample drift are lacking as it is currently not possible to study this with the data as provided.

Likewise, for platforms with spatial organization of partitions (e.g. chamber or crystal based), the omission of spatial information may eliminate the possibility of detecting and/or correcting for spatial effects, as may be present from e.g. thermal heterogeneity, architecture-based heterogeneity (e.g. heterogeneity in the size of droplet forming channels in crystal digital PCR or heterogeneity in chamber size in chamber digital PCR), ...(Spiess et al., 2016). As with drift, information on spatial organization may allow to reduce technical variability of intensity measurements.

In summary: making the raw measurements available to the scientist may result in the development of methods to deal with these issues and may thus improve the performance (in terms of e.g. variability) of the platform.

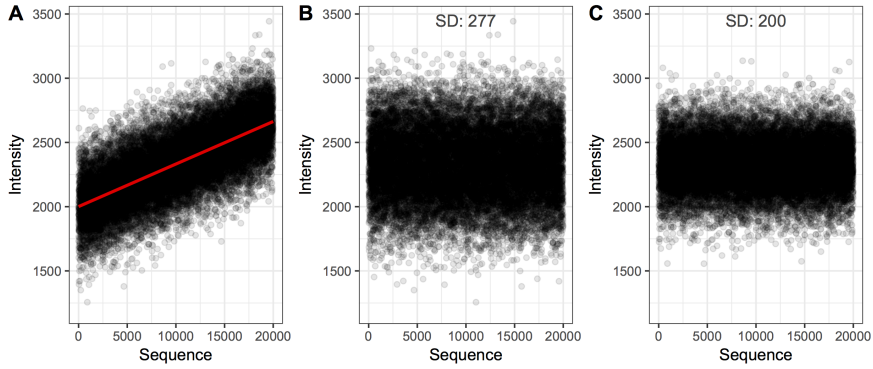


FIGURE 8.1: When the sequence of intensity measurements is available, correction for drift may result in reduced technical variability. A: intensity measurements with clear signal drift. B: intensity measurements when drift correction is not possible (random reordering of intensities in panel A). C: intensity measurements with drift correction by regressing the intensity on the sequence (red line in panel A). SD: standard deviation of intensity measurements.

Data availability in the community is an additional issue. Unlike for other types of biological data such as RNA-seq data, there are currently no data repositories for digital PCR data, nor is there a tendency to provide the data underlying manuscripts as supplementary material. This impedes (re)assessment of the data and assessment of data reliability/quality.

Availability of such a data repository may result in an organized digital PCR database and an increased incentive to provide data along with manuscripts. This would not only be beneficial in terms of experiment (re)assessment but may also result in increased focus on data analysis procedures and their impact on the final results. Indeed, much of the data analytical research performed to this date has been based on data simulation and in some cases limited validation on real datasets. Widespread availability of data from a single source may allow more involved assessment of the performance of different data analysis procedures.

A third issue is the lack of a digital PCR data standard. The development of a data repository may be hampered by the lack of a uniform format in which digital PCR data are presented. While different data types may be processed to be subsequently stored in such a data repository, a data standard poses several advantages such as easy data exchange and compatibility with different data analysis tools

built to support this data standard. For example, the RDML data standard has been proposed for quantitative PCR data (Lefever et al., 2009) and its advantages are acknowledged by many researchers (RDML, 2017). It is likely that digital PCR data may benefit from a similar data standard.

8.2.5 Software for data analysis

In addition to a data repository, the community could benefit from dedicated, open source digital PCR data analysis software. Many data analysis methods have been suggested (see e.g. Chapter 2), yet the adoption of these data analysis methods by those performing digital PCR experiments has been slow. The reason for this slow adoption is likely multifaceted.

Firstly, some of the methods lack an implementation (e.g. Strain et al. (2013)). It is unlikely that experimentalists will implement these methods themselves. Not providing an implementation is likely to prevent application.

Secondly, when an implementation is available, it should ideally be open source so that implementation details can be verified, or the method further improved. When source code is unavailable, authors need to make sure that their data analysis procedures are described in sufficient detail so that these are reproducible. There are several examples for which, sadly, this is not the case (e.g. Strain et al. (2013); Dobnik et al. (2016)).

Thirdly, while most of the data analysis procedures have been implemented in the R environment, this is not the case for all methods (e.g. Jones et al. (2014)). This hampers integration of the different steps in the data analysis pipeline, or the collection of alternatives for a single step in the data analysis pipeline and providing the user with a choice of method. For example, it is unlikely that someone wishing to go through the different steps in the analysis pipeline will be willing to import/export data and switch programming languages or software tools multiple times. At the very least, it is indeed not encouraging the adoption of perhaps more appropriate data analysis procedures.

Fourthly, and last, while there has been some effort in collecting different data analysis procedures (e.g. the dpcR package), these efforts have been incomprehensive and development has been slow.

A concerted and focused effort towards collecting the different data analysis procedures and making them available as a user-friendly package is likely to increase their adoption. An open data standard may simplify the development of such software and improve compatibility of the software with different digital PCR platforms.

8.3 Conclusion

While the digital PCR concept has been around for several decades, rigorous data analysis approaches for the design and analysis of digital PCR experiments were lacking until a few years ago. The commercial availability of an increasing number of digital PCR platforms has led to an increased adoption of digital PCR for the unbiased and precise quantification of nucleic acids. With this increasing use came awareness of several potential issues that may arise in the quantification of nucleic acids with digital PCR, and the usefulness of appropriate data analysis procedures to tackle these issues.

In this thesis we have described several methods for tackling some previously outstanding issues in digital PCR data analysis. Concurrently, other research groups have described other data analysis methods, aiming to tackle other or identical issues. These developments have led to better or more realistic estimates of nucleic acid quantities and the imprecision of these estimated quantities.

In addition, we have provided means to design digital PCR experiments, and for the post-hoc verification of the quality of an experiment. This allows for a more realistic view of the capabilities of current digital PCR platforms, the appropriate design of experiments, and the verification of the usefulness of digital PCR for the application under consideration.

Notwithstanding this growing interest in, and the development of many digital PCR data analysis procedures, several questions remain unanswered. Further addressing these questions will aid in further extending the application domain of digital PCR, and the refinement of the current digital PCR data analysis pipeline.

Despite these remaining challenges, the current state of the field is such that for several applications unprecedented trueness and precision of nucleic acid quantity measurements is obtained with state-of-the-art digital PCR platforms and data analysis.

Bibliography

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2008). Molecular Biology of the Cell, 5th ed. Garland Science, New York, NY.

Arvia, R., Sollai, M., Pierucci, F., Urso, C., Massi, D. and Zakrzewska, K. (2017). Droplet digital PCR (ddPCR) vs quantitative real-time PCR (qPCR) approach for detection and quantification of Merkel cell polyomavirus (MCPyV) DNA in formalin fixed paraffin embedded (FFPE) cutaneous biopsies. Journal of Virological Methods, 246, 15–20.

Attali, D., Bidshahri, R., Haynes, C. and Bryan, J. (2016). ddpcr: an R package and web application for analysis of droplet digital PCR data. F1000Research, 5.

Baker, M. (2012). Digital PCR hits its stride. Nature Methods, 9(6), 541-544.

Bartlett, J. M. S. and Stirling, D. (2003). A short history of the polymerase chain reaction. Methods in Molecular Biology, 226, 3–6.

Bassham, L. E., Rukhin, A. L., Soto, J., Nechvatal, J. R., Smid, M. E., Barker, E. B., ... and Vo, S. (2010). A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. National Institute of Standards and Technology, Gaithersburg, MD.

Bates, D., Machler, M., Bolker, B. and Walker, S. (2015). lme4: Linear Mixed-Effects Models Using Eigen and S4. R package version 1.1-10.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289-300.

Bhat, S., Herrmann, J., Armishaw, P., Corbisier, P. and Emslie, K. R. (2009). Single molecule detection in nanofluidic digital array enables accurate measurement of DNA copy number. Analytical and Bioanalytical Chemistry, 394, 457–67.

Bickel, D. R. and Frühwirth, R. (2006). On a fast, robust estimator of the mode: comparisons to other robust estimators with applications. Computational Statistics and Data Analysis, 50(12), 3500-3530.

Bizouarn, F. (2014). Introduction to digital PCR. Quantitative Real-Time PCR: Methods and Protocols, 27-41.

Bornmann, L. and Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science and Technology, 66(11), 2215-2222.

Bosman, K. J., Nijhuis, M., Van Ham, P. M., Wensing, A. M., Vervisch, K., Vandekerckhove, L. and De Spiegelaere, W. (2015). Comparison of digital PCR platforms and semi-nested qPCR as a tool to determine the size of the HIV reservoir. Scientific Reports, 5, 13811.

Brunetto, G. S., Massoud, R., Leibovitch, E. C., Caruso, B., Johnson, K., Ohayon, J., ... and Jacobson, S. (2014). Digital droplet PCR (ddPCR) for the precise quantification of human T-lymphotropic virus 1 proviral loads in peripheral blood and cerebrospinal fluid of HAM/TSP patients and identification of viral mutations. Journal of Neurovirology, 20(4), 341-351.

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., ... and Vandesompele, J. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry, 55(4), 611-622.

Bustin, S. A. (2014). The reproducibility of biomedical research: sleepers awake!. Biomolecular Detection and Quantification, 2, 35-42.

Bustin, S. A. (2017). The continuing problem of poor transparency of reporting and use of inappropriate methods for RT-qPCR. Biomolecular Detection and Quantification.

Buysse, K., Delle Chiaie, B., Van Coster, R., Loeys, B., De Paepe, A., Mortier, G., ... and Menten, B. (2009). Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. European Journal of Medical Genetics, 52(6), 398-403.

Buzón, M. J., Massanella, M., Llibre, J. M., Esteve, A., Dahl, V., Puertas, M. C., ... and Palmer, S. (2010). HIV-1 replication and immune dynamics are affected

by raltegravir intensification of HAART-suppressed subjects. Nature Medicine, 16(4), 460-465.

Carroll, R. J. and Cline, D. B. (1988). An asymptotic theory for weighted least-squares with weights estimated by replication. Biometrika, 75(1), 35-43.

Chiu, A., Ayub, M., Dive, C., Brady, G. and Miller, C. J. (2017). twoddpcr: an R/Bioconductor package and Shiny app for droplet digital PCR analysis. Bioinformatics, 33(17), 2743-2745.

Clementi, M. and Bagnarelli, P. (2015). Are three generations of quantitative molecular methods sufficient in medical virology?. New Microbiologica, 38(4), 437-441.

Clinical and Laboratory Standards Institute. (2003). EP06-A: Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach. Clinical and Laboratory Standards Institute, Wayne, PA.

Clinical and Laboratory Standards Institute. (2014). EP05-A3: Evaluation of Precision of Quantitative Measurement Procedures, 3rd Edition. Clinical and Laboratory Standards Institute, Wayne, PA.

Coles, S., Bawa, J., Trenner, L. and Dorazio, P. (2001). An introduction to statistical modeling of extreme values. Springer, London.

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. Royal Society Open Science, 1(3), 140216.

Cooper, G. M. (2000). Chapter 14: The cell cycle. In: The cell: a molecular approach. Boston University, Sunderland.

Coudray-Meunier, C., Fraisse, A., Martin-Latil, S., Guillier, L., Delannoy, S., Fach, P. and Perelle, S. (2015). A comparative study of digital RT-PCR and RT-qPCR for quantification of Hepatitis A virus and Norovirus in lettuce and water samples. International Journal of Food Microbiology, 201, 17-26.

Coudray-Meunier, C., Fraisse, A., Martin-Latil, S., Delannoy, S., Fach, P. and Perelle, S. (2016). A novel high-throughput method for molecular detection of human pathogenic viruses using a nanofluidic real-time PCR system. PLoS One, 11(1), e0147832.

Corbisier, P., Bhat, S., Partis, L., Xie, V. R. D., and Emslie, K. R. (2010). Absolute quantification of genetically modified MON810 maize (Zea mays L.) by digital polymerase chain reaction. Analytical and Bioanalytical Chemistry, 396(6), 2143-2150.

Corbisier, P., Pinheiro, L., Mazoua, S., Kortekaas, A. M., Chung, P. Y. J., Gerganova, T., ... and Emslie, K. (2015). DNA copy number concentration measured by digital and droplet digital quantitative PCR using certified reference materials. Analytical and Bioanalytical Chemistry, 407(7), 1831-1840.

Dagata, J. A., Farkas, N. and Kramar, J. A. (2016). Method for measuring the volume of nominally 100 m diameter spherical water-in-oil emulsion droplets. NIST Special Publication, 260.

Dangla, R., Kayi, S. C. and Baroud, C. N. (2013). Droplet microfluidics driven by gradients of confinement. Proceedings of the National Academy of Sciences, 110(3), 853-858.

Debski, P. R., Gewartowski, K., Sulima, M., Kaminski, T. S. and Garstecki, P. (2015). Rational design of digital assays. Analytical Chemistry, 87(16), 8203-8209.

De Spiegelaere, W., Erkens, T., De Craene, J., Burvenich, C., Peelman, L. and Van den Broeck, W. (2008). Elimination of amplification artifacts in real-time reverse transcription PCR using laser capture microdissected samples. Analytical Biochemistry, 382(1), 72-74.

De Spiegelaere, W., Malatinkova, E., Kiselinova, M., Bonczkowski, P., Verhofstede, C., Vogelaers, D. and Vandekerckhove, L. (2013). Touchdown digital polymerase chain reaction for quantification of highly conserved sequences in the HIV-1 genome. Analytical Biochemistry, 439(2), 201-203.

Devonshire, A. S., Honeyborne, I., Gutteridge, A., Whale, A. S., Nixon, G., Wilson, P., ... and Huggett, J. F. (2015). Highly reproducible absolute quantification of Mycobacterium tuberculosis complex by digital PCR. Analytical Chemistry, 87(7), 3706-3713.

D'haene, B., Vandesompele, J. and Hellemans, J. (2010). Accurate and objective copy number profiling using real-time quantitative PCR. Methods, 50(4), 262-270.

Dingle, T. C., Sedlak, R. H., Cook, L. and Jerome, K. R. (2013). Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. Clinical Chemistry, 59(11), 1670-1672.

Dobnik, D., Štebih, D., Blejec, A., Morisset, D., and Žel, J. (2016). Multiplex quantification of four DNA targets in one reaction with Bio-Rad droplet digital PCR system for GMO detection. Scientific Reports, 6, 35451.

Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H. and Minamoto, T. (2015). Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PLoS One, 10(3), e0122763.

Dong, L., Meng, Y., Sui, Z., Wang, J., Wu, L. and Fu, B. (2015). Comparison of four digital PCR platforms for accurate quantification of DNA copy number of a certified plasmid DNA reference material. Scientific Reports, 5, 13174.

Dong, L., Meng, Y., Wang, J. and Liu, Y. (2014). Evaluation of droplet digital PCR for characterizing plasmid reference material used for quantifying ammonia oxidizers and denitrifiers. Analytical and Bioanalytical Chemistry, 406(6), 1701–1712.

Dorazio, R. M. and Hunter, M. E. (2015). Statistical models for the analysis and design of digital polymerase chain reaction (dPCR) experiments. Analytical Chemistry, 87(21), 10886-10893.

Dreo, T., Pirc, M., Ramšak, Ž., Pavšič, J., Milavec, M., Žel, J. and Gruden, K. (2014). Optimising droplet digital PCR analysis approaches for detection and quantification of bacteria: a case study of fire blight and potato brown rot. Analytical and Bioanalytical Chemistry, 406(26), 6513-6528.

Dube, S., Qin, J. and Ramakrishnan, R. (2008). Mathematical analysis of copy number variation in a DNA sample using digital PCR on a nanofluidic device. PLoS One, 3(8), e2876.

Dunlop, D. D. (1994). Regression for longitudinal data: a bridge from least squares regression. The American Statistician, 48(4), 299-303.

El Khattabi, L. A., Rouillac-Le Sciellour, C., Le Tessier, D., Luscan, A., Coustier, A., Porcher, R., ... and Mandelbrot, L. (2016). Could digital PCR be an alternative as a non-invasive prenatal test for trisomy 21: a proof of concept study. PLoS One, 11(5), e0155009.

Environmental Protection Agency (EPA), FEM Microbiology Action Team. (2009). Method Validation of U.S. Environmental Protection Agency Microbiological Methods of Analysis.

Evans, M. I., Wright, D. A., Pergament, E., Cuckle, H. S. and Nicolaides, K. H. (2012). Digital PCR for noninvasive detection of aneuploidy: power analysis equations for feasibility. Fetal Diagnosis and Therapy, 31(4), 244-247.

Fischer, M., Joos, B., Hirschel, B., Bleiber, G., Weber, R., Günthard, H. F. and Swiss HIV Cohort Study. (2004). Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding nef. The Journal of Infectious Diseases, 190(11), 1979-1988.

Food and Agricultural Organization. (2010). Guidelines on Performance Criteria and Validation of Methods for Detection, Identification and Quantification of Specific DNA Sequences and Specific Proteins in Foods CAC/GL 74-2010.

Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER). (2001). Guidance for Industry: Bioanalytical Method Evaluation.

Gerdes, L., Iwobi, A., Busch, U. and Pecoraro, S. (2016). Optimization of digital droplet polymerase chain reaction for quantification of genetically modified organisms. Biomolecular Detection and Quantification, 7, 9-20.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Full Conditional Distributions. Markov Chain Monte Carlo in Practice, Chapter 5, pp. 75–88.

Harwood, V. J., and Stoeckel, D. M. Chapter 2: Performance Criteria. In: Hagedorn, Charles, Anicet R. Blanch, and Valerie J. Harwood, eds. Microbial source tracking: methods, applications, and case studies. Springer Science and Business Media; 2011.

Hayden, R. T., Gu, Z., Ingersoll, J., Abdul-Ali, D., Shi, L., Pounds, S. and Caliendo, A. M. (2013). Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. Journal of Clinical Microbiology, 51(2), 540-546.

Hayden, R. T., Gu, Z., Sam, S. S., Sun, Y., Tang, L., Pounds, S. and Caliendo, A. M. (2015). Comparative evaluation of three commercial quantitative cytomegalovirus standards by use of digital and real-time PCR. Journal of Clinical Microbiology, 53(5), 1500-1505.

Haynes, R. J., Kline, M. C., Toman, B., Scott, C., Wallace, P., Butler, J. M. and Holden, M. J. (2013). Standard reference material 2366 for measurement of human cytomegalovirus DNA. The Journal of Molecular Diagnostics, 15(2), 177-185.

Henrich, T. J., Gallien, S., Li, J. Z., Pereyra, F. and Kuritzkes, D. R. (2012). Low-level detection and quantitation of cellular HIV-1 DNA and 2-LTR circles using droplet digital PCR. Journal of Virological Methods, 186(1), 68-72.

Higuchi, R., Dollinger, G., Walsh, P. S. and Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. Nature Biotechnology, 10(4), 413-417.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., ... and Tewari, M. (2013). Absolute quantification by droplet digital PCR versus analog real-time PCR. Nature Methods, 10(10), 1003-1005.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., ... and Kitano, T. K. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical Chemistry, 83(22), 8604-8610.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 65-70.

Hoshino, T. and Inagaki, F. (2012). Molecular quantification of environmental DNA using microfluidics and digital PCR. Systematic and Applied Microbiology, 35(6), 390-395.

Hosmer, D. W. and Hjort, N. L. (2002). Goodness-of-fit processes for logistic regression: simulation results. Statistics in Medicine, 21(18), 2723-2738.

Huggett, J. F., Foy, C. A., Benes, V., Emslie, K., Garson, J. A., Haynes, R., ... and Pfaffl, M. W. (2013). The digital MIQE guidelines: minimum information for publication of quantitative digital PCR experiments. Clinical Chemistry, 59(6), 892-902.

Huggett, J. F., Cowen, S. and Foy, C. A. (2015). Considerations for digital PCR as an accurate molecular diagnostic tool. Clinical Chemistry, 61(1), 79-88.

Huggett, J. F., Garson, J. A. and Whale, A. S. (2016). Digital PCR and its potential application to microbiology. In Molecular Microbiology (pp. 49-57). American Society of Microbiology.

Jacobs, B. K., Goetghebeur, E. and Clement, L. (2014). Impact of variance components on reliability of absolute quantification using digital PCR. BMC Bioinformatics, 15(1), 283.

Jacobs, B. K., Goetghebeur, E., Vandesompele, J., De Ganck, A., Nijs, N., Beckers, A., ... and Clement, L. (2017). Model-Based Classification for Digital PCR: Your Umbrella for Rain. Analytical Chemistry, 89(8), 4461-4467.

Johnson, P. C., Barry, S. J., Ferguson, H. M. and Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. Methods in Ecology and Evolution, 6(2), 133-142.

Jones, G. M., Busby, E., Garson, J. A., Grant, P. R., Nastouli, E., Devonshire, A. S. and Whale, A. S. (2016). Digital PCR dynamic range is approaching that of real-time quantitative PCR. Biomolecular Detection and Quantification, 10, 31-33.

Jones, M., Williams, J., Gärtner, K., Phillips, R., Hurst, J. and Frater, J. (2014). Low copy target detection by Droplet Digital PCR through application of a novel open access bioinformatic pipeline,'definetherain'. Journal of Virological Methods, 202, 46-53.

Kain, M. P., Bolker, B. M. and McCoy, M. W. (2015). A practical guide and power analysis for GLMMs: detecting among treatment variation in random effects. PeerJ, 3, e1226.

Kiselinova, M., Pasternak, A. O., De Spiegelaere, W., Vogelaers, D., Berkhout, B. and Vandekerckhove, L. (2014). Comparison of droplet digital PCR and semi-nested real-time PCR for quantification of cell-associated HIV-1 RNA. PLoS One, 9(1), e85999.

Kiselinova, M., De Spiegelaere, W., Verhofstede, C., Callens, S. F. and Vandekerckhove, L. (2014). Antiretrovirals for HIV prevention: when should they be recommended?. Expert Review of Anti-infective Therapy, 12(4), 431-445.

Kiselinova, M., Geretti, A. M., Malatinkova, E., Vervisch, K., Beloukas, A., Messiaen, P., ... and De Spiegelaere, W. (2015). HIV-1 RNA and HIV-1 DNA persistence during suppressive ART with PI-based or nevirapine-based regimens. Journal of Antimicrobial Chemotherapy, 70(12), 3311-3316.

Kiselinova, M., De Spiegelaere, W., Buzon, M. J., Malatinkova, E., Lichterfeld, M. and Vandekerckhove, L. (2016). Integrated and total HIV-1 DNA predict ex vivo viral outgrowth. PLoS Pathogens, 12(3), e1005472.

Kishida, N., Noda, N., Haramoto, E., Kawaharasaki, M., Akiba, M. and Sekiguchi, Y. (2014). Quantitative detection of human enteric adenoviruses in river water by microfluidic digital polymerase chain reaction. Water Science and Technology, 70(3), 555-560.

Košir, A. B., Divieto, C., Pavšič, J., Pavarelli, S., Dobnik, D., Dreo, T., ... and Žel, J. (2017). Droplet volume variability as a critical factor for accuracy of absolute quantification using droplet digital PCR. Analytical and Bioanalytical Chemistry, 1-9.

Kreutz, J. E., Munson, T., Huynh, T., Shen, F., Du, W. and Ismagilov, R. F. (2011). Theoretical design and analysis of multivolume digital assays with wide dynamic range validated experimentally with microfluidic digital PCR. Analytical Chemistry, 83(21), 8158-8168.

Kutner, M. H., Nachtsheim, C, J., Neter, J. and Li, W. Diagnostics and Remedial Measures. In Applied Linear Statistical Models (5th ed.):McGraw-Hill, New York; 2005, pp 100–153.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology, 10(3), R25.

Lau, B. T., Wood-Bouwens, C. and Ji, H. P. (2017). Robust multiplexed clustering and denoising of digital PCR assays by data gridding. Analytical Chemistry, 89(22), 11913-11917.

Lefever, S., Hellemans, J., Pattyn, F., Przybylski, D. R., Taylor, C., Geurts, R., ... and RDML consortium. (2009). RDML: structured language and reporting guidelines for real-time quantitative PCR data. Nucleic Acids Research, 37(7), 2065-2069.

Li, N., Ma, J., Guarnera, M. A., Fang, H., Cai, L. and Jiang, F. (2014). Digital PCR quantification of miRNAs in sputum for diagnosis of lung cancer. Journal of Cancer Research and Clinical Oncology, 140(1), 145-150.

Lievens, A., Jacchia, S., Kagkli, D., Savini, C. and Querci, M. (2016). Measuring digital PCR quality: performance parameters and their optimization. PLoS One, 11(5), e0153317.

Liszewski, M. K., Jianqing, J. Y. and O'Doherty, U. (2009). Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR. Methods, 47(4), 254-260.

Lizée, G., Aerts, J. L., Gonzales, M. I., Chinnasamy, N., Morgan, R. A. and Topalian, S. L. (2003). Real-time quantitative reverse transcriptase-polymerase chain reaction as a method for determining lentiviral vector titers and measuring transgene expression. Human Gene Therapy, 14(6), 497-507.

Lo, Y. D. and Chan, K. A. (2006). Introduction to the polymerase chain reaction. Clinical Applications of PCR, 1-10.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician, 54(3), 217-224.

Lun, F. M., Chiu, R. W., Chan, K. A., Leung, T. Y., Lau, T. K. and Lo, Y. D. (2008). Microfluidics digital PCR reveals a higher than expected fraction of fetal DNA in maternal plasma. Clinical Chemistry, 54(10), 1664-1672.

McKay, A. T. (1932). Distribution of the Coefficient of Variation and the Extended t Distribution. Journal of the Royal Statistical Society, 95(4), 695-698.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305-325.

Madic, J., Zocevic, A., Senlis, V., Fradet, E., Andre, B., Muller, S., Dangla, R. and Droniou, M. E. (2016). Three-color crystal digital PCR. Biomolecular Detection and Quantification, 10, 34-36.

Majumdar, N., Wessel, T. and Marks, J. (2015). Digital PCR modeling for maximal sensitivity, dynamic range and measurement precision. PLoS One, 10(3), e0118833.

Malatinkova, E., Kiselinova, M., Bonczkowski, P., Trypsteen, W., Messiaen, P., Vermeire, J., ... and De Spiegelaere, W. (2015). Accurate quantification of episomal HIV-1 two-long terminal repeat circles by use of optimized DNA isolation and droplet digital PCR. Journal of Clinical Microbiology, 53(2), 699-701.

Malec, P. and Schienle, M. (2014). Nonparametric kernel density estimation near the boundary. Computational Statistics and Data Analysis, 72, 57-76.

Mattiuzzo, G., Ashall, J., Doris, K. S., MacLellan-Gibson, K., Nicolson, C., Wilkinson, D. E., ... and Minor, P. D. (2015). Development of lentivirus-based reference materials for Ebola virus nucleic acid amplification technology-based assays. PLoS One, 10(11), e0142751.

Menten, B., Maas, N., Thienpont, B., Buysse, K., Vandesompele, J., Melotte, C., ... and Janssens, S. (2006). Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. Journal of Medical Genetics, 43(8), 625-633.

Menten, B., Swerts, K., Delle Chiaie, B., Janssens, S., Buysse, K., Philippé, J. and Speleman, F. (2009). Array comparative genomic hybridization and flow cytometry analysis of spontaneous abortions and mors in utero samples. BMC Medical Genetics, 10(1), 89.

Morisset, D., Štebih, D., Milavec, M., Gruden, K. and Žel, J. (2013). Quantitative analysis of food and feed samples with droplet digital PCR. PLoS One, 8(5), e62583.

Morley, A. A. (2014). Digital PCR: a brief history. Biomolecular Detection and Quantification, 1(1), 1-2.

Mu, D., Yan, L., Tang, H. and Liao, Y. (2015). A sensitive and accurate quantification method for the detection of hepatitis B virus covalently closed circular DNA by the application of a droplet digital polymerase chain reaction amplification system. Biotechnology Letters, 37(10), 2063-2073.

Mukaide, M., Sugiyama, M., Korenaga, M., Murata, K., Kanto, T., Masaki, N. and Mizokami, M. (2014). High-throughput and sensitive next-generation droplet digital PCR assay for the quantitation of the hepatitis C virus mutation at core amino acid 70. Journal of Virological Methods, 207, 169-177.

Mullis, K., Faloona, F., Scharf, S., Saiki, R. K., Horn, G. T. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In Cold Spring Harbor symposia on quantitative biology (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.

Noble, W. S. (2009). How does multiple testing correction work?. Nature Biotechnology, 27(12), 1135.

Nixon, G., Garson, J. A., Grant, P., Nastouli, E., Foy, C. A. and Huggett, J. F. (2014). Comparative study of sensitivity, linearity, and resistance to inhibition of digital and nondigital polymerase chain reaction and loop mediated isothermal amplification assays for quantification of human cytomegalovirus. Analytical Chemistry, 86(9), 4387-4394.

Opel, K. L., Chung, D. and McCord, B. R. (2010). A study of PCR inhibition mechanisms using real time PCR. Journal of Forensic Sciences, 55(1), 25-33.

Ota, S., Ishikawa, S., Takazawa, Y., Goto, A., Fujii, T., Ohashi, K. I. and Fukayama, M. (2012). Quantitative analysis of viral load per haploid genome revealed the different biological features of Merkel cell polyomavirus infection in skin tumor. PLoS One, 7(6), e39954.

Palmer, S., Wiegand, A. P., Maldarelli, F., Bazmi, H., Mican, J. M., Polis, M., ... and Mellors, J. W. (2003). New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. Journal of Clinical Microbiology, 41(10), 4531-4536.

Pavšič, J., Devonshire, A. S., Parkes, H., Schimmel, H., Foy, C. A., Karczmarczyk, M., ... and Milavec, M. (2015). Standardization of nucleic acid tests for clinical measurements of bacteria and viruses. Journal of Clinical Microbiology, 53(7), 2008-2014.

Pavšič, J., Žel, J. and Milavec, M. (2016). Digital PCR for direct quantification of viruses without DNA extraction. Analytical and Bioanalytical Chemistry, 408(1), 67-75.

Pekin, D., Skhiri, Y., Baret, J. C., Le Corre, D., Mazutis, L., Salem, C. B., ... and Link, D. R. (2011). Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab on a Chip, 11(13), 2156-2166.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Research, 29(9), e45-e45.

Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S. and Emslie, K. R. (2011). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. Analytical Chemistry, 84(2), 1003-1011.

Rabinow, P. (2011). Making PCR: a story of biotechnology. University of Chicago Press.

Rački, N., Morisset, D., Gutierrez-Aguirre, I. and Ravnikar, M. (2014). One-step RT-droplet digital PCR: a breakthrough in the quantification of waterborne RNA viruses. Analytical and Bioanalytical Chemistry, 406(3), 661-667.

Rački, N., Dreo, T., Gutierrez-Aguirre, I., Blejec, A. and Ravnikar, M. (2014). Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. Plant Methods, 10(1), 42.

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing.

The RDML Consortium. http://www.rdml.org. Retrieved on 17 Nov 2017.

Robertson, T. and Cryer, J. D. (1974). An iterative procedure for estimating the mode. Journal of the American Statistical Association, 69(348), 1012-1016.

Rödiger, S., Burdukiewicz, M., Blagodatskikh, K. A. and Schierack, P. R. (2015). R as an Environment for the Reproducible Analysis of DNA Amplification Experiments. The R Journal, 7(2), 127-50.

Rousseeuw, P. J. and Croux, C. Alternatives to the median absolute deviation. Journal of the American Statistical Association, 88(424), 1273–1283.

Sanders, R., Huggett, J. F., Bushell, C. A., Cowen, S., Scott, D. J. and Foy, C. A. (2011). Evaluation of digital PCR for absolute DNA quantification. Analytical Chemistry, 83(17), 6474-6484.

Sanders, R., Mason, D. J., Foy, C. A. and Huggett, J. F. (2013). Evaluation of digital PCR for absolute RNA quantification. PLoS One, 8(9), e75296.

Sedlak, R. H. and Jerome, K. R. (2013). Viral diagnostics in the era of digital polymerase chain reaction. Diagnostic Microbiology and Infectious Disease, 75(1), 1-4.

Hall Sedlak, R. and Jerome, K. R. (2014). The potential advantages of digital PCR for clinical virology diagnostics. Expert Review of Molecular Diagnostics, 14(4), 501-507.

Sedlak, R. H., Cook, L., Huang, M. L., Magaret, A., Zerr, D. M., Boeckh, M. and Jerome, K. R. (2014). Identification of chromosomally integrated human herpesvirus 6 by droplet digital PCR. Clinical Chemistry, 60(5), 765-772.

Sedlak, R. H., Kuypers, J. and Jerome, K. R. (2014). A multiplexed droplet digital PCR assay performs better than qPCR on inhibition prone samples. Diagnostic Microbiology and Infectious Disease, 80(4), 285-286.

Sedlak, R. H., Cook, L., Cheng, A., Magaret, A. and Jerome, K. R. (2014). Clinical utility of droplet digital PCR for human cytomegalovirus. Journal of Clinical Microbiology, 52(8), 2844-2848.

Simmonds, P., Balfe, P. E. T. E. R., Peutherer, J. F., Ludlam, C. A., Bishop, J. O. and Brown, A. J. (1990). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. Journal of Virology, 64(2), 864-872.

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A. and Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. Nature, 550(7676), 345-353.

Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bulletin Mathematique de l'Université de Moscou, 2(2).

Spiess, A. N., Rödiger, S., Burdukiewicz, M., Volksdorf, T. and Tellinghuisen, J. (2016). System-specific periodicity in quantitative real-time polymerase chain reaction data questions threshold-based quantitation. Scientific Reports, 6, 38951.

Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. The American Statistician, 56(1), 29-38.

Stephenson, A. G. (2002). evd: Extreme value distributions. R News, 2(2), 31-32.

Strain, M. C., Lada, S. M., Luong, T., Rought, S. E., Gianella, S., Terry, V. H., ... and Richman, D. D. (2013). Highly precise measurement of HIV DNA by droplet digital PCR. PLoS One, 8(4), e55943.

Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., and Friess, H. (2009). Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. Nature Protocols, 4(1), 37-43.

Stroup, W. W. Generalized linear mixed models: modern concepts, methods and applications. CRC press, Boca Raton, Florida, 2012.

Suslov, O. and Steindler, D. A. (2005). PCR inhibition by reverse transcriptase leads to an overestimation of amplification efficiency. Nucleic Acids Research, 33(20), e181-e181.

Tang, L., Sun, Y., Buelow, D., Gu, Z., Caliendo, A. M., Pounds, S. and Hayden, R. T. (2016). Quantitative assessment of commutability for clinical viral load testing using a digital PCR-based reference standard. Journal of Clinical Microbiology, 54(6), 1616-1623.

Taylor, S. C., Carbonneau, J., Shelton, D. N. and Boivin, G. (2015). Optimization of Droplet Digital PCR from RNA and DNA extracts with direct comparison to RT-qPCR: Clinical implications for quantification of Oseltamivir-resistant subpopulations. Journal of Virological Methods, 224, 58-66.

Te, S. H., Chen, E. Y. and Gin, K. Y. (2015). Comparison of quantitative PCR and droplet digital PCR multiplex assays for two genera of bloom-forming cyanobacteria, Cylindrospermopsis and Microcystis. Applied Environmental Microbiology, 81(15), 5203–5211.

Tellinghuisen, J. (2016). Partition volume variability in vigital polymerase chain reaction methods: polydispersity causes bias but can improve precision. Analytical Chemistry, 88(24), 12183–12187.

Trypsteen, W., De Neve, J., Bosman, K., Nijhuis, M., Thas, O., Vandekerckhove, L. and De Spiegelaere, W. (2015). Robust regression methods for real-time polymerase chain reaction. Analytical Biochemistry, 480, 34-36.

Trypsteen, W., Vynck, M., De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., ... and De Spiegelaere, W. (2015). ddpcRquant: threshold determination for single channel droplet digital PCR experiments. Analytical and Bioanalytical Chemistry, 407(19), 5827-5834.

Tzur, A., Moore, J. K., Jorgensen, P., Shapiro, H. M. and Kirschner, M. W. (2011). Optimizing optical flow cytometry for cell volume-based sorting and analysis. PLoS One, 6(1), e16053.

van der Sluis, R. M., van Montfort, T., Centlivre, M., Schopman, N. C., Cornelissen, M., Sanders, R. W., ... and Pollakis, G. (2013). Quantitation of HIV-1 DNA with a sensitive TaqMan assay that has broad subtype specificity. Journal of Virological Methods, 187(1), 94-102.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology, 3(7), research0034-1.

Verhaegen, B., De Reu, K., De Zutter, L., Verstraete, K., Heyndrickx, M. and Van Coillie, E. (2016). Comparison of Droplet Digital PCR and qPCR for the Quantification of Shiga Toxin-Producing Escherichia coli in Bovine Feces. Toxins, 8(5), 157.

Versluis, M., de Lange, M. J., van Pelt, S. I., Ruivenkamp, C. A., Kroes, W. G., Cao, J., ... and van der Velden, P. A. (2015). Digital PCR validates 8q dosage as prognostic tool in uveal melanoma. PLoS One, 10(3), e0116371.

Vogelstein, B. and Kinzler, K. W. (1999). Digital PCR. Proceedings of the National Academy of Sciences, 96(16), 9236-9241.

Vynck, M., Vandesompele, J., Nijs, N., Menten, B., De Ganck, A. and Thas, O. (2016). Flexible analysis of digital PCR experiments using generalized linear mixed models. Biomolecular Detection and Quantification, 9, 1-13.

Vynck, M., Trypsteen, W., Thas, O., Vandekerckhove, L. and De Spiegelaere, W. (2016). The future of digital polymerase chain reaction in virology. Molecular Diagnosis and Therapy, 20(5), 437-447.

Vynck, M., Vandesompele, J. and Thas, O. (2017). Quality control of digital PCR assays and platforms. Analytical and Bioanalytical Chemistry, 409(25), 5919-5931.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. The Annals of Mathematical Statistics, 11(2), 147-162.

Wang, J., Ramakrishnan, R., Tang, Z., Fan, W., Kluge, A., Dowlati, A., ... and Ma, P. C. (2010). Quantifying EGFR alterations in the lung cancer genome with nanofluidic digital PCR arrays. Clinical Chemistry, 56(4), 623-632.

Whale, A. S., Huggett, J. F., Cowen, S., Speirs, V., Shaw, J., Ellison, S., ... and Scott, D. J. (2012). Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation. Nucleic Acids Research, 40(11), e82-e82.

Whale, A. S., Bushell, C. A., Grant, P. R., Cowen, S., Gutierrez-Aguirre, I., O'Sullivan, D. M., ... and Garson, J. A. (2016). Detection of rare drug resistance mutations by digital PCR in a human influenza A virus model system and clinical samples. Journal of Clinical Microbiology, 54(2), 392-400.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: Journal of the Econometric Society, 817-838.

White, R. A., Quake, S. R. and Curr, K. (2012). Digital PCR provides absolute quantitation of viral load for an occult RNA virus. Journal of Virological Methods, 179(1), 45-50.

Yang, R., Paparini, A., Monis, P. and Ryan, U. (2014). Comparison of next-generation droplet digital PCR (ddPCR) with quantitative PCR (qPCR) for enumeration of Cryptosporidium oocysts in faecal samples. International Journal for Parasitology, 44(14), 1105-1113.

Yu, M., Carter, K. T., Makar, K. W., Vickers, K., Ulrich, C. M., Schoen, R. E., ... and Grady, W. M. (2015). MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics, 10(9), 803-809.

Zmienko, A., Samelak-Czajka, A., Goralski, M., Sobieszczuk-Nowicka, E., Kozlowski, P. and Figlerowicz, M. (2015). Selection of reference genes for qPCR- and ddPCR-based analyses of gene expression in senescing barley leaves. PLoS One, 10(2), e0118226.

A Introduction

A.1 Maximum likelihood estimation for digital PCR

In a digital PCR setup, the number of target copies per partition Y_i is assumed to follow a Poisson distribution with mean λ. The outcome of the experiment is a digital signal, i.e. a partition is said to be either negative or positive. The probability of observing a negative partition is

$$P(Y_i = 0) = \exp(-\lambda), \tag{A.1}$$

and the probabibility of observing a positive partition

$$P(Y_i = 1) = 1 - \exp(-\lambda). \tag{A.2}$$

Given r the number of negative partitions and n the total number of partitions, the likelihood is given by

$$L(\lambda) = \exp(-\lambda)^r (1 - \exp(-\lambda))^{n-r}. \tag{A.3}$$

The loglikelihood is

$$l(\lambda) = -r\lambda + (n-r)\log(1 - \exp(-\lambda)). \tag{A.4}$$

The derivative of the loglikelihood is

$$\frac{\partial l(\lambda)}{\partial \lambda} = -r + (n-r)\frac{\exp(-\lambda)}{(1 - \exp(-\lambda))}, \tag{A.5}$$

and the maximum likelihood estimator

$$\hat{\lambda} = -\log\left(\frac{\mathrm{r}}{\mathrm{n}}\right). \tag{A.6}$$

or, with k the number of positive partitions

$$\hat{\lambda} = -\log\left(1 - \frac{\mathrm{k}}{\mathrm{n}}\right), \tag{A.7}$$

i.e. corresponding to equation 1.4.

The variance of $\hat{\lambda}$ is given by

$$\mathrm{var}(\hat{\lambda}) = I^{-1} = -E\left[\frac{\partial^2 l(\lambda)}{\partial^2 \lambda}\right]^{-1}, \tag{A.8}$$

with I the Fisher information matrix, and from

$$\begin{aligned}\frac{\partial^2 l(\lambda)}{\partial^2 \lambda} &= \frac{-(n-r)\exp(-\lambda)(1-\exp(-\lambda)) - (n-r)\exp(-\lambda)\exp(-\lambda)}{(1-\exp(-\lambda))^2} \\ &= \frac{-(n-r)\exp(-\lambda)}{(1-\exp(-\lambda))^2},\end{aligned} \tag{A.9}$$

and with

$$E[r] = n\exp(-\lambda), \tag{A.10}$$

the variance is

$$\mathrm{var}(\hat{\lambda}) = -E\left[\frac{-(n-r)\exp(-\lambda)}{(1-\exp(-\lambda))^2}\right]^{-1} = \left(\frac{n\exp(-\lambda)}{1-\exp(-\lambda)}\right)^{-1} = \frac{1-\exp(-\lambda)}{n\exp(-\lambda)}. \tag{A.11}$$

A plug-in estimator for the variance of $\hat{\lambda}$ is, consequently

$$\widehat{\mathrm{var}}(\hat{\lambda}) = \frac{1-\exp(-\hat{\lambda})}{n\exp(-\hat{\lambda})}. \tag{A.12}$$

B Power

B.1 Derivations

B.1.1 Absolute quantification

B.1.1.1 Estimating the absolute quantity and its variance

Let A denote the target nucleic acid, N_{neg} is the count of negative partitions, N the total number of partitions and V_p the partition volume.

The fraction of negative partitions is

$$\hat{p}_A = \frac{N_{neg,A}}{N}. \tag{B.1}$$

Under the normal approximation for a binomial proportion

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right), \tag{B.2}$$

so that the variance of the fraction of negative partitions is

$$\mathrm{Var}(\hat{p}_A) = \frac{\hat{p}_A(1-\hat{p}_A)}{N}. \tag{B.3}$$

The absolute quantity is given by

$$\hat{c}_A = \frac{-\log \hat{p}_A}{V_p}. \tag{B.4}$$

A first order Taylor approximation gives

$$\mathrm{Var}(-\log(p_A)) = \mathrm{Var}(\log(p_A)) = \frac{\mathrm{Var}(p_A)}{\mathrm{E}^2(p_A)}, \tag{B.5}$$

so that the variance of the absolute quantity for a single replicate is given by

$$\mathrm{Var}(c_A) = \frac{p_A(1-p_A)}{N{p_A}^2 V_p^2}. \tag{B.6}$$

This Taylor approximation approximates the true variance well (Figure B.1).

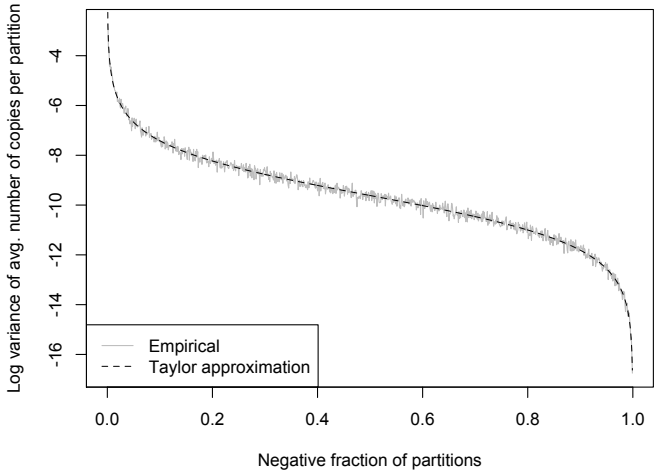


FIGURE B.1: Logarithm of the variance of the average number of copies per partition as a function of the negative fraction. "Empirical" reflects the estimated variance of the average number of copies per partition using 15000 partitions, and calculated from 100 simulated experiments for the given negative fraction. "Taylor approximation" is the variance as calculated for that negative fraction using Equation B.5.

B.1.1.2 Variance estimation

The law of total variance gives

$$\mathrm{Var}(C) = \mathrm{E}_R(\mathrm{Var}(C \mid R)) + \mathrm{Var}_R(\mathrm{E}(C \mid R)), \tag{B.7}$$

with C the absolute concentration and R a replicate-effect.

Thus, the total variance of the concentration $\mathrm{Var}(C)$ (further denoted by v_T) is the sum of two components: (i) $\mathrm{E}_R(\mathrm{Var}(C \mid R))$, which is the replicate-specific variance averaged over the replicates, and (ii) $\mathrm{Var}_R(\mathrm{E}(C \mid R))$, which is the variance of the replicate-specific expected concentrations. The first component originates from the Poisson variability, and is further denoted by v_P, while the second component quantifies the between-replicate variability, and is denoted by further denoted by v_B, i.e.

$$v_T = v_P + v_B. \tag{B.8}$$

An estimate of v_B is calculated as

$$v_B = v_T - v_P \tag{B.9}$$

with v_T the total variance as obtained from a preliminary experiment (i.e. the variance of the estimated absolute concentrations) and v_P as in (B.6).

For the power calculations, v_B is assumed to be a known constant and is required as an input parameter. When v_B is non-zero, this furthermore implies that the negative fraction of partitions is not constant but depends on the replicate. Relying on equation B.6, we estimate $\mathrm{E}_R(\mathrm{Var}(C \mid R))$ as

$$\mathrm{E}_R(\mathrm{Var}(C \mid R)) = \mathrm{E}_R\left(\frac{p_A(1-p_A)}{N{p_A}^2 V_p^2}\right) \propto \mathrm{E}_R\left(\frac{1-p_A}{p_A}\right) \approx \frac{1}{\mathrm{E}_R(p_A)} - 1, \tag{B.10}$$

i.e. using the first-order Taylor approximation

$$\mathrm{E}_R\left(\frac{1}{p_A}\right) \approx \frac{1}{\mathrm{E}_R(p_A)}. \tag{B.11}$$

$\mathrm{E}_R(p_A)$, i.e. the expected negative fraction, is required as an input parameter for the power calculations.

The first-order Taylor approximation in Equation B.11 is reasonable when the coefficient of variation of replicated reactions is not large (Figure B.2), which is usually the case in practice for digital PCR experiments: in the Lievens et al. data,

for example, the median coefficient of variation among replicate reactions for the fraction of negative partitions is 1%.

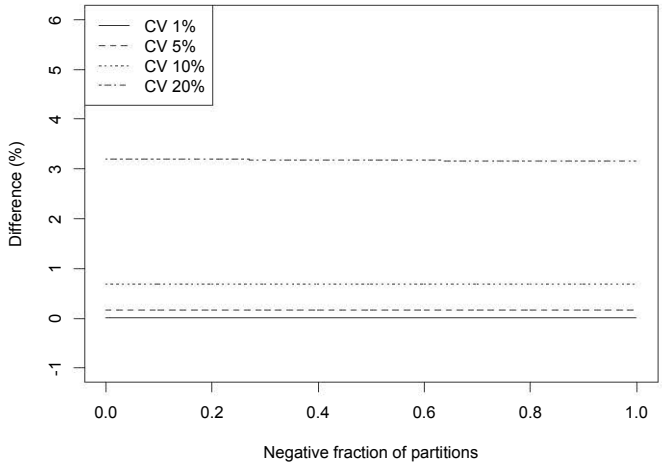


FIGURE B.2: Percentage difference in the expectation of the inverse when compared to the inverse of the expectation as a function of the negative fraction, with a normal distribution for the negative fractions in the replicate reactions and with given values of the coefficient of variation (CV). Results obtained through Monte Carlo simulation, calculating the expectation based on three replicate reactions, and averaged and smoothed over the 500 Monte Carlo simulations. The approximation works well for reasonably small values of the CV, but the approximation is less good when the CV is reasonably large.

Finally, the variance of the mean of absolute quantities $\hat{C}$ based on r replicates is given by

$$s^2 = \frac{v_P + v_B}{r}. \tag{B.12}$$

B.1.2 Copy number quantification

B.1.2.1 Copy number estimation

Let A denote the target nucleic acid and B the reference nucleic acid. For the reference, we know the expected number of copies N_B in the genome. N_{neg} is the count of negative partitions and N the total number of partitions.

$$\hat{p}_A = \frac{N_{neg,A}}{N} \tag{B.13}$$

$$\hat{p}_B = \hat{p_A}^{\frac{1}{\delta}} \tag{B.14}$$

Under the normal approximation

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \tag{B.15}$$

and

$$\mathrm{Var}(\hat{p}_A) = \frac{\hat{p}_A(1-\hat{p}_A)}{N} \tag{B.16}$$

$$\mathrm{Var}(\hat{p}_B) = \frac{\hat{p}_A^{\frac{1}{\delta}}(1-\hat{p}_A^{\frac{1}{\delta}})}{N} \tag{B.17}$$

The CN is then given by:

$$\hat{CN} = \frac{\hat{c}_A}{\hat{c}_B} N_B = \frac{\log \hat{p}_A}{\log \hat{p}_B} N_B \tag{B.18}$$

B.1.2.2 Total variance estimation

The variance of the CN estimator based on r replicates is given by

$$s^2 = \frac{v_P + v_B}{r} \tag{B.19}$$

where v_B is the between-replicate variance (see later) and v_P is the variance caused by the Poisson process (for a single replicate). The latter, a ratio of two random variables, say X and Y, is given by (delta method approximation):

$$\mathrm{Var}\left(\frac{X}{Y}\right) = \frac{1}{\mathrm{E}^2(Y)}\left(\mathrm{Var}(X) + \frac{\mathrm{E}^2(X)}{\mathrm{E}^2(Y)}\mathrm{Var}(Y)\right) \tag{B.20}$$

with

$$\begin{aligned} \mathrm{Var}(\log X) &= \frac{\mathrm{Var}(X)}{\mathrm{E}^2(X)}, \\ \mathrm{E}(\log X) &= \log \mathrm{E}(X) - \frac{\mathrm{Var}(X)}{2\mathrm{E}^2(X)}. \end{aligned} \tag{B.21}$$

In the copy number case this becomes:

$$\begin{aligned} \mathrm{Var}\left(N_B \frac{\log p_A}{\log p_B}\right) &= \frac{N_B^2}{\mathrm{E}^2(\log p_B)}\left(\mathrm{Var}(\log p_A) + \frac{\mathrm{E}^2(\log p_A)}{\mathrm{E}^2(\log p_B)}\mathrm{Var}(\log p_B)\right) \\ &= \frac{N_B^2}{\mathrm{E}^2(\log p_B)}\left(\frac{\mathrm{Var}(p_A)}{p_A^2} + \frac{\mathrm{E}^2(\log p_A)}{\mathrm{E}^2(\log p_B)}\frac{\mathrm{Var}(p_B)}{p_B^2}\right) \\ &= \frac{N_B^2}{\mathrm{E}^2(\log p_B)}\left(\frac{\mathrm{Var}(p_A)}{p_A^2} + \frac{\mathrm{E}^2(\log p_A)}{\frac{1}{\delta^2}\mathrm{E}^2(\log p_A)}\frac{\mathrm{Var}(p_B)}{p_B^2}\right) \\ &= \frac{N_B^2}{\mathrm{E}^2(\log p_B)}\left(\frac{\mathrm{Var}(p_A)}{p_A^2} + \frac{\delta^2\mathrm{Var}(p_B)}{p_B^2}\right) \\ &= \frac{N_B^2}{\mathrm{E}^2(\log p_A)}\left(\frac{\delta^2}{p_A^2}\mathrm{Var}(p_A) + \frac{\delta^4}{p_B^2}\mathrm{Var}(p_B)\right). \end{aligned} \tag{B.22}$$

and with

$$\mathrm{E}^2(\log p_A) = \left(\log p_A - \frac{\mathrm{Var}(p_A)}{2p_A^2}\right)^2 \approx \log {p_A}^2, \tag{B.23}$$

this yields an estimator of the copy number variance

$$v_P = \mathrm{Var}\left(N_B \frac{\log p_A}{\log p_B}\right) = (N_B \delta/(p_A * \log(p_A)))^2 * v_{PA} + (N_B \delta^2/(\log(p_A) * p_B))^2 * v_{PB}. \tag{B.24}$$

An estimate of v_B is calculated as

$$v_B = v_T - v_P \tag{B.25}$$

with v_T the total variance, as obtained from a preliminary experiment (i.e. the variance of the estimated copy numbers) and v_P as in (B.24).

B.1.3 Relative quantification

This corresponds to setting the parameter $N_B = 1$ in the formulae derived for the copy number scenario.

B.2 Supplementary figures

B.2.1 Small effect size

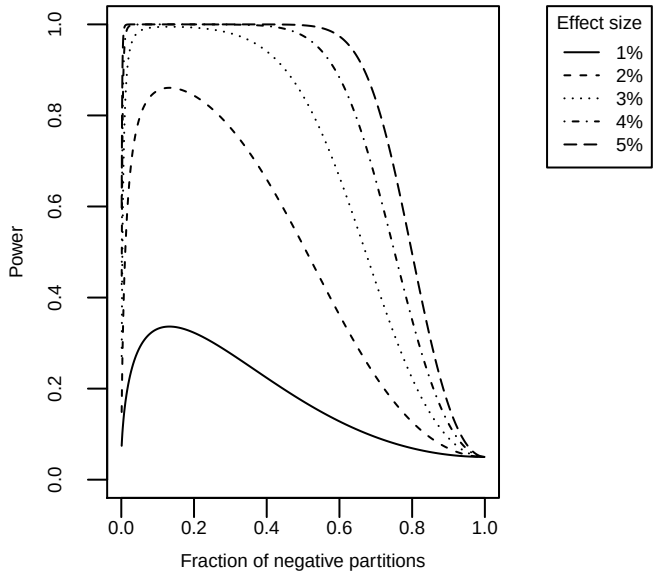


FIGURE B.3: Influence of the fraction of negative partitions for small effect sizes. Between-replicate variance set to 0.0001, number of replicates set to 3, number of partitions set to 15 000.

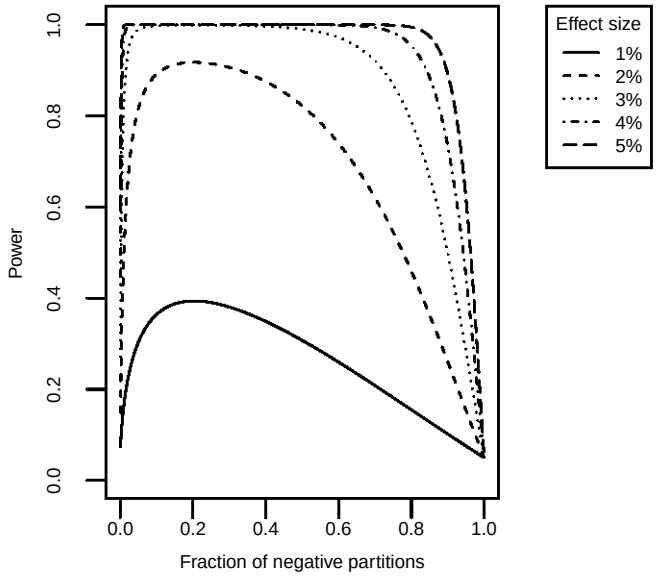


FIGURE B.4: Influence of the fraction of negative partitions for small effect sizes. Between-replicate variance set to 0, number of replicates set to 3, number of partitions set to 15 000.

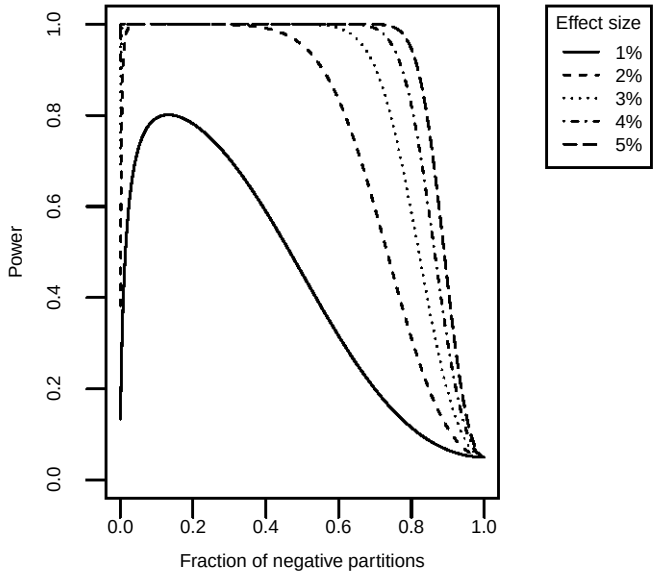


FIGURE B.5: Influence of the fraction of negative partitions for small effect sizes. Between-replicate variance set to 0.0001, number of replicates set to 10, number of partitions set to 15 000.

B.2.2 Relative contributions of variance sources

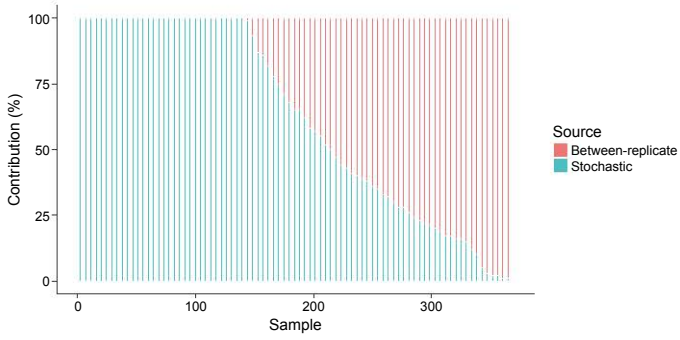


FIGURE B.6: Estimated contribution of the two sources of variance to the total variance for the absolute quantification experiments.

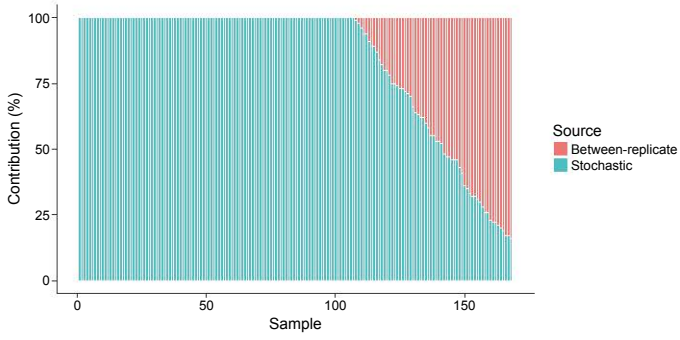


FIGURE B.7: Estimated contribution of the two sources of variance to the total variance for the relative quantification experiments.

B.3 Shiny web application

A Shiny web implementation of power calculation for the different scenarios is available at `http://statapps.ugent.be/dPCR/dPowerCalcR/`. This web application also allows to calculate the between-replicate variance based on a preliminary experiment. All R and R/Shiny code underlying this web application is available at `https://github.com/CenterForStatistics-UGent/dPowerCalcR`.

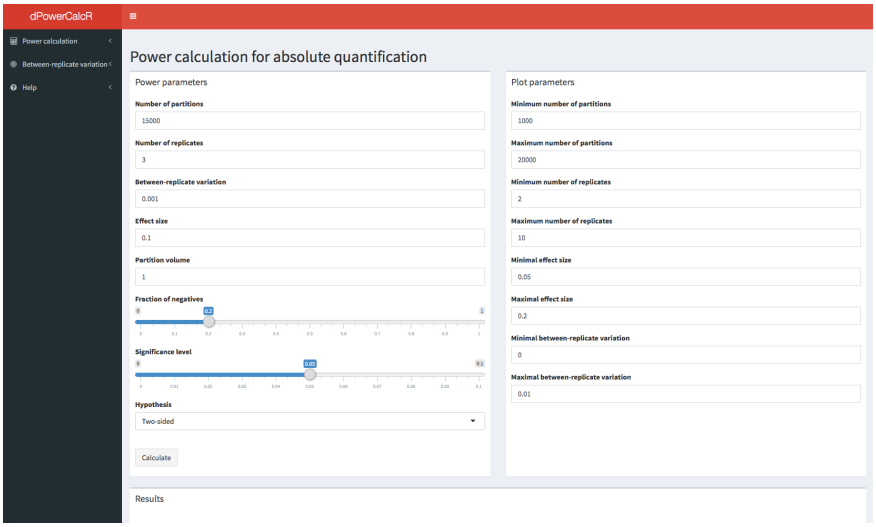


FIGURE B.8: Screenshot of the Shiny web application landing page.

C Thresholding

C.1 Primers, probes, amplicons and cycling conditions for the inhibition assay

The bold bases are mismatches inserted to inhibit probe binding on the template sequences.

PCR cycling was performed with the QX100 system (Bio-Rad, Pleasanton, CA) as previously described (Kiselinova et al., 2014a). Briefly, PCR mix consisted of 10 μl 2x ddPCRTM super mix for probes (Bio-Rad); 800 nM of the primers and 300 nM of the probes and a diluted amount of sequence specific oligo's. Droplets were generated using the QX100 droplet generator (Bio-Rad) cycling conditions included an initial denaturation at 95°C for 5 minutes and a subsequent 40 cycles of 15 seconds denaturation at 95°C and an annealing/elongation step for 60 seconds at 58°C. Subsequently, samples were run in the QX100 droplet reader (Bio-Rad) and data was analyzed using the QuantaSoft analysis software 1.3.2.0.

Name	Description	Sequence
F-LTR-gag(late RT)	forward	TGTGTGCCCGTCTGTTGTGT
R-LTR-gag(late RT)	reverse	GAGTCCTGCGTCGAGAGAGC
P-LTR-gag(late RT)	probe	CAGTGGCGCCCGAACAGGGA
LTR-gag oligo	match oligo	TGTGTGCCCGTCTGTTGTGTC AGTGGCGCCC GAACAGGGAG CTCTCTCGACGCAGGACTC
LTR-gag oligo probe G/A p18	mismatch oligo	TGTGTGCCCGTCTGTTGTGTC AGTGGCGCCCGAACAA**A**GAGCT CTCTCGACGCAGGACTC
LTR-gag oligo probe C/G p10	mismatch oligo	TGTGTGCCCGTCTGTTGTGTC AGTGGCGC**C**CGAACAGGGAGC TCTCTCGACGCAGGACTC

C.2 Analysis of ddPCR inhibition on fluorescence caused by primer/probe mismatches

A clear delineation of the population of positive droplets in a ddPCR reaction is often impeded by the occurrence of droplets with an intermediate fluorescence amplitude, referred to as (positive) rain. To show that (positive) rain can contain quantifiable information and is not always an artefact, a series of ddPCR reactions were performed with amplicons that contain mismatches in the probe region.

When sequence mismatches were present (Figure C.1B and Figure C.1C), an intermediate fluorescence was detected. Consequently, the exclusion of these positive rain droplets, as suggested by Strain et al. (2013) and Jones et al. (2014) can lead to experimental bias and an underestimation of the true quantity of DNA templates.

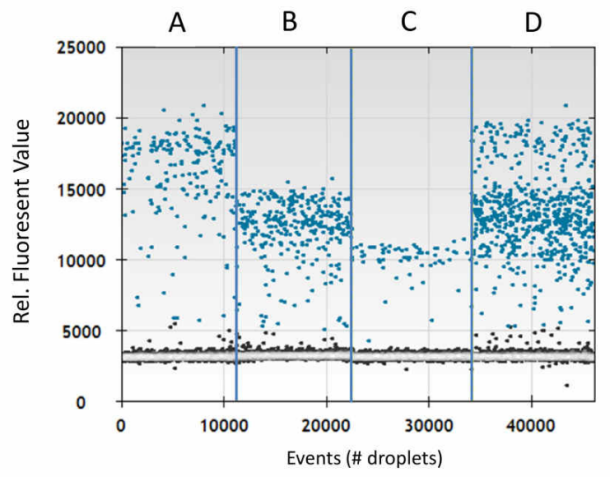


Figure C.1: QuantaSoft data view of ddPCR reactions with a template containing no mismatches (A) or one mismatch (B, C) or all templates together (D).

C.3 QuantaSoft dataview of fluorescent output with baseline shift

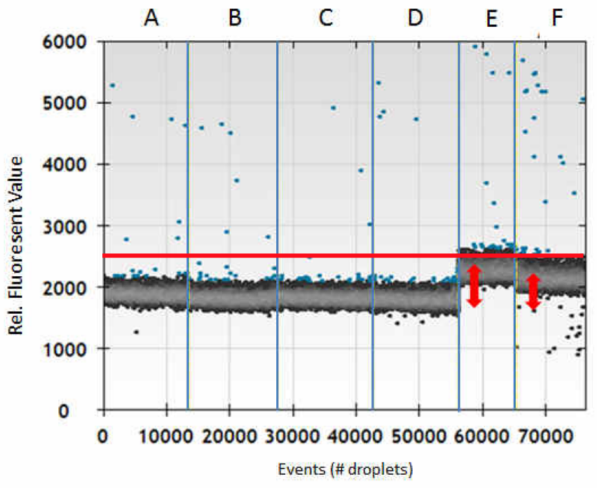


FIGURE C.2: Shift in baseline between samples is visible, especially for sample E and F (red arrows). When a threshold value is applied based on an NTC, for sample E and F inaccurate quantification can be the case (red line).

C.4 Implementation of extreme value theory

C.4.1 Glossary

For the researcher unfamiliar with extreme value theory, we provide an overview of keywords and how they relate to the ddPCR setting.

Block A block is a "subsample" of the original sample. If we for example have 14 000 droplets, we can divide this sample into 100 blocks containing 140 droplets each.

Block maximum The maximum of the observations in a given block. In this case the droplet within that block that has the highest fluorescence intensity.

Block size The block size is the number of observations (i.e. the number of droplets) within each block.

Generalized extreme value distribution (GEV) A family of distributions comprising all members to which the distribution of the block maxima can converge.

Fisher-Tippett theorem Theorem that proves that the distribution of block maxima is given by one of the distributions of the GEV distribution.

C.4.2 Summary

This document contains a discussion of the methodology used in the main manuscript on threshold determination for single channel droplet digital PCR experiments. We also demonstrate the methodology by means of an example. The document is divided into four parts:

1. General considerations for extreme value theory: we discuss how to select the block size and generalized extreme value distribution percentile. We further provide a resampling method to obtain a stable threshold estimate. These considerations will be applied in processing of the no template controls (NTCs).
2. NTC (pre)processing: we discuss a normalization strategy for the NTCs, and how to apply extreme value theory to obtain the threshold.
3. Sample processing: using the threshold determined from the NTCs, we describe how to set the threshold in samples containing both positive and negative droplets.
4. Example: we demonstrate the procedure on an example dataset.

C.4.3 General considerations for extreme value theory

Throughout this work we will use the "classical" extreme value (EV) theory (as termed in Coles et al. (2001)) based on the modeling of block maxima. Crucial in the use of this "classical" theory is the determination of the block size and, linked to the block size, the percentile selection which will eventually determine the threshold. We furthermore discuss the influence of creating random subsamples on the parameter and threshold estimates.

C.4.3.1 Block size

Classical theory relies on the maxima of blocks, on which a (generalized) extreme value distribution is fitted. Historically, often only these block maxima are available (e.g. yearly gauging station flow maxima), and the block size is fixed (i.e. one year in the given example). In the digital droplet case, we however have fluorescence intensities for all droplets and can thus freely choose the block size. From the resulting blocks, we will then determine maxima on which the generalized extreme value (GEV) distribution will be fitted.
Keeping in mind that the Fisher-Tippett theorem (the maxima of i.i.d. random variables can only converge to one of three extreme value distributions which are

all embedded in the GEV distribution) only holds asymptotically, the block size needs to be sufficiently large. Selecting the block size is however a bias-variance tradeoff. As block sizes become larger, the number of blocks will decrease, resulting in a more variable estimate of the model parameters. As block sizes become smaller, the variance will become smaller, though the estimates become more biased.

Since, to our knowledge, no research has been done in determining the optimal block size for "classical" EV models (as opposed to a whole body of research on this topic for threshold/k largest order statistics models) the block size is determined empircally. In order to select the block size, a simulation, based on empirical data, was performed: for the number of subsamples in the range of $[5, 2000]$ (corresponding to block sizes of as high as a couple of thousands to as low as below 10) we estimated the parameters of the GEV distribution (location, scale and shape) and some corresponding thresholds. The results are shown in Figure C.3. The figure demonstrates instability of parameter and thresholds estimates when the number of subsamples is particularly small (< 100), while these estimates stabilize for larger numbers of subsamples. Keeping in mind, however, that the Fisher-Tippett theorem only holds asymptotically, one does not want to end up with too large a number of subsamples (i.e. extremely small block size). For this reason, it is preferential to select a subsample corresponding to a block size as large as possible, though avoiding the instability of the extremely large block sizes. In this case, a number of blocks in the range $[100, 200]$ (corresponding to block sizes in the range of $[70, 140]$) seems appropriate.

Figure C.4 demonstrates the bias-variance tradeoff (more details on the construction of these graphs is given in Section C.4.3.3): if the number of blocks is small, the parameters and threshold estimates show more variability, while the estimates become more biased when the number of blocks becomes larger.

C.4.3.2 Threshold selection

The percentile chosen to be the threshold is directly linked to the number of blocks chosen. If the droplets of each of k samples are divided into m blocks, the $(1-\alpha)\%$ percentile would correspond to an expectation of one false positive droplet every $(mk\alpha)$ samples. For example, if the droplets of one sample are divided in 100 blocks, then the 99% percentile corresponds to an expectation of one droplet exceeding this threshold per sample. This interpretation is valid in the case only one no template control (NTC) was used to select the threshold. If,

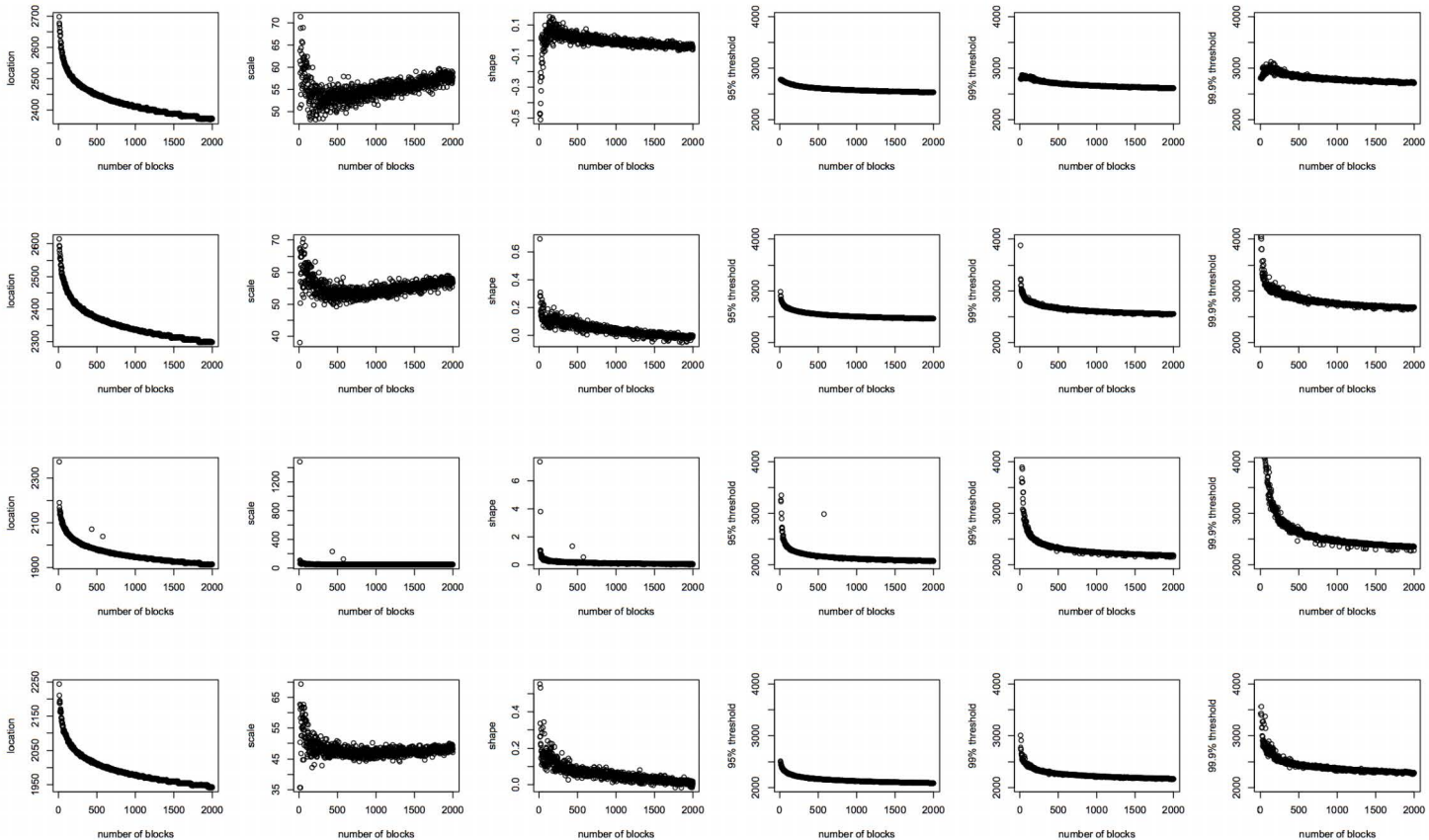


FIGURE C.3: Influence of the number of blocks/block size on parameter and threshold estimates. Each row represents a different NTC. As the number of blocks increases (moving from left to right on the individual plots), variability of the estimates decreases but the estimates becomes more biased. When choosing a very large block size (few blocks), the estimates display a very large variance.

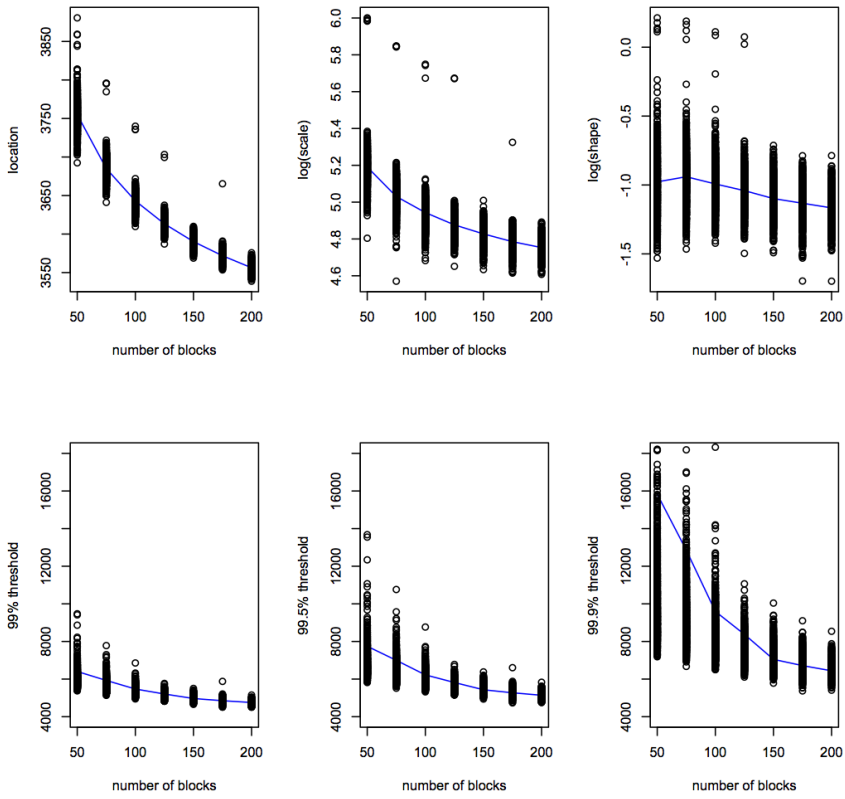


FIGURE C.4: Bias-variance tradeoff: a smaller number of blocks shows increased variability of the estimates compared to a larger number of blocks.

for example, 10 NTCs were pooled (after preprocessing, see Section C.4.4.1) and the number of blocks was chosen to be 100, the 99% percentile would corresponds to an expectation of one droplet exceeding the threshold every 10 samples. When selecting the number of blocks and given the number of no template controls, one should thus keep in mind to select an appropriate percentile for determining the threshold. One should furthermore keep in mind that there is always a tradeoff between false positives and false negatives: choosing too large of a percentile (consequently too large a threshold) may decrease the number false positives per sample, but it may well increase the number of false negatives.

C.4.3.3 Block creation

As mentioned in Section C.4.3, the block size is often fixed in the classical setting if e.g. only yearly historical data are available. In that case, the maxima are already determined for the given block size. In our case, the random assignment of the droplets to a given block may influence the parameter and threshold estimation. For example, suppose that there are two droplets with very high fluorescence intensities. If these, by chance, end up in the same block there will only be one very high maximum, as opposed to two very high maxima when they each are assigned to a different block. Given the nature of the problem, we can however do this sampling repeatedly. This will results in different parameter and threshold estimates, after which one can estimate a measure of central tendency (e.g. the mean of all estimates). This is demonstrated in Figure C.4: for a given number of blocks there is variation in the three GEV parameter estimates as well as in the threshold estimates.

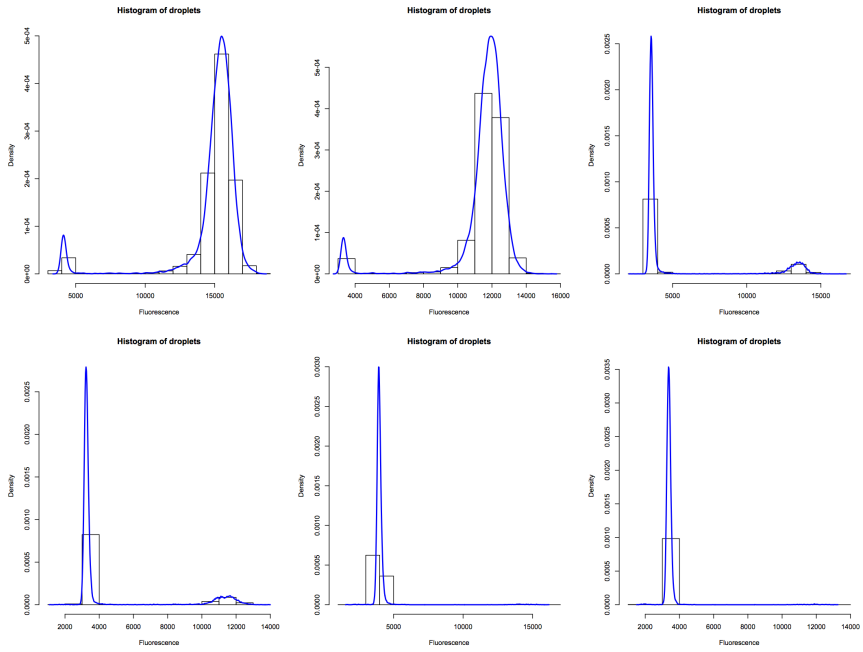


FIGURE C.5: Density estimates (blue) for samples from a dilution series. Depending on the initial concentration, the maximum density may be located in the fluorescence levels of the positive droplet droplet population (e.g. upper left panel, upper central panel), requiring that an initial threshold be set to obtain the maximum density of the fluorescence levels in the negative droplet population. Depending on the density bandwidth parameter, the density estimate may be quite peaked, which may result in an instable estimate of the mode (as can be seen in the positive droplet population of the upper right panel and the lower left panel).

C.4.4 NTC (pre)processing

C.4.4.1 Preprocessing

Prior to deriving a threshold using NTCs, the different NTCs need to be normalized to correct for NTC-to-NTC variation: different NTCs typically show slight shifts in baseline fluorescence. To normalize the NTCs one could subtract a measure of central tendency e.g. the median or the maximum density (which is a non-parametric estimator of the mode) of the fluorescence intensities. Using the density is problematic when determining the maximum density in samples where the number of negative droplets is relatively low and rain is quite prevalent, the maximum density may be shifted towards higher fluorescence intensities representing not only negative droplets, but also rain. This can be remediated by lowering the bandwidth of the density estimator, but choosing a bandwidth too small may result in very rough density estimates (see e.g. the blue jagged estimates in the upper right panel of Figure C.5), and thus a highly unstable estimate. This procedure furthermore requires the user to define a suitable bandwidth, possibly multiple times when analyzing a single run. This may in turn result in investigator-specific bias or unreproducibility. The median or robust mode does not require a parameter to be set. While the median may be more stable than the mode, a robust mode estimator has some benefits over the median: in samples containing negative and positive droplets, it will be quite independent of what droplets are selected to make up the distribution of the negative droplets, because the mode of a unimodal distribution does not depend on the range/support of the fluorescence (the rain). The median will however depend on the cutoff being set to select negative droplets, and especially when a large amount of rain (high outliers in the negative droplet population) is present it may be shifted towards higher fluorescence intensities. When applying the robust mode estimator suggested by Robertson and Cryer (1974) (half-sample mode) we expect to obtain a stable estimate that is not dependent on the (amount of) rain, a user-defined parameter, symmetry of the droplet fluorescence intensity distribution or the cutoff (Figure C.6). After normalizing the NTCs, they are pooled to represent a sample from the population of negative droplets.

C.4.4.2 Processing

Using the pooled NTCs, one can subsequently calculate the GEV parameter estimates as described previously. Depending on the number of NTCs the block size

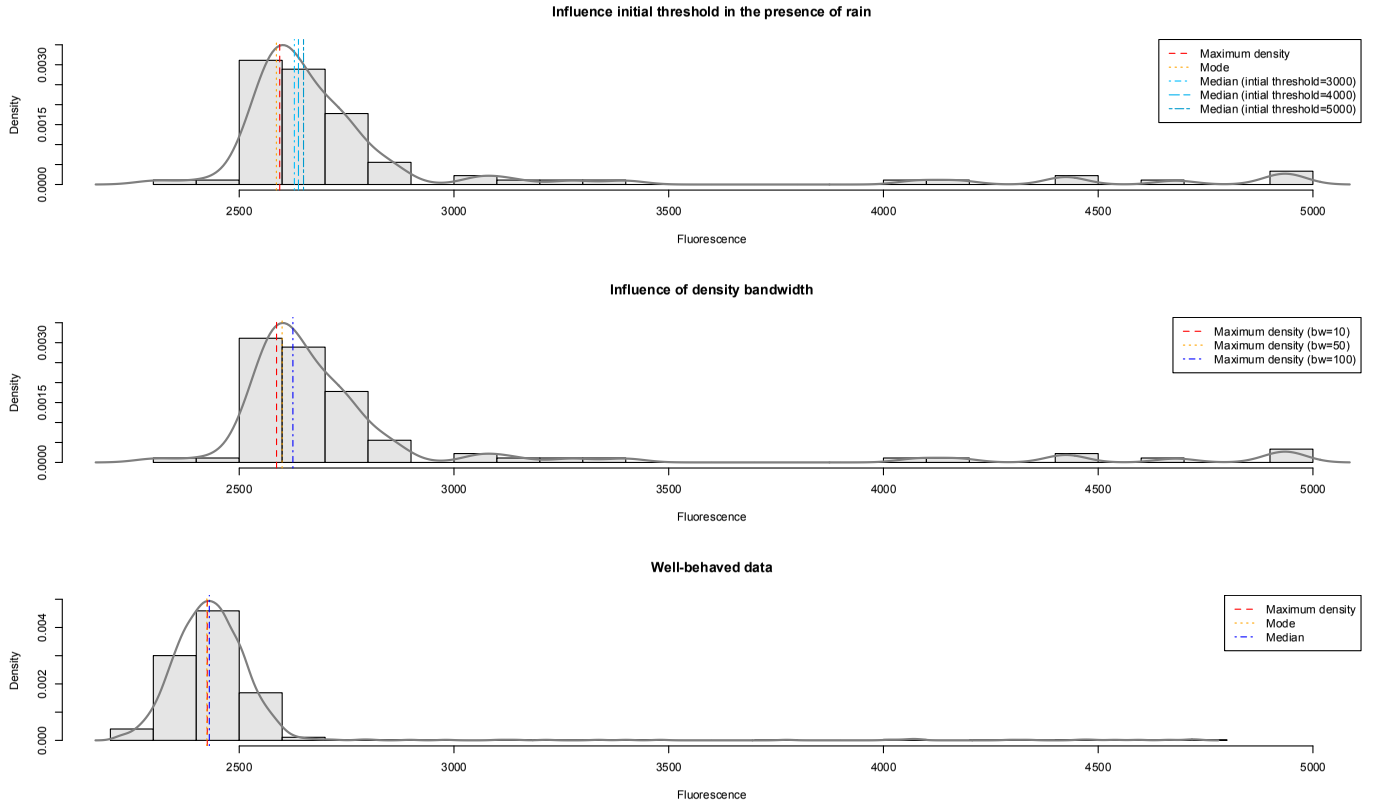


FIGURE C.6: Stability of central tendency measures in different scenarios. **Upper panel:** When a relatively large amount of rain is present, the median will be shifted towards a higher value. The more prevalent the rain (or the higher the initial threshold is set), the more right-shifted the estimate will become. **Middle panel:** In the presence of rain and/or a heavy right-tailed distribution, a higher bandwidth will result in the maximum density shifting right. Choosing a small bandwidth may result in multiple local maxima and thus an unstable estimate of the central tendency. **Lower panel:** When the data is well-behaved (symmetrical distribution, no or low amount of rain) the three measures coincide, no matter what was chosen to be the initial threshold or the bandwidth of the density estimator.

as well as an appropriate threshold has to be determined by the investigator. This will result in a single value for the threshold (relative to the normalized NTCs).

C.4.5 Sample processing

Since the fluoresence intensities in the samples also suffer from sample-to-sample variation, these need to be normalized as well before setting the threshold. This can again be done by calculating the Robertson-Cryer robust mode estimator for the negative droplets.
In order to do this, we first need to determine an initial cutoff discriminating negative and positive droplets: failing to set this initial cutoff may result in the Robertson-Cryer estimate being located in the positive droplets (in cases where positive droplets are more prevalent than negative droplets, e.g. the upper left panel in Figure C.5). When using the Robertson-Cryer estimator this initial cutoff can be determined visually by the investigator. Due to the robustness of the estimator, this will not influence the final result. Alternatively, using the average of Robertson-Cryer estimates of the NTCs and adding the calculated threshold value is a good starting point since major negative droplet fluorescence intensity shifts occur seldomly.
After setting the starting point, the Robertson-Cryer estimator can be calculated using the droplet fluorescence intensities below the starting point. The threshold is subsequently added to this estimate, resulting in the final threshold. Values below this threshold are negative droplets, values above the threshold are positive droplets.

C.4.6 Example

We will demonstrate the procedure described in the previous sections by means of an example. Parameters chosen in this example are for demonstrative purposes only and may not be realistic choices when analyzing experiments. The data we will be working with is from a real ddPCR run. For easiness of the example we will use three NTCs and one sample, but the extension to more NTCs or more samples is easy.

C.4.6.1 Step one: determine the robust mode and normalize the NTCs

Before we start building the model, we first normalize the data. For each of the three samples we determine the robust mode and subtract this robust mode from

all measured intensities. This is illustrated in Figures C.7 and C.8.

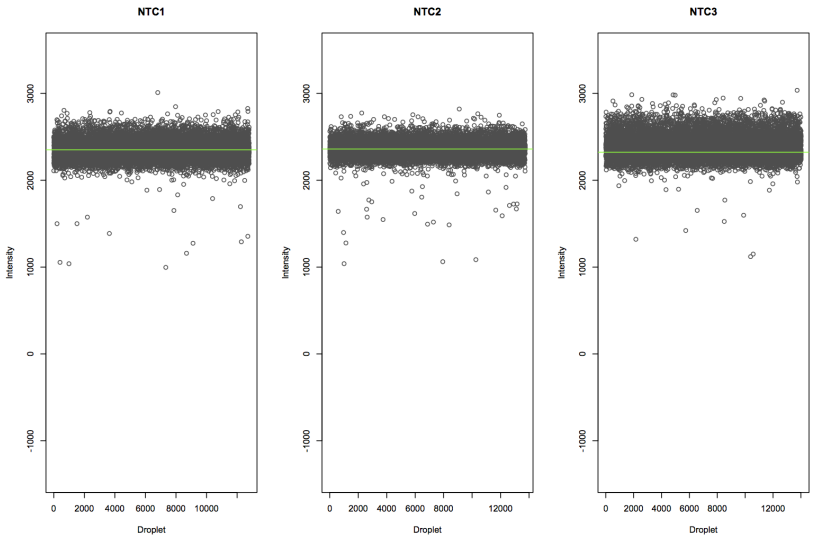


FIGURE C.7: Raw fluorescence intensities (gray) with robust mode (green).

C.4.6.2 Step two: merge the NTCs

Subsequently we merge the NTCs (Figure C.9).

C.4.6.3 Step three: fit the GEV on the block maxima and determine the threshold

On this merged NTC, we will apply the proposed strategy. We divide our sample into a certain amount of blocks, in this case 50 (we create 50 "subsamples" from the large sample, illustrated by the orange lines in Figure C.10). Within each of these blocks we then determine the maximum intensity (blue triangles in Figure C.10). These maxima will be used to model the threshold. For this we can rely on the convergence to a generalized extreme value (GEV) distribution.
The fitted GEV distribution is shown in Figure C.11: the black line is the density of the fitted GEV distribution, the blue triangles represent the 50 block maxima. We determine the threshold as a predetermined percentile of this fitted distribution, for example the 99% percentile in this case (indicated by the red vertical line).

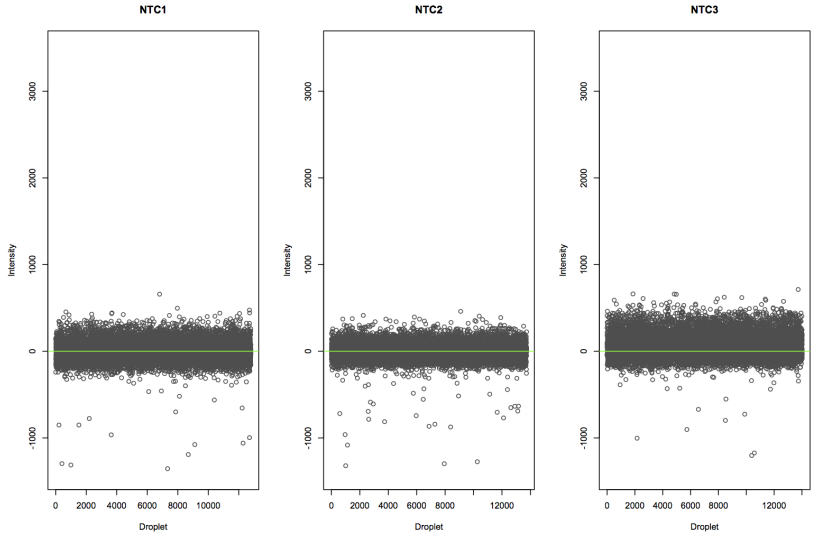


FIGURE C.8: Normalized fluorescence intensities (gray) with robust mode (green).

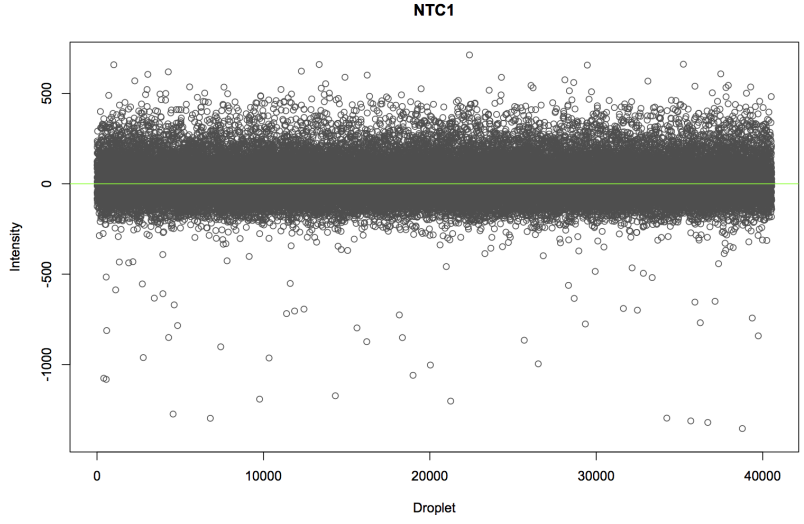


FIGURE C.9: Merged normalized fluorescence intensities (gray) with robust mode (green).

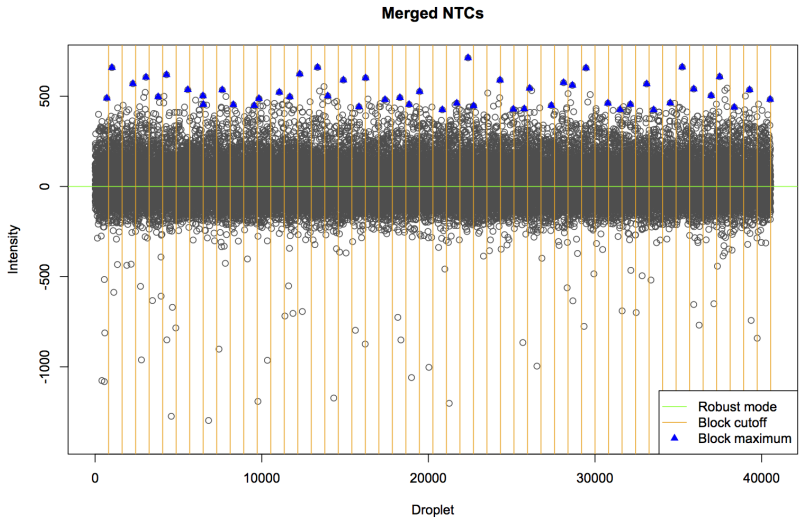


FIGURE C.10: Merged normalized fluorescence intensities (gray) with robust mode (green), cutoffs for the 50 blocks (orange) and block maxima (blue).

We have now obtained a threshold relative to our normalized NTCs. As described previously, this procedure is repeated many times, since sampling variability of the blocks may result in differences in estimates of the GEV distribution. After calculating many of these thresholds, we take the average of the obtained thresholds.

C.4.6.4 Step four: determine the robust mode of the sample and apply the threshold

We can then move on to our samples. We first need to determine the robust mode of the negative droplets in this sample. For this, we need to decide which droplets will be part of the negative and positive fraction. This can be done by averaging the robust modes of the three NTCs and adding the threshold. Everything below this value is then our negative subset. We can determine the robust mode of this subset (green line in Figure C.12). We then simply add the previously calculated threshold to this robust mode of the negative droplets of the sample, and obtain our final threshold (red line in Figure C.12). Everything below this line will be classified as negative droplets and everything above this line will be classified as positive droplets.

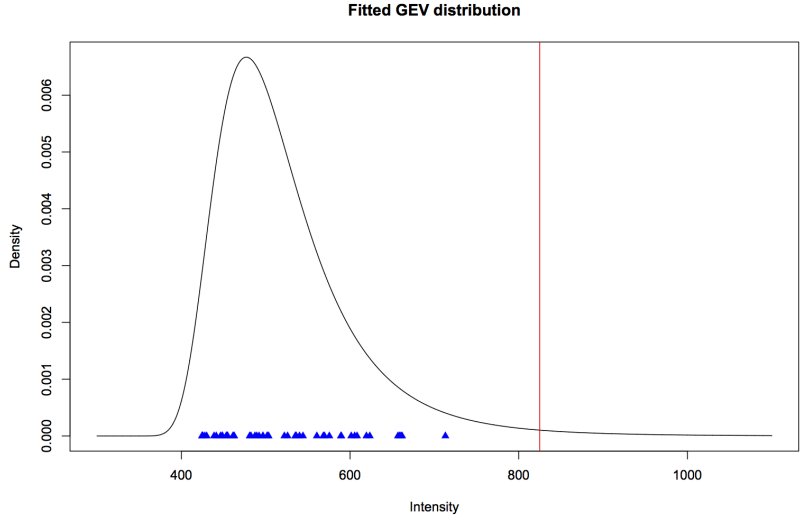


FIGURE C.11: Fitted GEV distribution (black), block maxima (blue) and 99% threshold (red).

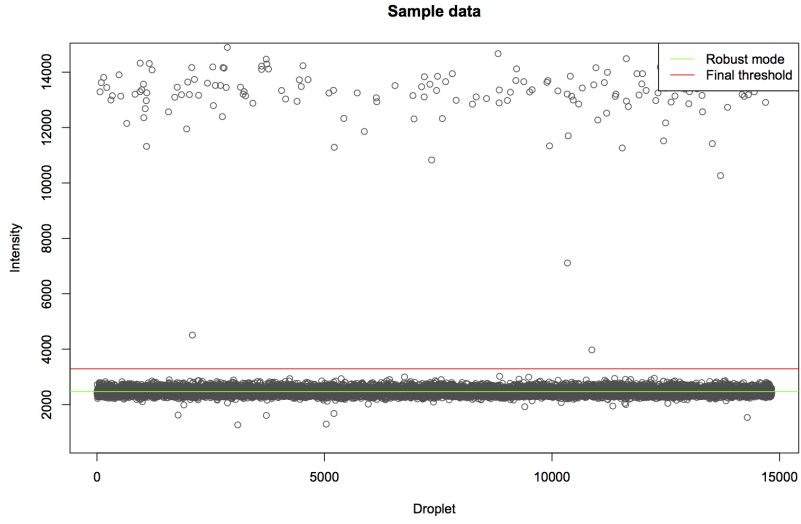


FIGURE C.12: Sample data: robust mode (green) and final threshold (red).

C.5 QuantaSoft export csv file

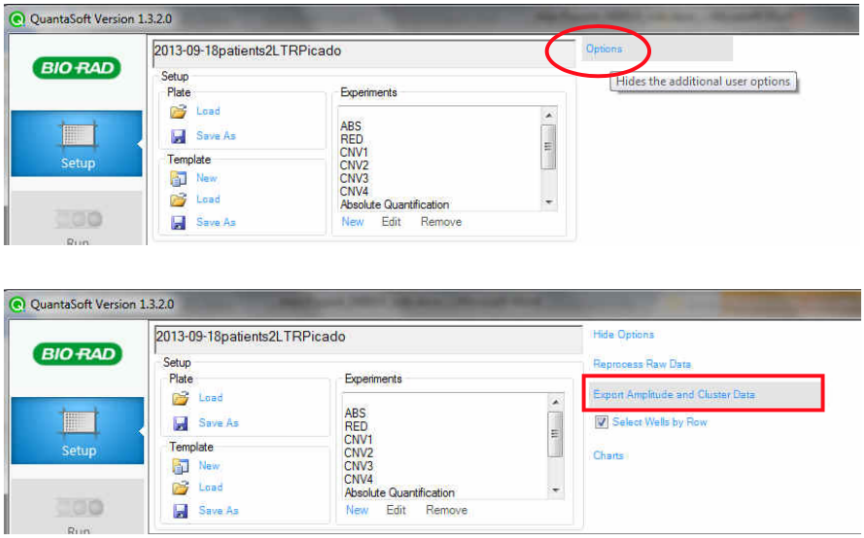


FIGURE C.13: Export of csv files form the Quantasoft software. Before starting the file export, make sure that the PC configurations for "decimal sign" is put to ".". In setup mode, select the wells of interest and click "options". Next, click "Export Amplitude and Cluster Data". Finally, select a ddPCR folder for storage of the csv files.

C.6 ddpcRquant annotation requirements before running a ddPCR experiment

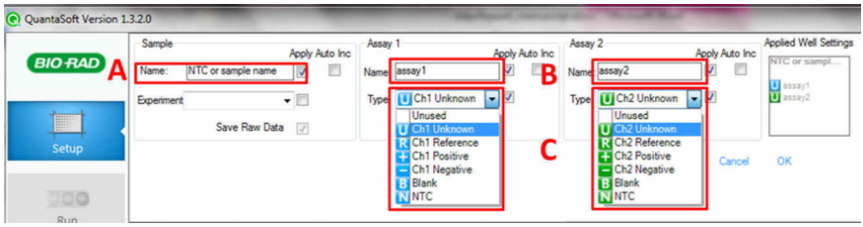


FIGURE C.14: Before performing a ddPCR run, the correct annotation requirements have to be followed to have a proper ddpcRquant analysis. (A) Name: Write your sample name or mention NTC or ntc. (B) Assay: Write your assay name. (C) Type: Always select "Ch1Unknown" or "Ch2Unkown" (don't select anything else).

C.7 QuantaSoft automated analysis methods vs. ddpcRquant

Information			Concentration			Positive and Negative Droplets					
Well	Sample	Assay	Single	Combined	ddcprQuant	Single		Combined		ddcprQuant	
H02	NTC DNA	total HIV DNA	No Call	1.99	NA	0	10496	19	10477	NA	NA
H04	NTC DNA	total HIV DNA	No Call	0	NA	0	12239	0	12239	NA	NA
H06	NTC DNA	total HIV DNA	No Call	0.0791	NA	0	13891	1	13890	NA	NA
B08	NTC DNA	total HIV DNA	2.38	0.461	NA	31	14278	6	14303	NA	NA
NTC merge	NTC DNA	total HIV DNA	No Call	0.561	0.1131	31	50904	26	50909	6	50929
A05	sample1	total HIV DNA	No Call	No Call	1.3695	23	8809	14	8818	11	8821
B05	sample2	total HIV DNA	2.57	1.81	1.52321	27	11524	19	11532	16	11535
C05	sample3	total HIV DNA	3.39	2.12	1.484	32	10342	20	10354	14	10360
D05	sample4	total HIV DNA	2.79	0.898	0.69806	28	10995	9	11014	7	11016
E05	sample5	total HIV DNA	No Call	No Call	0.79552	24	9649	16	9657	7	9666
F05	sample6	total HIV DNA	No Call	No Call	1.10476	22	9930	11	9941	10	9942
G05	sample7	total HIV DNA	3.92	2.31	2.31062	39	10911	23	10927	23	10927
H05	sample8	total HIV DNA	3.24	1.76	1.75995	35	11838	19	11854	19	11854

FIGURE C.15: Comparison of QuantaSoft auto analysis methods (single and combined well analysis) and ddpcRquant on an example dataset of an HIV total DNA quantification assay.

C.8 Automated analysis by Quantasoft for the NTCs of an HIV total DNA quantification assay

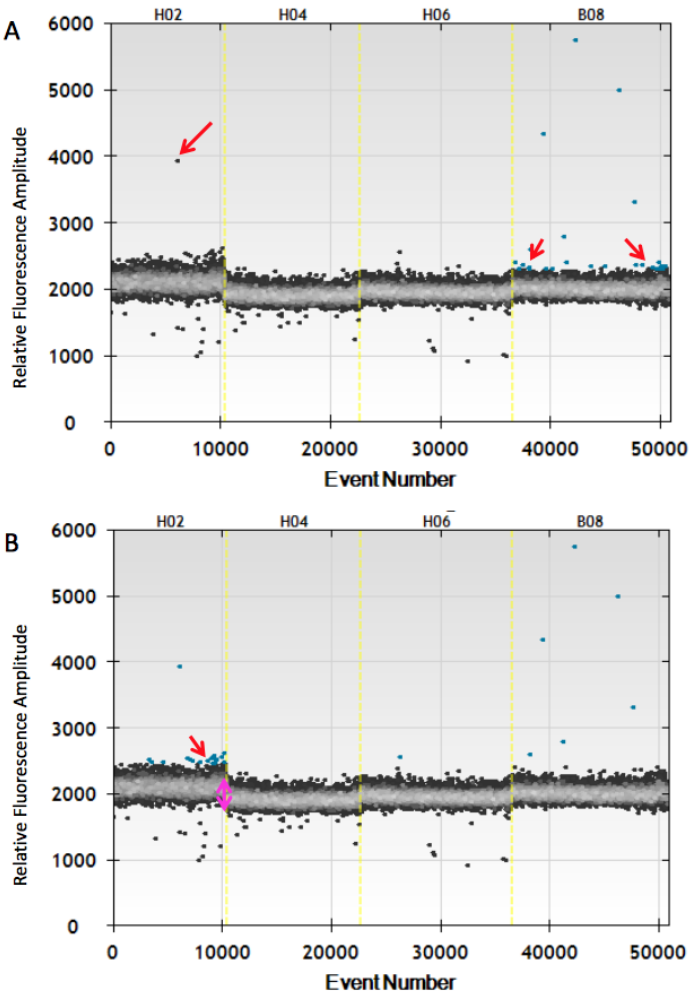


FIGURE C.16: (A) Single well analysis. In H02 a clear positive droplet is seen as negative (red arrow) and in B08 negative droplets are seen as positive. (red arrows) (B) Combined well analysis. In H02 negative droplets are seen as positive (red arrow). In addition, a baseline shift is observed between NTCs (pink double arrow)

C.9 Automated analysis by QuantaSoft for the samples of an HIV total DNA quantification assay

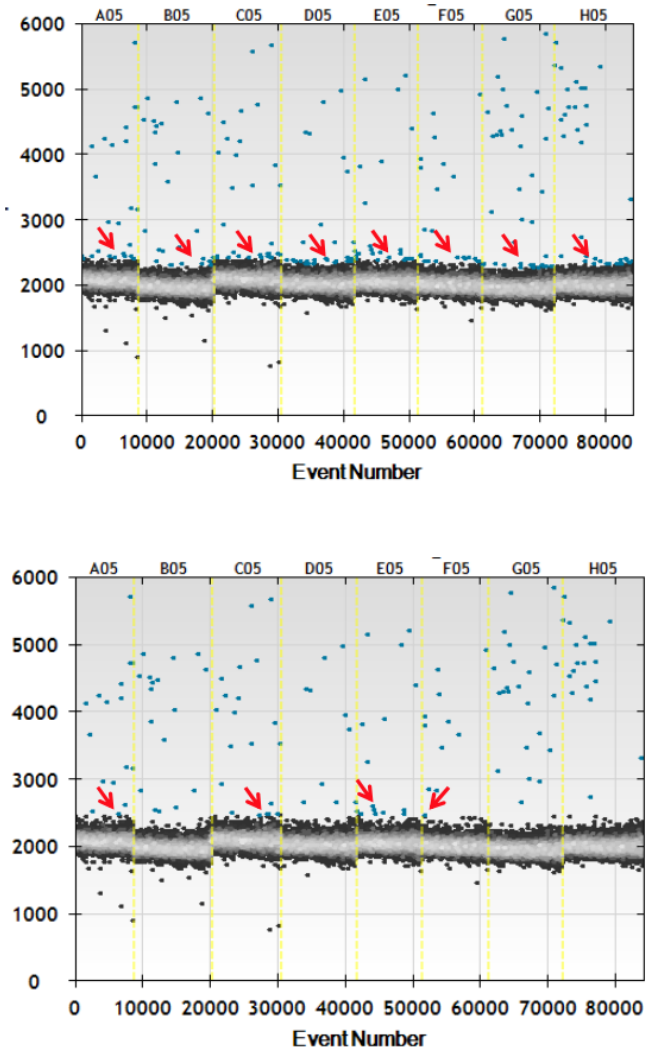


FIGURE C.17: (A) Single well analysis. In all conditions, droplets from the negative droplet population are marked positive (red arrows). (B) Combined well analysis. In column 1, 3, 5 and 6 droplets from the negative droplet population are marked positive (red arrows). Blue dots: positive droplets, grey dots: negative droplets.

C.10 Normality assessment

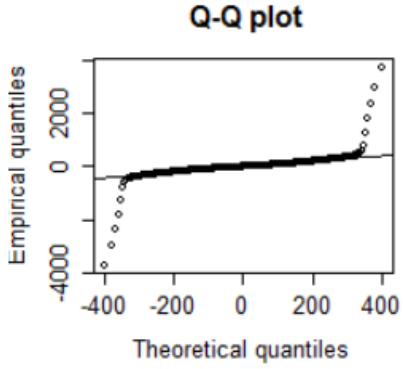


FIGURE C.18: Normality evaluation for the combined NTCs of an HIV total DNA quantification assay, the NTC analysis was performed using four merged NTC outputs.

C.11 ddpcRquant output for an HIV total DNA quantification assay

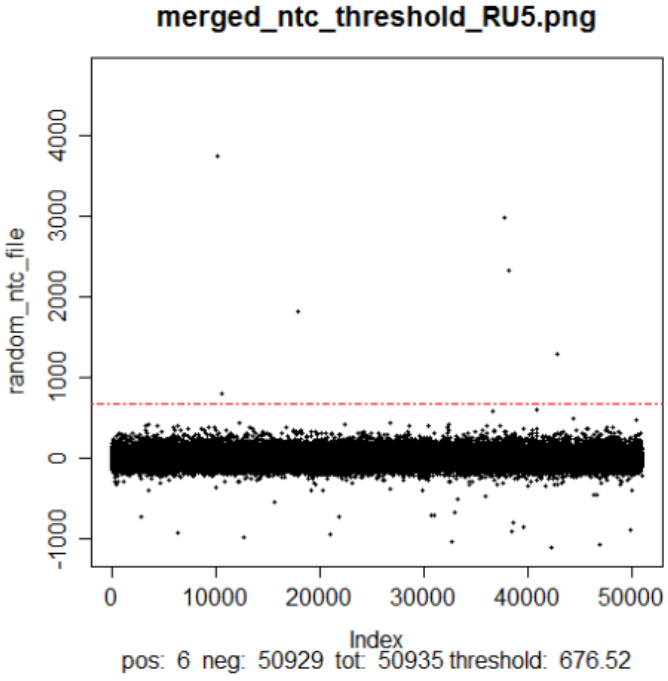


FIGURE C.19: Normality evaluation for the combined NTCs of an HIV total DNA quantification assay, the NTC analysis was performed using four merged NTC outputs.

D Mixed models

D.1 R manual

D.1.1 Data

The standard data input requires four columns and as many rows as there are samples:

1. **Gene name**: The name of the gene. If you have technical replicates, these need to have the same name. (Character.)
2. **Positive**: The number of positive partitions obtained. (Numeric.)
3. **Negative**: The number of negative partitions obtained. (Numeric.)
4. **Target**: Whether or not the gene is to be considered as a target or a reference gene. (Logical, TRUE = target, FALSE = reference.)
5. **Well:** For duplex (multiplex) experiments: genes analyzed in duplex (multiplex) need to have the same name. For singeplex experiments, this column can be left empty or can be a copy of the *Gene name* column. (Character.)

An example dataset is displayed in Table D.1.

Name	Positive	Negative	Target	Well
CLIC6	1666	10929	TRUE	CLIC6-SRY
CLIC6	1986	12757	TRUE	CLIC6-SRY
CLIC6	1861	12189	TRUE	CLIC6-SRY
DSCR3	2134	13173	TRUE	ITGBLA-DSCR3
DSCR3	1937	11939	TRUE	ITGBLA-DSCR3
SYNJ1	1788	11285	TRUE	SYNJ1-RPP30
SYNJ1	1856	11011	TRUE	SYNJ1-RPP30
SYNJ1	2035	12286	TRUE	SYNJ1-RPP30
AMELX	650	13335	TRUE	AMELX-LAMA3
AMELX	597	13712	TRUE	AMELX-LAMA3
AMELX	671	14352	TRUE	AMELX-LAMA3
AMELY	616	14312	TRUE	AMELY-C18ORF62
AMELY	630	14392	TRUE	AMELY-C18ORF62
IL13RA1	573	11036	TRUE	IL13RA-LNX2
IL13RA1	672	13822	TRUE	IL13RA-LNX2
IL13RA1	681	14621	TRUE	IL13RA-LNX2
SRY	543	12052	TRUE	CLIC6-SRY
SRY	676	14067	TRUE	CLIC6-SRY
SRY	586	13464	TRUE	CLIC6-SRY
C18ORF62	1370	13558	FALSE	AMELY-C18ORF62
C18ORF62	1392	13630	FALSE	AMELY-C18ORF62
LAMA3	1207	12778	FALSE	AMELX-LAMA3
LAMA3	1340	12969	FALSE	AMELX-LAMA3
LAMA3	1317	13706	FALSE	AMELX-LAMA3
LNX2	1075	10534	FALSE	IL13RA-LNX2
LNX2	1304	13190	FALSE	IL13RA-LNX2
LNX2	1389	13913	FALSE	IL13RA-LNX2

TABLE D.1: Example dataset with 7 target genes and 3 reference genes with two to three replicates.

D.1.2 Data analysis

The data for this tutorial is available at `http://users.ugent.be/~mvynck/glmm/tutorial.csv`. We will first load this example dataset:

```
> data.tutorial <- read.csv("tutorial.csv")
```

Make sure the data is in the correct format:

```
> head(data.tutorial)
   gene  pos   neg target         well
1 CLIC6 1666 10929   TRUE    CLIC6-SRY
2 CLIC6 1986 12757   TRUE    CLIC6-SRY
3 CLIC6 1861 12189   TRUE    CLIC6-SRY
4 DSCR3 2134 13173   TRUE ITGBLA-DSCR3
5 DSCR3 1937 11939   TRUE ITGBLA-DSCR3
7 SYNJ1 1788 11285   TRUE  SYNJ1-RPP30
```

The script requires the lme4 library. If this library is not installed on your system, you may install it by running the following command:

```
> install.packages("lme4")
```

Load some functions:

```
> source("http://users.ugent.be/~mvynck/glmm/glmmfunctions.R")
```

Some preprocessing to make sure the variables are of the correct type and have the correct name:

```
> colnames(data.tutorial)<-c("gene","pos","neg","target","well")
> data.tutorial$gene<-as.character(data.tutorial$gene)
> data.tutorial$pos<-as.numeric(as.character(data.tutorial$pos))
> data.tutorial$neg<-as.numeric(as.character(data.tutorial$neg))
> data.tutorial$target<-as.logical(data.tutorial$target)
> data.tutorial$well <- as.character(data.tutorial$well)
```

D.1.2.1 Absolute quantification

To fit a model, we will use the function `fit.set`. This function will automatically decide whether to account for interreplicate variation or interreference gene variation depending on the data you feed it: in the absolute quantification case, there are no reference genes. Depending on whether you have technical replicates or not, the function will automatically account for it. This will also be displayed in the output of the function.

Example 1

For example, if we feed the function only one sample of our data (the first row) it will see that there is only one target and no replicates, and will analyze the data as such:

```
> fit.example1 <- fit.set(data.tutorial[1,])
Analyzing single target ...
Converting data ...
Fitting model:
absolute quantification
without replicates ...
```

The object fit.example1 now contains the fitted model. You can process this manually, but it is easier to rely on the function `get.estimates` to get the estimates and a 95% confidence interval in a data frame object. For this, we need to supply the data once again, along with the fitted model:

```
> (example1 <- get.estimates(data.tutorial[1,],fit.example1))
  genenames       est   est.low  est.high
1     CLIC6 0.1418801 0.1352226 0.1488653
```

The average number of copies per partition is thus 0.14. To obtain a concentration estimate (number of copies per volume), simply divide this number by the partition volume (e.g. 0.85nl):

```
> (example1[2]/0.85)
```

```
       est
1 0.1669178
```

Example 2

A second scenario would be to analyze multiple technical replicates, for example the first three samples in our dataset are a triplicate experiment of the CLIC6 gene. The code is exactly the same, and the function will automatically recognize that there are technical replicates and account for interreplicate variability:

```
> fit.example2 <- fit.set(data.tutorial[1:3,])
Analyzing single target ...
Converting data ...
Fitting model:
absolute quantification
with replicates ...
> (example2 <- get.estimates(data.tutorial[1:3,],fit.example2))
  genenames       est   est.low  est.high
1     CLIC6 0.1429503 0.1392255 0.1467748
```

Example 3

A third possibility is to estimate multiple targets simultaneously, with or without replicates. Again, use the exact same code, and the software will recognize that there are multiple targets (different names with target set to TRUE). For demonstration purposes, let's calculate absolute quantities for CLIC6 with replicates, and DSCR3 without replicates (i.e. the first four rows of the dataset):

```
> fit.example3 <- fit.set(data.tutorial[1:4,])
Analyzing multiple targets ...
Analyzing target 1 of 2 (CLIC6) ...
Converting data ...
Fitting model:
absolute quantification
with replicates ...


Analyzing target 2 of 2 (DSCR3) ...
Converting data ...
Fitting model:
absolute quantification
without replicates ...


> (example3 <- get.estimates(data.tutorial[1:4,],fit.example3))
  genenames       est   est.low  est.high
1     CLIC6 0.1429503 0.1392255 0.1467748
2     DSCR3 0.1501410 0.1438982 0.1566545
```

D.1.2.2 Relative quantification / copy number variation

In this case we will need reference genes. The functions to use are again exactly the same, and the software will recognize whether to account for interreplicate variation (when target and/or reference genes have replicates) and/or to account for interreference gene variation (when multiple reference genes are present).

Example 4

For example, copy number calculation with a single reference gene and a single target gene, both having replicates:

```
> fit.example4 <- fit.set(data.tutorial[c(1:3,23:24),])
Analyzing single target ...
Converting data ...
Fitting model:
single reference gene
copy number variation
with replicates ...


> (example4 <- get.estimates(data.tutorial[c(1:3,23:24),],
                    fit.example4))
  genenames      est  est.low est.high
1     CLIC6 2.954934 2.823234 3.092777
```

The estimated copy number is 2.95. The 95% confidence interval is $[2.82, 3.09]$.

Example 5

Another scenario is copy number calculation with multiple reference genes and a single target gene:

```
> fit.example5 <- fit.set(data.tutorial[c(1:3,23:25),])
Analyzing single target ...
Converting data ...
Fitting model:
multiple (n=2) reference genes
copy number variation
with replicates ...
> (example5 <- get.estimates(data.tutorial[c(1:3,23:25),],
                    fit.example5))
  genenames      est  est.low est.high
1     CLIC6 3.039879 2.877905 3.210969
```

The estimated copy number is 3.03. The 95% confidence interval is $[2.88, 3.21]$.

Example 6

It is also possible to quantify multiple target genes at once by using a data frame with multiple targets:

```
> fit.example6 <- fit.set(data.tutorial[c(1:4,23:25),])
Analyzing multiple targets ...
```

```
Analyzing target 1 of 2 (CLIC6) ...
Converting data ...
Fitting model:
multiple (n=2) reference genes
copy number variation
with replicates ...


Analyzing target 2 of 2 (DSCR3) ...
Converting data ...
Fitting model:
multiple (n=2) reference genes
copy number variation
with replicates ...


> (example6 <- get.estimates(data.tutorial[c(1:4,23:25),],
                    fit.example6))
  genenames      est  est.low est.high
1     CLIC6 3.039879 2.877905 3.210969
2     DSCR3 3.192794 2.995193 3.403432
```

The estimated copy number is 3.03 for CLIC. The 95% confidence interval for CLIC is $[2.88, 3.21]$. (Identical to the results obtained in example 5.) The estimated copy number is 3.19 for DSCR3. The 95% confidence interval for DSCR3 is $[3.00, 3.40]$.

D.1.2.3 Determine reference gene stability

To allow an optimal quantification, using reliable reference genes is essential. Within-sample reference gene stability can be determined using the function `determine.stability`. This function will calculate all pairwise copy number/relative expression quantities and calculate a measure of deviance. The smaller this measure of deviance, the more reliable the reference gene (see equations 5.33-5.36 for details).

Example 7

```
> stability.tutorial <- determine.stability(data.tutorial)
Determining stability...
Reference gene  1  of  3

Analyzing single target ...
Converting data ...
Fitting model:
single reference gene
copy number variation
with replicates ...
=== OUTPUT TRUNCATED FOR CLARITY ===
```

The resulting objected can be plotted and will return an ordered barplot (from most stable to least stable reference gene):

```
> plot.stability(stability.tutorial)
```

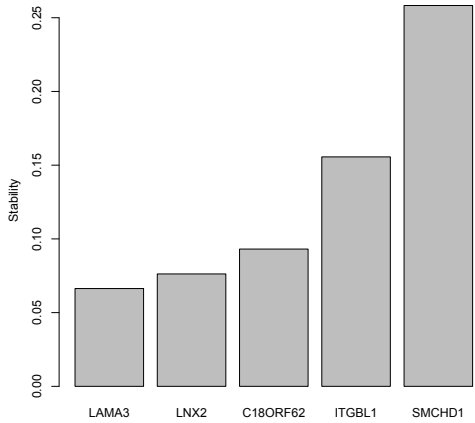


FIGURE D.1: Stability of candidate reference genes.

D.1.3 R code

For the user more familiar with R, we dig deeper into the R code in this section.

D.1.3.1 Arranging the data

In order to fit the models, the data need to be rearranged into a format that can be handled by the GLM(M) functions. The data format will have a row for each partition observation. In the most simple case of absolute quantification with no replicates, the data is just a vector with $0/1$ observations.
Data for absolute quantification with replicates, should be presented in a two column data frame: one column containing the $0/1$ data and a second column indicating the replicate from which it was taken (Table D.2).
When dealing with copy number variation determination of relative quantification one or more reference genes will be present. In that case, the data will be represented by a three column data frame (in the absence of replicates): one with the outcome, one detailing whether the gene is a target or a reference and one detailing which gene was measured (Table D.3). In this section we assume singeplex experiments. For multiplex experiments, an additional column detailing the genes measured in multiplex should be added as well (an example is given in Table D.1).

repl	Y
repl1	0
repl1	0
repl1	0
...	...
repl1	1
repl1	1
repl1	1
...	...
repl2	0
repl2	0
repl2	0
...	...
repl2	1
repl2	1
repl2	1
...	...

Table D.2: Data format for absolute quantification with replicates.

sample	target	Y
geneA	yes	0
geneA	yes	1
…	…	…
geneB	no	0
geneB	no	1
…	…	…
geneC	no	0
geneC	no	1
…	…	…

TABLE D.3: Data format for copy number variation without replicates.

repl	sample	target	Y
geneA:rep1	geneA	yes	0
geneA:rep1	geneA	yes	1
…	…	…	…
geneA:rep2	geneA	yes	0
geneA:rep2	geneA	yes	1
…	…	…	…
geneB:rep1	geneB	no	0
geneB:rep1	geneB	no	1
…	…	…	…
geneB:rep2	geneB	no	0
geneB:rep2	geneB	no	1
…	…	…	…
geneC:rep1	geneC	no	0
geneC:rep1	geneC	no	1
…	…	…	…
geneC:rep2	geneC	no	0
geneC:rep2	geneC	no	1
…	…	…	…

TABLE D.4: Data format for copy number variation with replicates.

If replicates are present in the copy number variation case the data will be represented by a four column data frame: one additional column detailing the replicate (Table D.4).
This procedure is implemented in the function `generate.df` and can be executed when sourcing the following script:

```
> source("http://users.ugent.be/~mvynck/glmm/glmmfunctions.R")
```

The function will only work with a data frame in the format as displayed in Table D.1, but can easily be adjusted to accomodate other data formats.

D.1.3.2 Absolute quantification

No replicates

In absence of both replicates and reference genes, there are no sources of variation to account for, and the model reduces to a simple generalized linear model with a complementary log-log link:

```
model.fit<-glm(Y~1,data=db,family=binomial(cloglog))
```

The estimate is obtained as:

```
(exp(coef(model.fit)))
```

An approximate Wald 95% confidence interval (more information in Appendix D.4.3) on the estimate can be obtained as follows:

```
(exp(coef(model.fit)-1.96*sqrt(vcov(model.fit))))
(exp(coef(model.fit)+1.96*sqrt(vcov(model.fit))))
```

With replicates

In presence of replicates, a random effect for interreplicate variation is added:

```
model.rand<-glmer(Y~1+(1|repl),data=db,family=binomial(cloglog),
                  nAGQ=1,verbose=F)
```

The estimate is obtained as:

```
(exp(model.fit@beta[1]+0.5*summary(model.fit)$varcor$repl[1]))
```

where Nb indicates the ploidy level.

An approximate Wald 95% confidence interval on the estimate can be obtained as follows:

```
(exp(model.fit@beta[1]+0.5*summary(model.fit)$varcor$repl[1]-
                1.96*sqrt(vcov(model.fit)[1,1])))
(exp(model.fit@beta[1]+0.5*summary(model.fit)$varcor$repl[1]+
                1.96*sqrt(vcov(model.rand)[1,1])))
```

D.1.3.3 Copy number variation/relative quantification

No replicates

In the presence of one or more reference genes, the model needs to indicate whether the observed outcome is for the target or for a reference gene by introducing the dummy `target`. A random effect for interreference gene variation is added as well. We give an example for calculating copy number variation with multiple reference genes.

```
model.rand<-glmer(Y~dummy(target,"no")+(-1+dummy(target,"no")|sample),
                  data=db, family=binomial(cloglog),nAGQ=1,verbose=F)
```

The estimate is obtained as:

```
(exp(-(model.fit@beta[2])+summary(model.fit)$varcor$sample[1]))*Nb
```

where Nb indicates the ploidy level.

An approximate Wald 95% confidence interval on the estimate can be obtained as follows:

```
(exp(-(model.fit@beta[2])+summary(model.fit)$varcor$sample[1]-
                1.96*sqrt(vcov(model.fit)[2,2])))*Nb
(exp(-(model.fit@beta[2])+summary(model.fit)$varcor$sample[1]+
                1.96*sqrt(vcov(model.fit)[2,2])))*Nb
```

With replicates

In addition to the previous situation, we now need to add a random effect for the replicates as well.

```
glmer(Y~dummy(target,"no")+(1|repl)+(-1+dummy(target,"no")|sample),
                data=db,family=binomial(cloglog),nAGQ=1,verbose=F)
```

The estimate is obtained as:

```
(exp(-(model.rand@beta[2])+summary(model.rand)$varcor$sample[1]/2+
                summary(model.rand)$varcor$repl[1]))*Nb
```

where Nb indicates the ploidy level.

An approximate Wald 95% confidence interval on the estimate can be obtained as follows:

```
(exp(-(model.rand@beta[2])+summary(model.rand)$varcor$sample[1]/2+
                  summary(model.rand)$varcor$repl[1]-
                  1.96*sqrt(vcov(model.rand)[2,2])))*2
(exp(-(model.rand@beta[2])+summary(model.rand)$varcor$sample[1]/2+
                  summary(model.rand)$varcor$repl[1]+
                  1.96*sqrt(vcov(model.rand)[2,2])))*2
```

D.2 Overview of trueness, precision and stability estimates

D.2.1 Stability plots

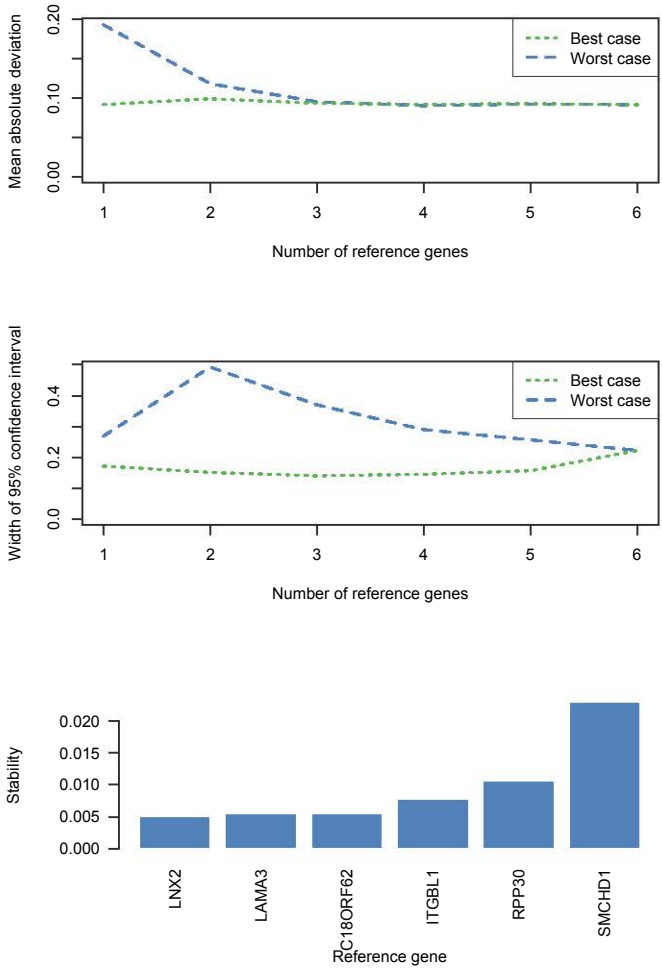


FIGURE D.2: Sample 5. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

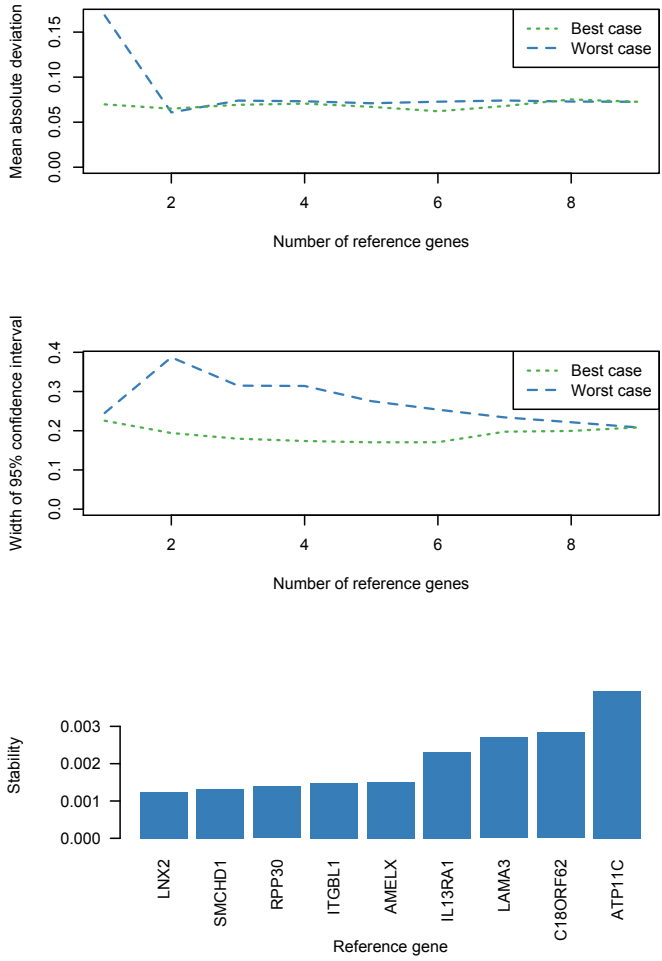


FIGURE D.3: Sample 6. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

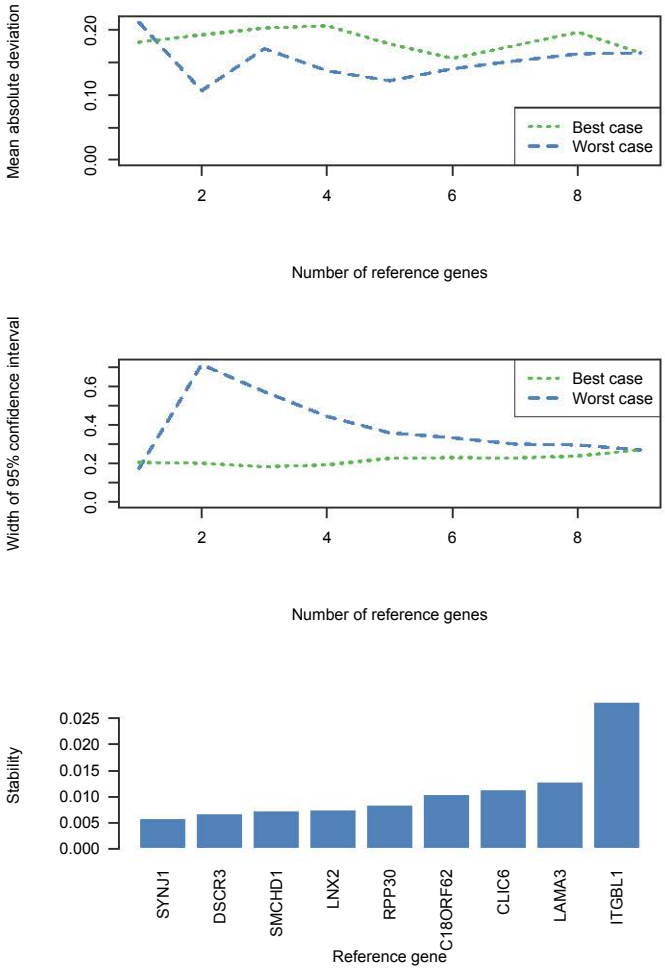


FIGURE D.4: Sample 9. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

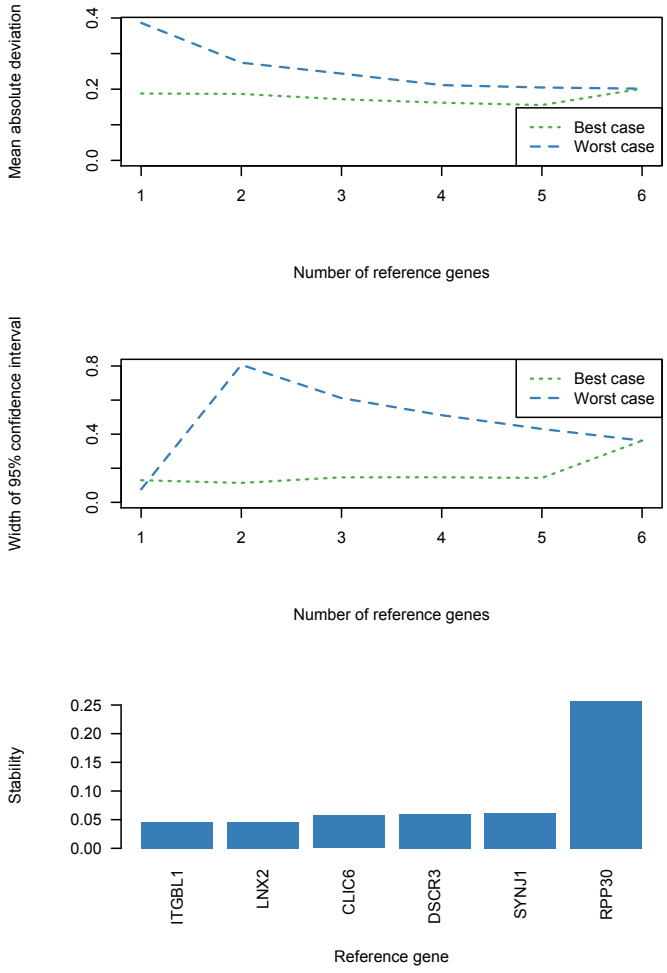


FIGURE D.5: Sample 13. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

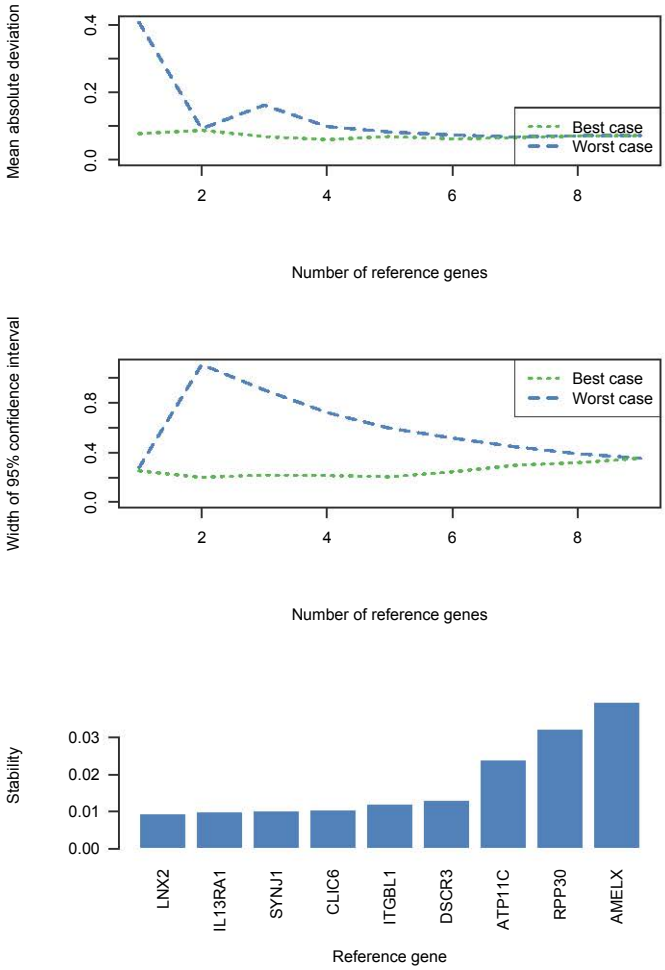


FIGURE D.6: Sample 15. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

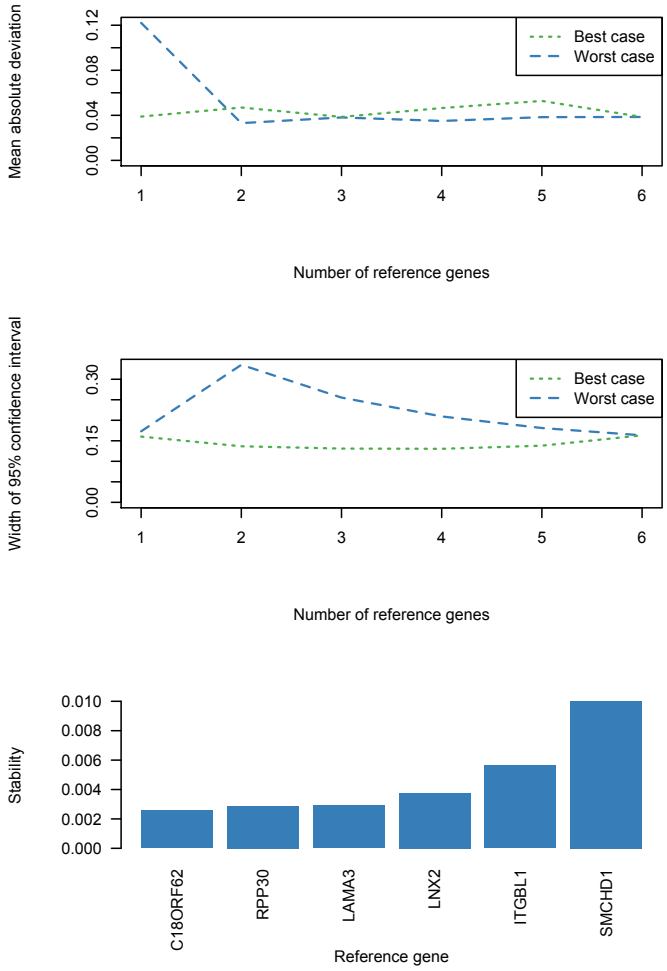


FIGURE D.7: Sample 16. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

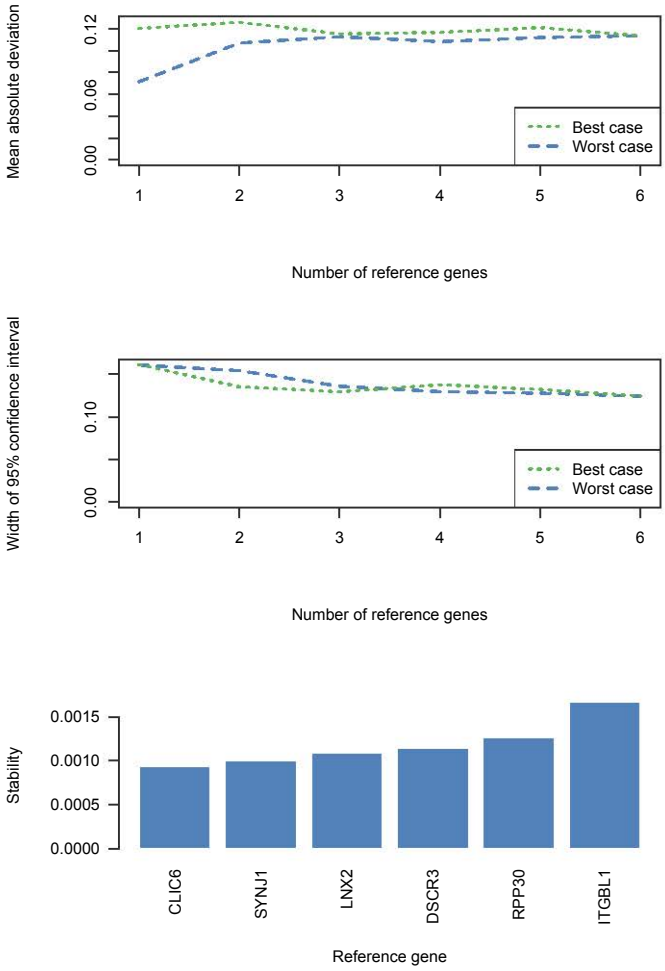


FIGURE D.8: Sample 18. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

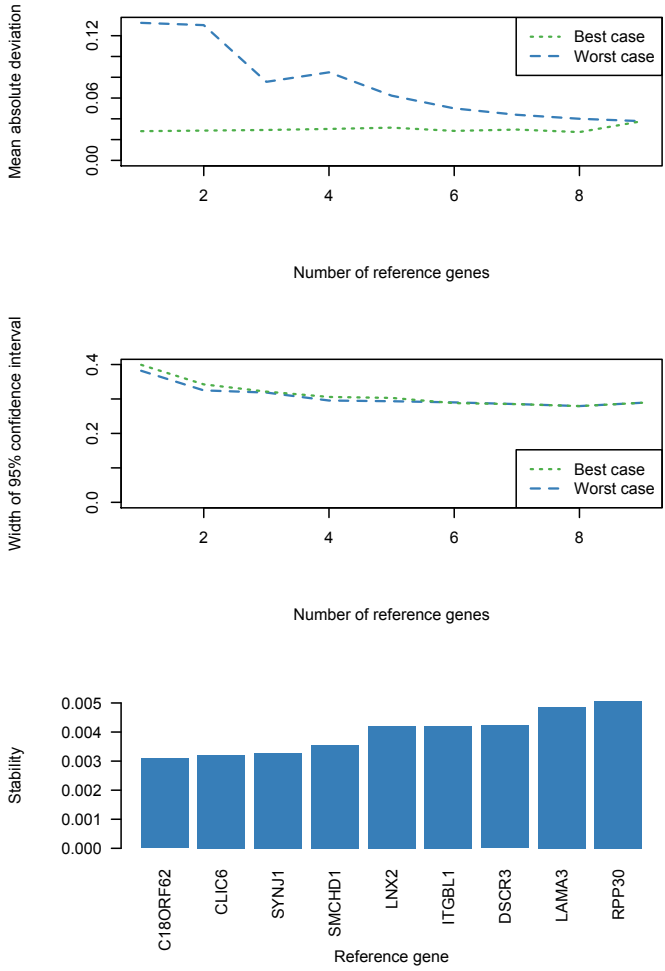


FIGURE D.9: Sample 19. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

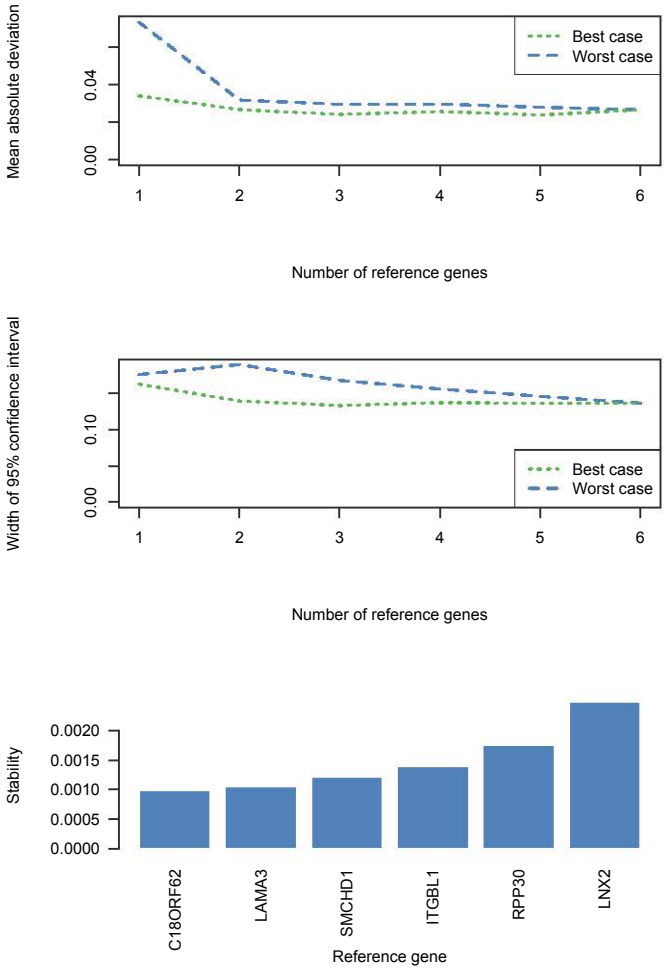


FIGURE D.10: Sample 20. **Top:** Mean absolute deviation from integer copy numbers as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Middle:** Uncertainty on target estimates, expressed as the width of a 95% confidence interval, as a function of the number of reference genes. The reference genes are added in the order of decreasing stability (best case) or increasing stability (worst case). **Bottom:** Stability of candidate reference genes.

D.3 Case study results

D.3.1 Introduction

The setup of the study was to assess chromosomal aberrations, i.e. deviations from diploid copy numbers on 13 target loci. The reference locus was located on the *RPP30* gene. Below we discuss the analysis of all samples in our case study, similar to the analysis presented in the "case study results" section in the main manuscript.

D.3.2 Sample 5

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.11. It is clear we are dealing with a sample belonging to a male (single AMELX, IL13RA1, ATP11C, AMELY, SRY copy), but there seems to be a small underestimation of the copy numbers, which may be due to aberrant quantification of the *RPP30* locus. It can furthermore be seen that the *CLIC6*, *DSCR3* and *SYNJ1* loci (all located on chromosome 21) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 21 trisomy is likely. We thus use the loci on chromosomes X, 13, 18 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.12) indicates that the loci on *LNX2*, *LAMA3*, *C18ORF62* are of similar stability, with rising instability thereafter for the loci on *ITGBL1*, *RPP30* and *SMCHD1*. The latter three are thus excluded as reference loci and the copy numbers of the three chromosome 21 loci recalculated using the three stable reference loci. The results are shown in Figure D.13: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the CLIC6 locus for a trisomy 21 case, with the DSCR3 and SYNJ1 loci displaying a copy number close to the expected copy number.

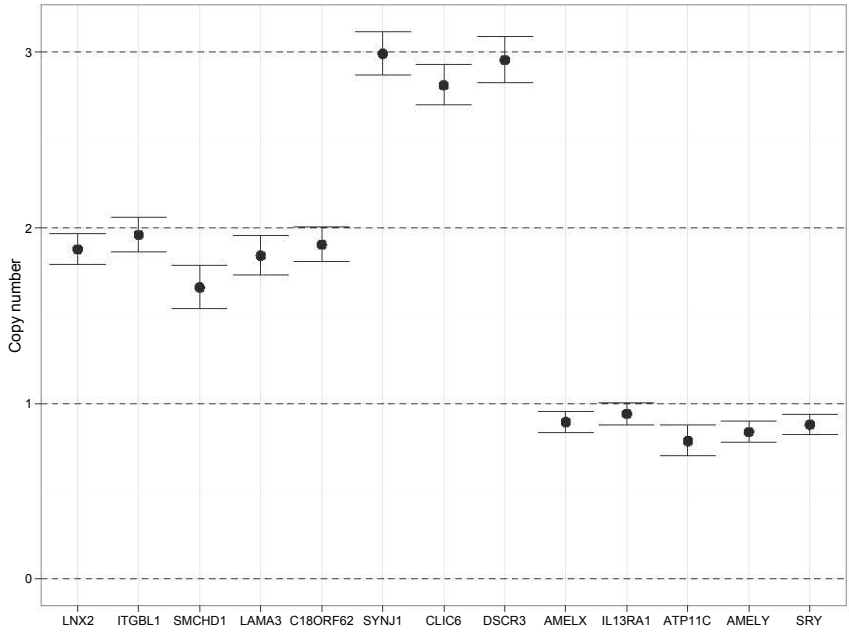


FIGURE D.11: Sample 5: RPP30 normalization.

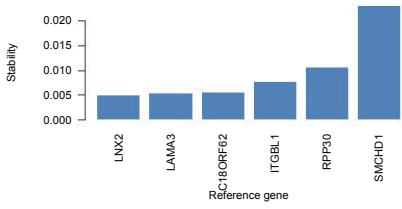


FIGURE D.12: Sample 5: Stability.

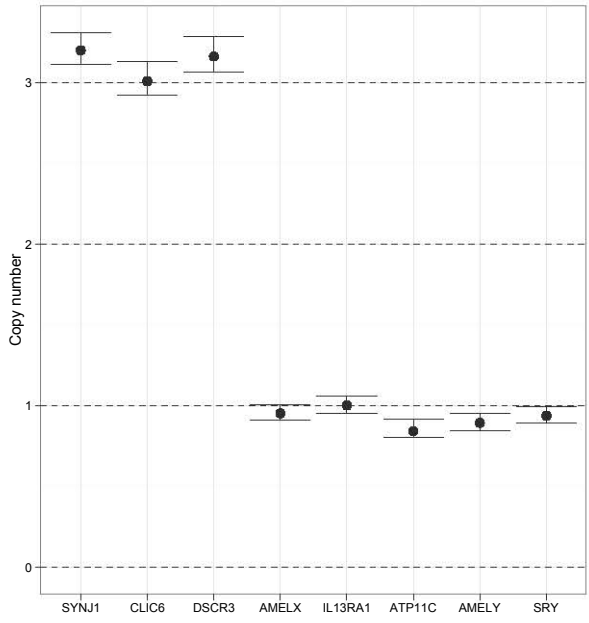


FIGURE D.13: Sample 5: stable reference gene normalization.

D.3.3 Sample 6

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.14. It is clear we are dealing with a sample belonging to a female (no AMELY/SRY copies), and the estimation of the copy numbers seem correct. It can furthermore be seen that the *CLIC6*, *DSCR3* and *SYNJ1* loci (all located on chromosome 21) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 21 trisomy is likely. We thus use the loci on chromosomes X, 13, 18 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.15) indicates that the *LNX2*, *SMCHD1*, *RPP30*, *ITGBL1* and *AMELX* loci are rather stable, with rising instability thereafter for the loci *IL13RA1*, *LAMA3*, *C18ORF62* and *ATP11C*. The stable loci are thus used as reference loci and the copy numbers of the three chromosome 21 loci recalculated using the stable reference loci. The results are shown in Figure D.16: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the DSCR3 and SYNJ1 locus for a trisomy 21 case, with the CLIC6 locus displaying a copy number close to the expected copy number.

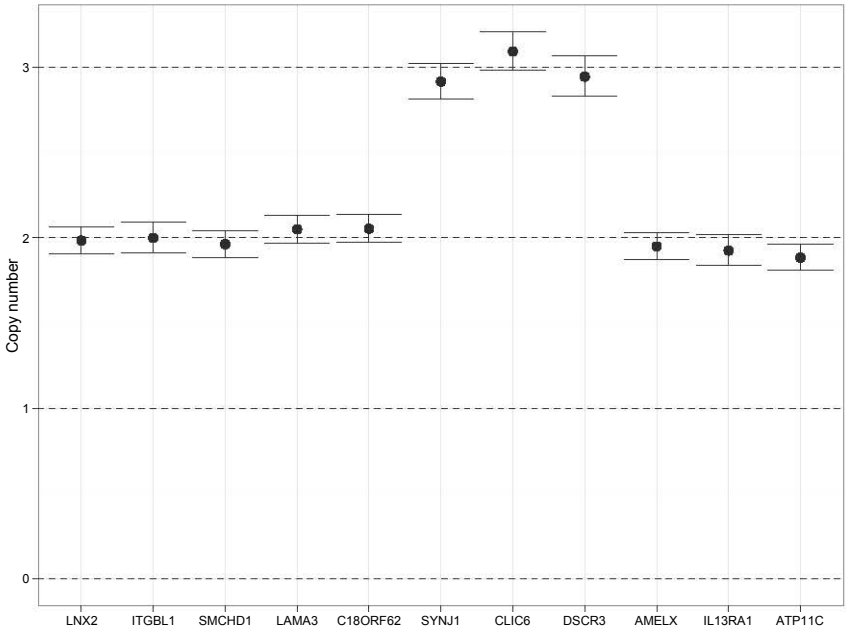


FIGURE D.14: Sample 6: RPP30 normalization.

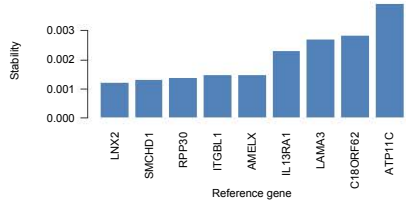


FIGURE D.15: Sample 6: Stability.

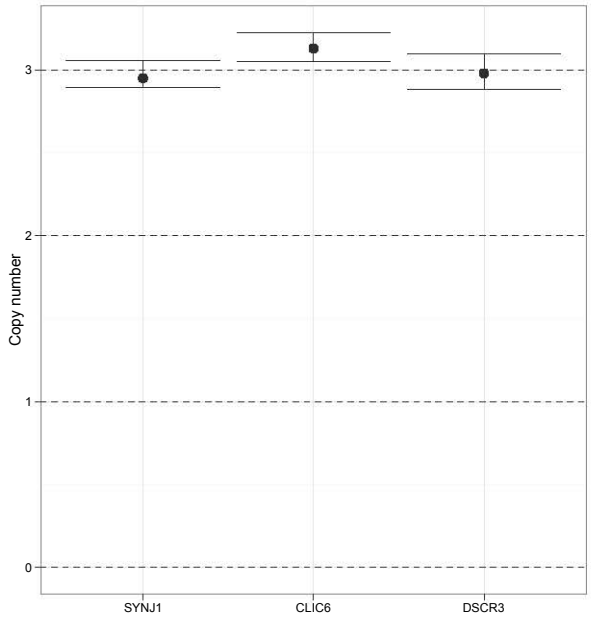


Figure D.16: Sample 6: stable reference gene normalization.

D.3.4 Sample 9

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.17. We are dealing with a sample belonging to a female (no AMELY/SRY copies), but there seems to be a small underestimation of the copy numbers, which may be due to aberrant quantification of the *RPP30* locus. It can furthermore be seen that the *ATP11C* and *IL13RA1* loci (both located on chromosome X) seem to have a higher copy number than loci located on the other chromosomes and that a isochromosome Xq is likely. We thus use the loci on chromosomes 13, 18, 21 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.18) indicates that the loci on *SYNJ1*, *DSCR3*, *SMCHD1* and *LNX2* are of similar stability, with rising instability thereafter for the loci on *RPP30*, *C18ORF62*, *CLIC6*, *LAMA3* and *ITGBL1*. The latter are thus excluded as reference loci and the copy numbers of the two chromosome X loci recalculated using the stable reference loci. The results are shown in Figure D.3.4: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *AMELX* locus for a isochromosome Xq case, with the *ATP11C* and *IL13RA1* loci displaying a copy number close to the expected copy number.

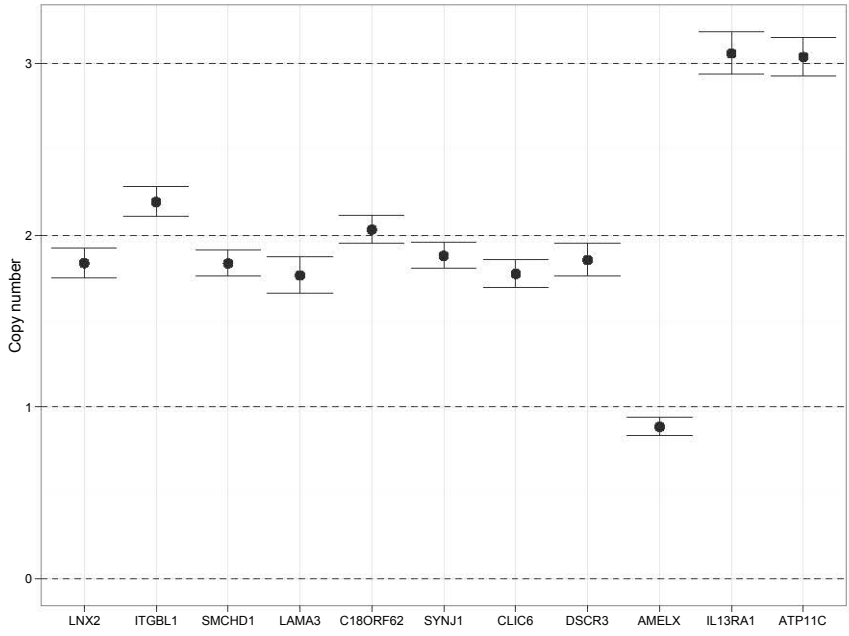


FIGURE D.17: Sample 9: RPP30 normalization.

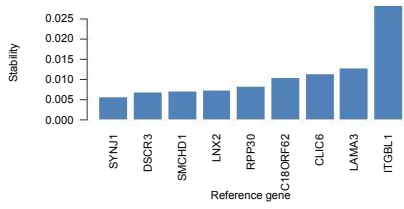


FIGURE D.18: Sample 9: Stability.

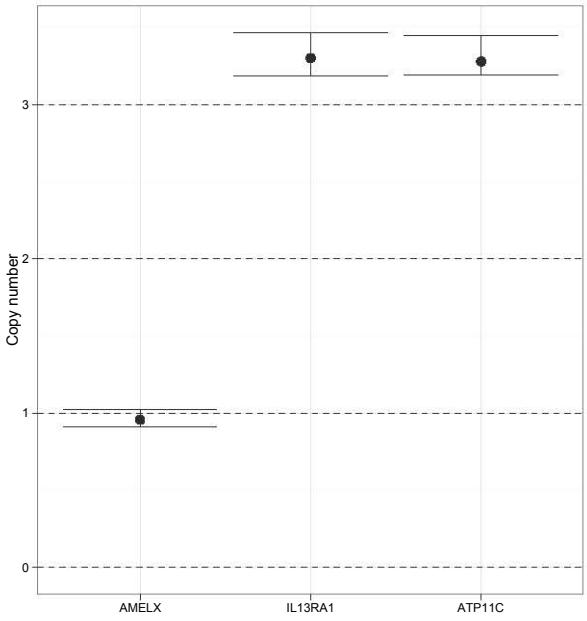


FIGURE D.19: Sample 9: stable reference gene normalization.

D.3.5 Sample 13

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.20. We are dealing with a sample belonging to a male (single AMELX copy, single AMELY/SRY copies), but there seems to be a large general underestimation of the copy numbers, which may be due to aberrant quantification of the *RPP30* locus. It can furthermore be seen that the *C18ORF62* and *LAMA3* loci (both located on chromosome 18) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 18 trisomy is likely. We thus use the loci on chromosomes 13, 21 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.21) indicates that the loci on *ITGBL1*, *LNX2*, *CLIC6*, *DSCR3* and *SYNJ1* are of similar stability, with a large instability for the locus on *RPP30*, confirming our suspicion of abberant *RPP30* quantification. The *RPP30* locus is thus excluded as reference locus and the copy numbers of the chromosome 18 loci recalculated using the stable reference loci. The results are shown in Figure D.22: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *C18ORF62* and *LAMA3* loci for a chromosome 18 trisomy case, with the *SMCHD1* locus displaying a copy number deviating from the expected copy number, but still at an elevated level.

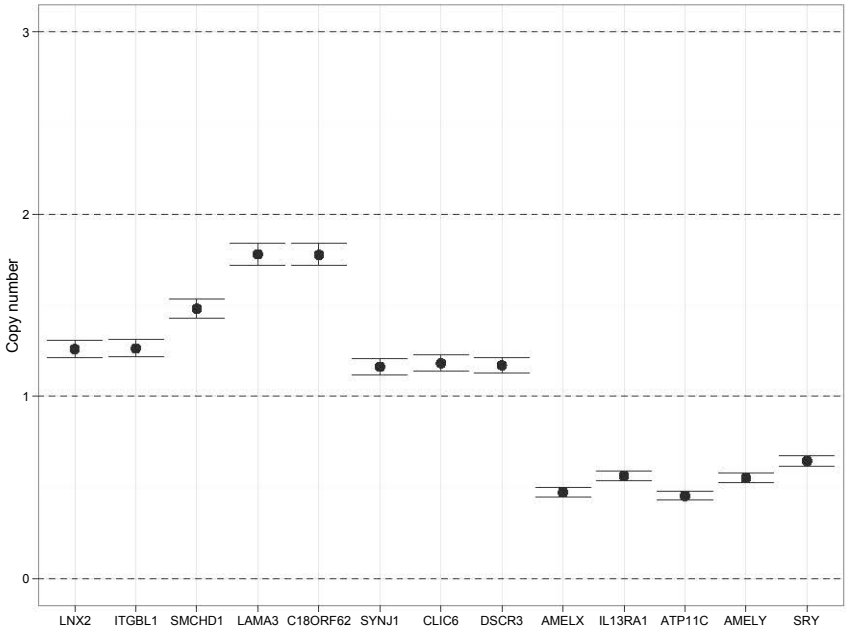


FIGURE D.20: Sample 13: RPP30 normalization.

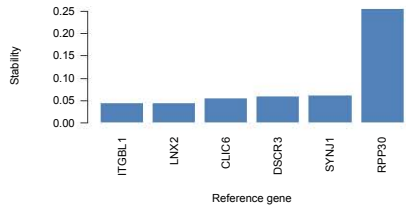


FIGURE D.21: Sample 13: Stability.

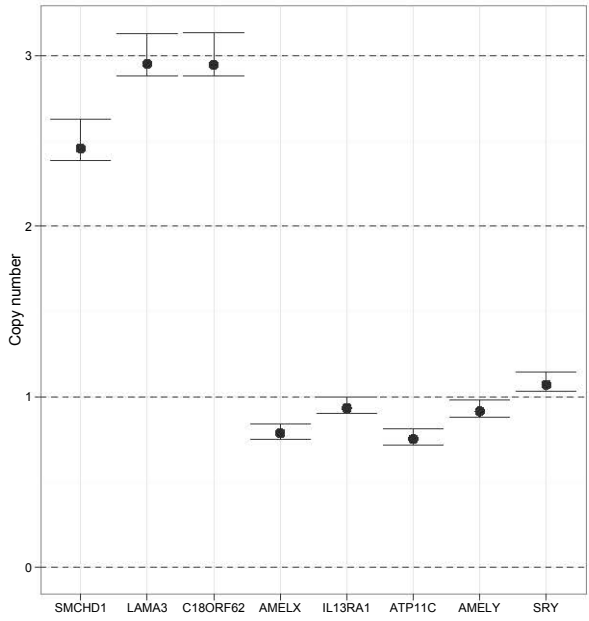


FIGURE D.22: Sample 13: stable reference gene normalization.

D.3.6 Sample 15

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.23. We are dealing with a sample belonging to a female (no AMELY/SRY copies), but there seems to be a general underestimation of the copy numbers, which may be due to aberrant quantification of the *RPP30* locus. It can furthermore be seen that the *C18ORF62*, *LAMA3* and *SMCHD1* loci (all located on chromosome 18) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 18 trisomy is likely. We thus use the loci on chromosomes 13, 21, X and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.24) indicates that the loci on *LNX2*, *IL13RA1*, *SYNJ1*, *ITGBL1* and *DSCR3* are of similar stability, with a larger instability for the loci on *ATP11C*, *AMELX* and *RPP30*, confirming our suspicion of abberant *RPP30* quantification. The unstable loci are thus excluded as reference loci and the copy numbers of the chromosome 18 loci recalculated using the stable reference loci. The results are shown in Figure D.25: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *C18ORF62*, *LAMA3* and *SMCHD1* loci for a chromosome 18 trisomy case.

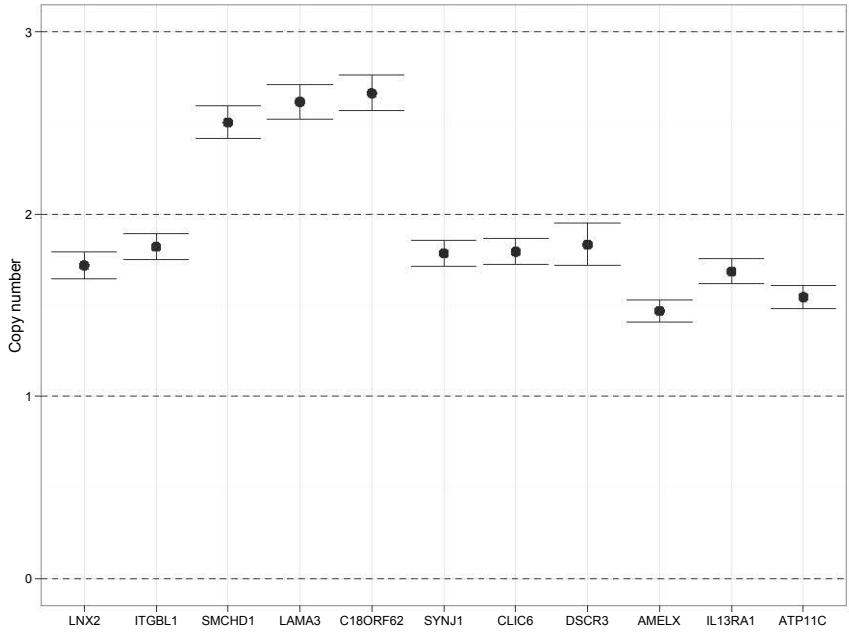


FIGURE D.23: Sample 15: RPP30 normalization.

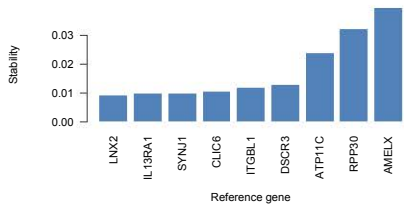


FIGURE D.24: Sample 15: Stability.

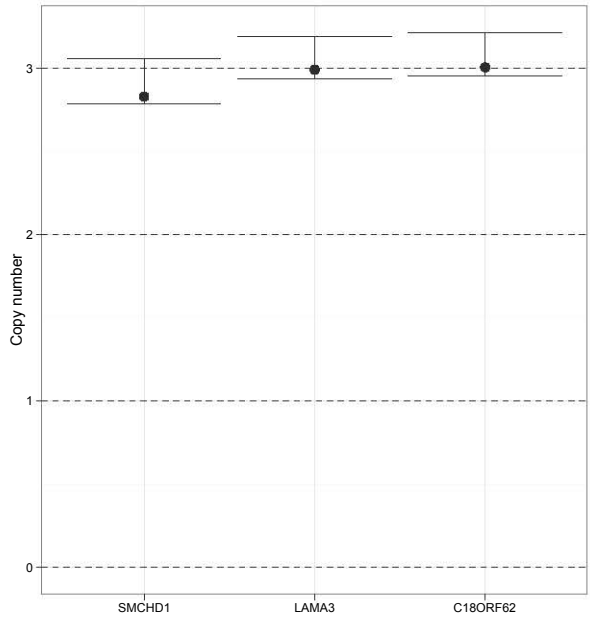


FIGURE D.25: Sample 15: stable reference gene normalization.

D.3.7 Sample 16

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.26. We are dealing with a sample belonging to a male (single AMELX, IL13RA1, ATP11C, AMELY and SRY copies), and the estimation of the copy numbers seems good. It can furthermore be seen that the *CLIC6*, *DSCR3* and *SYNJ1* loci (all located on chromosome 21) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 21 trisomy is likely. We thus use the loci on chromosomes 13, 18 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.27) indicates that the loci on *C18ORF62*, *RPP30* and *LAMA3* are of similar stability, with a larger instability for the loci on *LNX2*, *ITGBL1* and *SMCHD1*, confirming that the *RPP30* locus resulted in good quantification. The unstable loci are excluded as reference loci and the copy numbers of the chromosome 21 loci recalculated using the stable reference loci. The results are shown in Figure D.28: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *CLIC6*, *DSCR3* and *SYNJ1* loci for a chromosome 21 trisomy case.

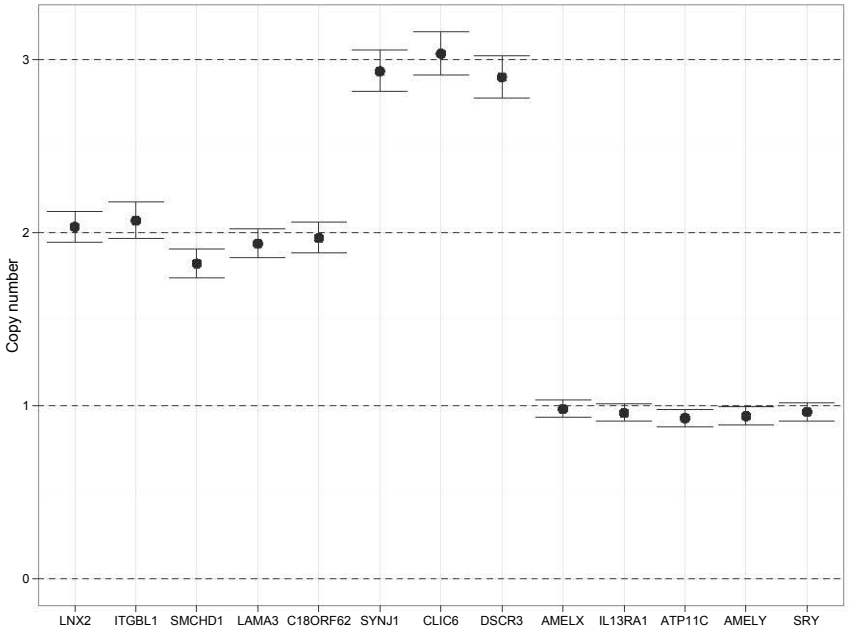


FIGURE D.26: Sample 16: RPP30 normalization.

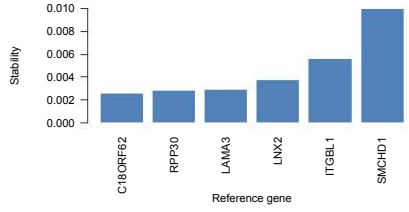


FIGURE D.27: Sample 16: Stability.

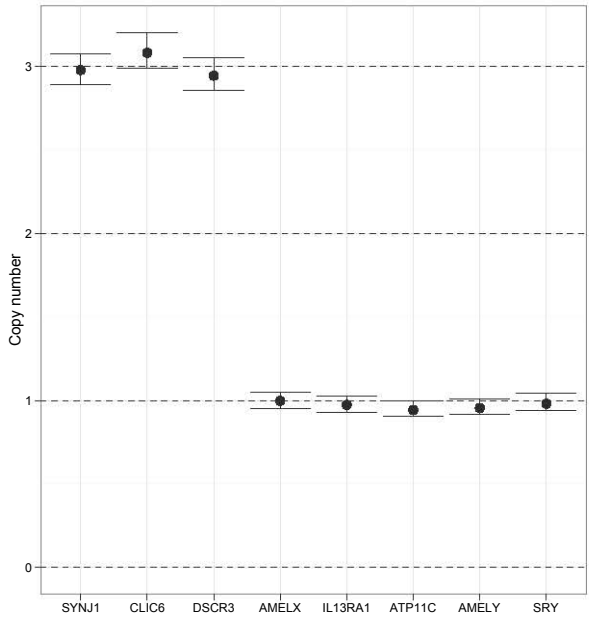


FIGURE D.28: Sample 16: stable reference gene normalization.

D.3.8 Sample 18

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.29. We are dealing with a sample belonging to a male (single AMELX, IL13RA1, ATP11C, AMELY and SRY copies), and the estimation of the copy numbers seems a bit too low. It can furthermore be seen that the *C18ORF62*, *LAMA3* and *SMCHD1* loci (all located on chromosome 18) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 18 trisomy is likely. We thus use the loci on chromosomes 13, 21 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.30) indicates that all loci display a very high stability. The copy numbers of the chromosome 21 loci are recalculated using all reference loci. The results are shown in Figure D.31: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *LAMA3* locus for a chromosome 18 trisomy case, with the *C18ORF62* and *SMCHD1* loci also showing an elevated copy number.

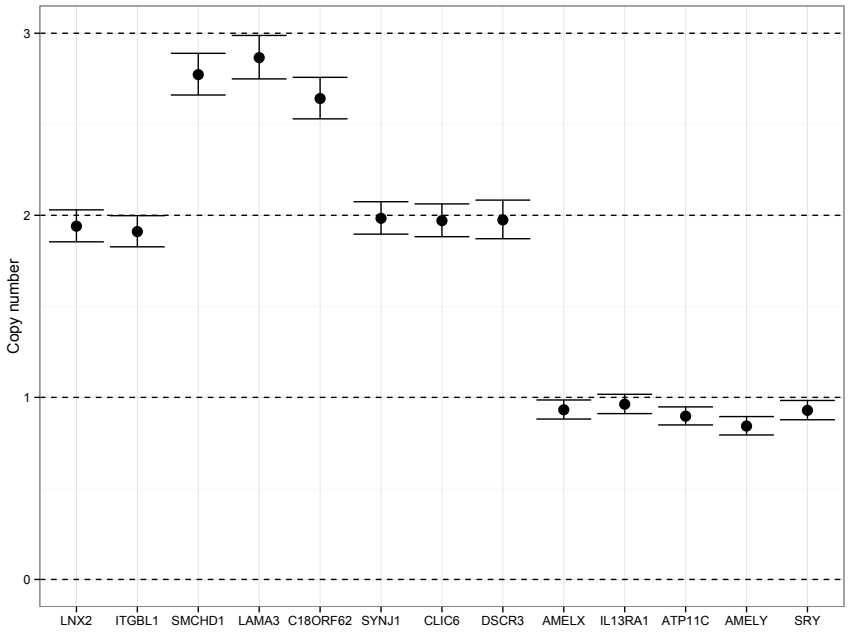


Figure D.29: Sample 18: RPP30 normalization.

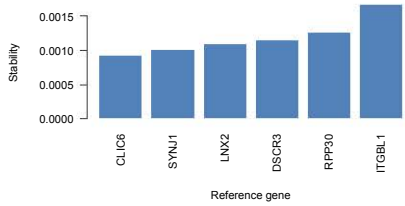


FIGURE D.30: Sample 18: Stability.

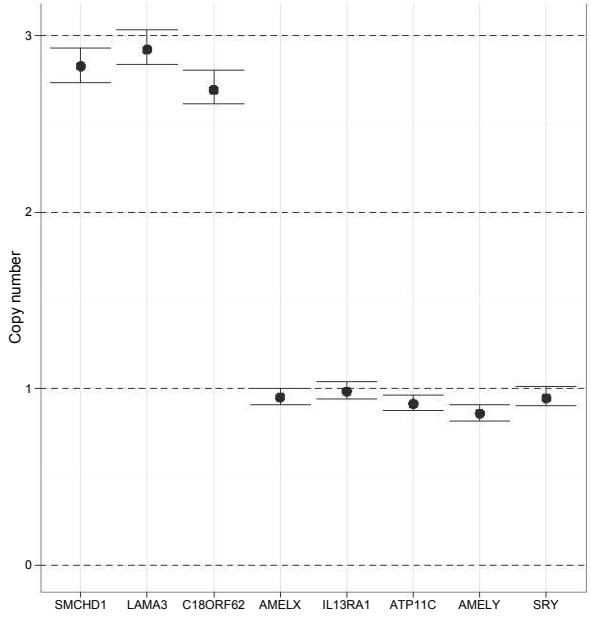


FIGURE D.31: Sample 18: stable reference gene normalization.

D.3.9 Sample 19

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.32. We are dealing with a sample belonging to a female (single AMELX, no AMELY and SRY copies), and the estimation of the copy numbers seems good. It can furthermore be seen that the *ATP11C* and *IL13RA1* loci (both located on chromosome X) seem to have a higher copy number than loci located on the other chromosomes and that a isochromosome Xq is likely. We thus use the loci on chromosomes 13, 18, 21 and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.33) indicates that all loci display a very high stability. The copy numbers of the chromosome X loci are recalculated using all reference loci. The results are shown in Figure D.34: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *AMELX*, *ATP11C* and *IL13RA1* loci for a isochromosome Xq case.

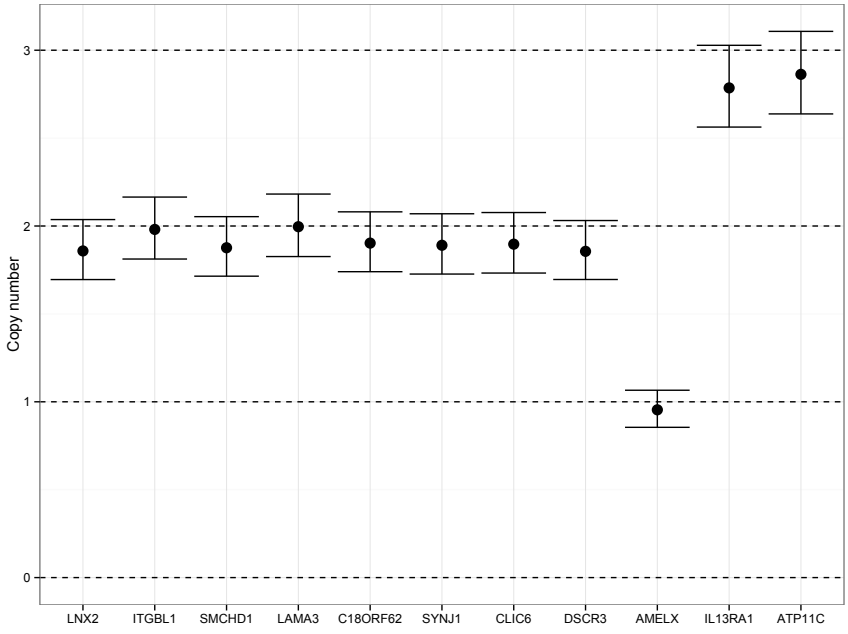


Figure D.32: Sample 19: RPP30 normalization.

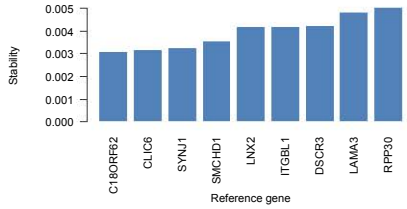


FIGURE D.33: Sample 19: Stability.

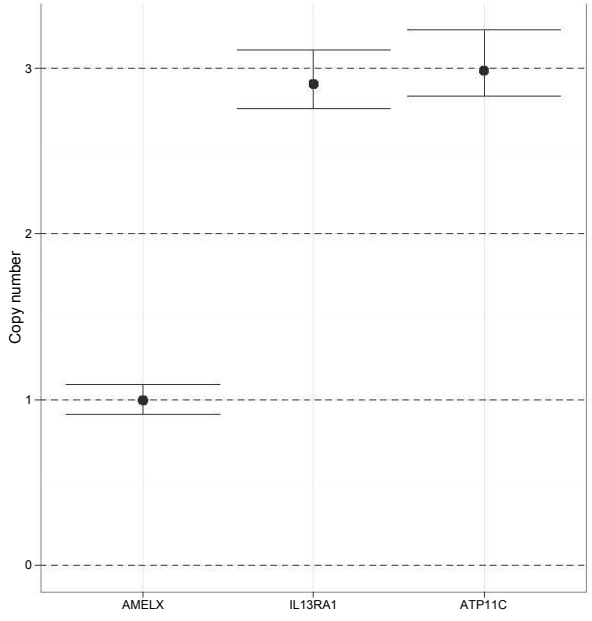


FIGURE D.34: Sample 19: stable reference gene normalization.

D.3.10 Sample 20

The results of normalisation using the *RPP30* locus (accounting for interreplicate variability) is displayed in Figure D.35. We are dealing with a sample belonging to a male (single AMELX, IL13RA1, ATP11C, AMELY and SRY copies), and the estimation of the copy numbers seems good. It can furthermore be seen that the *CLIC6*, *DSCR3* and *SYNJ1* loci (all located on chromosome 21) seem to have a higher copy number than loci located on the other chromosomes and that a chromosome 21 trisomy is likely. We thus use the loci on chromosomes 13, 18, and the locus on *RPP30* as candidate reference loci and determine the stability of these loci as described in the Methods section. The stability plot (Figure D.36) indicates that all loci display a very high stability. The copy numbers of the chromosome 21 loci are recalculated using all reference loci. The results are shown in Figure D.37: the 95% confidence intervals of the multiple reference locus normalized results encompass the expected integer copy number for the *CLIC6*, *DSCR3* and *SYNJ1* loci for a chromosome 21 trisomy case.

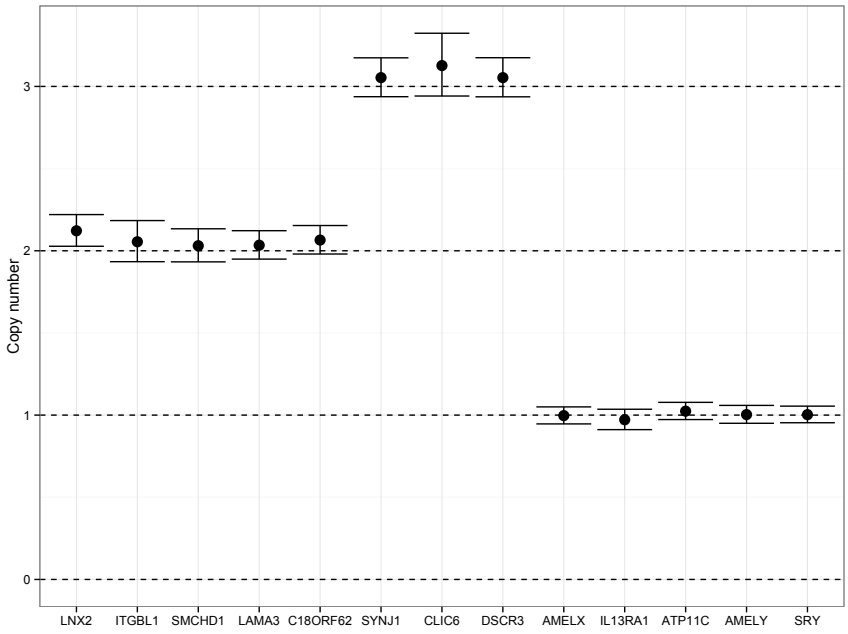


Figure D.35: Sample 20: RPP30 normalization.

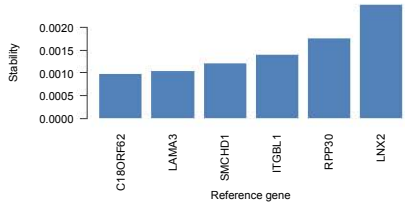


FIGURE D.36: Sample 20: Stability.

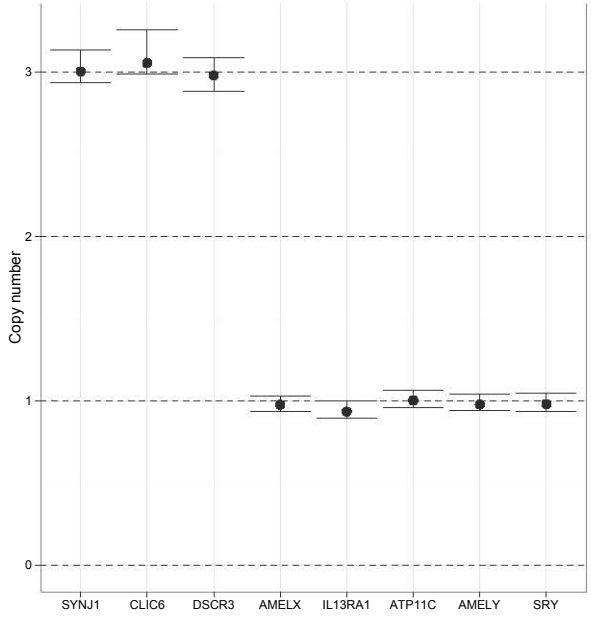


FIGURE D.37: Sample 20: stable reference gene normalization.

D.4 Statistical derivations and considerations

D.4.1 Absolute quantification

Equation 5.11 in the main manuscript suggests the model

$$\log \lambda_i = \beta + R_i,$$

or, equivalently,

$$\lambda_i = \exp(\beta)\exp(R_i),$$

where

$$R_i \sim N(0, \sigma^2).$$

Further

$$\mathrm{E}\{Y \mid R_i\} = \exp(\beta)\exp(R_i)$$

and using the law of iterated expectation

$$\mathrm{E}\{Y\} = \mathrm{E}\{\mathrm{E}\{Y \mid R_i\}\},$$

i.e the expected value of $\mathrm{E}\{Y \mid R_i\}$ over the distribution of R_i.
Further

$$\mathrm{E}\{Y\} = \mathrm{E}\{\mathrm{E}\{Y \mid R_i\}\} = \mathrm{E}\{\exp(\beta)\exp(R_i)\} = \exp(\beta)\mathrm{E}\{\exp(R_i)\}.$$

As R_i is normally distributed, $\exp(R_i)$ is log normally distributed, for which the expectation is given by

$$\mathrm{E}\{\exp(R_i)\} = \exp(0.5\sigma^2).$$

Combining the above, we obtain

$$\mathrm{E}\{Y\} = \exp(\beta)\exp(0.5\sigma^2) = \exp(\beta + 0.5\sigma^2).$$

i.e. equation 5.14 in the main manuscript.

D.4.2 Copy number variation

Similarly derivations can be made for calculating copy numbers according to the different designs as outlined in the manuscript. Concise derivations for each of

the suggested designs are given below.

D.4.2.1 Design A

$$\begin{aligned}
\mathrm{E}\{\mathrm{CNV}\} = \mathrm{E}\left\{\mathrm{E}\left\{\mathrm{CNV}_{i,i'} \mid R_i, R_i'\right\}\right\} &= \mathrm{E}\left\{\exp(-\beta_1 + R_i - R_i')N_b\right\} \\
&= \exp(-\beta_1)\mathrm{E}\left\{\exp(R_i - R_i')\right\}N_b \\
&= \exp(-\beta_1)\exp(\sigma^2)N_b \\
&= \exp(-\beta_1 + \sigma^2)N_b
\end{aligned}$$

D.4.2.2 Design B

$$\begin{aligned}
\mathrm{E}\{\mathrm{CNV}\} = \mathrm{E}\{\mathrm{E}\{\mathrm{CNV}_i \mid R_i\}\} &= \mathrm{E}\{\exp(-\beta_1)N_b\} \\
&= \exp(-\beta_1)N_b
\end{aligned}$$

D.4.2.3 Design C

$$\begin{aligned}
\mathrm{E}\{\mathrm{CNV}\} &= \mathrm{E}\left\{\mathrm{E}\left\{\mathrm{CNV}_{i,i';k} \mid S_k, R_i(0), R_{i'}(k)\right\}\right\} \\
&= \mathrm{E}\{\exp(-\beta_1 - S_k + R_i(0) - R_{i'}(k))N_b\} \\
&= \exp(-\beta_1)\mathrm{E}\{\exp(-S_k + R_i(0) - R_{i'}(k))\}N_b \\
&= \exp(-\beta_1)\exp\left(\frac{1}{2}\sigma_1^2 + \sigma_2^2\right)N_b \\
&= \exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2 + \sigma_2^2\right)N_b
\end{aligned}$$

D.4.2.4 Design D

$$\begin{aligned}
\mathrm{E}\{\mathrm{CNV}\} &= \mathrm{E}\{\mathrm{E}\{\mathrm{CNV}_{i;k} \mid S_k, R_i(k)\}\} \\
&= \mathrm{E}\{\exp(-\beta_1 - S_k)N_b\} \\
&= \exp(-\beta_1)\mathrm{E}\{\exp(-S_k)\}N_b \\
&= \exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2\right)N_b
\end{aligned}$$

D.4.2.5 Design E

$$\begin{aligned}
\mathrm{E}\{\mathrm{CNV}\} &= \mathrm{E}\left\{\mathrm{E}\left\{\mathrm{CNV}_{i,i';k} \mid S_k, R_i(0), R_{i'}(k)\right\}\right\} \\
&= \mathrm{E}\left\{\frac{\mathrm{E}\left\{CNV_{i;k|S_k,R_i}\right\}}{k} + \frac{(k-1)}{k}\mathrm{E}\left\{CNV_{i,i';k} \mid S_k, R_i(0), R_{i'}(k)\right\}\right\} \\
&= \frac{\exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2\right)}{k} + \frac{(k-1)}{k}\exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2 + \sigma_2^2\right) \\
&= \exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2\right) N_b \frac{(1+\exp(\sigma_2^2)(k-1))}{k}
\end{aligned}$$

D.4.2.6 Design F

$$\begin{aligned}
\mathrm{E}\{\mathrm{CNV}\} &= \mathrm{E}\{\mathrm{E}\{\mathrm{CNV}_{i;k} \mid S_k, R_i\}\} \\
&= \mathrm{E}\{\exp(-\beta_1 - S_k)N_b\} \\
&= \exp\left(-\beta_1 + \frac{1}{2}\sigma_1^2\right) N_b
\end{aligned}$$

D.4.3 Confidence intervals

Confidence intervals for the CNVs are calculated in two ways:

1. **Wald confidence intervals**: using the standard errors as estimated by the `glmer` function. A drawback of using this type of confidence interval is that it is not possible to account for the uncertainty of the random effect variance estimates, as no standard error for these are available in the `glmer` output. An approximate confidence interval taking only the standard error of the fixed effects into account can however be calculated. From our experience in dealing with dPCR data, this is typically a reasonable approximation of the more exact parametric bootstrap confidence intervals, with the Wald intervals typically being slightly too narrow and the bootstrap intervals being slightly asymmetric.
2. **Parametric bootstrap confidence intervals**: using the estimated model, parametric bootstraps are taken to estimate parameters in the model a large number (n) times. Confidence intervals for all model parameters can subsequently be calculated as the $[\frac{\alpha}{2}, 1-\frac{\alpha}{2}]$ quantiles of the n estimates of each of the parameters. This may result in better confidence intervals as confidence intervals for random effects are typically asymmetric (lower bound of 0, right skew), among other things. A major drawback is the computationally intensive nature of this approach.

Due to computational constraints, Wald confidence intervals have been used in our case study to calculate stability using a variable number of reference genes. Parametric bootstrap confidence intervals ($n = 200$) have been used to estimate confidence intervals once the reference genes were selected, i.e. for the final confidence intervals as listed in Appendix D.6.2.

D.4.4 Exchangeability

Introduction of the second random effect as in equation 21 (main paper) assumes that both random effects are independent, i.e. the two random effects are exchangeable / not correlated. We believe this is a valid assumption as we see no reason to assume that a given replicate of the target gene will be correlated (either positively or negatively) with a given replicate of one of the reference genes, i.e. that the slope will be changed. For example, if replicate i of the target has a lower than average target concentration, this does not tell anything about the concentration of replicate i' in the reference.

If one does not wish to make this assumption because one feels that this exchangeability may be violated, an alternative model omitting this random effect can be constructed: the reference loci can also be modelled by a set of fixed effect parameters $\beta_1, \ldots, \beta_K$, i.e. a fixed effect for each of the reference loci. Alternatively, one can simply allow for correlation between both random effects.

D.5 Shiny manual

D.5.1 Data management

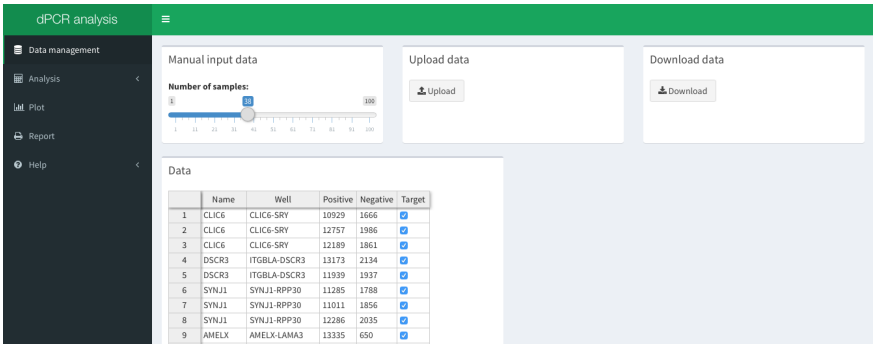


FIGURE D.38: Data management window.

The data management window (Figure D.38) consists of four parts:

- Sample slider input: this slider can be used to indicate the number of samples and will determine the number of rows in the data table for manual data input.
- Upload: the upload button will open the upload dialog, where the user can upload his or her data and load it into the data table. First upload the file from your local hard drive and subsequently click the load button. If loaded succesfully, this will be indicated and the dialog can be closed.

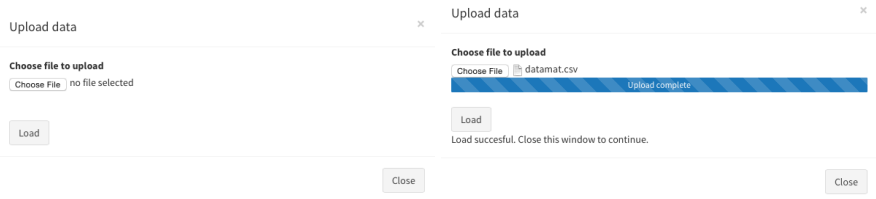


FIGURE D.39: **Left**: Click "choose file" to select a data file. **Right**: Data has been uploaded and loaded succesfully. Close the window.

- Download: data from the data input can be downloaded (e.g. for further use).

- Data: this table contains the data to be analyzed. Data can be typed manually into the table or can be uploaded (see above, click here for an example data file for upload). Each sample should:
    1. have a name (technical replicates of the same assay should have the same name),
    2. have a well name (for duplex/multiplex experiments this well name should be the same, can be left empty for singleplex experiments),
    3. have a number of positive and negative partitions,
    4. is either a target gene (check) or a reference gene (uncheck, for relative quantification or copy number variation).

D.5.2 Analysis

Currently three scenarios are implemented: absolute quantification, copy number variation (or relative quantification, for all cases described in the manuscript) and the within-sample reference gene stability (for copy number variation) procedure.

D.5.2.1 Absolute quantification

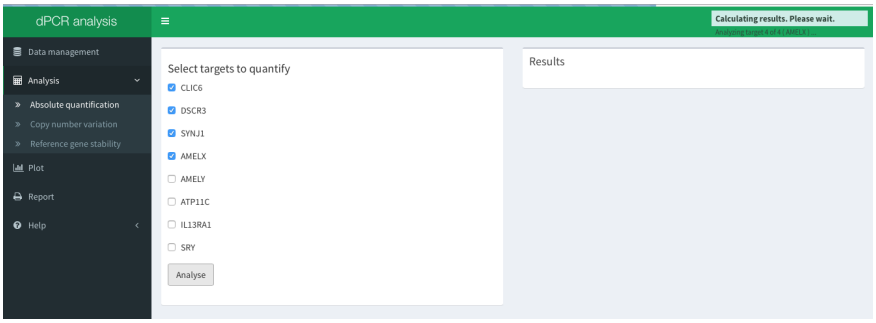


FIGURE D.40: Absolute quantification window.

Samples indicated as targets will appear as samples to be quantified. Samples that should not be quantified can be deselected. Upon clicking the "analyse" button, a progress bar will appear on top detailing the progress of the calculations (see Figure D.40).
If calculations have finished, the estimates (copies per partition) along with their 95% confidence intervals will appear in the right window (Figure D.41).

Results

	Name	Estimate	95% CI lower bound	95% CI upper bound
1	DSCR3	0.1502	0.1457	0.1549
2	SYNJ1	0.1520	0.1481	0.1561
3	CLIC6	0.1430	0.1392	0.1468
4	AMELX	0.0453	0.0431	0.0476

FIGURE D.41: Absolute quantification results.

D.5.2.2 Copy number variation or relative quantification

Using a suitable subset of reference genes (see reference gene stability), copy numbers or relative quantities can be determined. Targets to be quantified can

be selected, along with stable reference genes. All targets will then be quantified, relative to these reference genes. Clicking the "analyse" button will start the calculations and the progress will be indicated by a progress bar on top. When calculations have finished, a table displaying the estimated copy number or relative quantity along with a 95% confidence interval will be shown (Figure D.42).

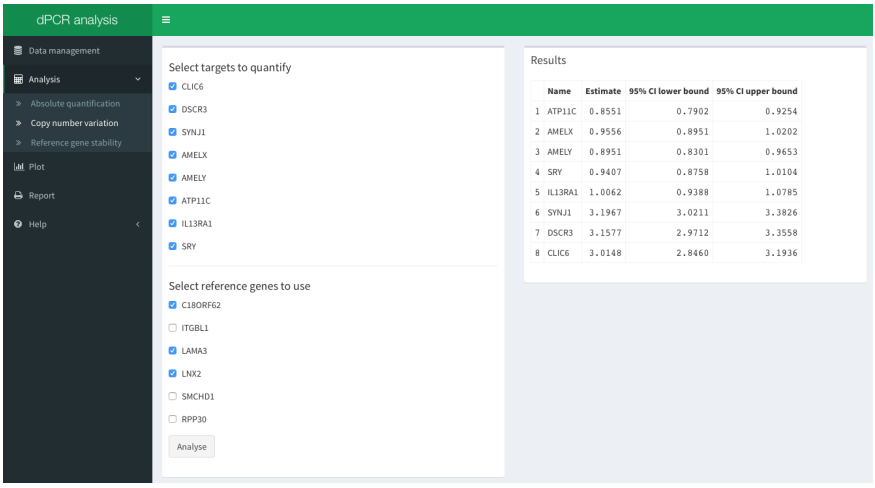


FIGURE D.42: Copy number/relative quantification window with results.

D.5.2.3 Reference gene stability

Reference gene stability can be determined as described in the main paper. All candidate reference genes or a subset of them can be used in the stability determination procedure. Select all candidate reference genes and press the "analyse" button. A progress bar will appear on the top of the window (Figure D.43).

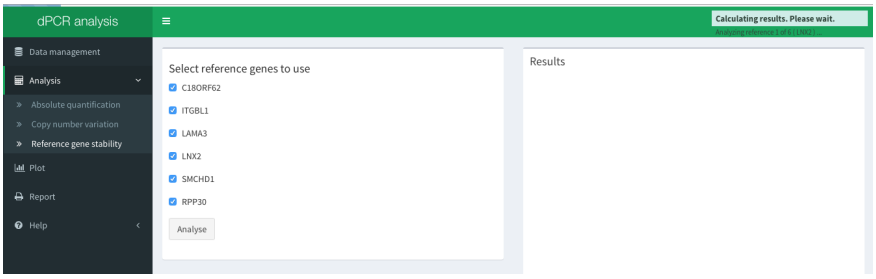


FIGURE D.43: Reference gene stability window.

When calculations have finished, the stability will be displayed in an ordered barchart (high stability (low deviance) to low stability (high deviance)). A table displaying the stability for each candidate reference gene is displayed (Figure D.44).

Results

	Gene	Stability
1	LNX2	0.0051
2	SMCHD1	0.0231
3	RPP30	0.0107
4	ITGBL1	0.0078
5	LAMA3	0.0056
6	C18ORF62	0.0056

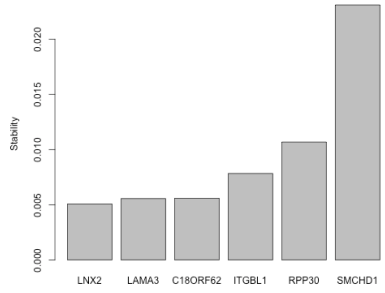


FIGURE D.44: Reference gene stability results (less is better, in this case LNX2 is the most stable reference locus).

D.5.3 Figures

After calculating absolute or relative quantities or a copy number, the obtained results can be plotted in the "plot window" (Figure D.45). The order of plotting is alphabetical. If you want a specific ordering, this can be done by adding a suitable prefix to the names.

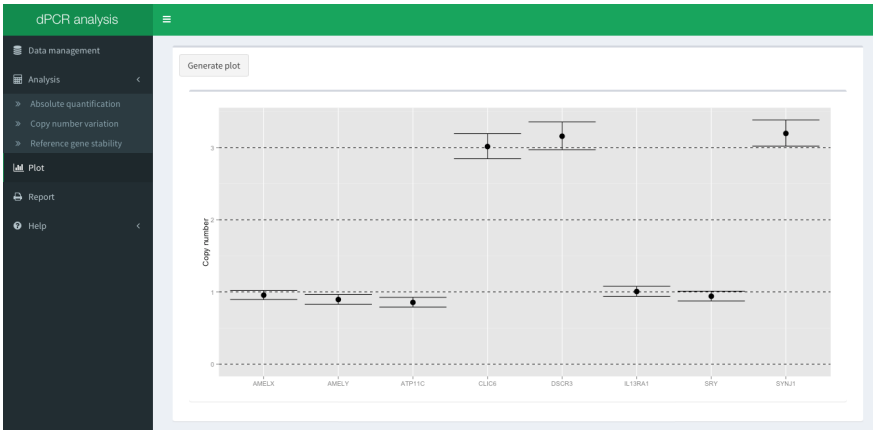


FIGURE D.45: Figures window.

D.5.4 Report

A report can be generated in the report window. The title of the report, along with its contents (data, results and/or figures) and the format of the report can be specified. Click the "download report" button to save a report to your hard drive (Figure D.46).

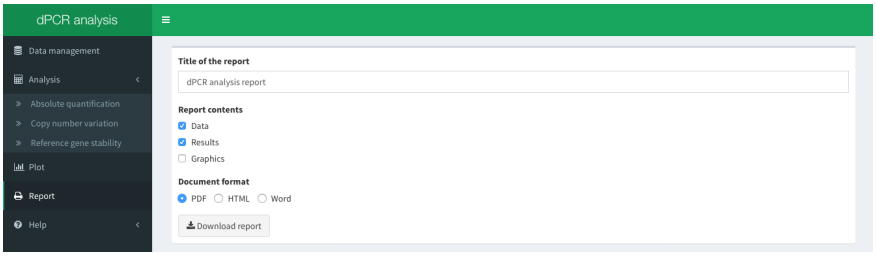


FIGURE D.46: Report window.

D.5.5 Help

A frequently asked questions (FAQ) page will be developed.

D.6 Supplementary tables

D.6.1 PCR assays

Gene symbol	Chromosomal position	Forward primer sequence	Reverse primer sequence	Hydrolysis probe sequence	Label	Quencher
AMELX	Xp	ACTCCATGACTCCAATCCAA	GCTTGGTCTTGTCTGTTGAT	ACCAGCCAAACCTCCCTCCG	FAM	Zen/Lowa Black
AMELY	Yp	TTGCTAAGTTAAGTGATTGTAAGC	CCAGAATATGAGACAGGAACTG	TGCTTGTGTCATCCCTGAAAGCA	FAM	Zen/Lowa Black
ATP11C	Xq	TCTTCATTAAGGTCTGATGTGT	ATGTTCACCGACTTCCTATC	TCCTGTCTCCATGTACGTCACAGT	HEX	Zen/Lowa Black
C18orf62	18q	AAATATGCTGTTGGCCCTTT	GTTTCTTCACATTGTTGGTTCT	TCTTCAGAAACCCCTTGTATCCTGCT	HEX	Zen/Lowa Black
CLIC6	21q	TCAGCCAGAATGTGTTCTTC	CACATACCACAACCAACTCT	TCGTACACATCTTACAGAAAACCCTCCT	FAM	Zen/Lowa Black
DSCR3	21q	AGCTCCACGCTTCTGAT	ATTTCTCCTTCGAGGACATC	CTCCACCACCAGCTCTCCCG	HEX	Zen/Lowa Black
IL13RA1	Xq	TGTTCTTCCTGATACTTTGAACA	CTCATTTCTTGGCTCCAATTAC	TCCAGAGTTTGTCATCCTCATAGCATAACT	FAM	Zen/Lowa Black
ITGBL1	13q	GGACTTGTGTATGTGGTGAA	CTCTACAGTCCCTCTCATCA	ACCTGTCACGATGTTGATCCGACT	FAM	Zen/Lowa Black
LAMA3	18q	CATATGGTGGTTACCTCACT	CTGTACATCCGGCTTCTT	TCCAGAAGAACCATGTCGCCAGG	HEX	Zen/Lowa Black
LNX2	13q	AACAATCTGGACCTCAAGTG	GCAGTTGCAATGCTGAAAG	CGTCCCCTGCTGTTGCCCTT	HEX	Zen/Lowa Black
SMCHD1	18p	CTACGACAACCATACTGGAAT	GATTTCAGCCGAACCATTCA	TGGGAAAGTTAGAACACTTGAATTCCCC	FAM	Zen/Lowa Black
SRY	Yp	CGCTGTTGAATACGCTTAAC	ACTCTGGCACCTTTCAATTT	TGTCGCACTCTCCTTGTTTTTGACA	HEX	Zen/Lowa Black
SYNJ1	21q	TGATAGTAACGGGCTCTTCT	AACCCATTTAGAGCCAAGTC	TCAGAGGCAACTTCATGGTTCTCCA	FAM	Zen/Lowa Black
RPP30	10q	GATTTGGACCTGCGAGCG	GCGGCTGTCTCCACAAGT	CTGACCTGAAGGCTCT	VIC	MGB

Table D.5: Details of the PCR assays used in the case study.

D.6.2 Final copy number estimates and confidence intervals

	Control 1	Control 2	Control 3	Control 4	Sample 5	Sample 5 RPP30	Sample 6	Sample 6 RPP30	Sample 9	Sample 9 RPP30	Sample 13
RPP30	ref	ref	ref	ref	no ref	ref	ref	ref	no ref	ref	no ref
LNX2	1.88 (1.78-1.98)	2.01 (1.92-2.11)	1.91 (1.82-1.99)	1.89 (1.79-1.98)	ref	1.88 (1.79-1.97)	ref	1.98 (1.90-2.06)	ref	1.84 (1.74-1.93)	ref
ITGBL1	1.92 (1.82-2.02)	1.95 (1.86-2.04)	1.95 (1.87-2.03)	1.89 (1.80-1.97)	no ref	1.97 (1.85-2.09)	ref	2.00 (1.91-2.09)	no ref	2.19 (2.11-2.27)	ref
SMCHD1	1.92 (1.83-2.01)	1.97 (1.89-2.06)	1.98 (1.91-2.06)	1.91 (1.82-1.99)	no ref	1.66 (1.56-1.75)	ref	1.96 (1.88-2.04)	ref	1.84 (1.75-1.93)	2.46 (2.38-2.63)
LAMA3	1.93 (1.83-2.02)	1.93 (1.84-2.02)	1.94 (1.86-2.02)	1.91 (1.82-2.00)	ref	1.83 (1.75-1.93)	no ref	2.05 (1.97-2.13)	no ref	1.77 (1.67-1.87)	2.95 (2.88-3.13)
C18ORF62	1.94 (1.84-2.04)	2.01 (1.92-2.09)	2.00 (1.92-2.08)	2.07 (1.97-2.16)	ref	1.91 (1.77-2.05)	no ref	2.05 (1.97-2.13)	no ref	2.03 (1.95-2.11)	2.95 (2.88-3.14)
SYNJ1	2.01 (1.90-2.11)	1.99 (1.90-2.08)	2.00 (1.92-2.08)	1.97 (1.87-2.06)	3.20 (3.11-3.31)	2.99 (2.91-3.07)	2.95 (2.89-3.05)	2.91 (2.84-2.99)	ref	1.88 (1.80-1.96)	ref
CLIC6	1.96 (1.87-2.05)	2.03 (1.94-2.11)	1.95 (1.87-2.03)	1.92 (1.82-2.01)	3.01 (2.93-3.13)	2.81 (2.73-2.89)	3.13 (3.05-3.23)	3.09 (3.01-3.16)	no ref	1.77 (1.68-1.87)	ref
DSCR3	2.02 (1.93-2.12)	2.01 (1.92-2.10)	1.92 (1.84-2.00)	1.86 (1.77-1.94)	3.16 (3.07-3.29)	2.96 (2.87-3.04)	2.98 (2.88-3.10)	2.94 (2.86-3.02)	ref	1.86 (1.77-1.94)	ref
AMELX	1.96 (1.87-2.06)	1.95 (1.86-2.04)	0.98 (0.88-1.07)	0.93 (0.83-1.04)	0.95 (0.91-1.01)	0.89 (0.78-1.00)	ref	1.98 (1.87-2.03)	0.96 (0.91-1.02)	0.88 (0.77-1.00)	0.79 (0.75-0.84)
IL13RA1	1.88 (1.77-1.98)	1.99 (1.89-2.08)	0.93 (0.82-1.05)	0.95 (0.84-1.06)	1.00 (0.95-1.06)	0.94 (0.83-1.05)	no ref	1.92 (1.84-2.00)	3.30 (3.18-3.47)	3.05 (2.98-3.14)	0.93 (0.90-1.00)
ATP11C	1.93 (1.83-2.02)	1.91 (1.82-1.99)	0.97 (0.87-1.06)	0.92 (0.82-1.02)	0.84 (0.80-0.92)	0.79 (0.67-0.91)	no ref	1.88 (1.80-1.96)	3.28 (3.19-3.44)	3.04 (2.96-3.11)	0.75 (0.72-0.81)
AMELY	-	-	1.01 (0.92-1.11)	1.02 (0.92-1.12)	0.89 (0.85-0.95)	0.84 (0.70-0.99)	-	-	-	-	0.91 (0.88-0.98)
SRY	-	-	0.99 (0.90-1.08)	0.95 (0.83-1.07)	0.94 (0.90-0.99)	0.88 (0.77-0.99)	-	-	-	-	1.07 (1.03-1.15)
	Sample 13 RPP30	Sample 15	Sample 15 RPP30	Sample 16	Sample 16 RPP30	Sample 18	Sample 18 RPP30	Sample 19	Sample 19 RPP30	Sample 20	Sample 20 RPP30
RPP30	ref	no ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
LNX2	1.26 (1.19-1.33)	ref	1.72 (1.64-1.80)	no ref	2.03 (1.95-2.12)	ref	1.94 (1.85-2.03)	ref	1.86 (1.67-2.04)	ref	2.12 (2.03-2.21)
ITGBL1	1.26 (1.19-1.34)	ref	1.82 (1.74-1.90)	no ref	2.07 (1.98-2.16)	ref	1.91 (1.82-2.00	ref	1.98 (1.80-2.16)	ref	2.06 (1.97-2.14)
SMCHD1	1.48 (1.41-1.55)	2.83 (2.79-3.06)	2.50 (2.43-2.58)	no ref	1.82 (1.73-1.91)	2.83 (2.73-2.93)	2.77 (2.69-2.86)	ref	1.88 (1.70-2.06)	ref	2.03 (1.95-2.11)
LAMA3	1.78 (1.71-1.85)	2.99 (2.93-3.19)	2.62 (2.54-2.69)	ref	1.94 (1.85-2.02)	2.92 (2.84-3.03)	2.87 (2.78-2.95)	ref	2.00(1.82-2.17)	ref	2.03 (1.95-2.12)
C18ORF62	1.78 (1.71-1.84)	3.00 (2.95-3.21)	2.66 (2.59-2.74)	ref	1.97 (1.88-2.06)	2.69 (2.61-2.80)	2.64 (2.55-2.73)	ref	1.90 (1.72-2.08)	ref	2.07 (1.98-2.15)
SYNJ1	1.16 (1.09-1.24)	ref	1.79 (1.71-1.86)	2.98 (2.89-3.08)	2.93 (2.85-3.01)	ref	1.98 (1.89-2.07)	ref	1.89 (1.71-2.07)	3.00 (2.93-3.14)	3.05 (2.98-3.13)
CLIC6	1.18 (1.11-1.26)	ref	1.79 (1.71-1.87)	3.08 (2.99-3.21)	3.03 (2.95-3.11)	ref	1.97 (1.88-2.06)	ref	1.90 (1.72-2.08)	3.05 (2.99-3.26)	3.12 (3.05-3.20)
DSCR3	1.17 (1.10-1.25)	ref	1.83 (1.75-1.91)	2.95 (2.86-3.06)	2.90 (2.81-2.98)	ref	1.97 (1.89-2.06)	ref	1.86 (1.67-2.04)	2.98 (2.88-3.09)	3.05 (2.98-3.13)
AMELX	0.47 (0.37-0.58)	no ref	1.47 (1.39-1.55)	1.00 (0.95-1.05)	0.98 (0.88-1.09)	0.95 (0.91-1.00)	0.93 (0.82-1.04)	1.00 (0.91-1.09)	0.95 (0.73-1.18)	0.98 (0.93-1.03)	1.00 (0.89-1.10)
IL13RA1	0.56 (0.47-0.66)	ref	1.69 (1.61-1.77)	0.97 (0.93-1.02)	0.96 (0.85-1.06)	0.98 (0.94-1.04)	0.96 (0.85-1.07)	2.91 (2.76-3.11)	2.79 (2.62-2.95)	0.93 (0.89-1.00)	0.97 (0.86-1.09)
ATP11C	0.45 (0.35-0.56)	no ref	1.54 (1.46-1.63)	0.94 (0.90-1.00)	0.93 (0.82-1.03)	0.91 (0.88-0.96)	0.90 (0.79-1.01)	2.99 (2.83-3.23)	2.86 (2.70-3.03)	1.00 (0.96-1.06)	1.02 (0.92-1.13)
AMELY	0.55 (0.45-0.65)	-	-	0.96 (0.92-1.01)	0.94 (0.83-1.05)	0.86 (0.82-0.91)	0.84 (0.72-0.96)	-	-	0.98 (0.94-1.04)	1.00 (0.90-1.11)
SRY	0.65 (0.56-0.73)	-	-	0.98 (0.94-1.04)	0.96 (0.86-1.07)	0.95 (0.91-1.01)	0.93 (0.81-1.04)	-	-	0.98 (0.93-1.05)	1.00 (0.90-1.10)

TABLE D.6: Estimated copy numbers for all samples and all loci, using either multiple stable reference lco or the RPP30 locus. In case of multiple reference loci, "ref" indicates a stable reference locus (i.e. a locus used as reference), while "no ref" indicates an unstable reference locus (i.e. a locus not used as reference).

E Volume

E.1 Algorithm for obtaining an unbiased estimate for absolute quantification

1. Obtain a large number of volumes v_i using rejection sampling from $\hat{g}$. Set $k = 1$.
2. Generate counts N_i from Poisson distributions with mean $\hat{\lambda}_{k-1} \frac{v_i}{\bar{v}}$. Digitalize the counts:

$$Y_i = \min(N_i, 1) = \begin{cases} 0 & \text{if } N_i = 0 \\ 1 & \text{otherwise.} \end{cases}$$

3. Using these Y_i, obtain an estimate $\hat{\lambda}_{\text{biased}}$ of λ_{k-1} from Equation 3 (main text).
4. Just as $\hat{\lambda}_0$ is an underestimate of λ, $\hat{\lambda}_{\text{biased}}$ is an underestimate of $\hat{\lambda}_{k-1}$. Unlike λ, we know the value of $\hat{\lambda}_0$, so that a bias b_k can be estimated as $b_k = \hat{\lambda}_0 - \hat{\lambda}_{\text{biased}}$.
5. b_k will be an underestimate of the true bias of $\hat{\lambda}_0$ (as an estimate of λ) as $\hat{\lambda}_{k-1} < \lambda$, so that $\hat{\lambda}_k = \hat{\lambda}_0 + b_k$ is a better, yet not unbiased estimate of λ.
6. Iterate through steps 2-5 until convergence ($|b_k| < \varepsilon$, with ε very small). Denote $\hat{\lambda}_{\text{final}} = \hat{\lambda}_k$. If convergence is not attained in K iterations, $\hat{\lambda}_{\text{final}}$ is set to the average of those $\hat{\lambda}_k$ with minimal bias.

E.2 A numerical example of obtaining an unbiased estimate

Suppose that the true value of λ is 1. A first estimate $\hat{\lambda}_0$ of λ is obtained, but as we ignore the volume variability, $\hat{\lambda}_0$ will have a downward bias, e.g. $\hat{\lambda}_0 = 0.9$. We cannot compute this bias, as in real life we do not know the value of λ.

Next, we obtain an estimate of $\hat{\lambda}_0$, say $\hat{\lambda}_{\text{biased}}$. Again, we ignore the volume variability, so that $\hat{\lambda}_{\text{biased}}$ will be an underestimate of $\hat{\lambda}_0$, just as $\hat{\lambda}_0$ was an underestimate of λ. For example, $\hat{\lambda}_{\text{biased}} = 0.81$.

We do know $\hat{\lambda}_0$, so that we can calculate the bias b_1 of $\hat{\lambda}_{\text{biased}}$ as an estimate of $\hat{\lambda}_0$, this is $b_1 = \hat{\lambda}_0 - \hat{\lambda}_{\text{biased}} = 0.90 - 0.81 = 0.09$. Suppose that we set $\varepsilon = 0.001$, then the algorithm has not converged, as $|\hat{\lambda}_0 - \hat{\lambda}_{\text{biased}}| \geq \varepsilon$.

A better estimate of λ is $\hat{\lambda}_1 = \hat{\lambda}_0 + b_1 = 0.99$.

We reiterate the procedure: we obtain an estimate of $\hat{\lambda}_1$, ignoring volume variability, so that we obtain a biased estimate $\hat{\lambda}_{\text{biased}} = 0.89$. We know the value of our initial estimate $\hat{\lambda}_0$, so to see how close $\hat{\lambda}_1$ is to λ, we can again calculate a bias b_2 as $b_2 = \hat{\lambda}_0 - \hat{\lambda}_{\text{biased}} = 0.90 - 0.89 = 0.01$. We update our estimate of λ: $\hat{\lambda}_2 = \hat{\lambda}_1 + \hat{\lambda}_{\text{biased}} = 0.99 + 0.01 = 1.00$. Still, $|\hat{\lambda}_0 - \hat{\lambda}_{\text{biased}}| \geq \varepsilon$, so convergence has not been attained.

We reiterate the procedure: we obtain an estimate of $\hat{\lambda}_2$, ignoring volume variability, so that we obtain a biased estimate $\hat{\lambda}_{\text{biased}} = 0.90$. We know the value of our initial estimate $\hat{\lambda}_0$, so to see how close $\hat{\lambda}_2$ is to λ, we can again calculate a bias b_3 as $b_3 = \hat{\lambda}_0 - \hat{\lambda}_{\text{biased}} = 0.90 - 0.90 = 0.00$. As $|b_3| < \varepsilon$, the algorithm has converged at the value λ_2 that will be identical or close to the true value of λ.

E.3 Approach for relative quantification

E.3.1 Algorithm for obtaining unbiased estimates

The procedure is as for absolute quantification, but the process is performed twice: once for the target nucleic acid and once the reference nucleic acid. This yields unbiased estimates for both the target ($\hat{\lambda}_{\text{final,target}}$) and reference ($\hat{\lambda}_{\text{final,reference}}$) nucleic acid copies per average volume.

E.3.2 Obtaining conditional densities

Based on $\hat{\lambda}_{\text{final,target}}$ and $\hat{\lambda}_{\text{final,reference}}$ estimates of

- $f(v \mid Y_{\text{target}} = 0, Y_{\text{reference}} = 0)$,
- $f(v \mid Y_{\text{target}} = 1, Y_{\text{reference}} = 0)$,
- $f(v \mid Y_{\text{target}} = 0, Y_{\text{reference}} = 1)$,
- $f(v \mid Y_{\text{target}} = 1, Y_{\text{reference}} = 1)$,

i.e. the four conditional densities of the volume may be obtained as follows:

1. generate counts $N_{i,\text{target}}$ and $N_{i,\text{reference}}$ from Poisson distributions with mean $\hat{\lambda}_{\text{final,target}} \frac{v_i}{\bar{v}}$ and $\hat{\lambda}_{\text{final,reference}} \frac{v_i}{\bar{v}}$ respectively.
2.
   - v_i resulting in $N_{i,\text{target}} = 0$ and $N_{i,\text{reference}} = 0$ make up $f(v \mid Y_{\text{target}} = 0, Y_{\text{reference}} = 0)$,
   - v_i resulting in $N_{i,\text{target}} > 0$ and $N_{i,\text{reference}} = 0$ make up $f(v \mid Y_{\text{target}} = 1, Y_{\text{reference}} = 0)$,
   - v_i resulting in $N_{i,\text{target}} = 0$ and $N_{i,\text{reference}} > 0$ make up $f(v \mid Y_{\text{target}} = 0, Y_{\text{reference}} = 1)$,
   - v_i resulting in $N_{i,\text{target}} > 0$ and $N_{i,\text{reference}} > 0$ make up $f(v \mid Y_{\text{target}} = 1, Y_{\text{reference}} = 1)$,

Volumes for each of the observed outcomes may again be obtained using rejection sampling, and used as an offset in a generalized linear model as described next.

E.3.3 A model for relative quantification estimates and associated variance

For relative quantification the model for the estimation of the absolute concentration is supplemented with an additional parameter to allow estimation of both the target and reference concentration:

$$\log(\lambda) = \beta_0 + X_1\beta_1 + \log(v),$$

where X_1 is a dummy variable that encodes the target ($X_1 = 1$) or reference ($X_1 = 0$) nucleic acid.
An estimate of the relative quantity (RQ) is then

$$\widehat{RQ} = \exp(\hat{\beta}_1),$$

and an estimate of the variance on the RQ, by application of the delta method, is

$$\mathrm{Var}(\widehat{RQ}) \approx \widehat{RQ}^2 \mathrm{Var}(\hat{\beta}_1).$$

E.4 Location - shape - scale

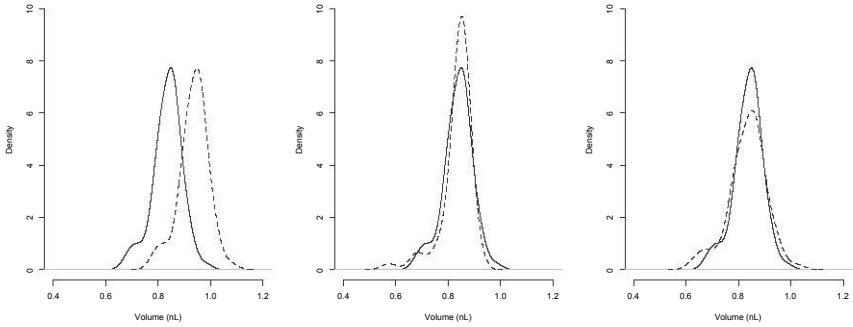


FIGURE E.1: Graphical representation of the concepts location - shape - scale. (Left) The distributions have the same scale and shape, but a different location. (Center) The distributions have the same scale and location, but a different shape. (Right) The distributions have the same shape and location, but a different scale.

E.5 Simulation results

E.5.1 First distribution

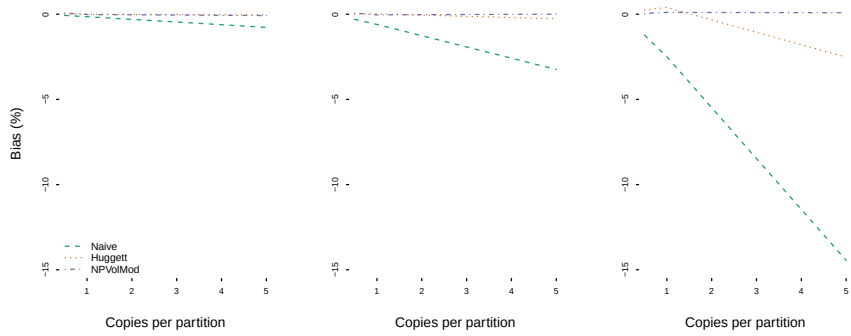


FIGURE E.2: Bias on concentration estimate for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

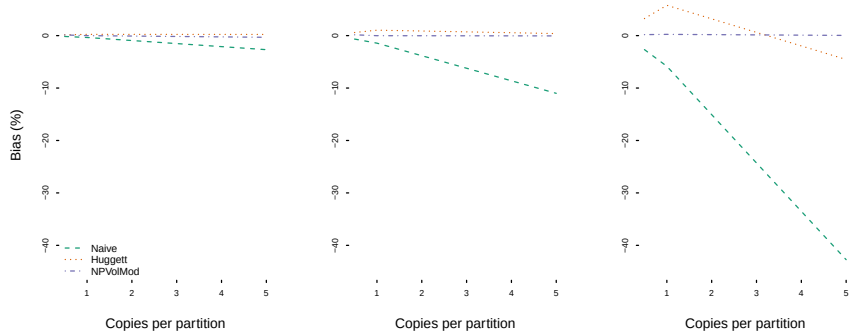


FIGURE E.3: Bias on variance of concentration estimate when ignoring volume variability for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

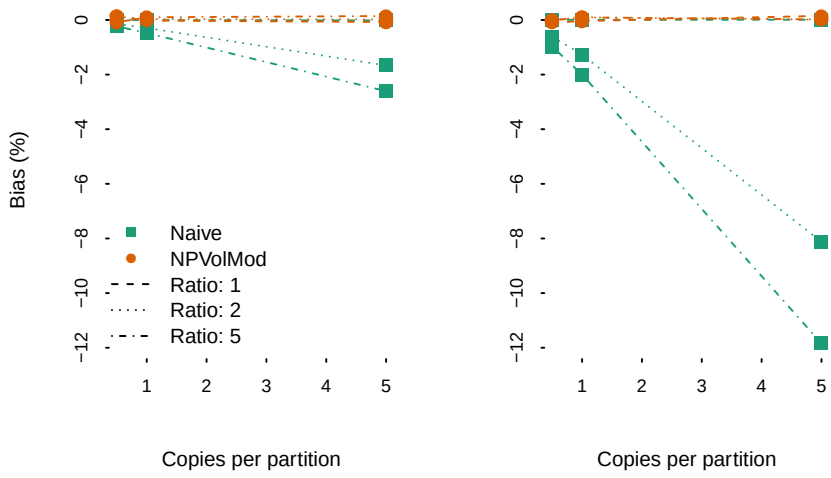


FIGURE E.4: Bias on copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

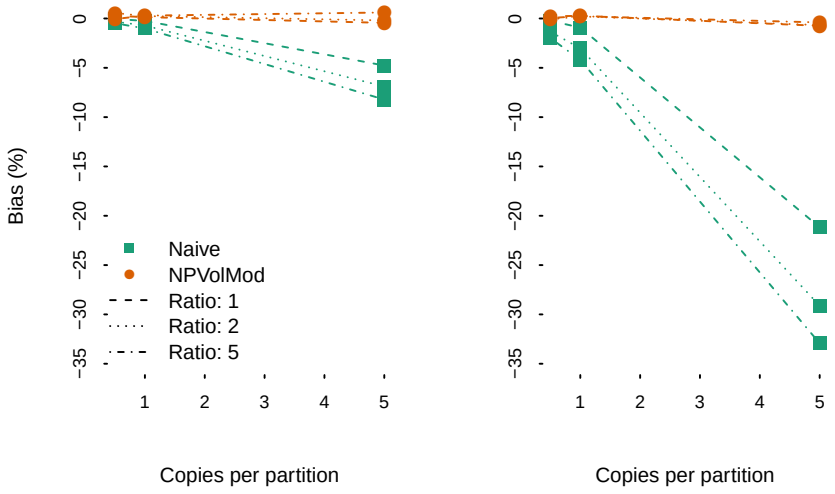


FIGURE E.5: Bias on variance of copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

E.5.2 Second distribution

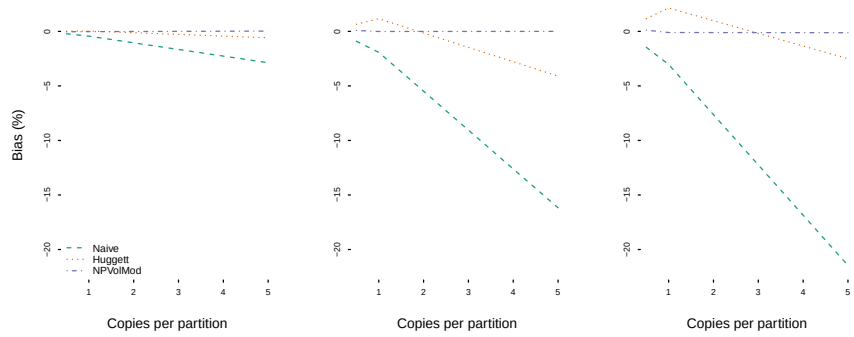


FIGURE E.6: Bias on concentration estimate for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

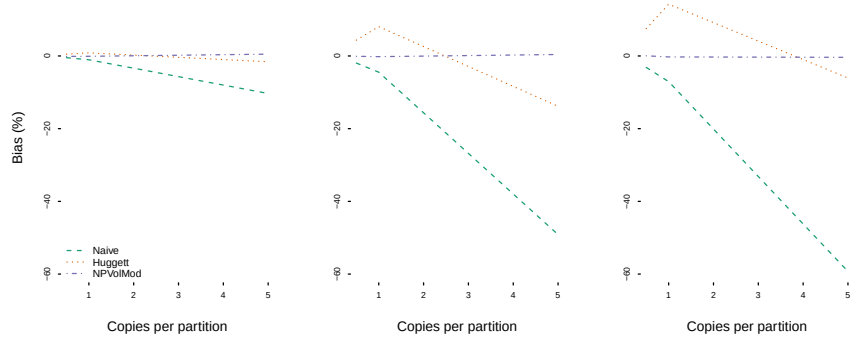


FIGURE E.7: Bias on variance of concentration estimate when ignoring volume variability for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

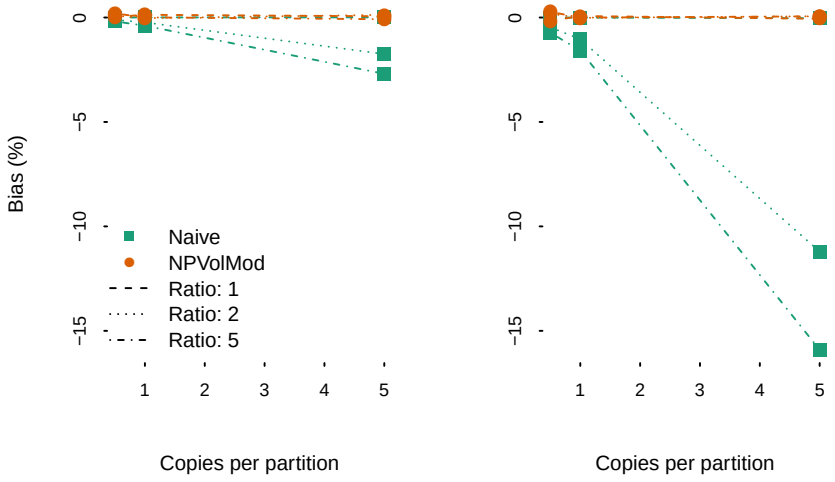


FIGURE E.8: Bias on copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

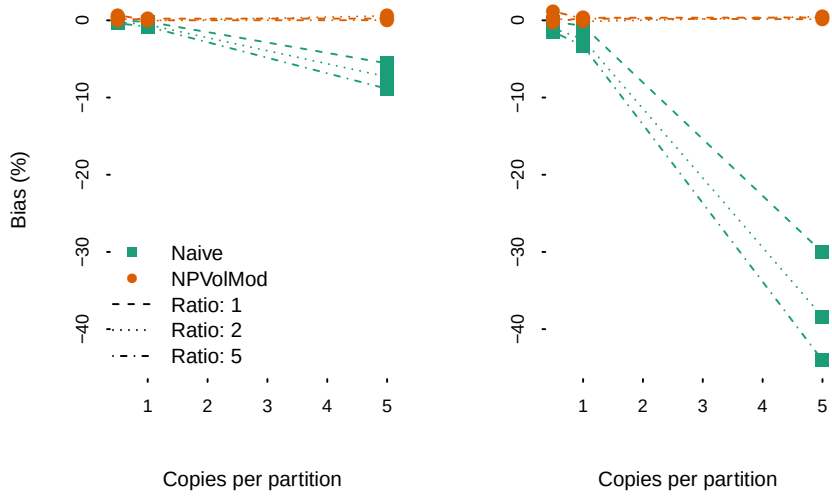


FIGURE E.9: Bias on variance of copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

E.5.3 Third distribution

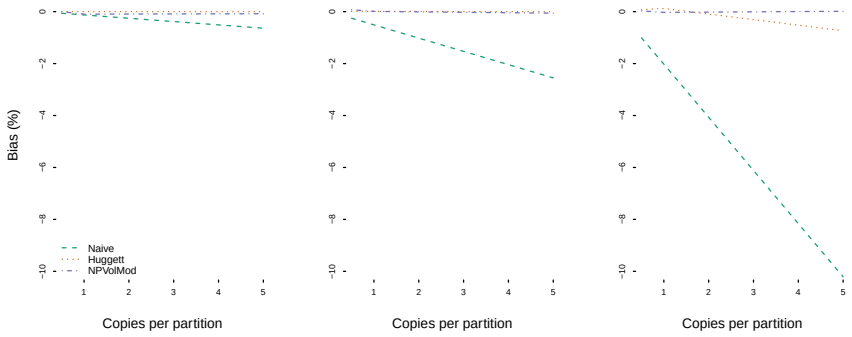


FIGURE E.10: Bias on concentration estimate for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

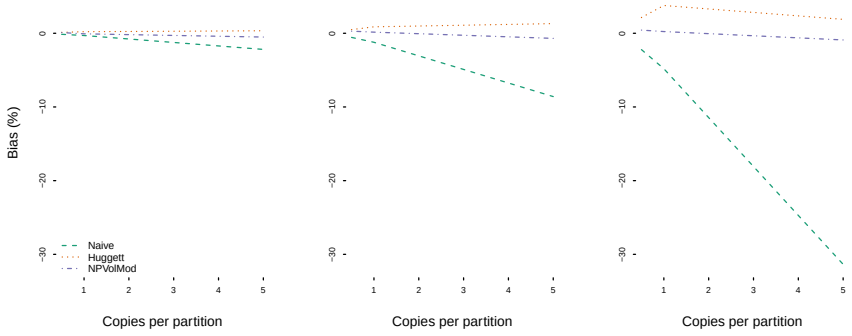


FIGURE E.11: Bias on variance of concentration estimate when ignoring volume variability for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

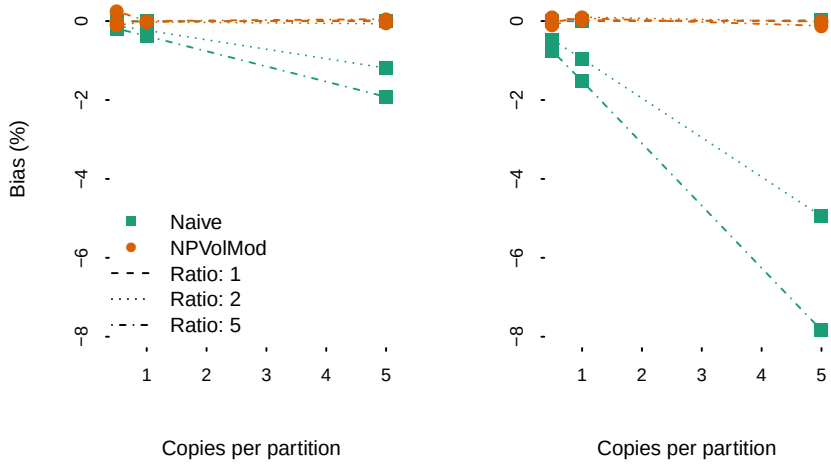


FIGURE E.12: Bias on copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

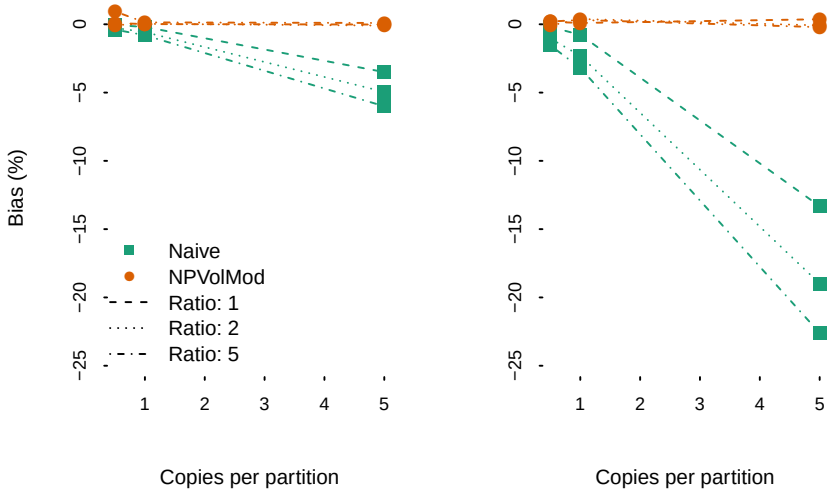


FIGURE E.13: Bias on variance of copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

E.5.4 Fourth distribution

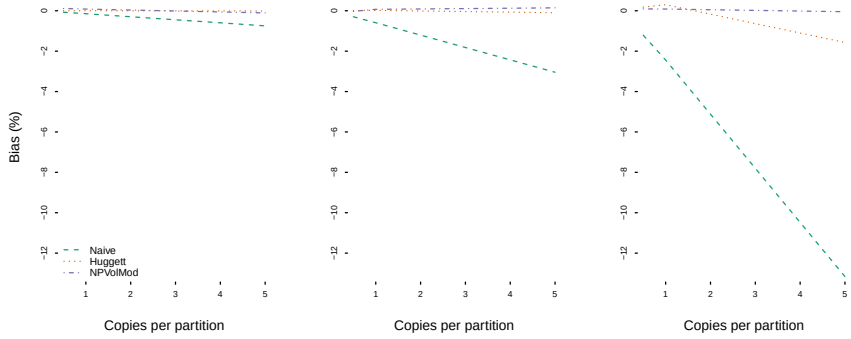


FIGURE E.14: Bias on concentration estimate for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

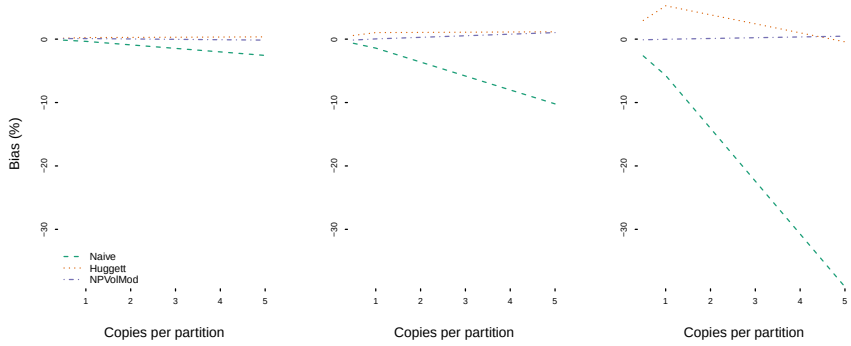


FIGURE E.15: Bias on variance of concentration estimate when ignoring volume variability for absolute quantification. Simulation results are presented for volume variability of 5%, 10% and 20% relative standard deviation. The naive approach ignores volume variability, Huggett's approach models the volume variability as a gamma distribution, and the new approach uses a non-parametric estimate of the volume variability.

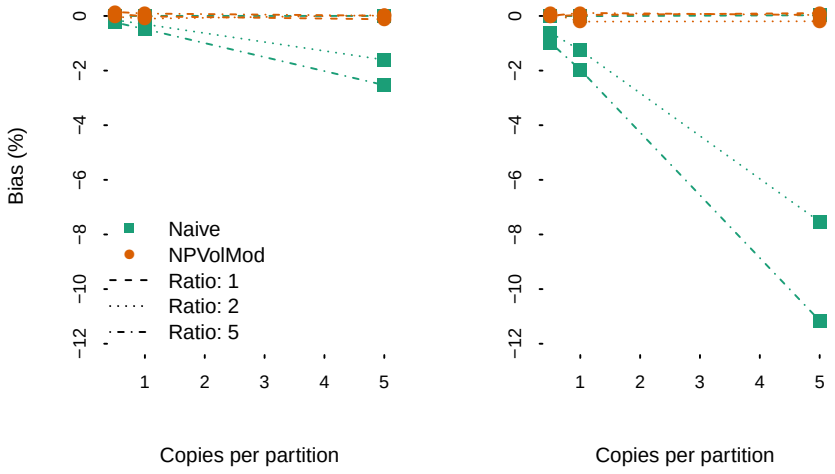


FIGURE E.16: Bias on copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

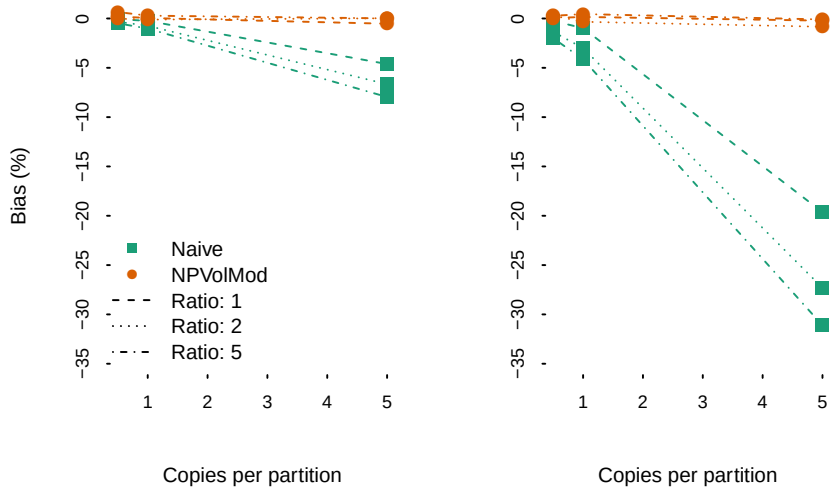


FIGURE E.17: Bias on variance of copy number estimate when ignoring volume variability for relative quantification. Simulation results are presented for volume variability of 10% and 20% relative standard deviation. The naive approach ignores volume variability, and the new approach uses a non-parametric estimate of the volume variability.

F Quality

F.1 Data simulation

Concentrations from an HIV genome fragment dilution series were obtained from a publicly available source (Jones et al., 2016). The aim of the experiment was to determine the dynamic range of an HIV assay on the RainDance RainDrop dPCR platform. The dilution series consisted of $I = 7$ ten-fold dilutions, ranging from $c_1 = 50$ million copies down to $c_7 = 50$ copies.
At each of these dilution levels $i = 1, \ldots, I$, $J = 5$ replicates ($j = 1, \ldots, J$) were quantified, thus resulting in a total of $I * J = 35$ concentration estimates.
The sample variances σ_i^2 of the five ($j = 1, \ldots, J$) observed concentrations at a given dilution point i were calculated. This was done for each of the seven ($i = 1, \ldots, I$) dilution levels, thus resulting in an estimate of the variability $\hat{\sigma}_i^2$ of measurements at each of the seven ($i = 1, \ldots, I$) dilution levels.
The expected concentrations $c_1, \ldots, c_7$ and the estimated sample variances $\hat{\sigma}_1^2, \ldots, \hat{\sigma}_7^2$ observed at those concentrations then served as a model from which dilution series data were simulated, assuming a normal distribution at each of the dilution levels: a given number of replicates k ($k = 2, \ldots, 8$) were generated from a normal distributions with mean parameter c_i and variance $\hat{\sigma}_i^2$.

F.2 Ordinary versus (robust) weighted least squares

Issues with common approaches

The type I error rate control of the ordinary least squares (OLS) approach and the weighted least squares (WLS) approach, with weights estimated as the inverse sample variances, was investigated (Figure 1).
The results indicate that

- the OLS approach is not suitable, because it suffers from an incorrect type I error rate for various measures, as a consequence of high leverage of points

with the highest concentration and violation of the homoscedasticity assumption. As a bias often occurs at these highest concentrations, this is problematic, and hence, an OLS approach should be discouraged.

- the WLS approach downweights these problematic observations, as a consequence of their higher variability. Additionally, this is a more correct model as it deals with the heteroscedasticity. However, it suffers from an inflated type I error rate (although often less severely than the OLS approach), because the exact weights are unknown, and they need to be estimated from the data (see e.g. Carroll and Cline (1988)). This inflation of the type I error rate is most severe for the very small sample sizes, which are most relevant in lab practice.

Alternative approaches

An alternative is the use of heteroscedasticity consistent covariance matrices (first introduced by White (1980)). These give consistent covariance estimates in the presence of heteroscedasticity of an unknown form. These heteroscedasticity consistent covariance matrices have also been improved for the case of small (finite) samples (MacKinnon and White (1985)). The small-sample adaptations give influential observations less weight. Long and Ervin (2000) have shown, using Monte Carlo simulations, that one of the small-sample adaptations (termed "HC3") is indeed to be preferred.

Results

From the methods outlined above, the following approaches were implemented and their type I error rate assessed:

1. OLS: the OLS approach.
2. WLS: the WLS approach with weights estimated as the inverse of within-group variances.
3. WLS.HCCM: the WLS approach with weights estimated as the inverse of within-group variances, and with a small-sample consistent covariance matrix estimate. (The small-sample consistent covariance matrix estimator used is HC3.)

Results are presented in Figure 1. The results indicate that the WLS approach with the HC3 covariance estimator consistently outperforms both the OLS and WLS approach. As such, the WLS with HC3 covariance estimator is used throughout

the analyses and compared with the OLS approach that is commonly used for these assessments.

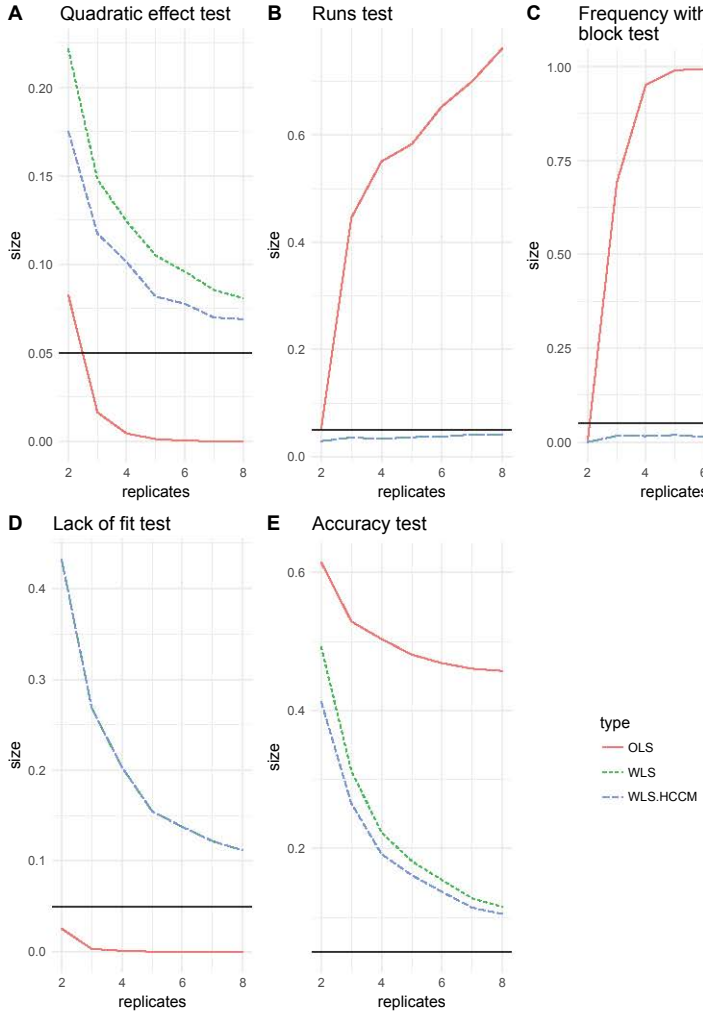


FIGURE F.1: Comparison of type I error rate control for (A) the quadratic term in a quadratic regression model, (B) detection of linearity deviation using a runs test, (C) detection of linearity deviation using a frequency within a block test, (D) detection of linearity deviation using a F test for lack of fit, (E) detection of deviation from perfect trueness. Of note, the runs test (B), frequency within a block test (C) and lack of fit test (D) rely on residuals only. Consequently, there is no difference in performance between WLS and WLS.HCCM for these assessments.

F.3 Linearity

TABLE F.1: Type I error rate (no bias) and power to observe deviation from linearity using an OLS approach and for different biases and replicate sizes (Rep). Methods studied are a quadratic regression (Quad), a runs test (Runs), a frequency within a block test (Freq) and an F test for lack of fit (LoF).

		Type I	Power			
Quad	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.08	0.52	0.89	0.57	0.05
	3	0.02	0.50	0.96	0.57	0.01
	4	0.00	0.55	0.99	0.66	0.00
	5	0.00	0.63	1.00	0.74	0.00
	6	0.00	0.72	1.00	0.83	0.00
	7	0.00	0.81	1.00	0.90	0.00
	8	0.00	0.87	1.00	0.95	0.00
Runs	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.05	0.00	0.00	0.34	0.15
	3	0.45	0.82	0.80	0.75	0.57
	4	0.55	1.00	1.00	0.95	0.85
	5	0.58	1.00	1.00	0.99	0.96
	6	0.65	1.00	1.00	0.99	0.98
	7	0.70	1.00	1.00	1.00	1.00
	8	0.76	1.00	1.00	1.00	1.00
Freq	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.00	0.00	0.00	0.00	0.00
	3	0.69	1.00	1.00	1.00	0.99
	4	0.95	1.00	1.00	1.00	1.00
	5	0.99	1.00	1.00	1.00	1.00
	6	0.99	1.00	1.00	1.00	1.00
	7	0.99	1.00	1.00	1.00	1.00
	8	1.00	1.00	1.00	1.00	1.00
LoF	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.03	0.02	0.02	0.47	0.84
	3	0.00	0.00	0.00	0.57	0.99
	4	0.00	0.00	0.00	0.71	1.00
	5	0.00	0.00	0.00	0.84	1.00
	6	0.00	0.00	0.00	0.93	1.00
	7	0.00	0.00	0.00	0.97	1.00
	8	0.00	0.00	0.00	0.99	1.00

TABLE F.2: Type I error rate (no bias) and power to observe deviation from linearity using a robust WLS approach and for different biases and replicate sizes (Rep). Methods studied are a quadratic regression (Quad), a runs test (Runs), a frequency within a block test (Freq) and an F test for lack of fit (LoF).

		Type I	Power			
Quad	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.18	0.82	0.93	0.81	0.40
	3	0.12	0.96	0.99	1.00	0.59
	4	0.10	0.99	1.00	1.00	0.82
	5	0.08	1.00	1.00	1.00	0.95
	6	0.08	1.00	1.00	1.00	0.99
	7	0.07	1.00	1.00	1.00	1.00
	8	0.07	1.00	1.00	1.00	1.00 0
Runs	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.03	0.07	0.14	0.30	0.38
	3	0.04	0.09	0.20	0.58	0.75
	4	0.03	0.14	0.34	0.86	0.96
	5	0.04	0.17	0.44	0.95	0.99
	6	0.04	0.21	0.55	0.98	1.00
	7	0.04	0.25	0.64	1.00	1.00
	8	0.04	0.30	0.70	1.00	1.00
Freq	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.00	0.00	0.00	0.00	0.00
	3	0.02	0.13	0.28	0.65	0.85
	4	0.02	0.22	0.52	0.96	0.99
	5	0.02	0.39	0.77	1.00	1.00
	6	0.01	0.44	0.86	1.00	1.00
	7	0.01	0.60	0.94	1.00	1.00
	8	0.01	0.72	0.98	1.00	1.00
LoF	Rep	No bias	10%	20%	20% + 10%	20% + 20%
	2	0.43	0.35	0.35	1.00	1.00
	3	0.27	0.22	0.21	1.00	1.00
	4	0.20	0.16	0.17	1.00	1.00
	5	0.15	0.14	0.13	1.00	1.00
	6	0.14	0.12	0.12	1.00	1.00
	7	0.12	0.11	0.10	1.00	1.00
	8	0.11	0.10	0.10	1.00	1.00

F.4 Trueness

TABLE F.3: Type I error rate (no bias) and power to detect deviation from a slope of one using a robust weighted least squares and an ordinary least squares approach for different replicate sizes (Rep).

	Robust WLS			OLS		
	Type I	Power		Type I	Power	
Rep	No bias	5%	10%	No bias	5%	10%
2	0.41	0.99	1.00	0.61	0.96	1.00
3	0.27	1.00	1.00	0.53	0.99	1.00
4	0.19	1.00	1.00	0.50	1.00	1.00
5	0.16	1.00	1.00	0.48	1.00	1.00
6	0.14	1.00	1.00	0.47	1.00	1.00
7	0.11	1.00	1.00	0.46	1.00	1.00
8	0.11	1.00	1.00	0.46	1.00	1.00

F.5 R code

R code underlying all analyses presented in the main text and appendices is available at `https://github.com/CenterForStatistics-UGent/dPCalibRate`.

G Curriculum vitae

G.1 Education

PhD in Statistical Data Analysis
Expected 2018 | Ghent University, BE

MSc in Statistical Data Analysis
Major Statistical Science
2017 | Ghent University, BE

MSc in Bioscience Engineering: Biotechnology
Major Computational Biology
2013 | Ghent University, BE

BSc in Bioscience Engineering: Biotechnology
2011 | Ghent University, BE

G.2 Employment

PhD Student and Teaching Assistant
Sep 2013 – Present | Ghent University, BE

G.3 International peer-reviewed journal publications

1. Buelens, S., Claeys, T., Dhondt, B., Poelaert, F., **Vynck, M.**, Yigit, N., Thas, O., Ost, P., Vandesompele, J., Lumen, N. and Kumps, C. (2018). Prognostic and therapeutic implications of circulating androgen receptor gene copy number in prostate cancer patients using droplet digital PCR. *Clinical Genitourinary Cancer.*

2. **Vynck, M.**, Vandesompele, J. and Thas, O. (2017). Quality control of digital PCR assays and platforms. *Analytical and Bioanalytical Chemistry*, 409(25), 5919-5931.
3. Pipelers, P., Clement, L., **Vynck, M.**, Hellemans, J., Vandesompele, J. and Thas, O. (2017). A Unified Censored Normal Regression Model for qPCR Differential Gene Expression Analysis. *PLoS One*, 12(8), e0182832.
4. Lamote, K., **Vynck, M.**, Thas, O., Van Cleemput, J., Nackaerts, K. and Van Meerbeeck, J. (2017). Exhaled breath to screen for malignant pleural mesothelioma: a validation study. *European Respiratory Journal*, *50(6)*, 1700919.
5. Lamote, K., Brinkman, P., Vandermeersch, L., **Vynck, M.**, Sterk, P., Van Langenhove, H., Thas, O., Van Cleemput, J., Nackaerts, K. and Van Meerbeeck, J. (2017). Breath analysis by gas chromatography-mass spectrometry and electronic nose to screen for pleural mesothelioma: a cross-sectional case-control study. *Oncotarget*, 21335.
6. **Vynck, M.**, Vandesompele, J., Nijs, N., Menten, B., De Ganck, A. and Thas, O. (2016). Flexible analysis of digital PCR experiments using generalized linear mixed models. *Biomolecular Detection and Quantification*, 9, 1-13.
7. **Vynck, M.**, Trypsteen, W., Thas, O., Vandekerckhove, L. and De Spiegelaere, W. (2016). The future of digital polymerase chain reaction in virology. *Molecular Diagnosis and Therapy*, 20(5), 437-447.
8. Lamote, K., **Vynck, M.**, Van Cleemput, J., Thas, O., Nackaerts, K. and Van Meerbeeck, J. (2016). Detection of malignant pleural mesothelioma in exhaled breath by multicapillary column/ion mobility spectrometry (MCC/IMS). *Journal of Breath Research*, 10(4), 046001.
9. Trypsteen, W., **Vynck, M.**, De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., Vervisch, K., Thas, O., Vandekerckhove, L. and De Spiegelaere W. (2015). ddpcRquant: Threshold determination for single channel droplet digital PCR experiments. *Analytical and Bioanalytical Chemistry*, 407(19), 5827-5834.

G.3.1 In preparation, under review or revision

1. **Vynck, M.** and Thas, O. Appropriate modelling of volumetric variability reduces bias in digital PCR quantification experiments. *Under review.*
2. **Vynck, M.**, Vandesompele, J. and Thas, O. On the design of digital PCR experiments. *In preparation.*

G.4 Conference contributions

1. **Vynck, M.** and Thas, O. A digital PCR data analysis workflow: from sample to effective result. 6th International Biometric Society Channel Network Conference, Diepenbeek, Belgium, April 24-26, 2017.
2. Buelens, S., Claeys, T., Kumps, C., Dhondt, B., Poelaert, F., Yigit, N., **Vynck, M.**, Thas, O., Ost, P., Vandesompele, J. and Lumen, N. The occurrence and therapeutic consequences of androgen receptor copy number gain in prostate cancer patients using Droplet Digital PCR. 32nd Annual congress of the European Association of Urology (EAU-17), London, United Kingdom, March 24-28, 2017.
3. Buelens, S., Claeys, T., Kumps, C., Dhondt, B., Poelaert, F., Yigit, N., **Vynck, M.**, Thas, O., Ost, P., Vandesompele, J. and Lumen, N. Androgen receptor gene copy number detection using droplet digital PCR and its therapeutic consequences in prostate cancer patients. 5th OncoPoint research seminar, Ghent, Belgium, March 15, 2017.
4. **Vynck, M.** and Thas, O. Threshold setting in duplex digital PCR experiments. 28th International Biometric Conference, Victoria, BC, Canada, July 10-15, 2016.
5. **Vynck, M.**, Vandesompele, J. and Thas, O. Improving accuracy and precision of dPCR experiments by relying on multiple reference genes. The 3rd qPCR and Digital PCR Congress, London, United Kingdom, October 20-21, 2015.
6. Trypsteen, W., **Vynck, M.**, De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., Vervisch, K., Thas, O., Vandekerckhove, L. and De Spiegelare W. ddpcRquant: Threshold determination for single channel droplet digital PCR experiments. The 3rd qPCR and Digital PCR Congress, London, United Kingdom, October 20-21, 2015.
7. **Vynck, M.**, Trypsteen, W., De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., Vervisch, K., Thas, O., Vandekerckhove, L. and De Spiegelare W. Threshold determination in droplet digital PCR experiments using extreme value theory. 23rd Annual Meeting of the Belgian Statistical Society, Antwerp, Belgium, October 14-16, 2015.
8. Lamote, K., **Vynck, M.**, Van Cleemput, J., Thas, O., Nackaerts, K. and Van

Meerbeeck, J. Volatile organic compounds increase the likelihood of detecting malignant pleural mesothelioma. ERS (European Respiratory Society) International Congress, Amsterdam, The Netherlands, September 26-30, 2015.

9. Lamote, K., **Vynck, M.**, Van Cleemput, J., Thas, O., Nackaerts, K. and Van Meerbeeck, J. Breath analysis by ion mobility spectrometry allows discrimination of pleural mesothelioma patients from controls. 16th World conference on Lung Cancer, Denver, CO, USA, September 06-09, 2015.
10. **Vynck, M.**, Trypsteen, W., De Neve, J., Bonczkowski, P., Kiselinova, M., Malatinkova, E., Vervisch, K., Thas, O., Vandekerckhove, L. and De Spiegelare W. Extreme value theory for threshold determination in droplet digital PCR experiments. Advances in qPCR and dPCR, Singapore, Singapore, May 21-22, 2015.
11. Thas, O., **Vynck, M.** and Vandesompele, J. and GLMMs for nucleic acid concentration estimation in digital droplet PCR. Advances in qPCR and dPCR, Singapore, Singapore, May 21-22, 2015.
12. **Vynck, M.**, Vandesompele, J. and Thas, O. GLMMs for nucleic acid concentration estimation in digital droplet PCR. International Biometric Society Channel Network Meeting, Nijmegen, The Netherlands, April 20-22, 2015.
13. Lamote, K., Van Cleemput, J., Hiddinga, B., Nackaerts, K., **Vynck, M.**, Thas, O., Van Meerbeeck, J. Exhaled Breath for the Diagnosis of Malignant Pleural Mesothelioma. European Society for Medical Oncology (ESMO) Meeting. Madrid, Spain, September 26-30, 2014.
14. Lamote, K., Van Cleemput, J., Hiddinga, B., Nackaerts, K., **Vynck, M.**, Thas, O., Van Meerbeeck, J. Exhaled Breath for the Diagnosis of Malignant Pleural Mesothelioma. European Respiratory Society (ERS) Congress 2014. Munich, Germany, September 6-10, 2014.
15. Lamote, K., Van Cleemput, J., Nackaerts, K., **Vynck, M.**, Thas, O., Van Meerbeeck, J. Breath Analysis for the Diagnosis of Malignant Pleural Mesothelioma. Breath Summit 2014. Torún, Poland, July 6-9, 2014.
16. De Meyer, T., Trooskens, G., Mensaert, K., Steyaert, S., **Vynck, M.**, Denil, S. and Van Criekinge, W. Applications of large-scale genome-wide DNA-methylation profiling. International Conference on Genomics Europe, Ghent, Belgium, June 25-28, 2013.
17. De Meyer, T., Trooskens, G., Mensaert, K., Steyaert, S., **Vynck, M.**, Denil,

S. and Van Criekinge, W. Applications of large-scale genome-wide DNA-methylation profiling. Benelux Bioinformatics Conference, Brussels, Belgium, December 9-10, 2013.

G.5 Teaching

Courses ($\sim$ 600 in-class hours)

1. Statistical Data Analysis (Fall 2013 – 2017)
   This is an introductory course in statistics, taught to 3rd year Bioscience Engineering students. For this course I was responsible for teaching exercises, writing exercise notes, writing and grading home work assignments, grading papers and exams.
2. Applied Statistics (Fall 2013 – 2017)
   This is an introductory course in statistics, taught to 4th year Bioscience Engineering students. For this course I was resopnsible for teaching exercises, writing exercise notes, writing and grading home work assignments, grading papers and exams.
3. Genome Analysis (Spring 2014 – 2018)
   This is an advanced course in statistical methods for the analysis of genomic data, taught to 4th year Bioscience Engineering students. For this course I was responsible for teaching exercises, writing exercise notes, coaching students in a data analysis project, grading oral presentations and exams.

Guest lectures ($\sim$ 25 in-class hours)

1. Digital PCR Specialist Course (doctoral schools). Ghent University, 2016, 2017.
   Short introductions to thresholding in digital PCR and the use of mixed models for the analysis of complex digital PCR experiments.
2. Regression Analysis. UGent Academy for Engineers, 2017.
   Exercises in linear regression analysis for doctoral students and industry employees.
3. Digital PCR: Data Analysis. Lecture for the Course Applied High Throughput Analysis, Master of Science in Bioinformatics, Ghent University, 2015 – 2017.
   Lecture on the use of mixed models for the analysis of complex digital PCR experiments (theory and exercises).

4. High Dimensional Data Analaysis. StatsWeek. University of Wollongong. 1-2 February, 2016.
   Exercises on multivariate methods such as (sparse) principal components analysis, (sparse) linear discriminant analysis, regularized regression, ...
5. qPCR data analysis. Marie Curie ITN: Mermaid Project. 2014.
   Half day session on the analysis of qPCR data (theory and exercises).
6. Summer School Statistics. 2014.
   Exercises on probability and exploratory statistics (half day session).

Supervision of dissertation students

1. Machteld Varewyck. Master's dissertation in Statistical Data Analysis. 2017.
   Analysis of digital PCR data with R.
2. Michael Niklaus. Master's dissertation in Statistical Data Analysis. 2015.
   Statistical analysis of RT-qPCR data for differential expression.

G.6 Awards

Best Poster Presentation. International Biometric Society: Channel Meeting, 2017.

Paper "Quality control of digital PCR assays and platforms" (2017) selected as Paper in Forefront by Analytical and Bioanalytical Chemistry. (*"Guided by the peer reviews, the Editors select a number of exceptional papers for very rapid publication as Papers in Forefront."*)

G.7 Other

Digital PCR: Specialist Course. 14, 15, and 19 September, 2017. Ghent University. Member of the scientific and organizing committee.

Reviewer for Analytical Chemistry, Bioinformatics, PLoS One.