	Development and verification of the use of reference samples in the DAS lab	ID	210228RW19030
		status	Final
		version	1.2
		effective date	20-Oct-2020

1 Table of contents

2 Objectives

Documentation of the development and verification of an Internal Assay Control (IAC) in DAS measurements.

3 Introduction

METHOD is an in vitro diagnostic assay service that uses kinase activity profiles to generate a prediction for a cancer patient's response to immunotherapy [1]. Hereto, a pre-calibrated multivariate model is used to calculate a predictive score from a measured kinase activity profile.

During the Analytical Performance Evaluation for METHOD an internal control sample was applied to assess and correct for sources of systematic variation such as PamChip batches, detection antibody batches, and PS12 run-run variation [2], and to meet the requirements for Precission of Measurement [3]. This report documents the development and verification of the use of this IAC.

Drift of a measurement system, with respect to the state it was in when calibration was performed, is a general problem that is likely to compromise measurement results. To counter this, internal control samples or *in line reference samples* can be used to monitor and correct for measurement drift. Methods applying reference samples for drift correction for analytical measurements using multivariate calibration were reported in the scientific literature, e.g. in [4]–[7].

It is by now generally recognized that batch effects related to e.g. chip batches, reagent batches, and day-to-day variation have a critical impact on the measurement results with microarray and other *omics* technologies [8]. For typical *discovery* studies in which multiple samples are analysed for correlation with (a) covariate(s) of interest (e.g. treatment response) basic experimental design principles should be applied to avoid confounding of covariates of interest with batch effects. This may then be combined with a correction of the measurement data for any observed batch effects [8][9]. Multiple methods for correcting batch effects with microarray technologies may be employed [10]. Wwhen the batch effect can be assigned to a known factor (i.e. a batch variable), the ComBat method [11] has been shown to be simple and effective [10].

An example of the use of (gene expression) microarray technology for predicting outcome of cancer patients is MammaPrint, a test for which both FDA and CE-IVD approval has been obtained [12]. It applies a 70–gene multivariate model to predict the risk of recurrence for breast cancer patients [13]. Glas et al. [14] describe how this microarray signature was converted to a simplified diagnostic product. It involves the use of a reference sample that is co-hybridized to the microarray with a clinical sample. Hereto a large amount of reference sample was obtained by amplifying and pooling RNA of relevant patient samples.

The specific nature of the PamChip kinase activity profiling technology does not allow methods such as co-hybridization with a reference and/or (quantile) normalization that are commonly used for controlling variation with gene expression micro-arrays[1]. In contrast, simple batch correction methods such as per batch centering have been previously applied in *discovery*-type studies with PamChip arrays, see for instance [15][16]. Recently, the ComBat method was also applied, e.g. when processing the data presented at the 2018 ASCO conference [17].

It was recognized early on, however, that for application in a diagnostic setting additional methods would be necessary to account for systematic variation between e.g. PamChip batches, reagent batches and PS12 runs. Mainly, because in a diagnostic setting the clinical samples are not measured in batches, at least not in batches that are large enough to allow unequivocal estimation and correction for batch effects. Therefore, the development and verification of a correction method using a reference sample was started. This reference sample is typically measured in the same PS12 run, or even on the same chip, but on a different array than the clinical samples under investigation.

[1] Due to, respectively, the fact that the assay is an activity-based assay rather than an analyte binding assay and the focused nature of the PamChip array.

Section 4 describes the pilot and proof of principle studies, section 5 the considerations that lead to the selection of the present reference sample, and section 6 is concerned with the verification of the use of the reference sample. When applicable this report will refer to previous reports and presentations.

4 Pilot and Proof of Principle studies

4.1 Pilot study with clinical tissue samples

Pilot work on using control samples to correct for systematic variation in PamChip measurements was reported in [18]. In this report, measurements with colon carcinoma and renal cell carcinoma tissue samples were performed on several PamChip batches using the CRL-2264 (CRL) cell line as a reference sample. This work showed the feasibility of using CRL as a reference sample to reduce variation in clinical tissue measurements (applying a simple multiplicative correction). It also showed the importance of using a reference sample that is representative for the clinical samples used, i.e. the batch-batch variation of the clinical tissue samples was not in all cases properly represented by the CRL reference sample.

4.2 Proof of Principle study with PBMC samples

In a next phase a Proof of Principle (PoP) study was conducted to show the feasibility of using a reference sample to reduce batch effects when computing a diagnostic score from kinase activity profiles. Project and experimental planning was documented in [19] and [20], results in a presentation [21]. This study was designed with the METHOD application in mind, which uses Peripheral Blood Mononuclear Cell (PBMC) samples and is centered around the calculation of a diagnostic score as the final result.

4.2.1 Design

For the PoP study a data set of kinase activity profiles of samples in two groups (“Grp0” and “Grp1”, representing e.g. responders and non-responders) was simulated with group means and variance-covariance matrix extracted from real measurement data with patient PBMCs. This data set was then used to train a model for predicting the class of “Grp0” and “Grp1” measurements.

Next, *real* measurements were performed with PBMC samples, taken to represent “Grp0” and “Grp1” observations, subject to PamChip batch effects and PS12 run-run variation. The measured kinase activity profiles were then used to calculate a “diagnostic score” using the simulated predictive model. An additional PBMC sample was included in the run and used as an IAC.

Finally, the variation of the diagnostic score with and without applying a correction using the IAC was evaluated.

4.2.2 Correction Method

Figure 1 shows the general scheme for applying a *drift* or *batch effect* correction based on IAC measurements. First, “@Calibration”, a predictive model is calibrated using a set of training samples. During this calibration IAC measurements are added to the measurements. At a later time, “@Prediction”, a new sample is measured for which a diagnostic score is calculated using the predictive model obtained during the calibration. If the measurement system has drifted (due to e.g. batch effects) relative to its state during calibration this may

cause an error in the calculated diagnostic score. The drift, however, may be observed by comparing the IAC measured "@Prediction" with the IAC measured "@calibration". Based on this comparison a *drift model (*for the systematic differences between calibration and prediction) is derived and used to correct the new measurements. As long as the drift measured with the IAC is representative of the drift experienced by the new samples the measurements will be corrected towards the measurements performed during calibration.

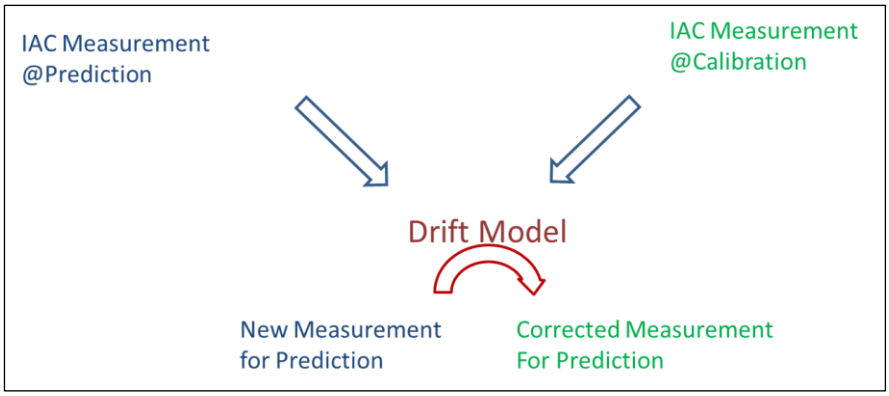


Figure 1: Scheme for drift correction based on IAC measurements

The actual correction may be a simple *per peptide* multiplicative correction as used in [18]. More refined methods aim at reducing the noise that is associated with measuring the drift model itself by looking at all spots simultaneously rather than in isolation.

With the ComBat method the drift model is stabilized by using Empirical Bayes methods to "borrow information" between spots [11]. However, we have adapted this method to allow for a drift model to be estimated on the (independent) IAC samples and applied to new samples, as depicted in Figure 1. Code for this adaptation is contained in the internal R-package *pgBatch* (v 1.3 , Rik de Wijn 2018) and may be applied using a BioNavigator app.

With Dynamic Orthogonal Projection (DOP e.g. [7]) the first *k* principal components of the drift space are used to calculate a correction. Hence, a noise reduction is achieved by omitting higher principal components from the calculation. Strictly, the method deviates from the scheme in Figure 1 because the correction is applied by making the *calibration data* orthogonal to the main drift components and re-calibrating the predictive model. This *rotated* predictive model is robust against the observed drift.

A DOP implementation was tested during this PoP study. During later phases the ComBat approach was used. ComBat is well known as a microarray tool and straightforward to implement using IAC samples.

4.2.3 Results

Here a summary of the PoP results is provided. For more details see [21]. Calculation scripts are available in I:\210-228 Diagnostic Assay Service\Control\Classifier simulation.

Figure 2 shows PCA scores for the measured activity profiles without correction. It shows that measurements for Grp1, Grp0 and the IAC sample performed with the same PamChip batch cluster together, i.e. the variation between batches is much larger than the separation between Grp0 and Grp1. At the same time Figure 2 also shows that the IAC sample (Ref, blue) experiences variation with PamChip Batch which is indicative for that experienced by the Grp0 and Grp1n sample.

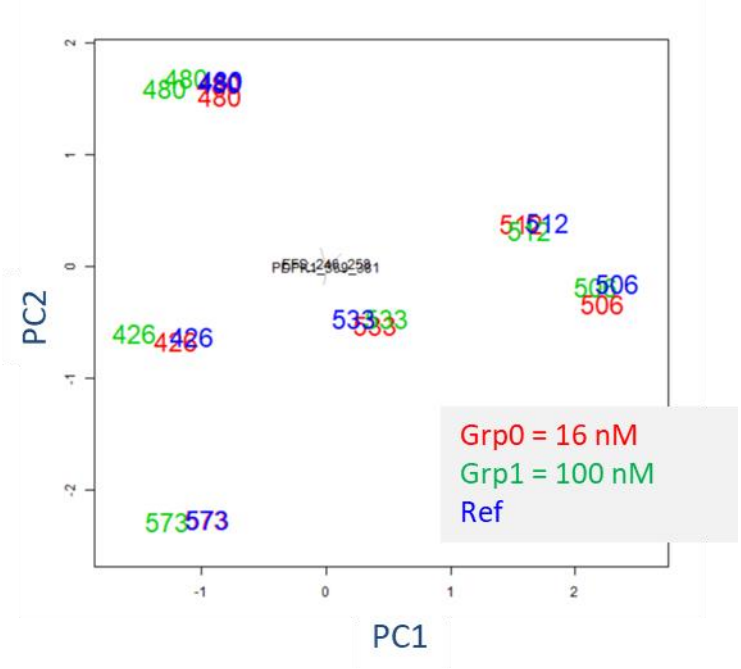


Figure 2: PCA Scores to PC1-2 for the measurements with different PamChip batches as described in section 4.2. The numbers indicate the different batches. Grp0, Grp1 and the IAC sample (Ref) are indicated in red, green, and blue, respectively.

Figure 3 shows the result of applying the trained predictive model to the measured kinase activity profiles in subsequent measurement runs: (1) without correction, (2) after applying the ComBat correction, and (3) after applying the DOP correction. Ideally, the diagnostic score does not vary over subsequent measurements and remains equal to the value obtained under the same conditions as the calibration (indicated by a square in Figure 3A). Similar to the situation for the PCA scores depicted in Figure 2, the diagnostic score without applying a correction (Figure 3A) is more indicative for run / batch variation than for Grp0 / Grp1 discrimination. For both ComBat (Figure 3B) and DOP (Figure 3C) correction this is much improved. The scores for Grp0 are clearly comparable throughout the runs. Also for Grp1 the variation is clearly reduced although the scores are not consistent for all runs. This remaining variation is attributed to experimental variation (or problems) not attributed to batch or run variation.

The reduction of the variation of the Diagnostic Score by applying a correction is also evident from the standard deviations in Table 1. Note that the IAC sample (Ref) was used for calculating the correction model, as a result it has the lowest residual variation as shown in Table 1.

Standard Deviation	A (No)	B (ComBat)	C (DOP)
Grp0	1.3	0.1	0.12
Grp1	1.4	0.46	0.55
Ref (IAC)	1.3	0.03	0.06

Table 1. Standard deviations corresponding to the variation of the Diagnostic Score shown in Figure 3 A-C.

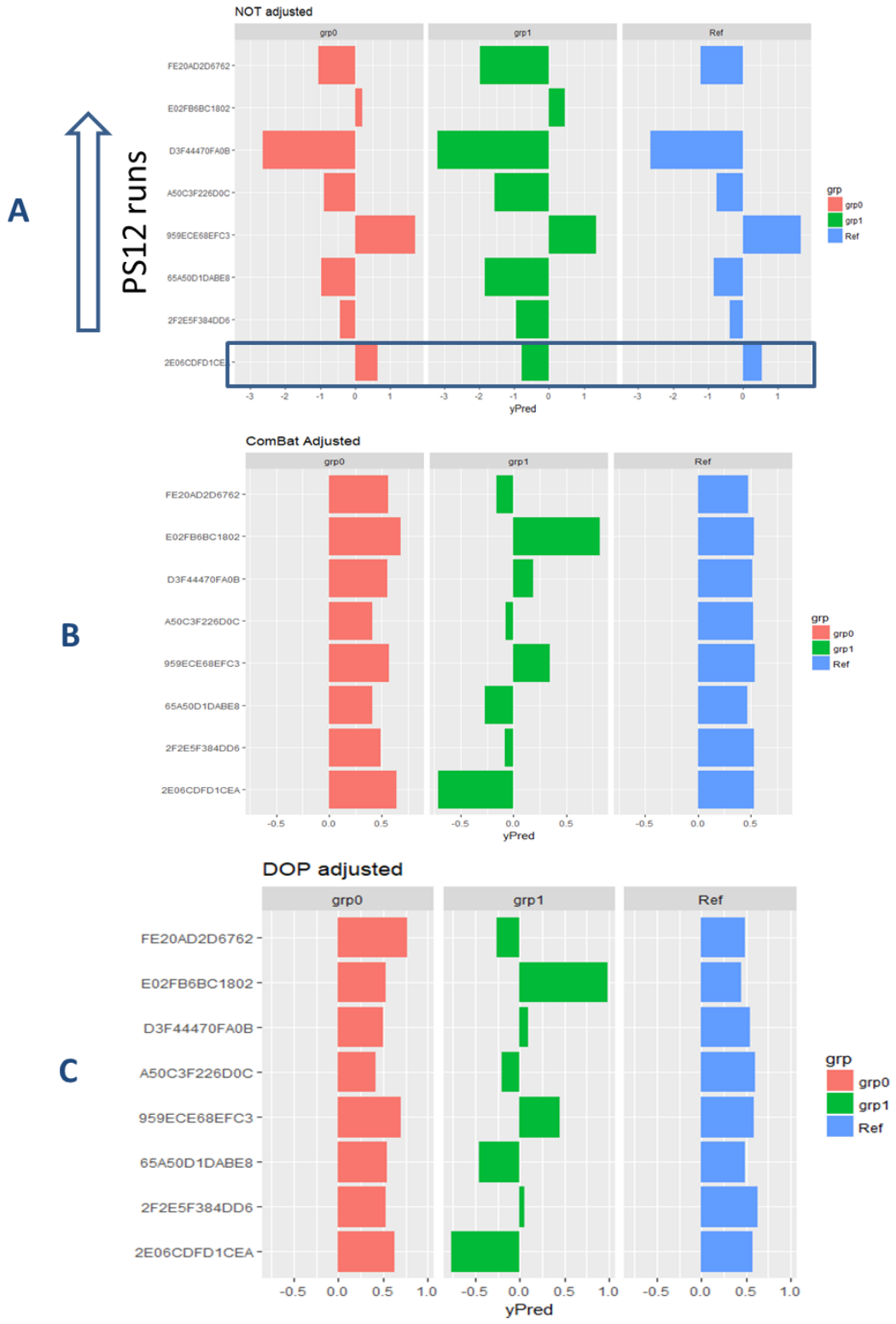


Figure 3: Diagnostic score (yPred) calculated over subsequent PS12 runs (A) without correction, (B) after applying ComBat correction, (C) after applying DOP correction (see section 4.2). Note the difference in x-axis scaling between A, B, and C. On the y-axis run identifiers are displayed. The lowest run ID (indicated by the box in A) refers to the run performed under the same conditions as the calibration.

4.2.4 Conclusion

Using an IAC sample to correct for systematic differences in PS12-PTK measurements with PBMC samples can considerably reduce the variation on a Diagnostic Score calculated based on the measurement results.

The ComBat and DOP correction are similar in performance. ComBat is well known as a microarray tool and straightforward to implement using IAC samples. Therefore, it has been used in the design of METHOD and will be used in the remainder of this report. For the purpose of developing the DAS software necessary for applying the correction in routine use the ComBat method [11], [22] has been documented in Appendix A and B of this report.

It may also be concluded that with the current batch / run variation such an IAC correction will be a necessary step in reducing the (long-term) variation on the future Diagnostic Assay Service result.

5 Selection and production of the reference sample

The PoP studies have shown the feasibility of using an IAC sample for correcting systematic differences between experiments. The next step was to select samples that can be used for the METHOD application.

For a sample to be practically usable in the METHOD application the following requirements have to be met:

1. The effect of systematic variation on the IAC sample has to be representative for that of the diagnostic samples.
2. The IAC sample(s) must be obtained in large amounts and stored in aliquots such that the same batch of IAC samples can be used for a long period of time

For METHOD PBMCs isolated from patient blood are measured. For the first requirement defined above it seems logical to use a PBMC sample for the IAC as well. Large amounts of PBMCs (typically 500 $*10^6$ cells per sample or more) can be obtained by isolating PBMCs from buffy coats that were previously isolated from healthy donor blood. The procedure for isolating these PBMCs differs from that used for METHOD patient samples in that the buffy coats are isolated from bags of donor blood, stored (typically overnight) with citrate as an anti-coagulant. The use of these buffy coat samples as an IAC for METHOD was verified as described in section 6 of this report.

A total of six these buffy coat PBMCs (bcPBMC) samples were obtained from Sanquin Blood Supply Foundation (Sanquin order 180911 Buffy PBMCs Sanquin/ quote SQN 2018 056). Sample lysis and protein determination were performed essentially according to standard methods (WI7603-Lysis and Storage Instructions for PBMC pellets, WI7608-Standard Protein Assay Bradford). The lysates were stored in a -80°C freezer in aliquots of 10 µl. The minimum number of aliquots obtained per sample was 340. For further details see the preparation lab journals: "210226LHB18040" and "210226LHB18050".

6 Verification of the reference sample

6.1 Verification of the buffy coat PBMC sample

Initially, four bcPBMC samples were tested in a PTK assay using 2 µg protein per array.

Functional verification of the bcPBMC was performed according to the protocol in [23], results were documented in a presentation [24]. Most importantly, the bcPBMC were measured together with two healthy donor PBMC samples “Donor 1” and “Donor 2”, obtained by isolating PBMCs according to the standard diagnostic protocol (WI7324 – Instruction PBMC isolation form whole blood). These measurements were performed using 2 µg protein per array on three PamChip batches, with 2 PS12 runs for each combination.

Figure 4 shows that the PamChip batch variation observed for Donor 1 and 2 is similar to that observed for the four bcPBMC samples. As an example, all samples show a signal for batch 480 that is low relatively to that observed for batch 479. Hence, it appears that the variation observed with the bcPBMC samples is representative of that observed with Donor 1 and 2, and that the bcPBMC are suitable for use as IAC sample.

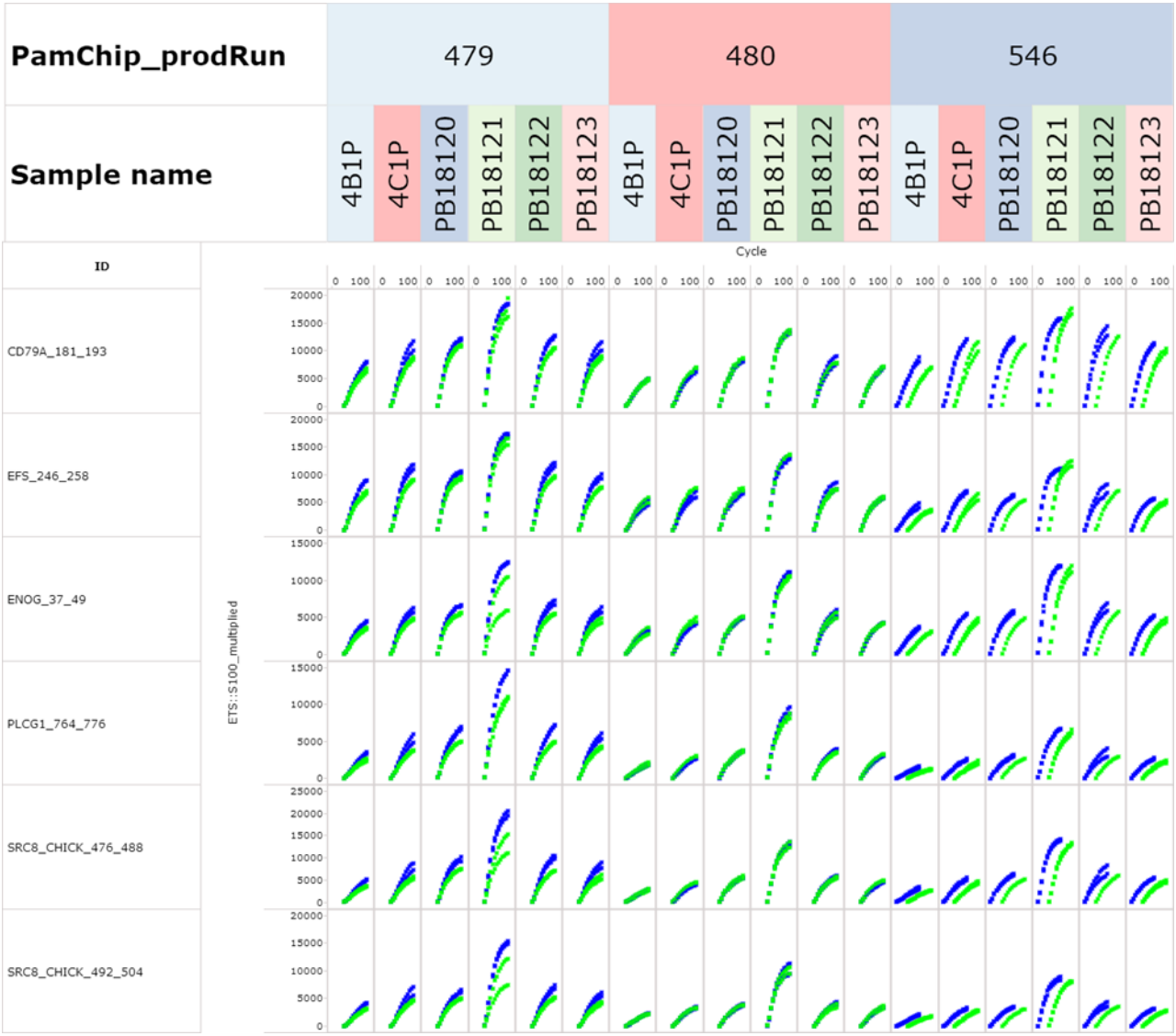


Figure 4: Verification measurements with samples from Donor 1 (4B1P), Donor 2 (4C1P) and four bcPBMC samples (PB18120-23). Phosphorylation time courses of the measurements are shown, using three PamChip batches (479, 480, 546), and two PS12 runs (blue: run 1, green: run 2).

Figure 5 shows the standard deviation of the repeated measurements with Donor 1 and Donor 2 sample with and without applying a correction based on the bcPBMC measurements. Applying the correction results in a 3-fold reduction of the high signal standard deviation compared to the standard deviation without applying the correction.

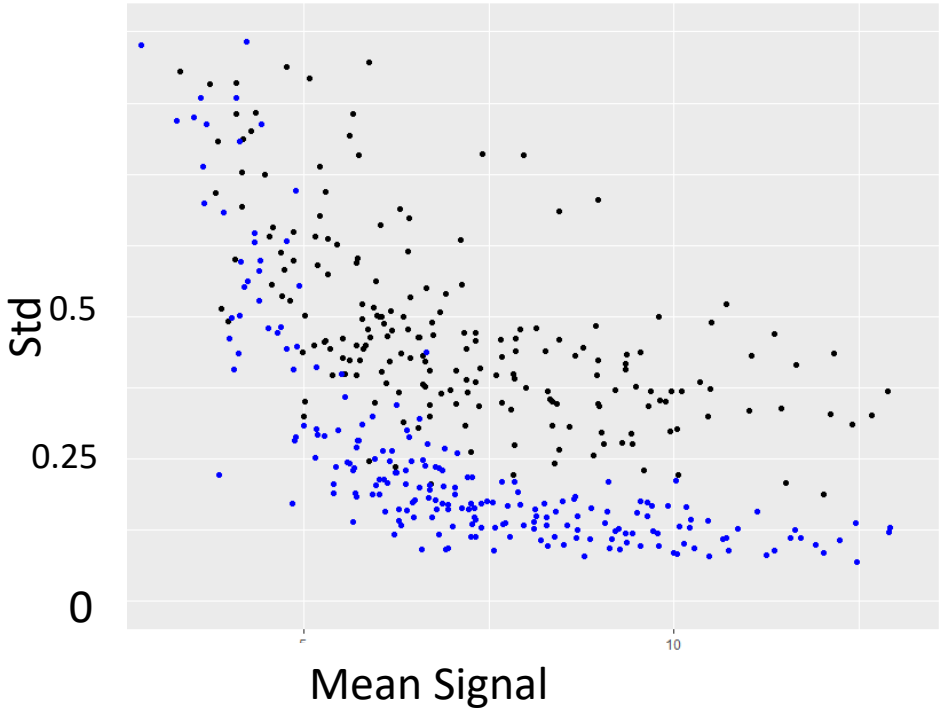


Figure 5: Standard deviation of the repeat measurements with two donor samples as a function of mean signal. Black: without correction, blue: after applying a correction using the bcPBMC samples.

The PCA scores in Figure 6A show that before applying the correction the measurements are dominated by technical variation, i.e. here the clustering of the scores is dominated by variation between batches and runs. After applying the correction (Figure 6B) the variation is dominated by biological variation, i.e. the variation between Donor 1 and Donor 2. Hence, applying the correction based on the bcPBMC sample increases the capability to measure differences between biological subjects as required for METHOD.

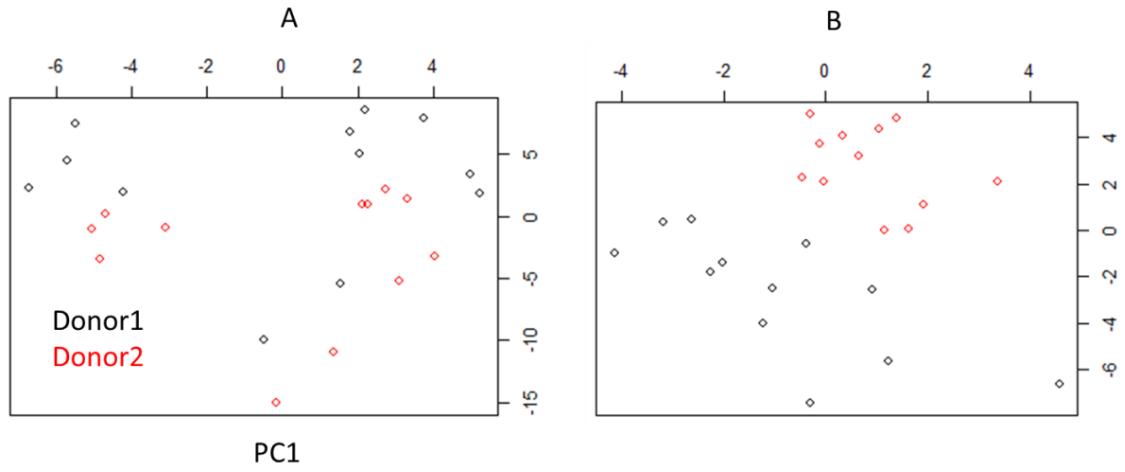


Figure 6: PCA scores to PC1-2 of the measurements with Donor 1 (black) and Donor 2 (red) samples as described in section 6.1. A: without correction B: with correction.

Based on these results it is concluded that the bcPBMC samples are suitable for as IAC samples for measurements with the diagnostic PBMC samples that will be used for METHOD.

6.2 Verification of the pooled buffy coat PBMC sample

For the measurements in section 6.1 four bcPBMC samples were measured separately. Using multiple bcPBMC samples as IAC is beneficial because (1) a lower amount of individual sample is needed per measurement, hence an increased number of measurements can be performed with the same batch of samples, (2) using multiple samples as IAC makes the calculated correction more representative of the general systematic variation between measurements, i.e. it averages out sample specific effects.

However, for execution of the experiments it is more practical to use a single bcPBMC sample. Therefore, it was verified that a pooled reference sample can be used as IAC sample:

- It was verified that pooling bcPBMC lysates (i.e. from three different donors) prior to performing an assay did not result in stability issues with the pooled lysate (e.g. due to potential cross-reactions between the individual samples) [25].
- It was verified that when using pooled bcPBMC sample a similar response to systematic variation was observed as with a single sample and that the pooled sample could be used to correct for systematic variation similar as shown in section 6.1 [26].

6.3 Verification during the Analytical Performance Evaluation

The final implementation and design for the use of the pooled bcPBMC sample was documented in the DAS work instruction for performing the PTK assay (WI7604 – PTK Assay Instructions). Final verification of this implementation was part of the Analytical Performance Evaluation reported in [2]. The results of this study show that by using the pooled bcPBMC sample as an IAC the requirements for Precision of Measurement (PoM) were met and variation was significantly reduced in comparison to the uncorrected measurements.

7 Discussion

This document reports the development and verification of the bcPBMC sample as an IAC that allows correction for systematic variation in the METHOD assay. The results are likely to be relevant for other applications of the PamChip assay as well.

The bcPBMC samples are effective in mitigating variation due to e.g. PamChip batches, run-to-run variation etc. It is important, however, that the reference samples can be prepared and measured in a reproducible manner. For this, the implementation and design in WI7604 will be further optimized.

In addition, routine use of a diagnostic assay such as METHOD in the DAS lab will require periodic verification of the long-term stability of the assay. For this relevant *monitoring* samples must be selected that can be used to check to the long-term stability the measured Diagnostic Score.

Implementation of the IACs in the DAS lab must also involve verification of the (in-use) stability of the bcPBMC samples, supply control of the bcPBMC samples and the development of a procedure to change to a new batch of IAC samples before the current batch is finished.

The current batch of IAC is expected to last for at least a year. Specifications and the procedure for preparation (production) and criteria for release of a new batch of IAC will be defined and documented in a new Work Instruction.

8 Conclusion

bcPBMC samples can be effectively used to mitigate the effects of systematic variation in the METHOD assay and can be implemented as IAC the DAS lab. The use of an IAC is a necessary part of the METHOD / DAS measurement design.

For the purpose of software development for implementation of the ComBat method in the DAS process a detailed description of the ComBat methodology is added to this report in Appendix A and Appendix B.

9 Appendices

- A: 210228RW19030 Appendix A methodology combat
- B: 210228RW19030 Appendix B: re-engineered r-code for the Combat method.

10 References

[1] PamGene International, “210-200RW18017 METHOD product description,” 2018.

[2] PamGene International, “210228RW19019 Analytical Performance Report,” 2019.

[3] PamGene International, “210200KS18003 product Design Requirements METHOD DAS,” 2019.

[4] M. Dabros, M. Amrhein, P. Gujral, and U. von Stockar, “On-Line Recalibration of Spectral Measurements Using Metabolite Injections and Dynamic Orthogonal Projection,” *Appl. Spectrosc.*, vol. 61, no. 5, pp. 507–513, May 2007.

[5] P. Gujral, M. Amrhein, B. M. Wise, and D. Bonvin, “Special Issue Article Framework for explicit drift correction in multivariate calibration models,” 2010.

[6] P. Gujral, M. Amrhein, and D. Bonvin, “Drift correction in multivariate calibration models using on-line reference measurements,” *Anal. Chim. Acta*, vol. 642, no. 1–2, pp. 27–36, May 2009.

[7] M. Zeaiter, J. M. Roger, and V. Bellon-Maurel, “Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations,” *Chemom. Intell. Lab. Syst.*, vol. 80, no. 2, pp. 227–235, Feb. 2006.

[8] J. T. Leek *et al.*, “Tackling the widespread and critical impact of batch effects in high-throughput data,” *Nat Rev Genet*, vol. 11, no. 10, 2010.

[9] E. M. Price and W. P. Robinson, “Adjusting for Batch Effects in DNA Methylation Microarray Data, a Lesson Learned.,” *Front. Genet.*, vol. 9, p. 83, 2018.

[10] C. Chen *et al.*, “Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods.,” *PLoS One*, vol. 6, no. 2, p. e17238, Feb. 2011.

[11] W. E. Johnson, C. Li, and A. Rabinovic, “Adjusting batch effects in microarray expression data using empirical Bayes methods,” *Biostatistics*, vol. 8, no. 1, pp. 118–127, Jan. 2007.

[12] Agendia, “MammaPrint.” [Online]. Available: https://www.agendia.com/our-tests/mammaprint. [Accessed: 25-Jun-2019].

[13] L. J. van ’t Veer *et al.*, “Gene expression profiling predicts clinical outcome of breast cancer,” *Nature*, vol. 415, no. 6871, pp. 530–536, Jan. 2002.

[14] A. M. Glas *et al.*, “Converting a breast cancer microarray signature into a high-throughput diagnostic test.,” *BMC Genomics*, vol. 7, no. 1, p. 278, Oct. 2006.

[15] A. Tahiri *et al.*, “Differential Inhibition of Ex-Vivo Tumor Kinase Activity by Vemurafenib in BRAF(V600E) and BRAF Wild-Type Metastatic Malignant Melanoma,” *PLoS One*, vol. 8, no. 8, 2013.

[16] S. Arni *et al.*, “Ex vivo multiplex profiling of protein tyrosine kinase activities in early stages of human lung adenocarcinoma,” *Oncotarget*, vol. 8, no. 40, pp. 68599–68613, Sep. 2017.

[17] D. Hurkmans *et al.*, “Blood-based multiplex kinase activity profiling as a predictive marker for clinical response to checkpoint blockade in advanced melanoma.,” *J. Clin. Onc*, vol. 36, p. Suppl. Abstracts 9579, 2018.

[18] Chantal Laurijsen, “Graduation report: ‘Kwantitatieve correctie PamChip tyrosine kinase assay,’” 2015.

[19] PamGene International, “210-226RW17003 PamDx QC Calibration,” 2017.

[20] PamGene International, “210-226RW18012 reference pop,” 2018.

[21] PamGene International, “210-226RW18012 Reference PoP.docx (pptx).” PamGene International B.V., 2018.

[22] Y. Zhang, D. F. Jenkins, S. Manimaran, and W. E. Johnson, “Alternative empirical Bayes models for adjusting for batch effects in genomic studies,” *BMC Bioinformatics*, vol. 19, no. 1, p. 262, Jul. 2018.

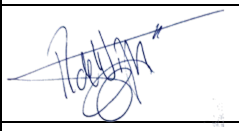
[23] PamGene International, “210-226RW18026 Verification and implementation of reference samples,” 2018.

[24] PamGene International, “210226RW18030 reference verification prelim report.pptx,” 2018.

[25] PamGene International, “210228RW19012 DAS Pre-verification of analytical parameters and control samples. Report 01 - 03,” 2019.

[26] PamGene International, “210228RW19024 DAS pre verification of analytical parameters and control samples 04,” 2019.

11 Signatures and approval

Role	Name	Function Title	Signature	Date
Author	Rik de Wijn	Head of DAS		8-Oct-2020
Reviewer	Kristiane Schmidt	Program Manager IVD		20-Oct-2020
Reviewer	Dianne van den Heuvel	Clinical Project Manager		20-Oct-20
Authorizer	Rik de Wijn	Head of DAS		20-Oct-2020
QARA	Theo van der Leij	QARA Manager		20-Oct-2020

12 History

Version	Date	Remarks/Changes	Document Owner
1.0	3-Oct-2019	Initial Version	Rik de Wijn
1.1	19-Nov-2019	Changed Doc ID (old doc ID same as other doc)	Rik de Wijn
1.2	14-Sep-2020	Added Appendix A and Appendix B	Rik de Wijn