RNA & DISEASE 2017; 4: e1183. doi: 10.14800/rd.1183; © 2017 by Sohila Zadran, *et al.*

http://www.smartscitech.com/index.php/rd

ARTICLE

Microfluidic chip with molecular beacons detects miRNAs in Human CSF to reliably characterize CNS-specific disorders

Sohila Zadran[1,2], Francoise Remacle[3], Raphael D. Levine[4]

[1]*Institute of Molecular Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA*

[2]*Department of Neurology, University of California, San Francisco San Francisco, CA, 94158, USA*
[3]*Department of Chemistry, B6c, University of Liege, B4000 Liege, Belgium*
[4]*Crump Institute for Molecular Imaging and Department of Molecular and Medical Pharmacology, David Geffen School of Medicine and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA*

Correspondence: Raphael D. Levine
E-mail: rafi@chem.ucla.edu
Received: January 10, 2016
Published: February 17, 2017

RNA profiling in biofluids holds promise as both diagnostic and prognostic markers. High expression levels of distinctive cell free circulating miRNAs in serum, plasma and cerebral spinal fluid (CSF), have been utilized as classifiers to detect and characterize disorders of the central nervous system (CNS). We formulated the quantitative theory showing how the results of surprisal analysis enable a reliable inference if tumor cells are present in the sample from a single measurement. Subsequently, we develop a molecular beacon-based microfluidic chip that enables for fluorescence detection of miRNAs without amplification in low volumes of human CSF. Using surprisal analysis, we identified a miRNA classifier that enables high fidelity detection and characterization of human brain tumors. We anticipate that this micro-fluidic platform will provide a critical translational tool with point of care potential for CNS disorders.

Keywords: FRET based chip; molecular beacons; microRNAs; CNS disorders; surprisal analysis

To cite this article: Sohila Zadran, *et al*. Microfluidic chip with molecular beacons detects miRNAs in Human CSF to reliably characterize CNS-specific disorders. RNA Dis 2017; 4: e1183. doi: 10.14800/rd.1183.

Introduction

MicroRNAs (miRNAs) are increasingly utilized as biomarkers for diagnostic and prognostic purposes. By binding to messenger RNAs (mRNAs) miRNAs guide the RNA-induced silencing complex (RISC) that post-transcriptionally represses the expression of protein-coding genes. MiRNAs are estimated to control about one third of all gene expression [1] and they have been shown to play crucial regulatory roles in several cellular processes, including proliferation, metabolism, development, and apoptosis [2, 3]. Given miRNAs extensive regulatory function within the cell, the aberrant expression of miRNAs has been studied in several human diseases, such as diabetes, arthritis, kidney disease, neurodegenerative disorders and cancer [4, 5]. Furthermore, cell free circulating miRNAs stably packaged in microvesicles can be detected in human serum, urine, plasma and cerebral spinal fluid [6]. In terms of their

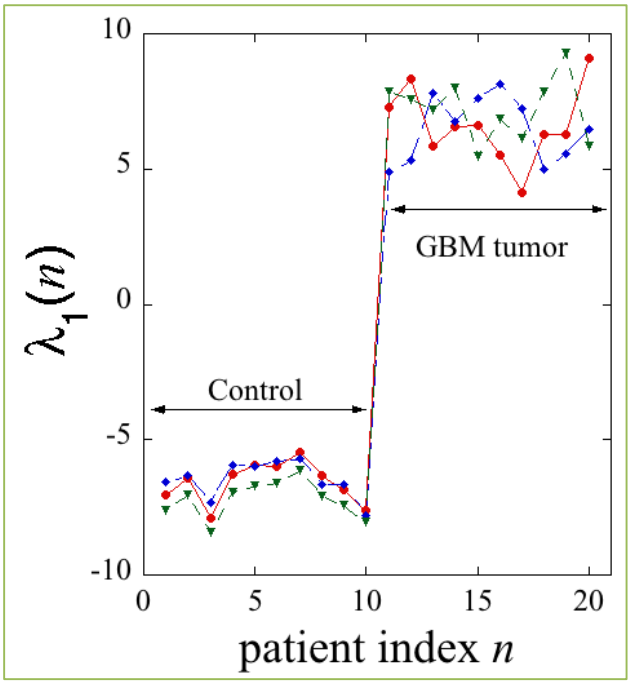


Figure 1. miRNA Analysis of Patient Samples. MiRNA analysis of 101 patient tumor samples and non-tumor samples reveal a subset of miRNAs unique to the tumor, and capable of distinguishing non-tumor and tumor samples, see Supplementary Theory 1 for details. Shown as an example are results of surprisal analysis for three groups of 20 patients, 10 control and 10 tumor samples. The analysis identifies those miRNAs that are significantly down-regulated, meaning a negative value of the multiplier, $\lambda_1(n)$, see equation (1) , in the non tumor samples (left side). The same miRNAs are significantly up-regulated in the tumor samples (right side). $\lambda_1(n)$ is the multiplier for the constraint that distinguishes between tumor and non-tumor samples where *n* is the index of the patient. See [10, 11, 14] for more details about surprisal analysis.

binding, miRNAs are short, about 22 bases in length, single stranded RNA molecules. The molecular beacons (MBs) that are used for identifying the miRNAs have a loop of 15-30 nucleotides that bind to the target sequence.

In recent studies, it has been demonstrated that expression levels of distinctive miRNAs are detectable in patient biofluids and can be correlated with different types of disorders of the central nervous system (CNS) and to disease prognosis [7, 8]. Cerebral spinal fluid (CSF), a clear fluid that cushions and delivers nutrients to the CNS and is in direct contact with the brain and spine, is a more effective mode of understanding symptoms and disorders of the CNS. We developed a micro-fluidic technology that employs molecular beacon sensing to detect and monitor miRNAs without amplification in patients' CSF, sustaining a linear dynamic range of 0.2 to 5 fmol. Utilizing this technology, we were able to identify a miRNA classifier unique to primary brain tumors, and with high fidelity distinguish patient samples using low volumes of CSF. Furthermore, we were able to characterize the molecular subtypes of these tumors, as defined by the Tumor Cancer Genome Atlas [9], using miRNAs present in patient CSF.

The microRNA classifier was identified by surprisal analysis of global microRNA expression levels in tumor and control of patients' tissue samples, as previously described [10, 11]. Shown in Figure 1 are results of surprisal analysis for three groups of 20 patients, 10 control and 10 tumor samples. Altogether we analyzed 101 samples. We identified a tumor-specific miRNA classifier capable of distinguishing tumor and non-tumor samples (Table 1). Furthermore, we validated this miRNA classifier in an independent cohort of patient tumor samples (Supplemental Table 1). Patients were selected as described in the Supplementary Materials (Materials and Methods). An important component of our work is showing quantitatively how we use surprisal analysis for a reliable reading by our chip. The theory is described in detail in a section of the Supplementary Materials. The essential point is that surprisal analysis [10, 11] enables us to quantitatively express the probability to detect the i'th miRNA in a tumor and non-tumor sample as

$$P(i\,|\,tumor) = \exp\left(-\lambda_0 G_{i0} - \lambda_1(tumor) G_{i1}\right)$$
$$P(i\,|\,non\ tumor) = \exp\left(-\lambda_0 G_{i0} - \lambda_1(non\ tumor) G_{i1}\right) \quad (1)$$

The expressions for the probabilities contain parameters that we determine by a fit to the input patient data. The large number of patients sampled and the large number, 534, of miRNAS that were detected is statistically sufficient to determine all these parameters. Among all the parameters, only the multiplier lambda1 differs for tumor and non-tumotr samples, see figure 1. As seen in figure 1 the sign of λ_1 is opposite for the tumor and nontumor samples. The value of lambda0 is the same for all patients and all miRNAs. The G's depend on the miRNA but not on the categories. Those miRNAs for which G_{i1} is particularly high or particularly low are listed in table 1. These are the miRNAs that are particularly useful as markers because it is for these miRNAs that their expression value is as different as it can be. This is because the value of G_{i1} is the same for either type of sample but their value of λ_1 is opposite. Quantitavely, equation (1) implies that the ratio of the expression values of the same i'th miRNA differ exponentially $P(i\,|\,tumor)/P(i\,|\,non\ tumor) = \exp\left(-\left[\lambda_1(tumor) - \lambda_1(non\ tumor)\right] G_{i1}\right)$. Using molecular beacons (MBs) targeting the miRNAs in the classifier we designed a high-sensitivity and high-affinity micro-fluidic chip. Briefly, MBs are fluorescent-labeled oligonucleotide chain, typically composed of 25-35 nucleotides. MBs have three distinct structural components. The, generally constructed by 15-30 nucleotides, specifically bind the target miRNAs. The stem portion consists of 5-8 base pairs that reversibly dissociates during binding to the

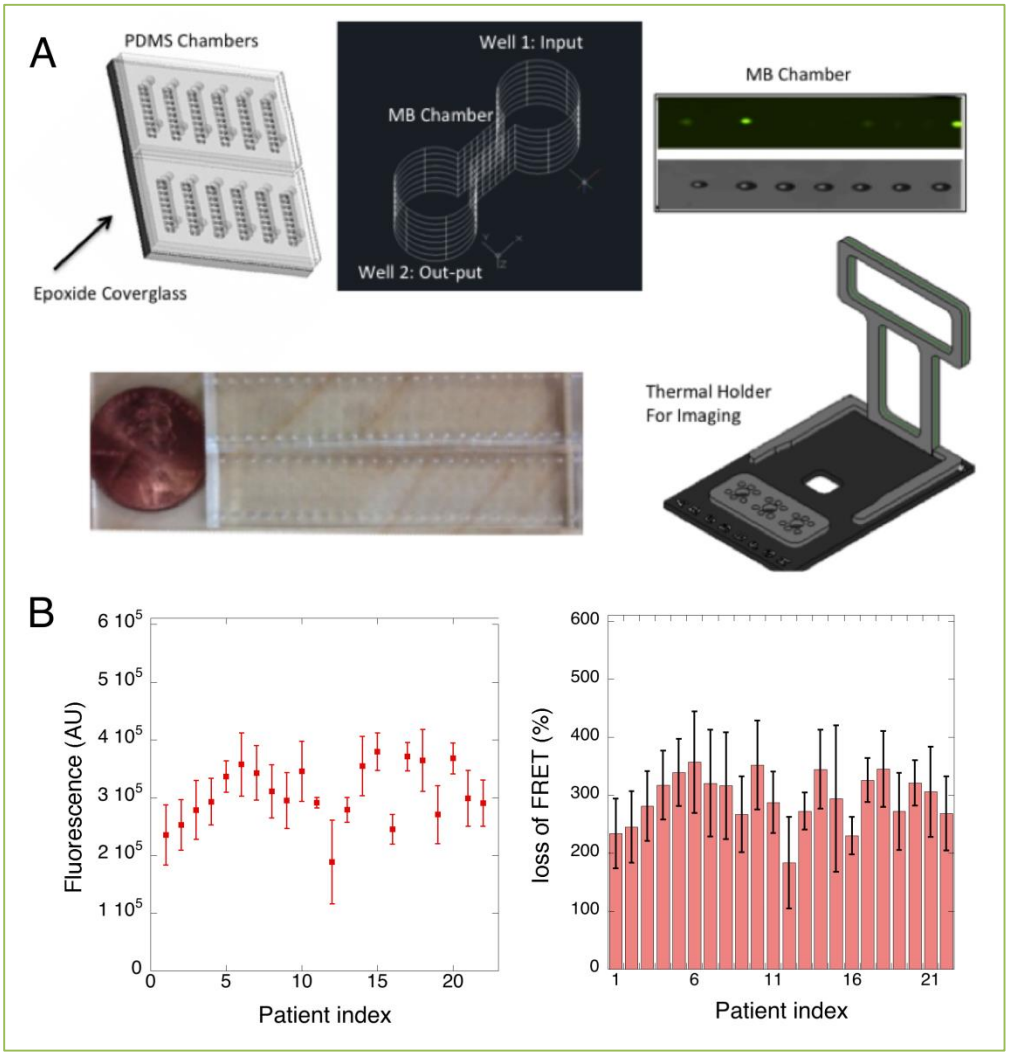


Figure 2. A Microfluidic Chip for miRNA sensing in CSF. A. The micro-chamber array consists of 2 layers: (i) a poly-D-lysine coated borosilicate glass chips substrate and (ii) a Polydimethylsiloxane (PDMS) channel layer (3X8 array of cutouts defines 24 micro-chambers), each for sample loading and unloading. MBs were printed onto chips and were excited at 495 nm and emission was monitored at 670 nm. B. Analysis of CSF samples from 22 tumor patients were also conducted following extraction of RNA using GFF filtration of CSF. Fluorescence emission was scanned from triple replicates and averaged across MBs. A loss of FRET behavior (increase in fluorescence activity) was detected consistently across MBs from micro-chambers loaded with CSF samples from patients with tumors. FRET loss was calculated as fluorescence units observed following CSF administration from baseline fluorescence and plotted as percent FRET loss over control (non-tumor patient biopsy). All CSF samples from tumor patients exhibited 2 to 3 fold increase in FRET loss. Error Bars=standard deviation of triple replicates.

target miRNAs. The thermodynamic equilibrium relations between the stem portion and double-stranded structure of loop portion increases the specificity of MBs compared to conventional linear probes. Finally, in the absence of the target miRNA, a fluorescence resonance energy transfer (FRET) procrss is enabled and observed. Upon target miRNA binding, the MBs can no longer undergo FRET due to the increased FÖrster radius and as a result an increase in fluorescence intensity is observed [12, 13]. These MBs will assist in distinguishing the targeted miRNA from unintended potential targets; the stem-loop structure destabilizes hybridization to larger non-target RNAs. MBs were designed to detect miRNAs in the miRNA classifier. To determine the number of miRNAs in the classifier necessary to distinguish patient samples, we apply a unique theoretical development to determine with high probability whether a patient sample contains tumor cells. This method of inference is described in detail in Supplementary Theory 1. In particular it is shown why and how our method of characterization of the expression levels, known as surprisal analysis [10, 11], is especially suited to identifying those miRNAs that are most relevant to distinguish healthy and diseased patients.

To determine the specificity of our MBs, synthetic human miRNAs with high sequence homology to the miRNAs in our classifier were labeled and cross hybridization was observed. As expected, a less than 2% cross-hybridization was observed for miRNAs differing by more than one nucleotide (data not shown).

Each MB was coupled to FRET fluorophores, Cy3 and Cy5. MBs were then spotted onto epoxide substrate coated borosilicate glass chips, to enable MB coupling to glass surface (L 75 X W 2 m, 1 mm thick) substrate, using a Sprint Inkjet Microarray (ArrayJet). MBs emission was monitored (Figure 2). The miRNA classifier consisted of miRNAs with expected increased expression in patient tumor samples and not in control samples. Since miRNA levels can span between two to four orders of magnitude, we also determined the linear dynamic range of our array. Briefly, varying mixtures of synthetic miRNAs were tested and as expected the fluorescence FRET signals were proportional to the input miRNA concentration over the entire range examined, providing a detection limit of 0.5 fmol and a linear dynamic range of nearly two orders of magnitude (data not shown).

Material and Methods

miRNA expression data. Level 3 miRNA expression data were obtained from The Cancer Genome Atlas (TCGA) data portal.

Surprisal analysis

Surprisal analysis as previously described [10, 11, 14] was utilized to directly compute the probabilities that a particular sample is from a diseased or a control patient as shown in equations (1).

Patient sample preparation

Patient samples were collected by the UCLA School of Medicine and the UCLA Department of Pathology for their Tissue Repository. None of the authors collected these directly from patients. No identifying patient information was collected alongside the tissue. All samples were deidentified as per HIPAA instructions. The patient samples (CSF and biopsy) were provided to this study under a protocol approved by the Western Institutional Review Board (protocol # 12-001039). In particular, the authors were provided with an exempt IRB by the UCLA Human Research Review Board specifically for the purpose of this study. The patient samples (WHO Grade IV) were collected from patients following informed written consent at the David Geffen School of Medicine, University of California, Los Angeles. Tissue samples and CSF were for patients who did not undergo prior radiation therapy using therapeutic subtotal or total resection that were performed with image guidance. See [11. 14] for additional details. Patient miRNA expression analysis will be deposited in GeoAsscession and NCBI.

RNA extraction

RNA was isolated using the same procedures and checks that we used in earlier work [11.14] on RNA in disease. Further quality control and normalization was performed using GeneSpring GX 11 (Agilent) following their recommended procedures for data handling.

Molecular beacon design and synthesis

MBs were designed using miRBase database. FRET fluorophores were coupled to the 3' and 5', respectively, and MBs were subjected to HPLC purification prior to use. The miRNA designs were as previously described. [13]

Microfluidic chip design and fabrication

The microfluidic chip consists of 2 layers: (i) an epoxide coated borosilicate glass chips (L 75 X W 2 cm, 1 mm thick) substrate and (ii) a Polydimethylsiloxane (PDMS) channel layer (3X8 array of cutouts defines 24 microchambers, each with length 9 mm, width 1mm, and volume 50 uL. Drilled inlet and outlet holes (a pair of 1mm-diameter holes for each microchamber) were also constructed for sample loading and unloading. MBs were printed onto chips using a Sprint Inkjet Microarray (ArrayJet). MBs were excited at 495 nm and emission was monitored at 670 nm using and confirmed by microscopy.

Sample handling

Prior to loading onto the chip, the mirVana miRNA isolation kit (Ambion, Austin, TX) was used to isolate RNA. GFF filtration was used for CSF samples. Samples were allowed to incubate for 2 hours at 37C. At high temperatures the helical order of the MB stem gives way to a random-coil configuration. Fluorescence versus temperature profiles of the molecular indicate that the molecular beacon is suitable for assays that are performed below 55 °C (42`C), because below 55 °C the free molecular beacons remain dark, yet the probe-target hybrids form spontaneously and are stable. Florescence emission was averaged across MBs in the signature and compared to the fluorescence emission from controls.

Summary, materials and methods

Patient samples from 53 tumor and 48 non-tumor brain tissue biopsies were homogenized and subjected to RNA

extraction preparation, as described in the Supplemental Material (Materials and Methods) prior to loading onto the PDMS micro-chambers to prevent cross-hybridizing to pri-miRNA, pre-miRNA and other small non-coding RNAs (Supplemental Figure 1). Non-tumor patient biopsy samples were extracted from individuals not matched to the tumors in this study. Patient selection was conducted as described in Supplementary Section (Material and Methods). Samples were then loaded onto the micro-chambers in triplets. The array fluorescence was scanned thrice and the response averaged across scans. Fluorescence scans of the MB-based array showed classifier-specific FRET behavior from tumor biopsy samples with more than two-fold increase in fluorescence emission, or loss of FRET, seen only from tumor patient samples. Only residual low florescence activity from non-tumor biopsy samples was detected, and no changes in FRET were observed (Supplemental Figure 1).

Results

Recent studies have reported the presence of miRNAs in cerebral spinal fluid (CSF) [6]. Subsequently, we determined whether we could use the small amounts of RNA that can be recovered from the volumes of CSF available clinically (300-500uL), usually following lumbar puncture or surgical resection, on this microfluidic chip. We isolated RNA from CSF samples using glass fiber filters (GFF) prior to loading into the micro-fluidic chip. Approximately 50 µl of CSF provided 10 to 15 ng of total RNA. The GFF was placed in inlets fabricated prior to patient samples entering micro-chambers, (at a flow rate of 50 µl min^{-1} to separate RNA from the CSF patient sample). Residual CSF was disposed. GFF-mediated extracted RNA from CSF samples was then sent into micro-chambers to monitor for MB hybridization and subsequent FRET loss. Fluorescence scans of the MBs from 22 individual patient CSF samples, showed tumor signature-specific FRET behavior (loss of FRET) from tumor CSF samples, similar to the scans of the tumor patient biopsy samples, with an increase in fluorescence emission, or loss of FRET in MBs and no change in FRET behavior from non-tumor patient CSF samples (Figure 2B).

Recently, an effort lead by the Cancer Genome Atlas (TCGA) network identified four molecular subclasses of tumor-specific genes capable of possibly characterizing sub-phenotypes of astrocytomas: Classical, Proneural, Neural and Mesenchymal. Each of these subclasses displayed unique gene expression profiles [8]. We anticipated that the miRNA classifier we identified provided predictive value for further characterization of the heterogynous tumor composition and cancer cell population dynamics. The miRNAs that compose the classifier have regulatory functions targeting genes critical to each of the four molecular subclasses. Specifically, miRNAs of our classifier have conserved binding site on the 3' UTR of many of the selected genes that are critical for sustaining these subclasses of astrocytomas (Supplemental Table 2). In the Proneural subclass, PDGFR was up regulated in patient tumor biopsy samples. We predict a down-regulation of miR-7, a miRNA in our cancer-specific classifier, in the Proneural subclass of tumor samples. MiR-7 has a conserved binding site on the 3' UTR of the PDGR transcript, unveiling possible miRNA and mRNA networks that regulate the molecular processes underlying these distinct subclasses. Similar networks were identified for both the Classical and Neural subtypes (Supplementary Table 2). Conversely, genes that are down regulated in a specific tumor subclass are correlated with an up-regulation of their corresponding miRNAs, which have conserved binding sites to their 3'UTR. In the Mesenchymal subtype for example, NFI is down regulated and our miRNA classifier shows an up-regulation of miR-25 and mir-27a, as expected. Both miR-25 and miR-27a have conserved binding sites of on the 3'UTR of NFI. With additional characterization of the four subclasses of these brain cancers, we anticipate that miRNA classifiers in the CSF may illuminate tumor composition and predict resistance to therapeutics.

Discussion

Sensing biofluids from a patient is typically done once or at best a few times. Yet it is clearly essential that the inference made about the sample being from a tumor or a non-tumor is as reliable as possible. To recognize that there is an inherent problem consider the far simpler decision as to whether a coin is honest or tempered with. A single toss of this coin will not provide a reliable answer. If it is important to know about the coin, many people will guesstimate that at least a dozen tosses are required. There is a similar problem for us, namely how can a single measurement by the chip get us a reliable inference. For both the chip and the coin we need to make one binary decision, tumor/non-tumor in our case, based on one measurement. There is however a key difference that enables us to make a reliable decision. Unlike the case of the single coin toss we have the essential advantage that in a single sensing we can read the fluorescence from several molecular beacons. To design a chip we need to determine how many is several and we need to quantify reliable. It is one of the key advantages of our approach to the analysis of miRNA expression levels, known as surprisal analysis [10, 11, 14], that it is optimally suited to deliver the required answers on both issues. The mathematical formulation is developed in the supplementary materials (Supplementary Theory 1). What we show is that using our miRNA classifier we can get a reliable inference, 98% confidence, using the chip design as described above. The key is the opposite change in expression levels of certain miRNAs in tumor and non-tumor patients as shown in figure 1. Moreover, as shown in Table I of the supplementary

materials, there are miRNAs that similarly have a high deviation in different cohorts of patients.

The mathematical result for the reliability when we measure the fluorescence from a single beacon is given by equation (7) of Supplementary Theory 1. For a single miRNA the difference in expression levels of non-tumor and tumor patient samples, while large, is not sufficient to generate a reliable inference. One cannot rely on monitoring just one miRNA to provide a reliable inference. But by reading several beacons we extend the reliability to the value given by equation (8) of the SI. The miRNAs most useful for increasing the reliability of the sensing are listed in Supplementary Table I. It is shown that there are enough miRNAs that consistently differ to a high extent in different cohorts of patients to insure a reliable inference from a single sensing.

In summary, it is the observation that more than one miRNA is consistently very different in the expression level in non-tumor and tumor patient samples that enables us to design a reliable chip.

MB-based sensing in biofluids, such as cerebral spinal fluid, can provide a point-of care platform for detecting CNS associated diseases, monitoring disease progression and characterizing diseases of the CNS using a less invasive and less high-risk module through analysis of CSF in lieu of patient biopsy sample extraction.

Conflicting interests

The authors have declared that no conflict of interests exist.

Acknowledgements

We thank the Institute of Molecular Medicine (IMED) and the David Geffen School of Medicine of the University of California, Los Angeles for reagents, patient samples, microarray analysis and early career support to SZ. A Prostate Cancer Foundation Creativity award to RDL also supported this work. The work on reliable inference was supported by FP7-funded BAMBI Project 618024. FR is a director of research with FNRS (Fonds National de la Recherche Scientifique), Belgium.

Abbreviations

FRET: fluorescence resonance energy transfer; miRNA: microRNA; CSF: Cerebral spinal fluid; CNS: Central nervous system.

Author contributions

SZ carried out the molecular genetic studies, participated in the surprisal analysis and in drafting the manuscript. FR carried out the surprisal analysis and participated in drafting the manuscript. RDL participated in the surprisal analysis and in drafting the manuscript. All authors read and approved the final manuscript.

References

1. Alvarez-Garcia I, Miska EA. MicroRNA Functions in Animal Development and Human Disease. Development 2005; 132: 4653-4662.
2. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, *et al*. MicroRNA expression profiles classify human cancers. Nature 2005; 435: 834-838.
3. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nature Rev Cancer 2006; 6: 857-866.
4. Kosaka N, Nobuyoshi HI, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. Cancer Sci 2010; 101: 2087-2092.
5. Kocerha J, Kauppinen S, Wahlestedt C. MicroRNAs in CNS disorders. Neuromol Med 2009; 11: 162-172.
6. Baraniskin A, Kuhnhenn J, Schlegel U, Schmiegel W, Hahn S, Schroers R. MicroRNAs in cerebrospinal fluid as biomarker for disease course monitoring in primary central nervous system lymphoma. J Neurooncol 2012; 109: 239-244.
7. De Smaele E, Ferretti E, Gulino A. MicroRNAs as biomarkers for CNS cancer and other disorders. Brain research 2010; 1338: 100-111.
8. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, *et al*. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in< i> PDGFRA, IDH1, EGFR and NF. Cancer Cell 2010; 17: 98-110.
9. Σανβορν JZ, Benz,SC, Craft B, Szeto C, Kober KM, Meyer L, *et al*. The UCSC cancer genomics browser: update 2011. Nucl Acids Res 2011; 39.suppl 1: D951-D959.
10. Remacle F, Kravchenko-Balasha N, Levitzki A, Levine RD. (2010) Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. Proc Natl Acad Sci USA 2010; 107: 10324-10329.
11. Zadran S, Remacle F, Levine RD. (2013) miRNA and mRNA cancer signatures determined by analysis of expression levels in large cohorts of patients. Proc Natl Acad Sci USA 2013; 110: 19160-19165.
12. Zadran S, Standley S, Wong K, Otiniano E, Amighi A, Baudry M. Fluorescence resonance energy transfer (FRET)-based biosensors: visualizing cellular dynamics and bioenergetics. Appl Microbiol Biotechnol 2012; 96: 895-902.
13. Tyagi S, Kramer FR, Molecular beacons: probes that fluoresce upon hybridization. Nature biotechnology 1996; 14: 303-308.
14. Zadran S, Remacle F, Levine RD. Surprisal Analysis of Glioblastoma Multiform (GBM) MicroRNA Dynamics Unveils Tumor Specific Phenotype. PLoS ONE 2015; 9: e108171. doi:10.1371/journal.pone.0108171

Supplementary Data

Supplementary Theory 1

In this supplement we demonstrate that the results of surprisal analysis of real patient data enable us to design a chip that reliably determines the probability that a particular patient is diseased or is not. The issue is that we only perform a single reading of the chip or at best, a few replicated. It is therefore necessary to define an 'event' that is sure to take place with a very high probability so that a single experiment will suffice. Technically, the design desideratum is that the decision that a patient is diseased is made with as high a reliability as is realistically possible. The input to knowing the probability that a patient is diseased is the reading of the chip for the patient's biopsy sample. In the main text we reported that surprisal analysis provided a clear signature that distinguishes between tumor and non tumor patients, see Figure 1 of the main text. In detail it means that many miRNAs are up regulated in the tumor patients and down regulated in the non tumor patients AND that many other miRNAs are down regulated in the tumor patients and up regulated in the non tumor patients. In other words, for tumor patients we generate a list of miRNAs where those at the top are most up regulated and those at the bottom are most down regulated. Surprisal analysis further shows that there is an exactly complementary list for the non tumor patients. Keeping the list the same, for non tumor patients the miRNAs at the top of the list are most down regulated and those at the bottom are most up regulated.

We show below that the essential observation that we derive from surprisal analysis is that the gap between the up- and the down-regulated miRNAs is wide enough that with reading the level of just a few miRNAs we can already provide a reliable reading for the probability that a patient is or is not diseased. These few miRNA's constitute the classifier and are listed in Table 1 of the main text. It is therefore practical to construct a chip as discussed in the main text. In this section we assume that, as discussed in the main text, the chip reads the level of several miRNAs. We need to show that the number of miRNAs whose level needs to be determined is small enough that it can be read by a single chip. Fortunately the results for tumor samples are such there is much scope for a margin of safety and one can readily read more than double the number of miRNAs that is strictly necessary. We therefore begin with a list of miRNAs whose level in the biopsy sample of the patient has been read. This we call the chip data for the patient. Given this data we want to compute the probability that the patient is diseased. The point is to read enough miRNAs until this probability is very high, as near to one as possible or it is very low, as near to zero as possible. In the notation of probability theory the probability we want to determine is written as. We compute the probability of disease given the chip read data from Bayes theorem

(1) $$P(tumor \mid \text{chip data of patient}) = P(tumor \mid \text{without patient data}) \frac{P(\text{chip data of patient} \mid tumor)}{P(\text{chip data of patient})}$$

In equation (1), $P(\text{chip data of patient} \mid tumor)$ is the probability that we get the reading that we got given that the patient is diseased. The problem is in the denominator of equation (1), which requires as an input a probability that we do not have. To bypass this problem we write an equation completely analogous to (1) but for a non tumor patient:

(2) $$P(non\ tumor \mid \text{chip data of patient}) = P(non\ tumor \mid \text{without patient data}) \frac{P(\text{chip data of patient} \mid non\ tumor)}{P(\text{chip data of patient})}$$

Therefore, given the chip data we have that

(3) $$\frac{P(tumor \mid \text{chip data of patient})}{P(non\ tumor \mid \text{chip data of patient})} = \frac{P(tumor \mid \text{without patient data})}{P(non\ tumor \mid \text{without patient data})} \frac{P(\text{chip data of patient} \mid tumor)}{P(\text{chip data of patient} \mid non\ tumor)}$$

We here take the view that a patient is either non tumor or tumor so that $P(non\ tumor \mid \text{chip data of patient}) = 1 - P(tumor \mid \text{chip data of patient})$ and therefore (3) can be rewritten as

(4) $$\frac{P(tumor \mid \text{chip data of patient})}{1 - P(tumor \mid \text{chip data of patient})} = \frac{P(tumor)}{1 - P(tumor)} \frac{P(\text{chip data of patient} \mid tumor)}{P(\text{chip data of patient} \mid non\ tumor)}$$

We do not know exactly what is the probability that a randomly chosen person is diseased with tumor. From US population data the probability is about 3 in 100,000. It does not matter if the estimate is somewhat off because the final result will not be sensitive to the exact value. We put $\frac{P(tumor)}{1 - P(tumor)} = 3 \times 10^{-5}$ and therefore we can solve equation (4) for the probability of interest

(5) $$P(tumor \mid \text{chip data of patient}) = \frac{1}{1 + \left(\frac{1}{3}10^5\right) \frac{P(\text{chip data of patient} \mid non\ tumor)}{P(\text{chip data of patient} \mid tumor)}} = \frac{1}{1 + \left(e^{10.4}\right) \frac{P(\text{chip data of patient} \mid non\ tumor)}{P(\text{chip data of patient} \mid tumor)}}$$

We will compute the two probabilities that we need, $P(\text{chip data of patient} \mid tumor)$ and $P(\text{chip data of patient} \mid non\ tumor)$ from surprisal analysis of data of both non tumor and tumor patients (15). We take into consideration that the chip is designed to read a few miRNAs that are up regulated. The dominant pattern identified by surprisal analysis distinguishes between tumor and non tumor patients. Explicitly in a tumor and non tumor patients the probability of the *i*'th miRNA is (15)

$$P(i\,|\,tumor) = \exp\left(-\lambda_0 G_{i0} - \lambda_1(tumor) G_{i1}\right)$$
$$P(i\,|\,non\ tumor) = \exp\left(-\lambda_0 G_{i0} - \lambda_1(non\ tumor) G_{i1}\right) \quad (4)$$

where typically $\lambda_1(non\ tumor) \simeq -\lambda_1(tumor)$ see figure 1 of the main text. Then

$$\frac{P(i\,|\,tumor)}{P(i\,|\,non\ tumor)} = \exp\left(-\left(\lambda(tumor) - \lambda(non\ tumor)\right) G_{ia}\right) \quad (5)$$

We sort the miRNAs so that i =1 is the most up regulated. i =2 is the next most up regulated, etc. To collect enough evidence we need to measure the overexpression of m miRNAs

$$\frac{P(1\,|\,tumor)}{P(1\,|\,non\ tumor)}\frac{P(2\,|\,tumor)}{P(2\,|\,non\ tumor)}\cdots\frac{P(m\,|\,tumor)}{P(m\,|\,non\ tumor)}$$
$$= \exp\left(-\left(\lambda(tumor) - \lambda(non\ tumor)\right)\sum_{i=1}^{m} G_{i\alpha}\right) \quad (8)$$

This brings us to the result we need for equation (5)

$$\frac{P(\text{chip data of patient}\,|\,tumor)}{P(\text{chip data of patient}\,|\,non\ tumor)} = \exp\left(-\left(\lambda(tumor) - \lambda(non\ tumor)\right)\sum_{i=1}^{m} G_{ia}\right)$$

so that

$$P(tumor\,|\,\text{chip data of patient}) = \frac{1}{1+\left(e^{10.4}\right)\exp\left(\left(\lambda(tumor) - \lambda(non\ tumor)\right)\sum_{i=1}^{m} G_{ia}\right)} \quad (9)$$

We therefore need to include as many miRNAs on the chip such that, see equation (9), the number m satisfies

$$\left(-\left(\lambda(tumor) - \lambda(non\ tumor)\right)\sum_{i=1}^{m} G_{ia}\right) > 10.4$$

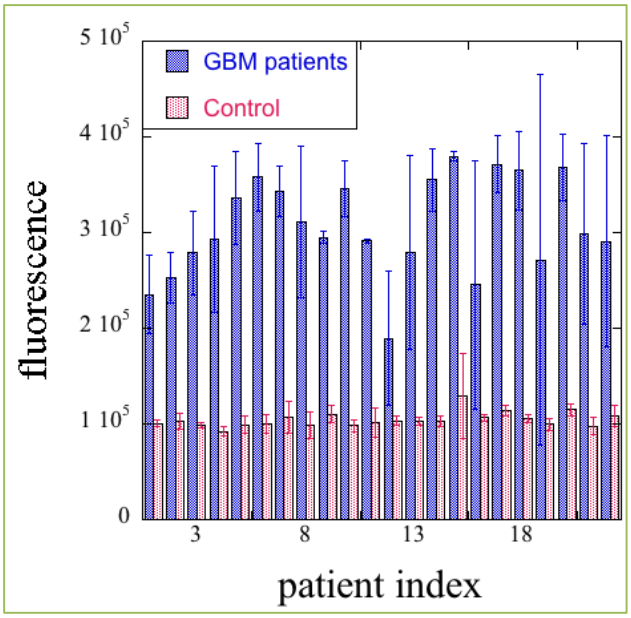
We need the exponent to be at least 13.4 if we need a probability of better than 95%. We need an exponent larger than 14.4 if the probability is to be larger than 98% etc. This is relatively easy to achieve even with few miRNAs because the difference $\left(\lambda(tumor) - \lambda(non\ tumor)\right)$ is a rather large negative number as shown in Figure 2B of the main text.

Supplemental Table 1. MiRNA analysis reveal consistent miRNA classifiers between TCGA and UCLA patient cohorts. Overlap of miRNAs between the two cohorts was color-matched to highlight the similarity

TGCA Patient Cohort	UCLA Patient Cohort
hsa-miR-21	hsa-miR-923
hsa-let-7b	hsa-let-7a
hsa-let-7a	hsa-miR-125b
hsa-miR-9*	hsa-let-7b
hsa-miR-125b	hsa-let-7f
hsa-let-7c	hsa-let-7c
hsa-miR-29a	hsa-miR-21
hsa-miR-26a	hsa-miR-29a
hsa-let-7f	hsa-miR-26a
hsa-miR-9	hsa-miR-9
hsa-let-7e	hsa-miR-451
hsa-miR-24	hsa-miR-9*
hsa-let-7d	hsa-miR-16
hsa-miR-22	hsa-miR-494
hsa-miR-126	hsa-let-7g
hsa-miR-16	hsa-let-7i
hsa-miR-195	hsa-miR-99a
hsa-let-7g	hsa-miR-100

Supplemental Table 2. Characterization of Tumor Subclasses. Genes that are up regulated in a specific subclass of the tumor, have 3'UTR sites that are conserved to miRNAs in the classifier, exhibiting a mRNA and miRNA network response. Conversely, genes that are down regulated in a subclass (blue*) are correlated with an up-regulation of miRNAs that have conserved binding sites to their 3' UTR.

Molecular Subtypes	Correlated miRNAs from Classifier
Classical	
EGFR	hsa-miR-7
GAS-1	hsa-miR-124a
Proneural	
PDGFRA	hsa-miR-137
SOX	hsa-miR-139
Neural	
SYTI	hsa-miR-7
SLC12A5	hsa-miR-137
Mesenchymal	
NF1*	hsa-miR-25
	hsa-miR27a



Supplemental Figure 1. Results of an analysis of 22 tumor and 22 control patient biopsy samples. Fluorescence emission was scanned and the data averaged across MBs. A significant loss of FRET behavior (increase in fluorescence activity) was detected from micro-chambers loaded with tumor patient biopsy samples. Red=control samples, Blue=tumor patient biopsy samples, Error Bars=standard deviation of triple replicates.