AFM detection enables multiplexed low-cycle-number quantitative PCR for biomarker assays

Andrey Mikheikin[1], Anita Olsen[1], Kevin Leslie[1], Bud Mishra[2], James K. Gimzewski[3,4], Jason Reed[1,5*]

[1] Department of Physics, Virginia Commonwealth University, Richmond, Virginia, USA
[2] Departments of Computer Science and Mathematics, Courant Institute of Mathematical Sciences, New York University, USA
[3] Department of Chemistry and Biochemistry, UCLA, Los Angeles, California, USA
[4] California NanoSystems Institute (CNSI) at UCLA, Los Angeles, California, USA
[5] VCU Massey Cancer Center, Richmond, Virginia, USA

*Correspondence should be addressed to jcreed@vcu.edu

Supplementary Information

This document contains the following supplementary information:

- Details of the 10-plex PCR primer set (**Table S1 and Fig. S1**)
- Data from the AFM, qPCR and Bioanalyzer experiments in tabular form (**Tables S2-S5**)
- Data from technical repeat experiments (**Table S6**)
- Supplementary Materials and Methods

Gene Name	Accesion Number	10-plex PCR				Single-plex qPCR
		[Primer], nM (each)	Amplicon size, bp (nm)	Direction	Primer sequences	Primer sequences
ABCA2	NM_001606.4	20	710 (234)	Forward Reverse	CTTCGAGGAGGAGCGGGCCCAGCTGTCCT GCTATGTACAGGCCCGGCTCCC	GATGGTGGGCTTCTTCTGTC AGCTGCTGCTCTGGAAGAAC
CAD	NM_004341.3	20	521 (172)	Forward Reverse	CCGCGGCACCAAGCAGGAGGA GGAATGAAGATGGGGCAGACTGGGC	AGCCCACACAGTTCTCATCC GCGAGTCAGCACAAAGATTG
CDK4	NM_000075.3	20	1469 (485)	Forward Reverse	TTGTCCGGCTGATGGACGTCTGTGC TTTTGACACAGAGTCTTGCTCTGTTGCCCAG GC	GTCGGCTTCAGAGTTTCCAC CCGAAGTTCTTCTGCAGTCC
CDK9	NM_001261.3	50	1250 (412)	Forward Reverse	GAAGGAGGGGTTCCCCATTACAGCC CACCACGCCGGGAGCTCTTAG	CAAGTTGACCACATTCTCGTG AAGGTGCTGATGGAAAACGA
CRLF	NM_004750.4	50	266 (88)	Forward Reverse	GCTCCAAGAAAGCCGGGATCTGG GATCCCCTCGTCCTGGTTGCG	TTCTTGAGCCAGCCCAG CACAGCCGCCTCCACTC
JAK3	NM_000215.3	20	2040 (673)	Forward Reverse	TGCCTGCCCTGCTGAGGTTCACG TGGACACAGACACTTCCCCAGCCCC	CAGCCTCCGTGGACAAGAG CTTCGAAAGTCCAGGGTC C
MMP2	NM_004530.4	20	2538 (837)	Forward Reverse	TGGGAGAAGGCCAAGTGGTCCGTG TGGAAAGGTTAAAATAGGTGACACGTGAAA AGTGCC	AAGAAGTAGCTGTGACCGCC TTGCTGGAGACAAATTCTGG
MYC	NM_002467.4	200	1064 (351)	Forward Reverse	CCACCACCAGCAGCGACTCTGAGGA TTTTTTAAGATTTGGCTCAATGATATATTTGC CAGT	CACCGAGTCGTAGTCGAGGT TTTCGGGTAGTGGAAAACCA
POLR2A	NM_000937.4	20	1629 (537)	Forward Reverse	CTGGTCCCCCAGTGTTGGGAGTGG ACCCTTTACCCGCCCCGCAT	CAGTCCGCTCAATCACCC TCCAGTTCGGAGTCCTGAGT
ZNF350	NM_021632.3	20	1811 (598)	Forward Reverse	CCTGGTGGCAGTGGGGTATCAAGCC AGCCCTTTCCCACCAACTGCCCC	AACCCCAAATCCACTTCCTC TAACGTTGAGGCCCTTCTTG

Table S1. PCR primers used in the AFM and qPCR assays.

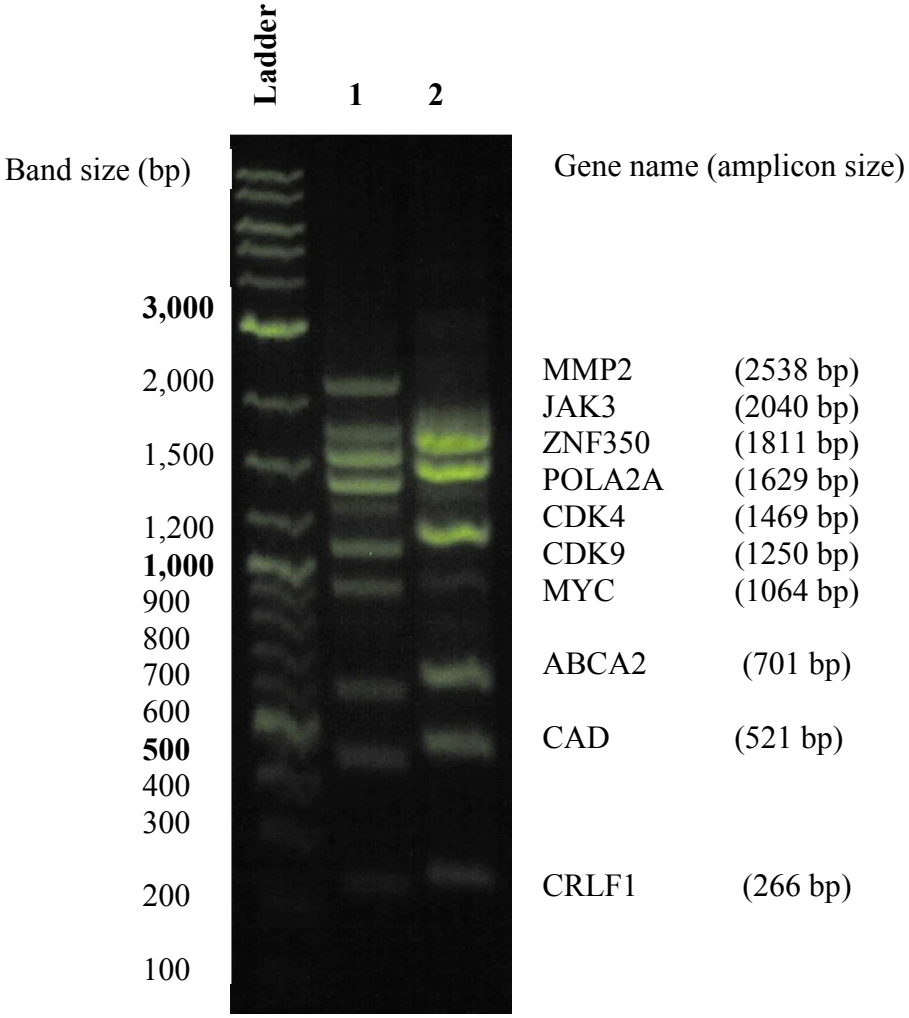

Figure S1 - Image of 10plex PCR products (30 amplification cycles) run on 1% agarose gel. cDNA obtained from Universal Total Human Reference RNA (Lane **1**) or FirstChoice Human Brain Reference RNA (Lane **2**) were used as templates. **Ladder** – 2-Log DNA Ladder (New England Biolabs). Band sizes and genes names (along with corresponding amplicon sizes) are shown on the left and right sides of the gel image, respectively.

Gene name	Ct values				Brain/Human Ratio	
	Universal Human		FirstChoice Brain			
	55C	60C	55C	60C	55C	60C
ABCA2	29.0	29.2	26.5	27.4	5.76	3.43
CAD	23.3	23.2	27.2	28.0	0.07	0.04
CDK4	19.2	19.3	23.2	24.1	0.06	0.04
CDK9	23.0	22.9	23.3	23.3	0.83	0.37
CRLF	33.1	31.3	35.1	32.3	0.25	0.49
JAK3	29.0	29.0	31.3	31.7	0.19	0.15
MMP2	19.8	18.8	27.8	29.3	<0.01	<0.01
MYC	21.8	21.4	29.0	30.0	0.01	<0.01
POLR2A	31.4	23.4	33.0	25.5	0.33	0.34
ZNF350	25.7	25.6	26.3	26.5	0.69	0.51

Table S2. qPCR Ct values at primer annealing temperatures 55°C and 60°C. Data represent average of two replicates for each sample. For replicates, the variance in Ct values was within the typical range for qPCR.

Gene name	Raw molecule counts		Brain/Human Ratio*
	Universal Human	FirstChoice Brain	
ABCA2	96	457	3.42
CAD	387	56	0.10
CDK4	42	10	0.17
CDK9	108	65	0.43
CRLF	40	10	0.27
JAK3	11	7	0.46
MMP2	12	2	0.12
MYC	352	14	0.03
POLR2A	105	36	0.25
ZNF350	14	13	0.67

Table S3. AFM-determined molecule counts. Sample concentration 10 pg/uL. *Ratio of molecule counts corrected for total number of images analyzed per sample (HR 3 mica surfaces, 9 images; HB 5 mica surfaces, 12.5 images).

Gene name	Bioanalyzer (pg / uL)		Brain/Human Ratio
	Universal Human	FirstChoice Brain	
ABCA2	43.2	299.6	6.93
CAD	132.9	20.9	0.16
CDK4	nd	nd	--
CDK9	70.4	64.9	0.92
CRLF	8.5	2.8	***0.33***
JAK3	nd	nd	--
MMP2	37.2	nd	--
MYC	169.0	12.7	0.08
POLR2A	87.2	53.0	0.61
ZNF350	41.5	32.3	0.78

Table S4. Bioanalyzer high sensitivity DNA kit. Sample concentrations 800 pg/uL. Data represent average of two replicates for each sample. Per sample concentration and sizing repeatability was within manufacturer specified values.

	qPCR	AFM	Bioanalyzer
qPCR	1	0.993	0.998
AFM		1	0.994
Bioanalyzer			1

Table S5. Correlations between relative expression data for qPCR, AFM and Bioanalyzer.

Gene name	Raw molecule counts			Mean	σ	CV	σ_{min}	CV_{min}
	Universal Human 1	Universal Human 2	Universal Human 3					
ABCA2	56	56	85	66	14	0.22	8	0.12
CAD	264	292	374	311	42	0.14	14	0.04
CDK4	30	38	10	27	15	0.56	5	0.19
CDK9	38	62	35	45	12	0.26	7	0.14
CRLF	24	36	48	36	8	0.24	6	0.16
JAK3	9	15	3	9	6	0.62	3	0.33
MMP2	12	8	6	9	5	0.51	3	0.33
MYC	171	307	243	237	38	0.16	13	0.05
POLR2A	45	44	35	42	11	0.26	6	0.15
ZNF350	12	16	11	13	2	0.19	4	0.27

Table S6. Technical repeats of AFM-determined molecule counts. Mean and standard deviations calculated after normalizing raw molecule counts such that the total number of corrected counts for each of the three samples are equal (normalizations: UH1 1.2, UH2 0.91, and UH3 0.94). Estimates of minimum variation due to statistical counting effects is calculated as the expected standard deviation of a binomial distribution: sqrt[counts * (1-p)] where p is sample fraction, given by counts per species/total molecules.

Supplementary Materials and Methods

Chemicals**.** Primers for multiplex PCR were design using Primer Blast software[1]. Sequences for qPCR primers were taken from qPrimerDepot[2]. For primer sequences see **Table S1**. All primers were ordered from IDT DNA Technology (Corallville, IA) and used without further purification. Universal Total Human Reference RNA (Stratagene, Agilent Technologies, Santa Clara, CA) solution was precipitated and re-suspended in RNAse-free water according to manufacturer's protocol, FirstChoice Human Brain Reference RNA (Life Technolgies, Carlsbad, CA) was used without purification. All RNA were aliquoted and stored at -80C. New England Biolab's AMV LongAmp® Taq RT-PCR Kit was used for reverse transcription and multiplex PCR amplification with LongAmp® *Taq* 2X Master Mix included in the kit replaced by LongAmp Hot Start *Taq* 2X Master Mix. 1-(3-aminopropyl)silatrane (APS) was a kind gift from Dr. Sergey Chassovskikh (Georgetown University, Washington, DC).

Multiplex RT-PCR. Reverse transcription was conducted according to the manufacturer's protocol with 2 ug amount of total RNA per reaction tube. $(dT)_{23}VN$ primer included in kit was used for reverse transcription. Multiplex PCR was conducted according to the manufacturer's recommendation except the extension time was increased to 10 min (for primer concentration see **Table S1**). Primer annealing

temperature was 60C. For Bioanalyzer and AFM analysis 20 PCR tubes with the amount of cDNA corresponding to 50 ng of total RNA per tube were used for 15 cycles of PCR, combined after the reaction in one tube, purified twice with ZYMO Research DCC-5 columns and quantified with Qubit fluorimetric assay. Typical yield was 800-1500 pg/micol in 20-50 microl elution volume. To check the quality of PCR products with 1% agarose gel, the amount of cDNA corresponding to 50 ng of total RNA was amplified (30 cycles) in one tube (**Fig. S1**).

Multiplex primer design and optimization of experimental conditions. Primers were designed using Primer-BLAST software with some constrains the same as for standard RT-qPCR: to avoid amplification of genomic DNA primers were designed either to span exon-exon junction or to be separated by at least one intron; primers binding sites do not contain known single nucleotide polymorphism; all primer pairs have been checked against human genome and trancriptome databases for non-specific binding. There are also some specific requirements for amplicons to be analyzed by AFM imaging: the amplicons should be longer than 200 bp to easily visualize them and have at least 100 bp difference in size to distinguish them from each other. This means that for tenplex, some amplicons should be 1000-2000 bp in length, which is significantly longer that those used in typical qPCR assays (100-150 bp). We analyzed the longest amplicon and found that increasing of extension time from 2.5 minutes as recommended by the kit's manufacturer for the amplcon of this size up to 10 minutes significantly improved the PCR yield of this amplicon (data not shown). In order to minimize primer-primer interactions and nonspecific primer binding to template the primer concentrations should be as low as possible. To optimize primer concentrations 30 amplification cycles of singleplexes were conducted for all primer pairs at primer concentration 20 nM and the amplicons were visualized in 1% agarose gel. It was found that most of primers work at this concentration. Other primers were taken at various concentrations up to 200 nM to identify the minimal working concentration. Primer sequences, concentrations and amplicon sizes are summarized in **Table S1**. All primer pairs were used in multiplex at concentrations determined in singleplexes without further optimization.

RT-qPCR. The same transcription products were used as templates for qPCR. ABI Prism 7900 Detection System (Applied Bioscience) was employed according to the manufacturer's protocol. SYBRGreen chemistry was used to detect amplicons. qPCR was conducted at two annealing temperatures – 55C and 60C, and good correspondence was found between these two datasets in most cases (see **Table S2**). For POLR2A gene, Ct values at 55C are shifted towards higher values probably due to primer dimer formation and/or secondary structures at this lower temperature; however, the brain/human reference ratio values are the same in both these cases.

Bioanalyzer assay. 1 microL of 800 pg/microL of each amplicon solution was used to run Agilent 2100 Bioanalyzer High Sensitivity DNA assay (Agilent Technologies) according to the manufacturer's protocol.

AFM sample preparation and imaging. Freshly cleaved mica was treated with 1-(3-aminopropyl)silatrane (APS) according to Shlyakhtenko et al.[3], then rinsed with 5 microL of water. One microL of 10 pg/microL of amplicons was deposited on derivatized mica, incubated for 20 min at room

temperature in a humidified environment, then rinsed with 3 ml of DI water. Images were acquired using a Bruker Dimension Icon AFM in 'soft' taping mode (RFESP cantilevers, 1-3 N/m spring constant) at a lateral speed of 0.02 mm/sec, and resolution of 2x2 nm per pixel. The double-stranded amplicon DNA can be distinguished from single stranded DNA and RNA by the shape and height (apparent height in AFM image 0.6 nm, apparent thickness 8-10 nm). A non-primer control was conducted at the standard experimental conditions and the abundance of DNA molecules is <1 per 10x10 micron AFM image (data not shown).

AFM data processing. Images were analyzed by an image processing program developed for this application, called AFMExplorer. Details of the image analysis procedure are given in Sundstrom et al.[4] Briefly, AFM images are flattened and pre-filtered to reduce noise, followed by adaptive thresholding based on pixel height to recognize regions corresponding to DNA molecules; a binary skeletonization procedure is used to determine the best backbone contour for each molecule, and the molecule length in nanometers is calculated by a cubic spline fit to the backbone pixel set. The program is manually queued to ignore crossed molecules, otherwise the molecule identification and measurement is fully automatic. The relative dispersion (CV) in measured contour lengths for a population of like DNA molecules is better than +/- 3% under these imaging conditions [5].

For the Bioanalyzer data, chromatogram peaks are associated with corresponding mRNA species by standard calibration; for AFM by expected amplicon length (**Table S1**; AFM, 0.33 bp/nm [pitch of bDNA]). In the AFM experiments, counts for each target represent the sum of molecules detected with lengths equal to the expected amplicon length, +/- 3%, which represents a conservative estimate of AFM sizing error. [5] Relative abundances are calculated as the ratio (cnts brain)/ (cnts hum ref) for each species. The qPCR abundances were calculated from the difference in threshold amplification cycle number (Ct value) between the samples. The number of amplicon molecules measured by AFM from the Human Reference and Brain total RNA samples is 2,535 (**4.16 zeptomole**) and 1,533 (**2.55 zeptomole),** respectively.

Technical repeat experiments. Products of three individual amplification reactions were quantified with AFM to study reaction-to-reaction variability of the developed technique. Universal Human Reference cDNA was taken in the amount, corresponding to 50 ng of total RNA per PCR tube. PCR was conducted using a protocol described in the manuscript. The products from each PCR tube were purified with MinElute Qiagen PCR CleanUp kit, dialyzed against deposition buffer and analyzed with AFM as described above. Note this scheme of this replicate experiment is different from one used to compare AFM-PCR to Bioanalyzer analysis and qPCR; in the latter case twenty PCR tubes were combined for comparison of AFM and Bioanalyzer quantification (necessary due to sensitivity limits of Bioanalyzer), and the PCR-induced variability is thus greatly reduced. Good reaction-to-reaction reproducibility allows for quantify gene expression without repeats in contrast to conventional qPCR where the results is usually obtained at least in triplicate or quadruplicate. This property of the PCR-AFM would be important in cases where the initial amount of nucleic acids is limited.

Supplementary References

(1) Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L. *BMC Bioinformatics* **2012**, *13*, 134.

(2) Cui, W.; Taub, D. D.; Gardner, K. *Nucleic Acids Res* **2007**, *35*, D805-809.

(3) Shlyakhtenko, L. S.; Gall, A. A.; Filonov, A.; Cerovac, Z.; Lushnikov, A.; Lyubchenko, Y. L. *Ultramicroscopy* **2003**, *97*, 279-287.

(4) Sundstrom, A.; Cirrone, S.; Paxia, S.; Hsueh, C.; Kjolby, R.; Gimzewski, J. K.; Reed, J.; Mishra, B. *IEEE Trans Inf Technol Biomed* **2012**, *16*, 1200-1207.

(5) Reed, J.; Hsueh, C.; Lam, M. L.; Kjolby, R.; Sundstrom, A.; Mishra, B.; Gimzewski, J. K. *J R Soc Interface* **2012**, *9*, 2341-2350.