生体の科学　65(2):154-158, 2014

特集　細胞の少数性と多様性に挑む—シングルセルアナリシス
C．シングルセルアナリシスで見えること

単細胞技術に基づくiPS細胞の標準化

山根順子　丸山徹　藤渕航

これまでの生物学の常識を大きく覆した人工多能性幹細胞(iPS細胞；induced pluripotent stem cell)の発見がなされたのが2006年のことである[1]。iPS細胞はES細胞と異なり作製段階にヒト胚を破壊する必要がないことから，ES細胞を用いた研究において大きな障壁となっていた倫理問題が生じず，再生医療を一気に加速させる夢の細胞として登場した。体細胞にわずか数因子を導入するのみで多能性を持った細胞を生み出すことができるという報告はあまりにセンセーショナルであり，それ以降様々な細胞種由来のiPS細胞の樹立や，より安全かつ効率的な樹立法が次々に見いだされた。また，iPS細胞を用いた幹細胞生物学としての基礎研究や再生医療，創薬へ向けた応用研究など，多数の報告がなされている。世界中で樹立が試みられ報告されているiPS細胞は，樹立された数だけ質の異なる細胞になっている可能性が指摘され，今度は質の良いiPS細胞を選別する手法を開発するという新たな研究の方向性も生まれた。

iPS細胞は通常コロニー(細胞集団)として維持培養される。しかしながら，コロニーのなかでも均一な状態ではなく細胞の個性があることがわかっており，集団レベルでiPS細胞の解析を続けていくだけでは標準化を目指すことは難しい。そこでわれわれはより解像度を上げた解析が必要になると考え，従来のような“細胞集団”として遺伝子発現レベルを調べるのではなく，“個”としての細胞，つまり“シングルセルレベル”での遺伝子発現を調べ，標準化に向けた試みを行った。

Ⅰ．シングルセルトランスクリプトーム解析

2009年にTangら[2]によってシングルセルレベルでRNA-seqを行う方法が発表されてから，これまでに幾つかの手法が報告されてきた。われわれが行ったシングルセルトランスクリプトーム解析の手法は各細胞を判別するためのDNAバーコードをtemplate switchingにより導入するSTRT法[3]であり，同時に大量のシングルセルを解析できる利点がある。この方法では，mRNAの5'末端にDNAバーコードを導入するため，次世代シークエンサーで読まれたリードは5'末端側の配列であるという特徴がある。その他の手法で最近報告されたものでは，SMART-seq[4]やCEL-seq[5]，Quartz-seq[6]などがあり，そのうちCEL-seq，Quartz-seqは3'末端側のバイアスがあることが知られている。加えて，これまで問題とされてきたデータの精度についても，Quartz-seqでは細胞周期が区別できるまで改善されており，この先シングルセルトランスクリプトーム解析が大幅に普及することが予想される。

Standardization of iPS cells by single-cell transcriptome analysis

Yamane Junko：京都大学iPS細胞研究所(特定研究員：〒606-8507 京都市左京区聖護院川原町53)
Maruyama Toru：早稲田大学大学院 先進理工学研究科 生命医科学専攻(修士課程)，京都大学iPS細胞研究所(特別研究学生)
Fujibuchi Wataru：京都大学iPS細胞研究所(教授)

0370-9531/04/¥500/論文/JCOPY

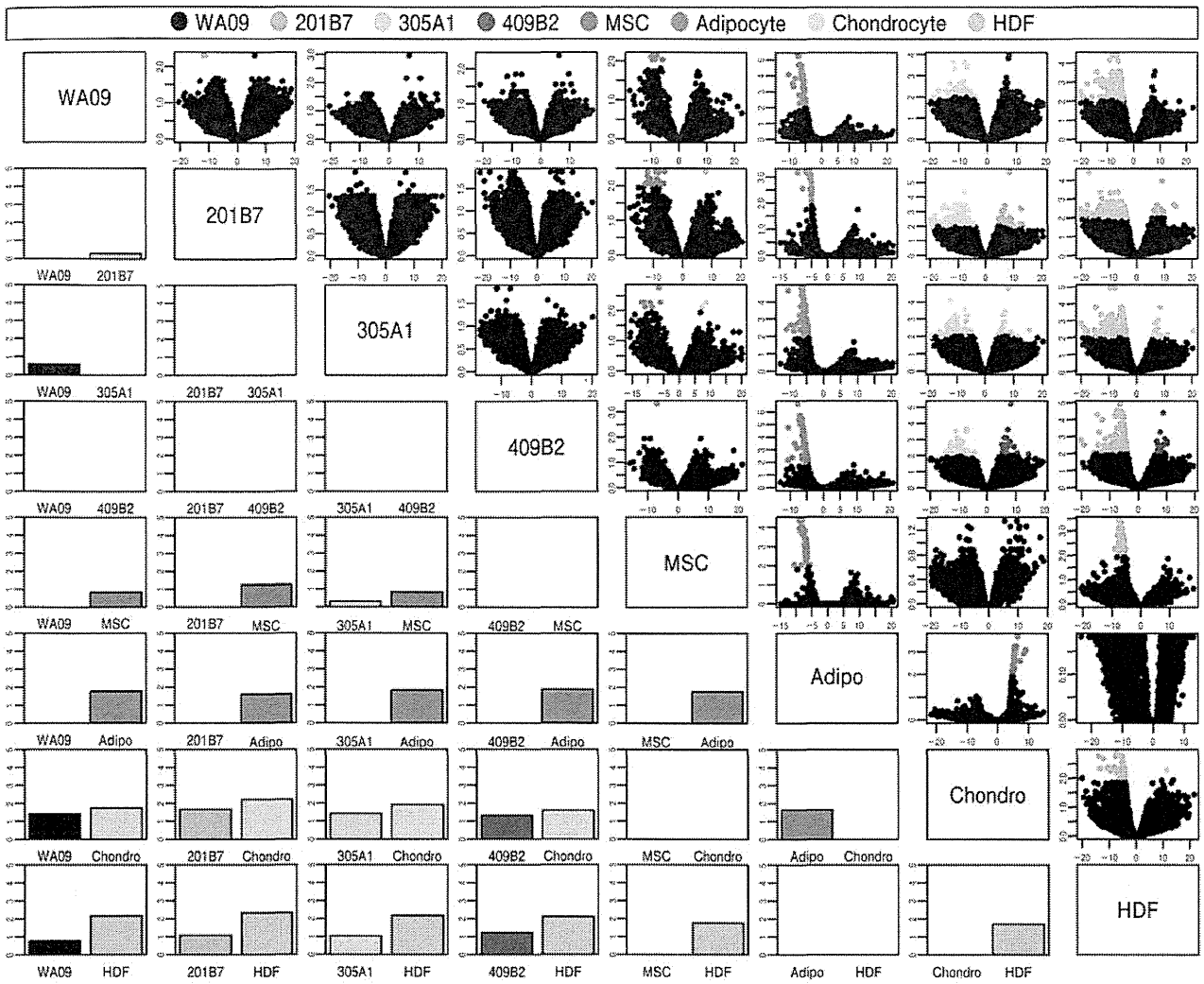

図 1　DEG 解析

対角線より上：Volcano plot（横軸：$\log_2$(FC)，縦軸：$\log_{10}$(FDR)），色がついたプロットは有意差（FDR<0.01）があった遺伝子。対角線より下：DEG の数を示す棒グラフ。プロット・棒の色は細胞種に対応している。例えば赤いプロットは Adipocyte で発現が高い遺伝子を，赤い棒は Adipocyte で発現が高い遺伝子の数を表す。

Ⅱ．シングルセルトランスクリプトーム解析により見えてきた特徴

われわれは異なる手法によって樹立された iPS 細胞 3 種類（201B7，305A1，409B2）に加えて ES 細胞 1 種類（WA09），ヒトの体細胞 4 種類（MSC：間葉系幹細胞，Chondro：軟骨細胞，adipo（誘導させた）脂肪細胞，HDF：皮膚線維芽細胞）の計 8 細胞種を対象にしてシングルセルレベルのトランスクリプトーム解析を行った。

まず，Bioconductor の edgeR パッケージを用いて 2 細胞種間で真の発現変動遺伝子（DEG）の数を調べたところ，ES 細胞と iPS 細胞間で発現変動がある遺伝子はほとんど見つからなかった（図 1）。しかし，本解析は細胞集団を対象にした手法を適用しているので，シングルセルレベルでの解析とは結果が異なる可能性がある。

そこでシングルセルレベルでの各種細胞における遺伝子発現パターンの特徴をより詳細に調べるため，8 細胞種の主成分分析（PCA）を行ったところ，おおよそ細胞種ごとにクラスターを形成していることが確認できた（図 2）。この結果から軟骨細胞と HDF のクラスターが重なっていることがわかったが，これらの細胞種間では遷移が起こることが知られている[7]。よって，単一細胞のトランスクリプトームには細胞系譜における細胞間の関係性や，細胞間で遷移が起こりうる可能性といった新たな情報が包含されている可能性が示唆

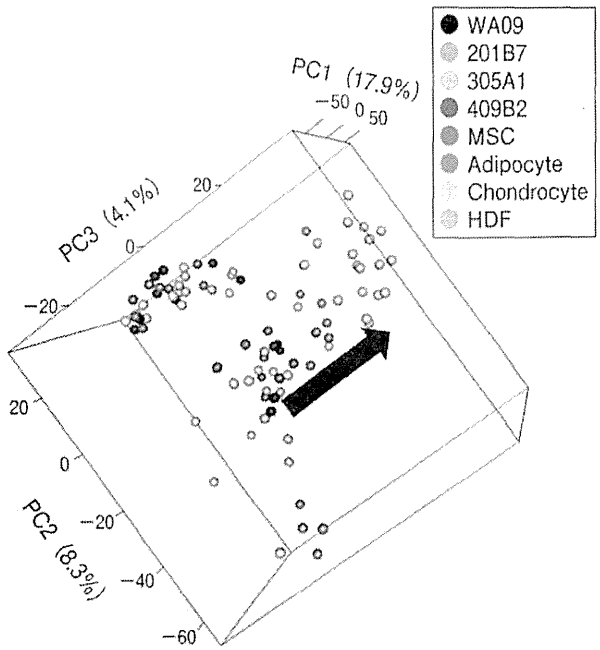

図 2　8細胞種の主成分解析(PCA)
細胞ごとにある程度クラスターができている。括弧内は各種成分の寄与率を示す。HDFとChondrocyteが混在しているが，これはHDFとChondrocyteの間でトランジションが起こりやすいことを示唆していると考えられる。また，矢印の方向に相転移が起こりうることを示唆している。

された。これまでの集団として解析してきた方法では平均化されてしまい見えていなかった細胞本来のばらつきの様相の一端が明らかになった。

また，遺伝子発現の揺らぎの大きさを評価するために，細胞種ごとに全遺伝子の変動係数(CV)を調べた(図3)。細胞集団での解析と一致して，すべての細胞でハウスキーピング遺伝子のCVは他の遺伝子と比較して遺伝子発現の揺らぎが小さいという結果が得られた。微量なサンプルを対象とするシングルセル解析を行ううえで，得られた結果が細胞由来の揺らぎによるものか，テクニカルな問題により生じる差を見ているのかを見極めることは非常に重要である。よって，指標の一つとなるハウスキーピング遺伝子の発現の揺らぎが他の遺伝子と比べて小さいという結果はシングルセルを扱う研究を進めるにあたり精度の良さを確認するための大きな情報となり得る。

最後に，全遺伝子の変動係数の解析において今回調べた細胞間でヒストグラム分布の異なる遺伝子について一部報告する。まず，DNAメチル化にかかわるDNMT3Bは既報のように[8]，ES細胞，iPS細胞の多能性幹細胞群ではほぼ共通し発現レ

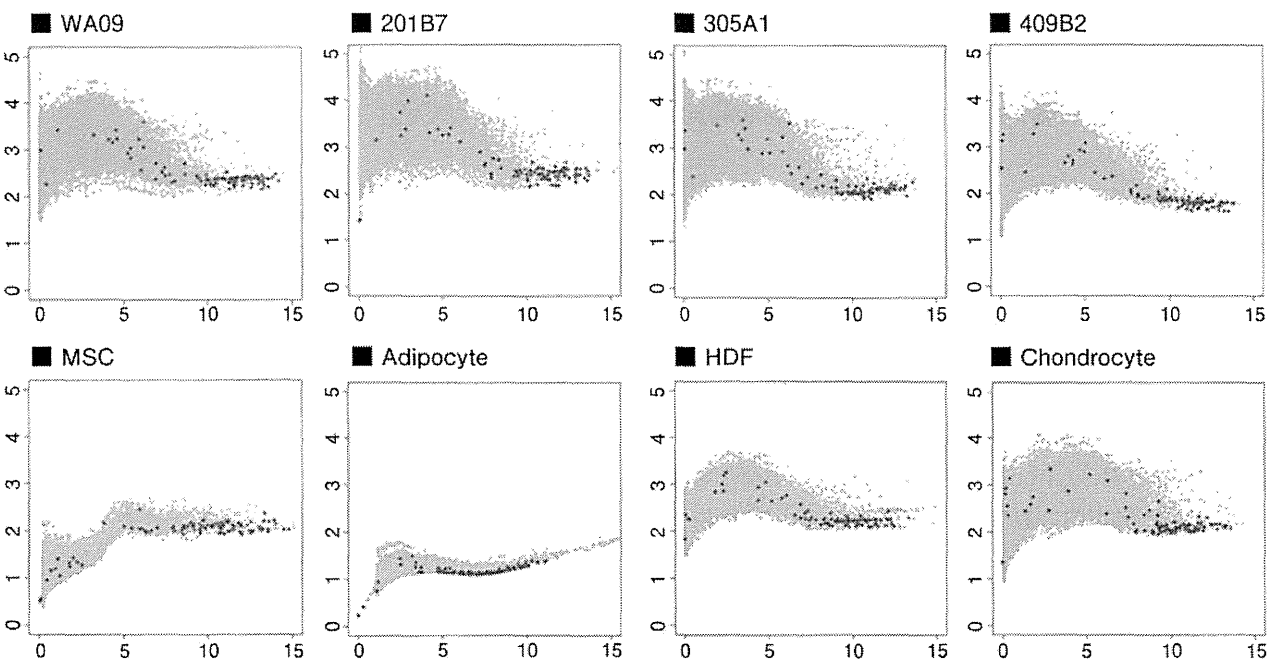

図 3　全遺伝子の変動係数(CV)
各点は遺伝子に対応しており，赤点はハウスキーピング遺伝子に対応している。横軸は$\log_2$(RPM)の平均，縦軸はedgeRによって算出した発現の変動(biological coefficient of variation)を表している。下に位置する遺伝子ほど安定に発現していることを意味する。

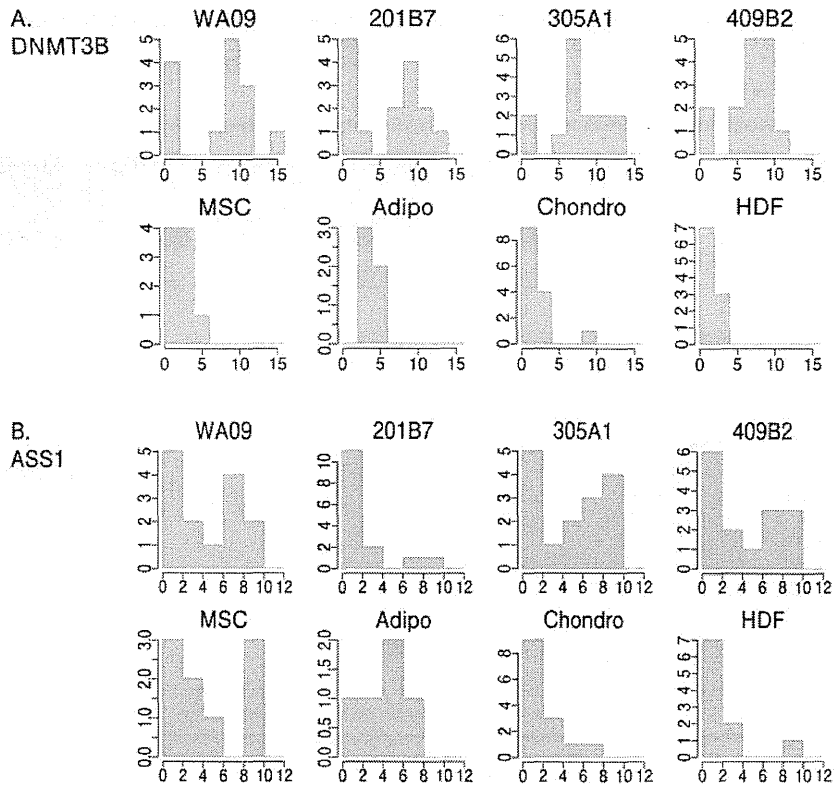

図 4　ヒストグラムの分布で差が見られた遺伝子例(横軸は $\log_2$(RPM+1))

A. は多能性幹細胞と分化細胞の間で顕著な差が見られた。B. は多能性幹細胞の中の比較で一部の株(201B7)で異なる分布が見られた。

ベルの高い分布が見られ，逆に分化細胞ではいずれも発現レベルの低い分布が傾向として見られた(図 4A)。また，アルギニン生合成経路にかかわる酵素の一つである ASS1 は多能性幹細胞群の中で比較をした場合，201B7 株でのみ低発現の分布を示している傾向があった。この遺伝子は分化細胞でも比較的低発現の分布を示している(図 4B)。このように平均で解析していると見えてこなかった遺伝子発現の分布が，個々のレベルで見ると差が見えてくるケースがある。

以上のように，われわれはシングルセルトランスクリプトーム解析によって，これまでの細胞集団を対象にする解析では見えなかった事象を確認することができた。

おわりに

シングルセル RNA-seq 解析が普及するに伴い，これまで 250 種類程度と言われてきた細胞がより細かく分類されるようになる可能性がある。そこで，われわれの研究室では細胞のデータベースである“SHOGoiN”を開発しており，細胞の遺伝子発現プロファイル，画像，形態計測情報，実験条件などのメタデータ，文献情報を貯蔵している(図 5)。ユーザーはウェブブラウザを通して本データベースにアクセスし，細胞に関する情報を検索することができる。

今後，われわれはシングルセル RNA-seq 解析と体系的に貯蔵された細胞のデータを用いて，細胞に普遍的に存在する理論を見つけることを目指している。その一つとして次のような問題に取り組んでいる。

iPS 細胞を作製する際には，元となる体細胞に *Oct4*, *Sox2*, *Klf4*, *c-Myc* などの遺伝子を導入し，強制的に発現させる必要がある。しかし，すべての細胞に同じ刺激を与えているにもかかわらず，なぜ iPS 細胞の作製効率は悪く，質の良し悪しがあるのだろうか。われわれはこの問題に対して以下のようなモデルを考えている。遺伝子の発現は同じ種類の細胞であっても均一でなく，揺らぎがある。揺らぎは転写因子の結合や転写反応，

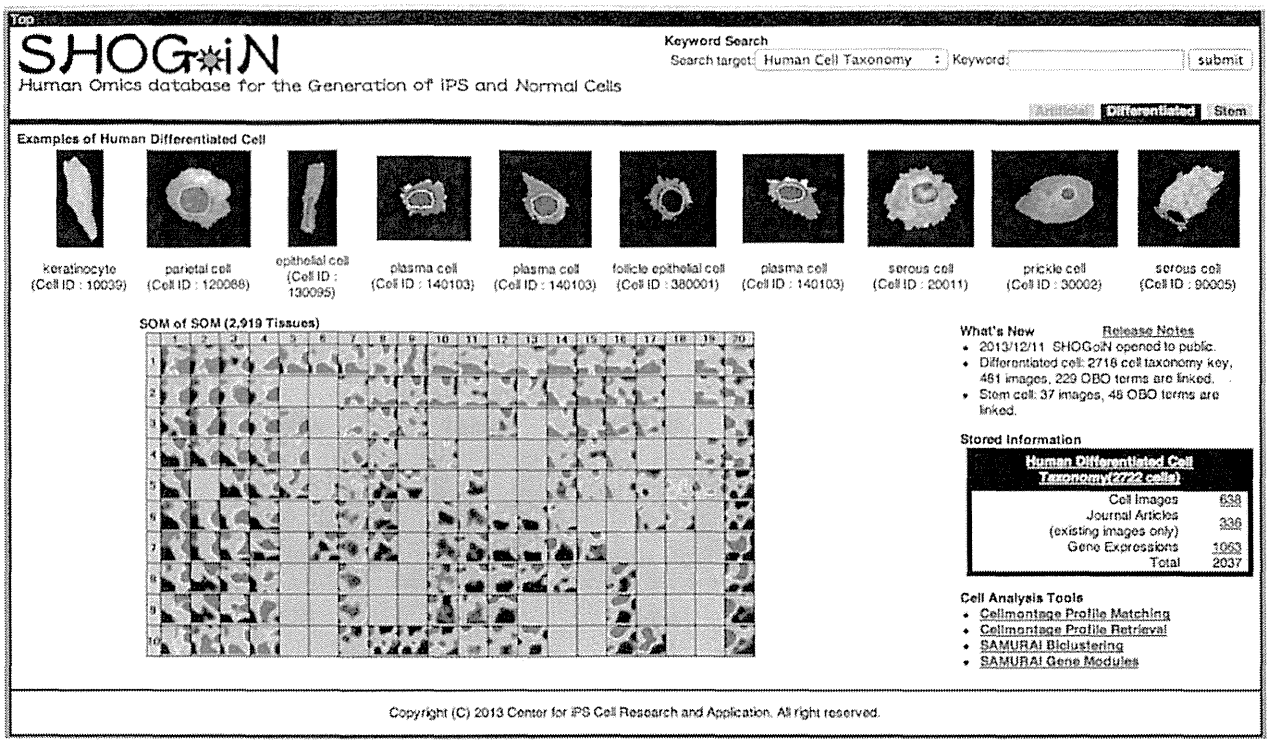

図5 ヒト細胞情報統合データベース“SHOGoiN”
http://shogoindb.cira.kyoto-u.ac.jp

RNAの分解といったイベントが確率的であることに起因している[9]。Huangらの論文[10]にもあるように，われわれはこういった遺伝子発現の揺らぎが細胞状態の変化に影響を与えているため，細胞分化やiPS細胞の作製が確率的なものになっているのではないかと考えている。よって，揺らぎが細胞分化に与える影響を解き明かすことができれば，質の良い細胞をより効率的に作製する手法が明らかになるかもしれない。また，シングルセル解析により得られた詳細な情報をデータベースに集約し，広く研究者らが利用することで，iPS細胞の標準化や，そこから分化させた各種分化細胞の標準化に向けた足がかりとなることを期待している。

謝辞 本研究は『内閣府最先端研究開発支援プログラム（FIRSTプログラム）』の支援のもと行いました。

●文 献

1) Takahashi K, Yamanaka S：*Cell* **126**：663-676, 2006
2) Tang F, Barbacioru C, Wang Y et al：*Nat Methods* **6**：377-382, 2009
3) Islam S, Kjällquist U, Moliner A et al：*Genome Res* **21**：1160-1167, 2011
4) Ramsköld D, Luo S, Wang Y-C et al：*Nat Biotech* **30**：777-782, 2012
5) Hashimshony T, Wagner F, Sher N et al：*Cell Rep* **2**：666-673, 2012
6) Sasagawa Y, Nikaido I, Hayashi T et al：*Genome Biol* **14**：R31, 2013
7) Outani H, Okada M, Yamashita A et al：*ProS One* **8**：e77365, 2013
8) Takahashi K, Tanabe K, Ohnuki M et al：*Cell* **131**：861-872, 2007
9) Wills QF, Livak KJ, Tipping AJ et al：*Nat Biotechnol* **31**：748-752, 2013
10) Chang HH, Hemberg M, Barahona M et al：*Nature* **453**：544-547, 2008

Analytical Biochemistry 472 (2015) 75–83

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

A set of external reference controls/probes that enable quality assurance between different microarray platforms

Hideo Akiyama [a,*], Yoji Ueda [a], Hitoshi Nobumasa [a], Hiroyuki Ooshima [b], Yohei Ishizawa [c], Koji Kitahiro [d], Isao Miyagawa [d], Kazufumi Watanabe [e], Takazumi Nakamura [f], Ritsuka Tanaka [g], Nobuko Yamamoto [h], Hiroki Nakae [h], Mitsuo Kawase [i], Nobuhiro Gemma [j], Yuji Sekiguchi [k], Wataru Fujibuchi [l], Ryo Matoba [c,*]

[a] New Projects Development Division, Toray Industries, Inc., Kamakura, Kanagawa 248-8555, Japan
[b] Yokohama Research Laboratories, Mitsubishi Rayon Co. Ltd., Yokohama, Kanagawa 230-0053, Japan
[c] DNA Chip Research Inc., Yokohama, Kanagawa 230-0045, Japan
[d] Technical Research Laboratory, Kurabo Industries Ltd., Neyagawa, Osaka 572-0823, Japan
[e] Hokkaido System Science Co. Ltd., Sapporo, Hokkaido 001-0932, Japan
[f] S-BIO Business Division, Sumitomo Bakelite Co. Ltd., Amagasaki, Hyogo 661-8588, Japan
[g] Department of Bio Research, Kamakura Techno-Science, Inc., Kamakura, Kanagawa 248-0036, Japan
[h] Japan Multiplex bio-Analysis Consortium (JMAC), Chiyoda-ku, Tokyo 102-0083, Japan
[i] Graduate School of Biomedical Engineering, Tohoku University, Sendai, Miyagi 980-8579, Japan
[j] Ricoh Institute of Technology, Ricoh Company, Ltd., Yokohama, Kanagawa 224-0035, Japan
[k] Bio-Measurement Research Group, Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, Ibaraki 305-8566, Japan
[l] Computational Biology Research Center, Advanced Industrial Science and Technology (AIST), Koto-ku, Tokyo 135-0064, Japan

ARTICLE INFO

Article history:
Received 18 August 2014
Received in revised form 16 November 2014
Accepted 19 November 2014
Available online 4 December 2014

Keywords:
External RNA standards
DNA microarray
Dynamic range
Cross-platform
Multiplatform calibration

ABSTRACT

RNA external standards, although important to ensure equivalence across many microarray platforms, have yet to be fully implemented in the research community. In this article, a set of unique RNA external standards (or RNA standards) and probe pairs that were added to total RNA in the samples before amplification and labeling are described. Concentration–response curves of RNA external standards were used across multiple commercial DNA microarray platforms and/or quantitative real-time polymerase chain reaction (RT–PCR) and next-generation sequencing to identify problematic assays and potential sources of variation in the analytical process. A variety of standards can be added in a range of concentrations spanning high and low abundances, thereby enabling the evaluation of assay performance across the expected range of concentrations found in a clinical sample. Using this approach, we show that we are able to confirm the dynamic range and the limit of detection for each DNA microarray platform, RT–PCR protocol, and next-generation sequencer. In addition, the combination of a series of standards and their probes was investigated on each platform, demonstrating that multiplatform calibration and validation is possible.

Recent advances in DNA microarray technology have opened up new applications in both basic and clinical research [1–4]. Consequently, new tests in many areas of biomedical science, including clinical pharmacogenetics, cancer genotyping, and cancer prognosis, have been developed [5–7].

Clinical applications of DNA microarray technology include gene expression analysis for early disease detection, disease classification and diagnosis, selection of treatment protocol, determination of changes in disease status, and the monitoring of therapeutic effects and side effects. A clinical application in which DNA microarray gene expression analysis has already been applied is the "MammaPrint," developed in the United States and Europe, used to select the optimal breast cancer treatment [5]. In addition, OncoType DX, a product based on quantitative real-time polymerase chain reaction (RT–PCR)[1], has also been used for analyzing the expression of multiple RNA targets as an indicator in the selection of optimal breast cancer treatment [6].

* Corresponding authors.
E-mail addresses: hideo_akiyama@nts.toray.co.jp (H. Akiyama), matoba@dna-chip.co.jp (R. Matoba).

[1] *Abbreviations used:* RT–PCR, real-time polymerase chain reaction; HURR, human universal reference total RNA; HBRR, human brain reference total RNA; JMAC, Japan Multiplex bio-Analysis Consortium; cDNA, complementary DNA; 3D, three-dimensional; aRNA, antisense amplified RNA; SSC, sodium saline citrate; SDS, sodium dodecyl sulfate; PBS, phosphate-buffered saline; mRNA, messenger RNA.

http://dx.doi.org/10.1016/j.ab.2014.11.012
0003-2697/

However, if DNA microarray data are to be routinely used for clinical applications, it is vital that the data are both reliable and reproducible and that errors or ambiguities in the interpretation of results are eliminated [8–10]. In particular, because gene expression is highly variable, quality assurance in the handling of specimens—storage conditions, transport conditions, and pretreatment protocols—must be robust (Fig. 1).

We report here the development of a set of unique RNA external standards (or RNA standards) and probe pairs that may be spiked into test samples to ensure equivalence across many microarray platforms. This suite of synthetic nucleotides is derived from unique non-mammalian sequences and designed to minimize cross-hybridization with common transcripts from humans, mice, and rats. Six microarray platforms were evaluated using this set of standards: 3D-Gene (Toray Industries, Tokyo), Agilent SurePrint (Agilent Technologies, Santa Clara, CA, USA), Genopal (Mitsubishi Rayon, Tokyo), GeneSQUARE (Kurabo Industries, Osaka, Japan), S-Bio (Sumitomo Bakelite, Tokyo), and NimbleGen (Roche NimbleGen, Basel, Switzerland). An RT–PCR protocol (Life Technologies, Foster City, CA, USA) and a next-generation sequencer GAII (Illumina, San Diego, CA, USA) were also tested. We compared performance across DNA microarray platforms and/or RT–PCR and next-generation sequencing by spiking a set of our standards into a commonly available commercial total RNA sample. A variety of standards can be added in a range of concentrations spanning high and low abundances, thereby enabling the evaluation of assay performance across the expected range of concentrations found in a clinical sample.

Using this approach, we show that we are able to confirm the dynamic range and the limit of detection for each DNA microarray platform, RT–PCR protocol, and next-generation sequencer. In addition, the combination of a series of standards and their probes was investigated on each platform, demonstrating that multiplatform calibration and validation is possible (Fig. 2).

Materials and methods

RNA external standard transcripts

Ten candidate external RNA standard clones (in pUC19 plasmid) were synthesized from artificial sequences designed to have the following characteristics: (i) a unique sequence that exhibits low similarity with any eukaryotic genome and EST sequence known to date, (ii) no nucleic acid homopolymer longer than three bases, (iii) a G+C content in the range of 40 to 60%, (iv) no repeated sequences such as a motif, and (v) no strong secondary structure within the sequence. The standard sequences were designed by using our original program software. Inserts for the clones are 500 to 1000 bp with a 30-bp polyadenylated tail and T7 promoter sequence. All candidate standards were prepared by in vitro transcription of linearized plasmids using a T7 RNA polymerase (MEGAScript Kit, Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Ten transcripts corresponding to the RNA external standards were purified using TURBO DNase (Life Technologies) and further purified by phenol–chloroform extraction and ethanol precipitation. The 10 standard transcripts were dissolved in RNase-free water and then quantified using a Quant-iT RNA Assay Kit (Life Technologies). The sequences of the external standards (R001-500 to R010-1000) have been deposited in the DDBJ/GenBank/EMBL databases under the accession numbers AB610939 to AB610950.

RNA external standard spiked total RNA cocktail

Human universal reference total RNA (HURR, Agilent Technologies) and human brain reference total RNA (HBRR, Agilent Technologies) controls were used. Ten external RNA standards were diluted using HURR or HBRR RNA solution at 50 ng/ml. The standard spiked total RNA cocktail (see Supplementary Tables S1 and S2 in online supplementary material) was prepared at the Japan Multiplex bio-Analysis Consortium (JMAC) central laboratory and delivered to each test site.

Design of probe for RNA external standards

For probe design, each external standard was divided into two regions as follows: 1- to 300-nt and 301- to 500-nt regions for 500-nt RNA and 1- to 500-nt and 501- to 1000-nt regions for 1000-nt RNA, numbering from their 3′ ends. All candidate sequences from the sense strand were extracted by moving 60-nt windows in each region.

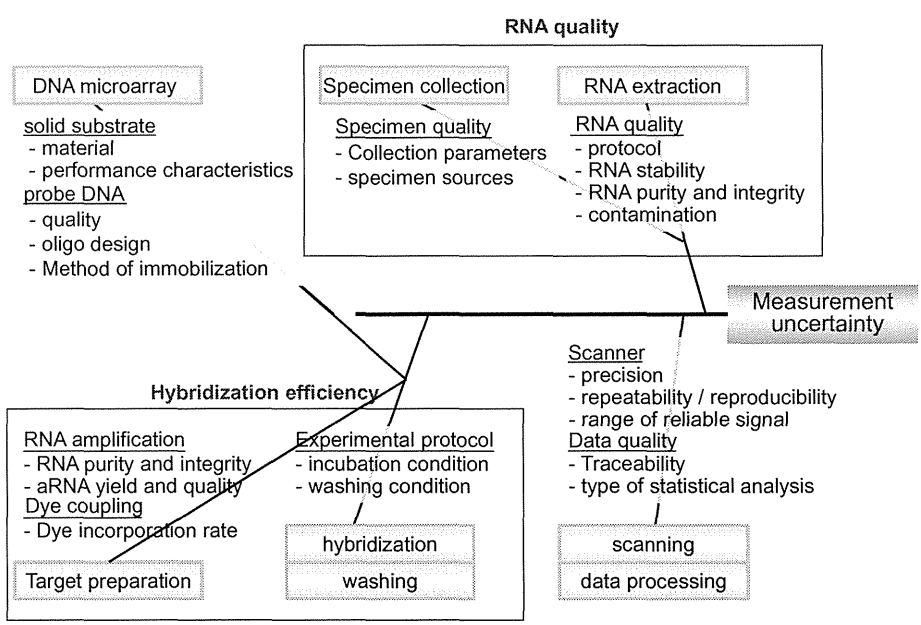

Fig.1. Measurement uncertainty of DNA microarray analysis. Unless the uncertainties of a measurement are being evaluated and stated, the fitness for the purpose of measurement cannot be judged properly. The uncertainties of a measurement using microarray are complicated and intertwined. The sources of uncertainties come from mainly the platform material, RNA quality, and hybridization efficiency and during data acquisition and processing.

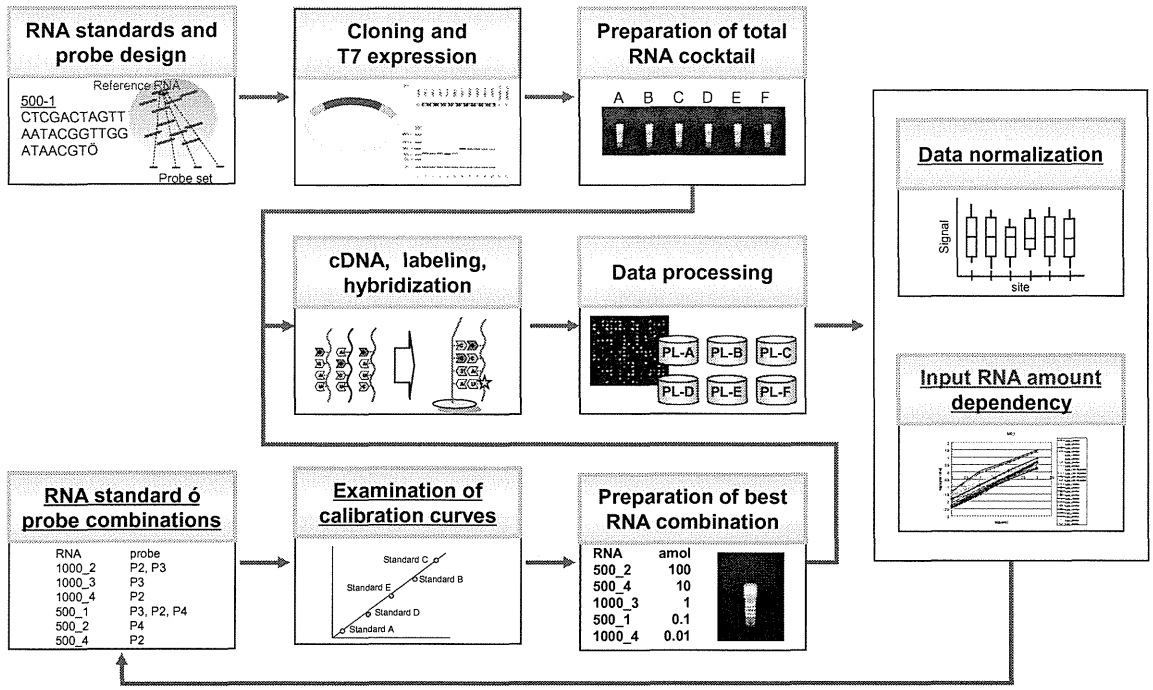

Fig.2. Experimental design of data comparison and quality assurance among different microarray platforms. Spiking RNA external standards into testing samples is an essential and effective method to monitor the quality of a microarray experiment, starting from sample preparation, hybridization, to data analysis.

First, the cross-hybridization potentials of candidate probes against off-target RNAs were analyzed. Because exact calculation based on thermodynamics requires a large computational cost, the search program in the FASTA package [11] (version 3.5) with default parameters to align candidate probes against both strands of 56,155 human complementary DNA (cDNA) sequences from the Ensembl database (http://www.ensembl.org, release 49) was used, and the top 100 off-target cDNA sequences that have the closest similarity to each probe based on the alignment scores were selected for potential cross-hybridization targets. To refine these results, the hybrid-min program in the UNAFold package [12] was performed to calculate the free energy change of hybridization, and then a program to calculate the cross-hybridization ratios for each probe based on Ref. [13] was coded. A cross-hybridization ratio $\geqslant$10–3 was removed from the candidate probes.

Second, T_m values using the nearest neighbor method [14] were calculated, and four candidate probes for each RNA standard that had the closest T_m to 80 °C were selected (Supplementary Table S3A). Potentials for dimerization and secondary structure formation were also calculated by hybrid-min and hybrid-ss in UNAFold. For thermodynamic calculations, 0.5 μM of primers, 2 mM Mg^{2+}, and 50 mM Na^{+} parameters were used.

DNA microarray platform analyses

3D-Gene

The custom DNA microarray was constructed using the 3D-Gene platform (Toray Industries) [15] and spotted with the DNA probes (140 probes) shown in Supplementary Tables S3A and S3B. The 3D-Gene platform has a three-dimensional (3D) array that is constructed within a well with the oligonucleotide probes on the top. A total RNA cocktail (0.5 μg) was amplified and labeled using an Amino Allyl MessageAmp II aRNA Amplification Kit (Life Technologies) according to the manufacturer's instructions. Each sample of aRNA (antisense amplified RNA) labeled with Cy5 was hybridized with 3D-Gene at 37 °C for 16 h. After hybridization, the DNA microarray was washed and dried. Hybridization signals derived from Cy5 were scanned using Scan Array Lite (PerkinElmer, Waltham, MA, USA). The scanned image was analyzed using GenePix Pro 6.0 software (Molecular Devices, Sunnyvale, CA, USA). Spots that might be associated with artifacts were eliminated using software- and visual-guided flags. In this study, the background (blank) average was subtracted from the median values of the foreground signals that are higher than the background (blank) average + 2 standard deviations to give a feature intensity.

Agilent SurePrint

The custom microarray used in this study was designed using the Agilent e-Array platform (Agilent Technologies). Total RNA cocktail (0.5 μg) was used as a starting material to prepare Cy3-labeled aRNA. Fluorescently labeled aRNA was produced using the Quick Amp Labeling Kit (Agilent Technologies) and purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany). The Cy3-labeled 600-ng aRNA was fragmented and hybridized at 65 °C for 17 h to microarray platform slides using the Agilent Gene Expression Hybridization Kit (Agilent Technologies). The microarray platform slides were washed and scanned with an Agilent scanner. The fluorescent intensities of individual spots were obtained with Feature Extraction (version 10.5.1.1, Agilent Technologies).

Genopal

The custom oligonucleotide microarray, Genopal (Mitsubishi Rayon), was made in the following manner. Plastic hollow fibers were bundled in an orderly arrangement, and hardened with resin to form a block. Oligonucleotide capture probes (140 probes) were chemically bonded inside each hollow fiber with hydrophilic gel [16]. The block was then sliced to make thin microarray platforms, each of which was set into a holder (for details, see http://www.mrc.co.jp/genome/e).

Total RNA cocktail was amplified using the MessageAmp II Biotin-Enhanced Amplification Kit (Life Technologies) according to the manufacturer's instructions, and was column purified. Biotinylated RNA (5 μg) was fragmented by incubation with fragmentation reagents (Life Technologies) at 94 °C for 7.5 min. Hybridization was carried out with DNA microarray in 150 μl of hybridization buffer (0.12 M Tris-HCl, 0.12 M NaCl, and 0.05% Tween 20) and

5 μg of fragmented biotinylated RNA at 65 °C overnight. For post-hybridization, the DNA microarray was washed twice in 0.12 M Tris–HCl/0.12 M NaCl/0.05% Tween 20 at 65 °C for 20 min, followed by washing in 0.12 M Tris–HCl/0.12 M NaCl for 10 min. The DNA microarray was then labeled with streptavidin–Cy5 (GE Healthcare Bio-Sciences, Tokyo). The fluorescent-labeled DNA microarray was washed for 5 min four times in 0.12 M Tris–HCl/0.12 M NaCl/0.05% Tween 20 at room temperature. Hybridization signal acquisition was performed using a DNA microarray reader adopting multi-beam excitation technology (Yokogawa Electric, Tokyo). The DNA microarray was scanned at multiple exposure times ranging from 0.1 to 40 s. The intensity values with the best exposure condition for each spot were then selected.

GeneSQUARE

The custom DNA microarray was designed using the GeneSQUARE Multiple Assay DNA Microarray platform. Alexa Fluor 555-labeled cDNA was prepared from total RNA cocktail (10 μg) by cDNA synthesis and in vitro transcription performed using the GeneSQUARE cDNA Direct Labeling System (Kurabo Industries) according to the manufacturer's instructions. Labeled cDNA was purified with a MinElute PCR Purification Kit (Qiagen) and added to hybridization buffer (5× sodium saline citrate [SSC, pH 7.0], 4× Denhardt's solution [Sigma–Aldrich, St. Louis, MO, USA], 1 μg of salmon sperm DNA [Life Technologies], and 0.5% [w/v] sodium dodecyl sulfate [SDS]). Hybridization was performed in a final volume of 8 μl per well on a GeneSQUARE Multiple Assay DNA Microarray (JMAC) in a hybridization chamber (Kurabo Industries) at 65 °C for 16 h in a water bath. After hybridization, the hybridized slides were washed by the following steps: (i) immersion in 1× SSC and 0.1% SDS solution for 5 min, (ii) immersion in 0.2× SSC and 0.1% SDS for 5 min, (iii) immersion in 0.2× SSC and 0.1% SDS at 55 °C for 5 min, (iv) rocking in 0.2× SSC, and (v) immersion in 0.05% SSC for 2 min. After they were dried by centrifuge, the slides were scanned with GenePix 4000B (Molecular Devices). Fluorescence intensities of scanned images were quantified with GenePix Pro 6.0 software (Molecular Devices).

S-Bio

The custom microarray was designed using the S-Bio plastic slide platform (Sumitomo Bakelite) and oligonucleotide probes spotted with ProbeBank [17,13] and EC amino linker [18]. The total amount of RNA required can be reduced by mechanical spotting using the GENESHOT Spotting Device (NGK Insulators, Nagoya, Japan).

Total RNA cocktail was amplified using the MessageAmp II Biotin-Enhanced Single Round aRNA Amplification Kit (Life Technologies). Briefly, the total RNA cocktail of each sample (1 μg each) was transcribed into double-stranded T7 RNA polymerase promoter-tagged cDNA and then amplified into single-stranded biotin-labeled aRNA by T7 polymerase. aRNA (3 μg) was fragmented at 94 °C for 15 min and hybridized on the microarray in the presence of formamide (final concentration 10%, v/v) at 37 °C for 16 h. The microarray was washed at room temperature for 5 min in 0.1× SSC and 0.1% SDS, followed by another 5 min wash in 0.05× SSC and 0.1% SDS at 43 °C. Finally, the microarray was rinsed in 0.05× SSC before drying by low-speed centrifugation. For staining, the microarray was immersed in a phosphate-buffered saline (PBS) solution containing 10 μg/ml streptavidin, R-phycoerythrin conjugate (Life Technologies), Tween 20 (0.05%, v/v), and bovine serum albumin (2 mg/ml) for 30 min. Washing was performed to remove the additional stain at room temperature in PBS buffer for 5 min, followed by another wash in a similar buffer prepared separately for 30 s. The microarray was rinsed in 0.05× SSC at room temperature before drying by low-speed centrifugation.

The microarray was scanned using an Agilent DNA Microarray platform scanner (Agilent Technologies, cat. no. G2565BA) at photo-multiplier tube 800 in a resolution of 10 μm. The intensity values of each feature of the scanned image were quantitated using Feature Extraction software (version 9.1, Agilent Technologies).

NimbleGen

The custom microarray was designed using the NimbleGen platform (Roche NimbleGen). Total RNA cocktail (0.5 μg) was used as a starting material to prepare double-stranded cDNA using a SuperScript Double-Stranded cDNA Synthesis Kit (Life Technologies). cDNA was labeled using the random priming method with Cy3-labeled random nonamer primers and Klenow DNA polymerase at 37 °C for 2 h using a NimbleGen One-Color DNA Labeling Kit (Roche NimbleGen.). The Cy3-labeled cDNA (4 μg) was hybridized to the DNA microarray using a NimbleGen Hybridization Kit (Roche NimbleGen) for 17 h at 42 °C. The microarray platform slides were washed and scanned with a NimbleGen MS200 microarray platform scanner (Roche NimbleGen).

Next-generation sequencer GAII

The RNA samples with standard RNA were prepared with an mRNA-Seq Sample Prep Kit (Illumina, San Diego, CA, USA). A 75-base single run was performed on the next-generation sequencer using a Single-Read Cluster Generation Kit (version 4) and TruSeq SBS v5-GA (Illumina GAII) with two samples. The read sequences were aligned against the human genome and the standard RNA sequences using a BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [19].

Taqman RT–PCR

A High-Capacity cDNA Reverse Transcription Kit (Life Technologies) was used for cDNA synthesis. The messenger RNA (mRNA) level was monitored with the 7500 Fast Real-Time PCR System (Life Technologies) and TaqMan Fast Universal PCR Master Mix (Life Technologies) following the manufacturer's instructions. TaqMan primer probes for standard combination number 183 except external standard 1000_4 were custom designed as follows: 500_2, 500_4, 1000_3, 500_1.

In addition, Taqman primer probes for ACTB (TaqMan Probe ID: Hs03023880_g1), B2M (TaqMan Probe ID: Hs00984239_m1), GAPDH (TaqMan Probe ID: Hs99999905_m1), GUSB (TaqMan Probe ID: Hs99999908_m1), HPRT1 (TaqMan Probe ID: Hs99999909_m1), PGK1 (TaqMan Probe ID: Hs00943178_g1), PPIA (TaqMan Probe ID: Hs99999904_m1), RPLP0 (TaqMan Probe ID: Hs99999902_m1), TBP (TaqMan Probe ID: Hs00920497_m1), and YWHAZ (TaqMan Probe ID: Hs01122451_m1) were purchased (Life Technologies). The RNA copy numbers were normalized to those of internal ACTB.

Results

Signal correction

In DNA microarray analyses, signal intensities may vary significantly between different platforms, even if measuring the same sample, making it difficult to compare cross-platform data sets. Therefore, signal normalization using internal standards was performed to account for this variation. Specifically, correction was performed by dividing the values obtained by subtracting background levels and by the median values of positive controls (Eq. (1)):

$$\frac{[500.1\_P1]_{\text{signal-BG}}}{[\text{ACTB}, \text{B2M}, \text{GAPDH}, \text{GUSB}, \text{HPRT1}, \text{PGK1}, \text{PPIA}, \text{RPLP0}, \text{TBP}, \text{YWHAZ}]_{\text{median}}}. \quad (1)$$

$[\text{geneX}]_{\text{signal-BG}}$ = Background (BG) subtracted signal intensity of gene X probe
$[\text{geneA, B, C,...}]_{\text{median}}$ = median of BG subtracted signal intensities of gene A, B, C,... probe

Background levels were calculated using a standard procedure for each microarray. In this correction procedure, 10 positive controls—ACTB, B2M, GAPDH, GUSB, HPRT1, PGK1, PPIA, RPLP0, TBP, and YWHAZ—were treated as internal standards. Although only a single RNA standard is used in typical chemical analyses, in DNA microarray platforms, where expression of the positive control results in considerably different signal intensities between platforms, multiple positive controls based on a large number of quantification targets are required. Their median value is then used for making corrections.

Confirmation of concentration dependency

First, confirmation was sought as to whether or not each microarray demonstrates concentration dependency. Here, a linear evaluation was conducted on four types of probe for each of 10 types of standard in order to select suitable combinations (1000_1–5, 500_1–5). The concentration of the standards ranged from 10 to 100,000 zmol, and serial dilutions were prepared in 10-fold increments. Measurements were repeated at least two times on the same solution.

Signal intensities obtained by hybridization using each DNA microarray platform were corrected using the median values of the positive controls (Eq. (1)). The relationship between those values and RNA concentration was plotted, linear regression was performed for each of the standards and probes (P1–P4), and their slopes and correlation coefficients were calculated. An example indicating the concentration dependency of standard substance candidate 500_1 is shown in Fig. 3.

Although no correlation was observed between concentration increases in the standards and the corrected signal values at a low amount (10 zmol), by selecting a higher amount (1 nmol) range correlation coefficients of 0.97 and above were obtained, thereby confirming concentration dependency. The observation of a direct association between signal and concentration strongly suggests that our probes are detecting our external RNA standards specifically.

Selection of standard substance candidate–probe combinations

The most linear concentration ranges were selected and linear regression was performed for each standard, each platform, and each probe using the results described in the previous section. Next, the slopes and correlation coefficients were extracted for those ranges. Microarray data were summarized for each probe, and average values for slopes and correlation coefficients were determined as shown in Table 1 and Supplementary Table S4 for all data:

$$\frac{|[\text{P1}]_{\text{Ave}} - [\text{P1, P2, P3, P4}]_{\text{Ave}}|}{\{[\text{P1, P2, P3, P4}]_{75\%} - [\text{P1, P2, P3, P4}]_{25\%}\} * 0.7413}. \quad (2)$$

Next, Z scores were calculated for each standard–probe combination (Eq. (2)), and combinations of standard and probe were selected by extracting and excluding outlier probes (probes having scores $\geqslant 2$). In other words, those probes that demonstrated bias were excluded from the P1 to P4 probes prepared for each standard.

Considering Z score and average and standard deviation of slope in each platform, the selected combinations of standard and probes consisted of the following nine types: 1000_2_P2, 1000_2_P3, 1000_3_P3, 1000_4_P2, 500_1_P2, 500_1_P3, 500_1_P4, 500_2_P4, and 500_4_P2.

Approach used to prepare calibration curves

Calibration curves indicate the relationship between signal and concentration as a result of preparing serially diluted standard solutions and measuring those solutions under the same conditions as samples.

DNA microarray platforms are characterized by being able to acquire a large amount of data using a single microarray. Thus, we devised a method in which mixtures of multiple standards were used and calibration curves were produced using a single DNA microarray by changing the concentration level of each. A graphical representation of this approach is indicated in Fig. 4. In general, although standards are prepared by sequentially diluting one type of standard to produce a single quantification target, in this study five different types of standard (RNA standards E1–E5) were prepared while changing their respective concentration levels. This results in a "relative" quantification of expression level instead of an "absolute" quantification.

Analysis of calibration curves

Combinations of the nine types of standard–probe combinations selected (see "Selection of standard substance candidate–probe combinations" section above) and at five concentration levels (10, 100, 1000, 10,000, and 100,000 zmol) were prepared based on the approach depicted in Fig. 4, and combinations having the lowest levels of error between linearity and microarray were selected.

Corrected values for the nine types of standard–probe pairs were calculated for each of the microarray platforms A to F and plotted. In those plots, a line corresponding to a slope of 1 and a line passing near the center of each concentration were drawn. In cases where all six microarray platforms fell between these two lines, the data for those microarray platforms were used and points falling outside these lines were excluded. A round-robin system was used until a standard probe pair that met the above criteria was found. A total of 186 combinations were evaluated.

Signals were corrected as defined above for each microarray platform A to F, for each combination of standard–probe pair, and for each concentration level, and linear regression was performed. Those combinations that demonstrated an average

$[\text{P1-PL-X}]_{\text{slope}}$: slope of probe P1 in platform X extracted for the most linear concentration ranges
$[\text{P1}]_{\text{Ave}}$: Average slope of $[\text{P1-PL-A}]_{\text{slope}}$, $[\text{P1-PL-B}]_{\text{slope}}$, $[\text{P1-PL-C}]_{\text{slope}}$, $[\text{P1-PL-D}]_{\text{slope}}$, $[\text{P1-PL-E}]_{\text{slope}}$, $[\text{P1-PL-F}]_{\text{slope}}$
[P1, P2, P3, P4]25% : 25% quintile of $[\text{P1}]_{\text{Ave}}$, $[\text{P2}]_{\text{Ave}}$, $[\text{P3}]_{\text{Ave}}$ and $[\text{P4}]_{\text{Ave}}$
[P1, P2, P3, P4]75% : 75% quintile of $[\text{P1}]_{\text{Ave}}$, $[\text{P2}]_{\text{Ave}}$, $[\text{P3}]_{\text{Ave}}$ and $[\text{P4}]_{\text{Ave}}$

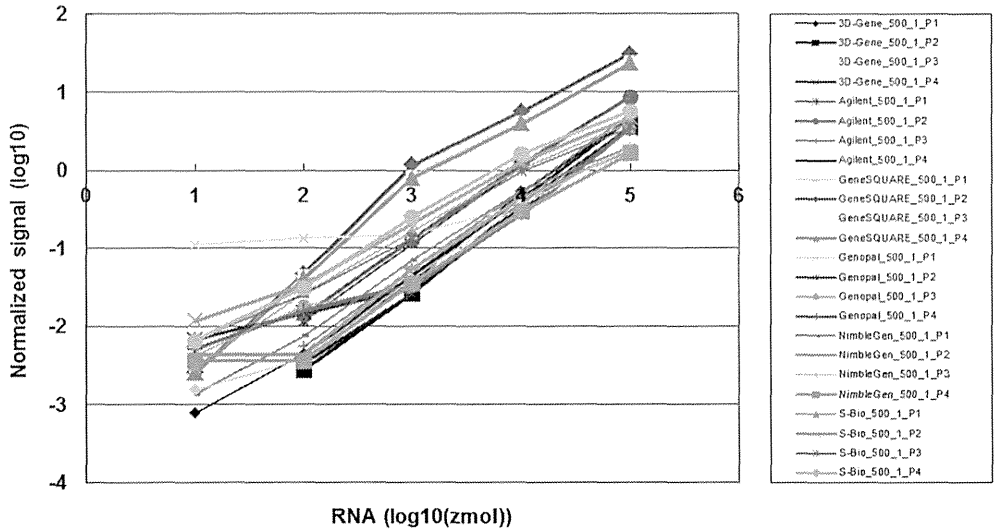

Fig.3. Procedures for selecting standard substance candidate–probe combinations using *Z* scores. The relationship between normalized signal intensities and RNA concentration of standard substance candidate 500_1 was plotted. Linear regression was performed for each of the standards and probes (P1–P4), and their slopes and correlation coefficients were calculated.

Table 1
Averages of slopes and correlation coefficients of six microarray platforms for P1 of standard substance candidate 500_1.

RNA	Probe name	Platform	RNA (log10 (zmol))		Slope	r^2
			Max	Min		
500_1	500_1_P1	3D-Gene	5	2	1.020	0.998
		Agilent	5	2	0.965	0.999
		GeneSQUARE	4	1	1.032	0.996
		Genopal	5	3	0.769	0.972
		NimbleGen	4	1	0.884	0.993
		S-Bio	5	1	0.777	0.994
		Average			0.908	0.992
		SD			0.117	0.010
		Z score			1.429	1.587
	500_1_P2	3D-Gene	5	2	1.047	0.999
		Agilent	5	2	0.935	0.995
		GeneSQUARE	4	1	1.071	0.998
		Genopal	5	3	0.993	0.999
		NimbleGen	5	2	0.881	0.915
		S-Bio	5	2	0.722	0.996
		Average			0.941	0.984
		SD			0.128	0.034
		Z score			2.137	1.189
	500_1_P3	3D-Gene	5	2	1.026	0.999
		Agilent	5	2	0.993	0.992
		GeneSQUARE	5	2	0.883	0.999
		Genopal	5	3	0.977	1.000
		NimbleGen	5	2	0.908	0.929
		S-Bio	5	2	0.723	0.997
		Average			0.918	0.986
		SD			0.110	0.028
		Z score			0.305	0.437
	500_1_P4	3D-Gene	5	2	1.023	0.998
		Agilent	5	2	1.002	1.000
		GeneSQUARE	5	2	0.888	0.997
		Genopal	5	3	0.981	0.998
		NimbleGen	5	2	0.889	0.942
		S-Bio	5	1	0.763	0.997
		Average			0.924	0.989
		SD			0.098	0.023
		Z score			0.305	0.437

Note: SD, standard deviation.

slope of microarray platforms A to F of 0.8 or higher, an average correlation coefficient of 0.95 or higher, and variability in the correlation coefficients of microarray platforms A to F of 5% or less were selected.

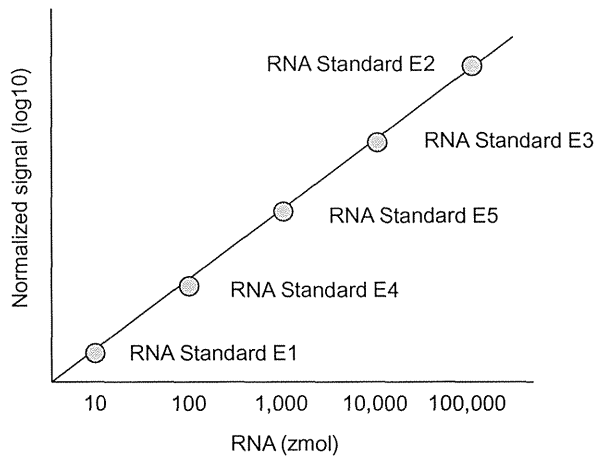

Fig.4. Approach used to prepare calibration curves: Evaluation of assay linearity at input levels for spike-in RNAs to the microarray platform. Linearity was assessed based on a five-point dilution series, ranging four orders of magnitude, of the spiked external RNA E1 to RNA E5. Each dilution series was measured, and the linearity of each of the 186 combinations was estimated by calculating the A round-robin system. The standard probe combination was determined to the extent that linearity was obtained.

Next, the variability of the signal correction values at each concentration was determined, and those combinations for which the variability between microarray platforms was 25% or less at three or more of the five concentration levels were selected (nos. 16, 33, and 183). The selected combinations are shown in Supplementary Table S5. A reproducibility experiment was conducted to confirm whether or not linearity is reproducible (Supplementary Table S6).

Best combination of probe sets for standard RNAs

Next, the standard combination numbers 16, 33, and 183 were added to HURR and HBRR, and microarray expression analyses were conducted using each platform. Standard combination number 183 demonstrated the best concentration linearity and reproducibility among all six platforms (Fig. 5). This consisted of the standard–probe combinations of 500_2_P4, 500_4_P2,

1000_3_P3, 500_1_P3, and 1000_4_P2. Thus, the use of this combination of standard substances and probes can ensure data compatibility between DNA microarray platforms. Although the dynamic range of compatibility varies according to each platform, it is thought to be at least three orders of magnitude.

Digital expression profiling (RNA-Seq)

Next, an RNA-Seq expression analysis was conducted using the model GAII manufactured by Illumina on the standard RNA combination number 183. The results were then compared with expression data obtained with DNA microarray in order to investigate the correlation with next-generation sequencers. In addition, RT–PCR was performed by using TaqMan primer probes for standard combination number 183. The results demonstrate a close correlation with respect to dynamic range, linearity, and the like (Fig. 6). We suggest that these results indicate that the use of standards can be expanded not only to include compatibility between DNA microarray platforms but also to evaluate the data obtained using next-generation sequencers.

Results of quantitative PCR comparisons

Quantitative PCR was conducted on four RNA external standards (500_2, 500_4, 1000_3, and 500_1) and 10 types of genes (ACTB, B2M, GAPDH, GUSB, HPRT1, PGK1, PPIA, RPLP0, TBP, and YWHAZ) for each of the HURR and HBRR samples to which a standard combination number 183 was added. Expression ratios for each sample were calculated using the ddCt method. When these results were compared with the expression ratio data obtained by three DNA microarray platforms and next-generation sequencer, significant correlations were observed for each DNA microarray platform and next-generation sequencer (Fig. 7). We conclude that we are able to obtain equivalent data across microarray platforms for both standard and real-time applications provided that the expression level is above a certain threshold level.

Discussion

Attempts to compare DNA microarray data across platforms have been made in the past, and data compatibility has been maintained by comparing expression fluctuation ratios. Signal intensity is not generally considered to be comparable directly across platforms due to differences in various factors such as the target preparation method, probe sequence, and detection method.

In this study, we were able to select standard–probe combinations that can be used across multiple platforms by developing methods for optimizing both the standard and the detection probe. When detection signal intensities from each platform were corrected using our RNA standards, the correlation with quantitative RT–PCR, considered to be the "gold standard," was shown to improve. We suggest that a calibration method based on our standards is effective and can contribute to improvements in data reliability. Moreover, due to the high level of correlation in signal

#183/HURR

probe	RNA (LOG10)	Normalized Signal (Log10)						max	min	ave	σ
		3D-Gene	Agilent	GAII	GeneSQUARE	NimbleGen	S-Bio				
500_2_4	5	0.628	0.714	0.664	0.308	0.182	0.841	0.841	0.182	0.556	0.255
500_4_2	4	-0.368	-0.036	-0.437	-0.684	-0.390	0.178	0.178	-0.684	-0.289	0.309
1000_3_3	3	-1.204	-1.684	-1.333	-1.567	-0.927	-0.711	-0.711	-1.684	-1.238	0.372
500_1_3	2	-2.347	-2.024	-2.410	-2.785	-2.275	-1.607	-1.607	-2.785	-2.241	0.397
1000_4_2	1	-2.903	-2.751	-3.586	-3.433	-2.884	-2.209	-2.209	-3.586	-2.961	0.497

#183/HBRR

probe	RNA (LOG10)	Normalized Signal (Log10)						max	min	ave	σ
		3D-Gene	Agilent	GAII	GeneSQUARE	NimbleGen	S-Bio				
500_2_4	5	0.931	1.191	0.562	0.782	0.471	1.058	1.191	0.471	0.832	0.281
500_4_2	4	0.019	0.466	-0.504	-0.251	-0.071	0.356	0.466	-0.504	0.002	0.365
1000_3_3	3	-0.828	-1.126	-1.366	-1.157	-0.637	-0.485	-0.485	-1.366	-0.933	0.339
500_1_3	2	-2.003	-1.532	-2.359	-2.572	-1.868	-1.370	-1.370	-2.572	-1.951	0.464
1000_4_2	1	-2.593	-2.268	-3.262	-3.047	-2.537	-2.392	-2.268	-3.262	-2.683	0.388

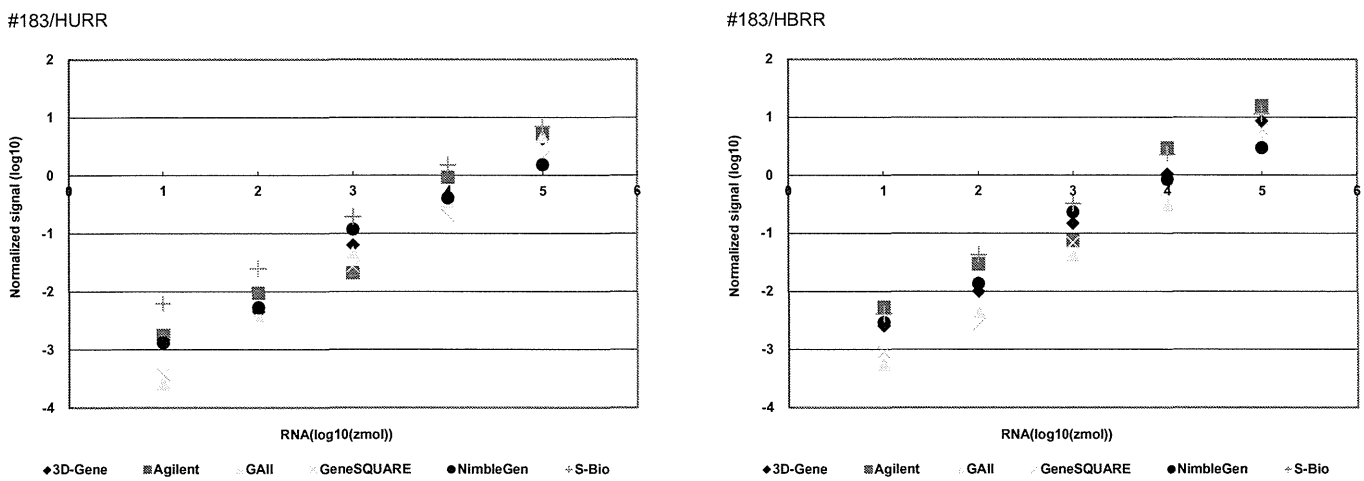

Fig.5. Linearity and reproducibility among all six microarray platforms of standard combination number 183. Among the three combinations, number 183 demonstrated the best concentration linearity and reproducibility (lowest standard deviation value) in all six platforms.

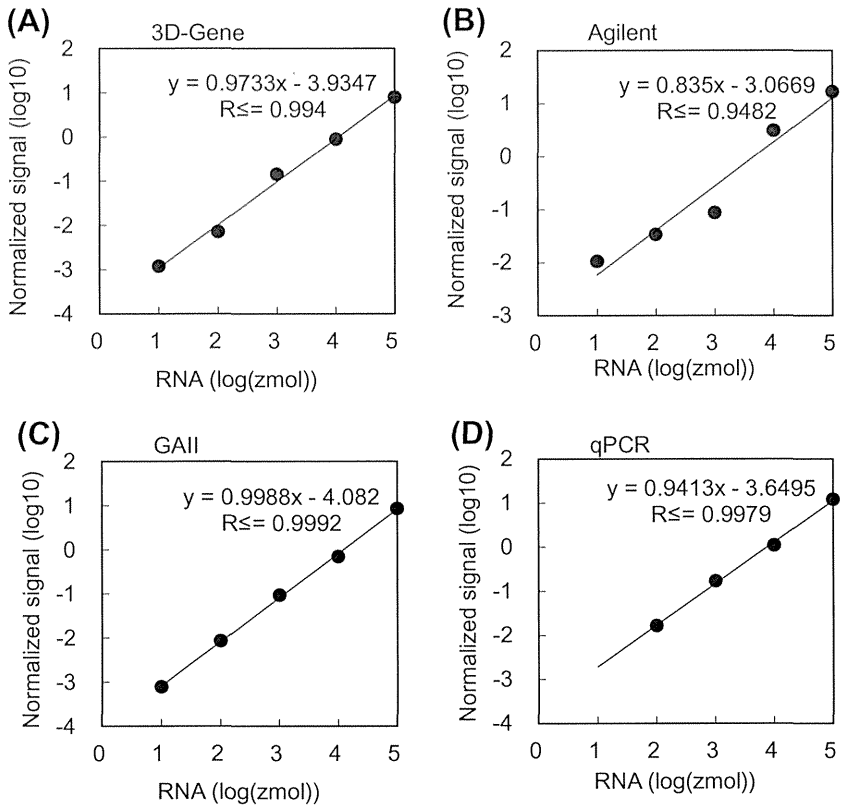

Fig.6. Comparison of quantitative RT–PCR data derived from microarray platforms and a next-generation sequencer: Assessment of the assay linearity between and microarray platforms and a next-generation sequencer. Using the standard RNA combination number 183 contents enabled us to obtain the high correlation with respect to dynamic range, linearity between microarray platforms, and a next-generation sequencer. qPCR, quantitative RT–PCR.

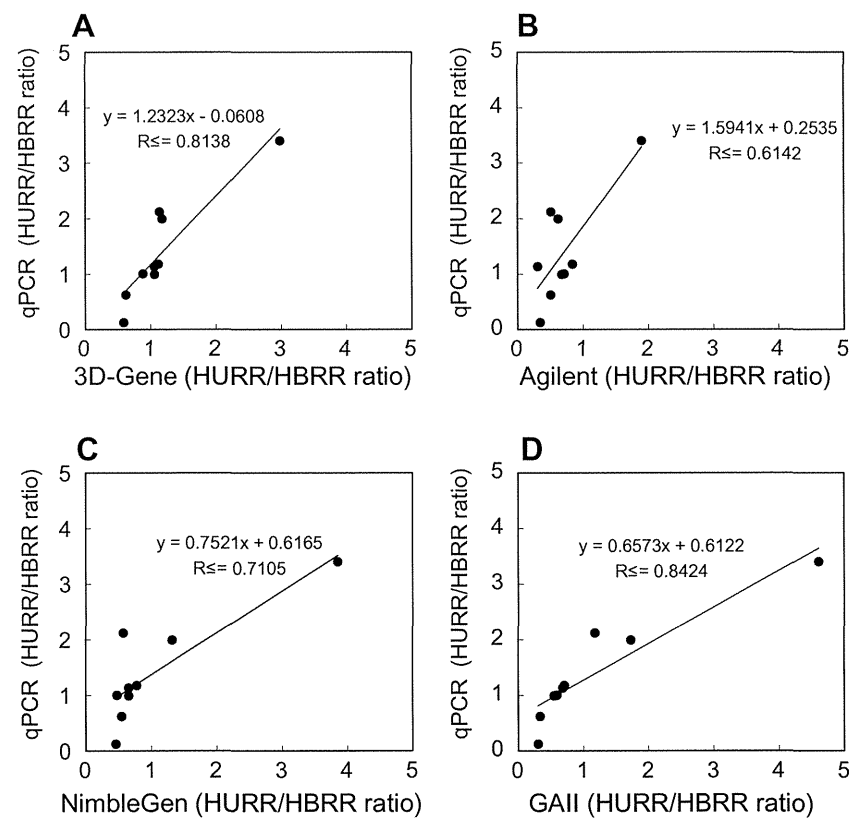

Fig.7. Comparison of fold change ratio (HURR/HBRR) among quantitative RT–PCR, microarray platforms, and next-generation sequencer. (A–C) Comparison of fold change ratio (HURR/HBRR) between microarrays and quantitative RT–PCR. (D) Comparison of fold change ratio (HURR/HBRR) between next-generation sequencer and quantitative RT–PCR. The dot plot indicates the fold change ratio (HURR/HBRR). Nine genes indicated that there is an expression difference. qPCR, quantitative RT–PCR.

intensity, the use of this calibration method makes it possible to directly compare detection across platforms.

Because this calibration method was also effective with next-generation sequencers employing different detection principles, we further suggest that our calibration method can be applied to various gene expression analysis techniques. In addition, we suggest that the standard selection method and calibration method we have developed is effective for detection methods other than DNA microarray platforms.

The Affymetrix DNA microarray was not assessed in this study because it is not possible to produce a custom microarray with this platform. The current pace of progress in the field of genetic diagnostics has resulted in DNA microarray platforms being increasingly used in patient management such as MammaPrint in breast cancer prognosis prediction. The calibration method we have developed enables evaluation of intra-run, inter-run, and cross-platform DNA microarray detection data, thereby making it possible to improve and maintain reliability [20–22].

In conclusion, we suggest that our set of validated nucleotide standards will enable direct comparison of data produced using multiple DNA microarray platforms provided with identical clinical samples, thereby ensuring the compatibility of detection results, inter-laboratory communication, and diagnoses. Moreover, because external RNA standards enable identification of failed steps during the assay process, it is possible to improve reliability and ensure compatibility between data sets, suggesting that similar results can be obtained in clinical diagnostic testing independent of the specific platform used.

Acknowledgments

We thank Ayumi Hasegawa (Sumika Chemical Analysis Service) for the design of data analysis and Tatsunobu Fukushima (Mitsubishi Chemical Holdings) for helpful discussion during this work. We also acknowledge support from a New Energy and Industrial Technology Development Organization (NEDO) grant (NEDO 10001779-0) funded by the Ministry of Economy, Trade and Industry – Japan. We also thank Dr. Keith Hart, Cardiff University and members of Japan MicroArray Consortium (JMAC) for helpful discussion in this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ab.2014.11.012.

References

[1] J. de Leon, M.T. Susce, E. Murray-Carmichael, The AmpliChip CYP450 Genotyping Test: integrating a new clinical tool, Mol. Diagn. Ther. 10 (2006) 135–151.

[2] T. Satoh, K. Matsumoto, T. Fujii, O. Sato, N. Gemma, M. Onuki, H. Saito, D. Aoki, Y. Hirai, H. Yoshikawa, Rapid genotyping of carcinogenic human papillomavirus by loop-mediated isothermal amplification using a new automated DNA test (CliniChip HPV), J. Virol. Methods 188 (2013) 83–93.

[3] C.I. Dumur, C.E. Fuller, T.L. Blevins, J.C. Schaum, D.S. Wilkinson, C.T. Garrett, C.N. Powers, Clinical verification of the performance of the pathwork tissue of origin test: utility and limitations, Am. J. Clin. Pathol. 136 (2011) 924–933.

[4] Z. Wang, A.P. Malanoski, B. Lin, C. Kidd, N.C. Long, K.M. Blaney, D.C. Thach, C. Tibbetts, D.A. Stenger, Resequencing microarray probe design for typing genetically diverse viruses: human rhinoviruses and enteroviruses, BMC Genomics 9 (2008) 577.

[5] M.J. van de Vijver, Y.D. He, L.J. van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde, H. Bartelink, S. Rodenhuis, E.T. Rutgers, S.H. Friend, R. Bernards, A gene-expression signature as a predictor of survival in breast cancer, N. Engl. J. Med. 347 (2002) 1999–2009.

[6] S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F.L. Baehner, M.G. Walker, D. Watson, T. Park, W. Hiller, E.R. Fisher, D.L. Wickerham, J. Bryant, N. Wolmark, A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer, N. Engl. J. Med. 351 (2004) 2817–2826.

[7] R. Salazar, P. Roepman, G. Capella, V. Moreno, I. Simon, C. Dreezen, A. Lopez-Doriga, C. Santos, C. Marijnen, J. Westerga, S. Bruin, D. Kerr, P. Kuppen, C. van de Velde, H. Morreau, L. van Velthuysen, A.M. Glas, L.J. van't Veer, R. Tollenaar, Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer, J. Clin. Oncol. 29 (2011) 17–24.

[8] J. Ambroise, B. Bearzatto, A. Robert, B. Govaerts, B. Macq, J.L. Gala, Impact of the spotted microarray preprocessing method on fold-change compression and variance stability, BMC Bioinformatics 12 (2011) 413.

[9] S. Mao, C. Wang, G. Dong, Evaluation of inter-laboratory and cross-platform concordance of DNA microarrays through discriminating genes and classifier transferability, J. Bioinf. Comput. Biol. 7 (2009) 157–173.

[10] S. Schwers, E. Reifenberger, M. Gehrmann, A. Izmailov, K. Bohmann, A high-sensitivity, medium-density, and target amplification-free planar waveguide microarray system for gene expression analysis of formalin-fixed and paraffin-embedded tissue, Clin. Chem. 55 (2009) 1995–2003.

[11] W.R. Pearson, Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms, Genomics 11 (1991) 635–650.

[12] N. Markham, M. Zuker, UNAFold: Software for nucleic acid folding and hybridization, Methods Mol. Biol. 453 (2008) 3–31.

[13] T. Yamada, H. Soma, S. Morishita, PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome, Nucleic Acids Res. 34 (2006) W665–W669.

[14] J. SantaLucia Jr., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, Proc. Natl. Acad. Sci. USA 95 (1998) 1460–1465.

[15] K. Nagino, O. Nomura, Y. Takii, A. Myomoto, M. Ichikawa, F. Nakamura, M. Higasa, H. Akiyama, H. Nobumasa, S. Shiojima, G. Tsujimoto, Ultrasensitive DNA chip: gene expression profile analysis without RNA amplification, J. Biochem. 139 (2006) 697–703.

[16] D. Okuzaki, T. Fukushima, T. Tougan, T. Ishii, S. Kobayashi, K. Yoshizaki, T. Akita, H. Nojima, "Genopal: a novel hollow fiber array for focused microarray analysis, DNA Res. 8 (2010) 1–11.

[17] T. Yamada, S. Morishita, Computing highly specific and noise tolerant oligomers efficiently, J. Bioinf. Comput. Biol. 2 (2004) 21–46.

[18] Y. Komatsu, N. Kojima, M. Sugino, A. Mikami, K. Nonaka, Y. Fujinawa, T. Sugimoto, K. Sato, K. Matsubara, E. Ohtsuka, Novel amino linkers enabling efficient labeling and convenient purification of amino-modified oligonucleotides, Bioorg. Med. Chem. 16 (2008) 941–949.

[19] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.

[20] MAQC Consortium, The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements, Nat. Biotechnol. 24 (2006) 1151–1161.

[21] MAQC Consortium, The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models, Nat. Biotechnol. 28 (2010) 827–838.

[22] T.A. Patterson, E.K. Lobenhofer, S.B. Fulmer-Smentek, P.J. Collins, T.-M. Chu, W. Bao, H. Fang, E.S. Kawasaki, J. Hager, I.R. Tikhonova, S.J. Walker, L. Zhang, P. Hurban, F. de Longueville, J.C. Fuscoe, W. Tong, L. Shi, R.D. Wolfinger, Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project, Nat. Biotechnol. 24 (2006) 1140–1150.