Analysing Magnetic Bead QuantiGene (R) Plex 2.0 Gene Expression Data in High Throughput Mode Using QGProfiler

De Wolf Hans

Griet Laenen

Eva Adriaensen

mcrabbe1

Bie Verbist

# Authors

Eva Adriaensen (OA), Griet Laenen (OA), Vikki Keersmaekers (JNJ) (?), Marjolein Crabbe (JNJ), Hans De Wolf (JNJ), Bie Verbist (JNJ)

Others? (Acknowledgement: Dea Putri?, Maaren Derks?)

# Abstract

# Introduction

Gene [1]expression[2] , through the quantification of mRNA is commonly used in biomedical research for patient diagnostics and/or therapeutics [3] [4] [5]. The quantification of mRNA is routinely performed using real-time quantitative PCR (qPCR), measuring gene expression levels in a highly sensitive and specific manner [6]. However, there are limitations to this technique, which relate to the need for RNA extraction and the enzymatic based reverse transcription and target mRNA amplification steps which are prone to errors [7] [8].  Branched chain DNA (bDNA) technology, in which the signal and not the mRNA target sequence is amplified, provides a non enzymatic alternative to qPCR [9] [10] [11]. The QuantiGene Plex 2.0 platform (Affymetrix) combines bDNA with the luminex/xMAP magnetic bead capturing technology. This platform does not require an RNA extraction step, as it measures mRNA levels directly from cultured cells [12], cell lysates [13], tissue homogenates [14], formalin-fixed tissues [15], to name only a few starting points. The amplification of the signal depends on the cooperative hybridization between the target mRNA and three oligonucleotide probes. These probes are capture extenders (CE), label extenders (LE) and blocking probes (BL), whose sequences depend on the mRNA target sequence [13].  The hybridized mRNA target sequence is immobilized on the bead via a capture probe that links the bead with parts of the CE sequence, which provides the specificity of the signal [11]. The signal is subsequently amplified by adding a pre-amplifier sequence, which partly overlaps with the LE sequence, and by adding several biotinylated amplifiers, which generate the branched DNA structure [13]. The bDNA binds to streptavidin conjugated R-phycoerythrin (SAPE) and  a luminex reader detects the individual beads by flow cytometry [13] [16]. The resulting fluorescent signal is proportional to the hybridized mRNA quantity [11] [16].  The luminex/xMAP magnetic bead capturing technology allows for multiplexing in a single well of a 94 or 384 multi-well plate and is thus able to quantify the expression of a series of genes in a high throughput mode [13]. As such, the QuantiGene Plex 2.0 platform (Affymetrix) does not only offer the possibility to quantify mRNA levels in the context of patient diagnostics and/or therapeutics, but can also be used in a high throughput drug discovery setting. Indeed, entire compound libraries could be tested, in single dose or dose response, against disease specific gene signatures in search for new disease relevant chemical starting points [12]. However, in order to analyse mRNA levels in high throughput mode, a proper data analysis frame work should be put in place. The data analysis flow proposed by Affymetrix is relatively straightforward and aims to translate and normalize gene expression, in the linear range of the assay, to fold change values. The latter is achieved by averaging all signals, subtracting the average background signal for each gene, normalizing against housekeeping genes and dividing the normalized values for the treated samples by the normalized value of the untreated sample. Hence, QuantiGene data is currently often processed in a local spreadsheet environment and normalized against two to four commonly used house keeping genes [17] [18] [19][16] [8].  The expression of generic housekeeping genes such as ACTB, GADPH, HPRT1 or B2M can, however, vary considerably across tissue types and/or under different experimental conditions, which make them less suited for normalisation  [20].  In addition to a proper housekeeping gene assessment, the optimal cell density, limit of quantification and number of beads will have to be analysed as well, prior to the start of a high throughput QuantiGene Plex 2.0 drug discovery campaign.

Against this background we present experimental data and introduce a newly developed open source available R based shiny application: QGprofiler, that allows for proper QuantiGene Plex 2.0 assay optimisation, choice of housekeeping genes and data pre-processing from raw gene expression to normalized fold change values. In addition, we propose a way to assess cytotoxicity and introduce a step-wise dose response fold change analysis. QGprofiler is available at URL and will accept both 96 and 384 multi well plate format in single and dose response .

bead number staat los van QGprofiler, maar geeft aan wanneer wells worden verwijderd op basis van bead number

cell density/linearity of the signal/LOQ

HKG with geomean and FC tresholds

FC for disease genes in d/r with FC treshold, horizontal fit, etc.

EC50 error distribution

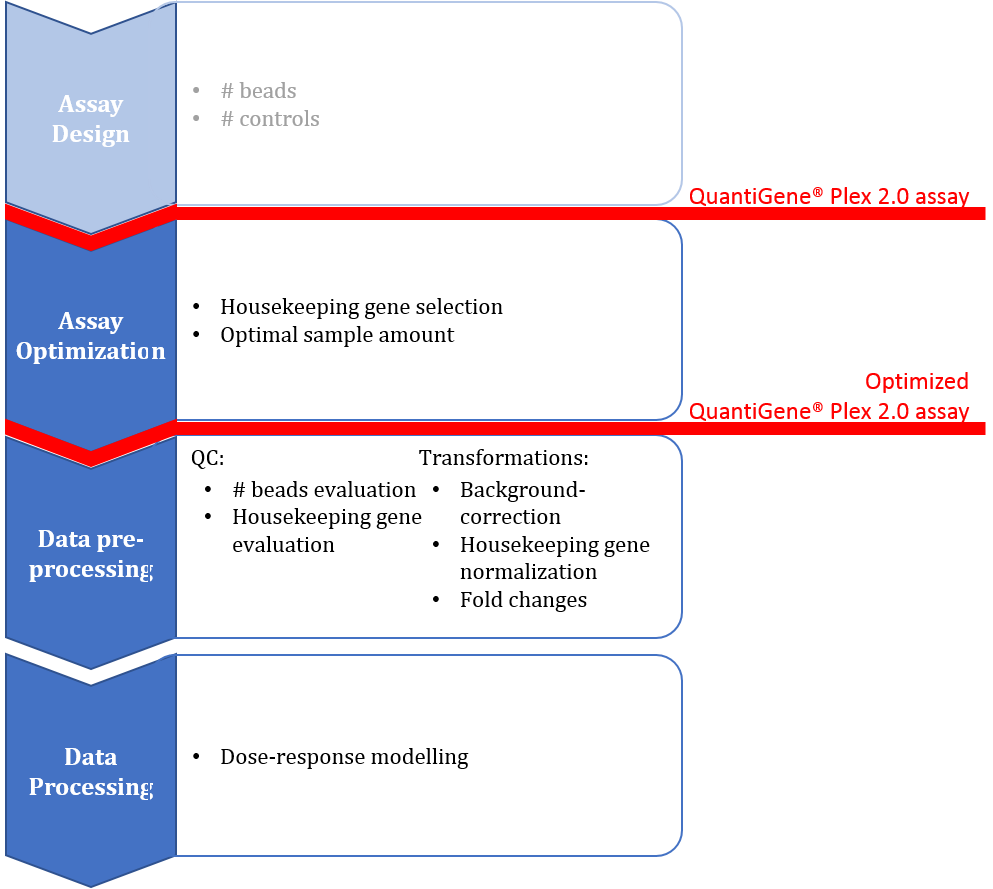
Cytotoxicity

# Material and Methods

# Results

## QGprofiler workflow

Overview of the  QuantiGene® Plex 2.0 workflow



QuantiGene® Plex 2.0 workflow

## Data analysis flow

The general analysis flow for QuantiGene® Plex 2.0 data encompasses three main steps which should be performed on a per-gene basis: background correction of the gene signal, normalization of the gene signal and calculation of fold change in expression (Affymetrix, manufacturer’s manual). In more detail: wells without loaded RNA sample can be summarized in an average background signal (recommendation of three background control wells). Background corrected gene signals are obtained by subtracting the average background signal from the raw median fluorescence intensity (MFI) values. Subsequently, dividing the background corrected values by a normalization gene signal (background subtracted) results in normalized gene signal values. Alternatively, if a set of normalization genes is used, one should divide by the geometric mean of the normalization gene signals. These normalized values are expression ratios (target/housekeeping gene) and will be further referred to as relative expression values. Finally, fold changes in gene expression versus untreated sample are calculated by dividing the relative expression value by the relative expression value of untreated sample (negative control).

Downstream analyses such as dose response modelling rely on qualitative fold change values for proper conclusions. Given the analysis flow as described above, this can only be warranted if accurate, high-quality signals are used throughout all the different data transformations steps (background subtraction, normalization and fold change calculation). QGprofiler includes clear guidelines at each step of the analysis flow in order to ensure maximal quality of the data at each transformation step. The selection of suitable normalization genes and the stability of their corresponding signal is of the utmost importance in this regard.

## Quality guidelines

### Bead numbers

The QuantiGene® Plex 2.0 platform utilizes the Luminex/xMAP magnetic bead array system to quantify multiple RNA targets simultaneously. Typically, QuantiGene output data is formatted as MFI values, median fluorescence across individual bead intensities. The individual bead intensities are usually not reported, which is why MFI values are considered raw QuantiGene data. The final number of magnetic capture beads per gene per well on readout is variable. Bead number values have been reported to range from 50-100 beads per gene per well (Ferrer, 2014) (Flagella, 2006), while the QuantiGene® 2.0 Plex assay user manual specifies an expected average bead count > 50 per gene (Affymetrix, manufacturer’s manual). Factors such as sample viscosity, washing steps throughout the assay and possible bead carryover across wells indeed affect the final number of beads per gene per well. Occasionally, this can drop to very low numbers in which case the corresponding MFI value cannot be considered a reliable median. Such values need to be discarded during the data analysis flow in order to ensure qualitative data. In a resampling experiment investigating 40 different genes from three different datasets we identified 40 as the minimum required number of beads per gene per well to guarantee stable and reliable MFI readout values (data currently not shown).

Raw MFI values obtained from a total number of beads < 40 are therefore not used while analysing expression levels. QGprofiler will list all gene-well combinations with a total number of beads < 40 and corresponding data values will be automatically discarded from downstream analysis.

### Background controls

Wells without loaded sample are referred to as background control wells. The average background signal is used for background correction. Additionally, gene-specific limits of detection (LOD) are calculated using the background signals. The LOD is defined as the average MFI value across background control wells plus three standard deviations of the background (Affymetrix, manufacturer’s manual). Any signals below this limit suggests absence of the corresponding gene and should not be used for the analysis of expression levels (should be set to zero?).

Three background control wells are suggested by the QuantiGene® 2.0 Plex user manual (Affymetrix, manufacturer’s manual). However, we recommend to increase this to at least six technical controls, based on experimental design (do we have data?). Moreover, QGprofiler will perform background correction and LOD calculation using the median background signal. This to ensure robustness of the calculated metric in case of outlying background signals. Additionally, plots on the background control wells are included in QGprofiler to facilitate visual quality control and to inspect variability and/or possible outlying values in the background control wells.

### Assay linearity

Each QuantiGene® 2.0 Plex setup is defined by its probe set and biological sample (cellular context or tissue type). Every setup requires the identification of the optimal cell density or amount of tissue in order to ensure signals in the linear range of the assay. That is, a doubling of the cell density (or amount of tissue) should result in a doubling of the signal intensity. To this end, a serial dilution of sample should be run and observed signal ratios of background corrected MFI values should be calculated. Ratios within the 20% range of the expected ratio of 100% are accepted (QuantiGene® 2.0 Plex user manual). The final amount of cells (or tissue) should result in signals within the linear range for all genes in the probe set. Additionally, signals should be above the gene-specific limit of quantification (see ‘Limit of Quantification’). In case the identified range is wide, expected effects on (the majority of) the genes could push the choice of optimal cell density (or amount of tissue) towards the lower or upper bound of the identified linear range.

*QGprofiler includes plots on cell density to visually evaluate the linear range and aide in the decision process on the cell density/amount of sample in the real assay.*

### Limit of quantification

The limit of quantification (LOQ) refers to the lowest MFI value within the assay’s linear range (QuantiGene® 2.0 Plex user manual) and as such it is also gene-specific. The luminex/xMAP bead technology typically results in MFI values that level of at low/high treatment values, depending on the gene-specific effect. In order to ensure assay linearity in function of the chosen treatment, the LOQ is intended as a cut-off for low MFI values. It makes a call on the acceptability of MFI values to estimate true expression values. Only MFI values > LOQ are deemed to be useful in quantitative analyses of gene expression levels. *However, given the throughput and the high number of measurable genes, determining the LOQ per gene, per experimental run is very cumbersome and in practice often not feasible*. In this regard we propose a pragmatic cut off in analogy to the LOD calculation. This cut off can be set in a more relaxed or stringent manner, i.e. the median MFI value in background control wells plus five or ten standard deviations of the background. These are referred to as LOQ5 and LOQ10, respectively.

*QGprofiler currently evaluates housekeeping genes against LOQ10 (see ‘Housekeeping genes‘) and visually indicates observations of disease genes < LOQ10.*

### Housekeeping genes

Although QuantiGene® Plex 2.0 overcomes several of the pitfalls associated with classical mRNA quantification techniques such as qPCR, the need for internal normalization using housekeeping genes remains. It reduces variability in the results due to sample preparation, sample input and well/plate/experimental differences (QuantiGene® 2.0 Plex user manual). Any method relying on housekeeping genes for normalization should dedicate a vast portion to identifying truly stable housekeeping genes. The selection of such housekeeping genes in the context of the envisioned experiment is crucial to limit noise in the downstream data analysis and to ensure reliable results. In the case unstable housekeeping genes are used, the data transformation will result in unreliable relative expression and corresponding fold change values.

Proper housekeeping gene assessment is thus an important step in identifying the final probe set to be used in the QuantiGene setup. A suitable housekeeping gene exhibits constant expression levels across a wide range of doses and treatments in the cellular context (or tissue) that is envisioned in the assay setup. Once the final set of housekeeping genes for a particular assay is determined, it should remain constant throughout different experimental runs.

*A housekeeping gene is said to be stable if its corresponding fold change values lie withing the [0.8; 1.2] interval, i.e. accepting a 20% variability. This is in analogy with the rationale to accept a 20% deviation from the expected 100% ratio when determining the linear range of the assay. A housekeeping gene should be stable across different experimental conditions (treatments and doses) in order to serve as normalization gene. We recommend the inclusion of at least three different stable housekeeping genes in the final probe set..*

*QGprofiler includes plot on FC values of housekeeping genes to evaluate stability using the 100% +/- 20% ruleof the assay linearity logic…*

# Discussion

# References

1. D’Ambrosio A, Mazzeo G, Iorio C, Siciliano R. A differential evolution algorithm for finding the median ranking under the Kemeny axiomatic approach. Computers & Operations Research [Internet]. Elsevier BV; 2017;82:126–38. Retrieved from: <https://doi.org/10.1016%2Fj.cor.2017.01.017>

2. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. Xia Y, editor. PLOS ONE [Internet]. Public Library of Science (PLoS); 2015;10:e0146021. Retrieved from: <https://doi.org/10.1371%2Fjournal.pone.0146021>

3. Pabinger S, Rödiger S, Kriegner A, Vierlinger K, Weinhäusel A. A survey of tools for the analysis of quantitative PCR (qPCR) data. Biomolecular Detection and Quantification [Internet]. Elsevier BV; 2014;1:23–33. Retrieved from: <https://doi.org/10.1016%2Fj.bdq.2014.08.002>

4. Oh EH, Jung S, Kim WJ, Kim KP, Kim SK. Microparticle-based RT-qPCR for highly selective rare mutation detection. Biosensors and Bioelectronics [Internet]. Elsevier BV; 2017;87:229–35. Retrieved from: <https://doi.org/10.1016%2Fj.bios.2016.08.057>

5. Hadi HE, Abdellaoui-Maane I, Kottwitz D, Amrani ME, Bouchoutrouch N, Qmichou Z, et al. Development and evaluation of a novel RT-qPCR based test for the quantification of HER2 gene expression in breast cancer. Gene [Internet]. Elsevier BV; 2017;605:114–22. Retrieved from: <https://doi.org/10.1016%2Fj.gene.2016.12.027>

6. Bustin S. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. Journal of Molecular Endocrinology [Internet]. Bioscientifica; 2000;25:169–93. Retrieved from: <https://doi.org/10.1677%2Fjme.0.0250169>

7. Yang W, Maqsodi B, Ma Y, Bui S, Crawford KL, McMaster GK, et al. Direct quantification of gene expression in homogenates of formalin-fixed, paraffin-embedded tissues. BioTechniques [Internet]. Future Science LTD; 2006;40:481–6. Retrieved from: <https://doi.org/10.2144%2F000112133>

8. Morten BC, Scott RJ, Avery-Kiejda KA. Comparison of the QuantiGene 2.0 Assay and Real-Time RT-PCR in the Detection of p53 Isoform mRNA Expression in Formalin-Fixed Paraffin-Embedded Tissues- A Preliminary Study. Bauer JA, editor. PLOS ONE [Internet]. Public Library of Science (PLoS); 2016;11:e0165930. Retrieved from: <https://doi.org/10.1371%2Fjournal.pone.0165930>

9. Zhang A, Pastor L, Nguyen Q, Luo Y, Yang W, Flagella M, et al. Small Interfering RNA and Gene Expression Analysis Using a Multiplex Branched DNA Assay without RNA Purification. Journal of Biomolecular Screening [Internet]. SAGE Publications; 2005;10:549–56. Retrieved from: <https://doi.org/10.1177%2F1087057105277414>

10. Flagella M, Bui S, Zheng Z, Nguyen CT, Zhang A, Pastor L, et al. A multiplex branched DNA assay for parallel quantitative gene expression profiling. Analytical Biochemistry [Internet]. Elsevier BV; 2006;352:50–60. Retrieved from: <https://doi.org/10.1016%2Fj.ab.2006.02.013>

11. Tsongalis GJ. Branched DNA Technology in Molecular Diagnostics. American Journal of Clinical Pathology [Internet]. Oxford University Press (OUP); 2006;126:448–53. Retrieved from: <https://doi.org/10.1309%2F90bu6kdxanfln4rj>

12. Severyn B, Nguyen T, Altman MD, Li L, Nagashima K, Naumov GN, et al. Development of a High-Throughput Gene Expression Screen for Modulators of RAS-MAPK Signaling in a Mutant RAS Cellular Context. Journal of Biomolecular Screening [Internet]. SAGE Publications; 2016;21:989–97. Retrieved from: <https://doi.org/10.1177%2F1087057116658646>

13. Ferrer M, Corneo B, Davis J, Wan Q, Miyagishima KJ, King R, et al. A Multiplex High-Throughput Gene Expression Assay to Simultaneously Detect Disease and Functional Markers in Induced Pluripotent Stem Cell-Derived Retinal Pigment Epithelium. STEM CELLS Translational Medicine [Internet]. Wiley; 2014;3:911–22. Retrieved from: <https://doi.org/10.5966%2Fsctm.2013-0192>

14. Metzger DC, Luckenbach JA, Dickey JT, Beckman BR. Development of a multiplex gene expression assay for components of the endocrine growth axis in coho salmon. General and Comparative Endocrinology [Internet]. Elsevier BV; 2013;189:134–40. Retrieved from: <https://doi.org/10.1016%2Fj.ygcen.2013.04.022>

15. Knudsen BS, Allen AN, McLerran DF, Vessella RL, Karademos J, Davies JE, et al. Evaluation of the Branched-Chain DNA Assay for Measurement of RNA in Formalin-Fixed Tissues. The Journal of Molecular Diagnostics [Internet]. Elsevier BV; 2008;10:169–76. Retrieved from: <https://doi.org/10.2353%2Fjmoldx.2008.070127>

16. Hall JS, Usher S, Byers RJ, Higgins RC, Memon D, Radford JA, et al. QuantiGene Plex Represents a Promising Diagnostic Tool for Cell-of-Origin Subtyping of Diffuse Large B-Cell Lymphoma. The Journal of Molecular Diagnostics [Internet]. Elsevier BV; 2015;17:402–11. Retrieved from: <https://doi.org/10.1016%2Fj.jmoldx.2015.03.010>

17. Yim HW, Song BJ, Jung SS, Kim H-J, Choi Y-J, Lee K-Y, et al. Branched-chain Assay for ER PR and HER2 RNA Levels is a Useful Adjunct in the Evaluation of ER PR and HER2 in Breast Cancer. Journal of Breast Cancer [Internet]. Korean Breast Cancer Society (KAMJE); 2010;13:267. Retrieved from: <https://doi.org/10.4048%2Fjbc.2010.13.3.267>

18. Hall JS, Leong HS, Armenoult LSC, Newton GE, Valentine HR, Irlam JJ, et al. Exon-array profiling unlocks clinically and biologically relevant gene signatures from formalin-fixed paraffin-embedded tumour samples. British Journal of Cancer [Internet]. Springer Nature; 2011;104:971–81. Retrieved from: <https://doi.org/10.1038%2Fbjc.2011.66>

19. Sun Y, Liu H, Yu G, Chen X, Liu H, Tian H, et al. Gene expression analysis of leprosy by using a multiplex branched DNA assay. Experimental Dermatology [Internet]. Wiley; 2011;20:520–2. Retrieved from: <https://doi.org/10.1111%2Fj.1600-0625.2011.01270.x>

20. Jonge HJM de, Fehrmann RSN, Bont ESJM de, Hofstra RMW, Gerbens F, Kamps WA, et al. Evidence Based Selection of Housekeeping Genes. Lichten M, editor. PLoS ONE [Internet]. Public Library of Science (PLoS); 2007;2:e898. Retrieved from: <https://doi.org/10.1371%2Fjournal.pone.0000898>