

1-2-2016

High-throughput profiling for discovery of non-coding RNA biomarkers of lung disease.

Paul J. McKiernan

Royal College of Surgeons in Ireland

Catherine M. Greene

Royal College of Surgeons in Ireland, cmgreene@rcsi.ie

Citation

McKiernan PJ, Greene CM. High-throughput profiling for discovery of non-coding RNA biomarkers of lung disease. *Expert Review of Molecular Diagnostics*. 2016;16(2):173-85

This Article is brought to you for free and open access by the Department of Medicine at e-publications@RCSI. It has been accepted for inclusion in Medicine Articles by an authorized administrator of e-publications@RCSI. For more information, please contact epubs@rcsi.ie.

— Use Licence —



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Title: High-throughput profiling for discovery of non-coding RNA biomarkers of lung disease

Running title: RNA biomarkers of lung disease

Authors: Paul J. McKiernan and Catherine M. Greene*

Respiratory Research, Department of Medicine, The Royal College of Surgeons in Ireland, Education and Research Centre, Beaumont Hospital, Dublin, Ireland

*Corresponding author: cmgreene@rcsi.ie

Abstract:

In respiratory medicine there is a need for clinical biomarkers for diagnosis, prognosis and assessment of response to therapy. Non-coding RNA is expressed in all human cells: two major classes, ncRNA and microRNA, are detectable extracellularly in the circulation and other biofluids. Altered ncRNA expression is associated with lung disease; collectively this indicates that ncRNA represents a potential biomarker class. This article presents and compares existing platforms for detection and quantification of ncRNA, specifically hybridization, qRT-PCR and RNA sequencing, and outlines methods for data interpretation and normalization. Each approach has merits and shortcomings, which can affect the choice of method when embarking on a biomarker study. Biomarker properties and pre-analytical considerations for ncRNA profiling are also presented. Since a variety of profiling approaches are available, careful study and experimental design are important. Finally challenges and goals for reliable, standardized high-throughput ncRNA profiling in biofluids as lung disease biomarkers are reviewed.

Keywords: microRNA , long non-coding RNA , biomarker , profiling , lung disease

1. Introduction:

Three major aspects of clinical medicine that are central to all human diseases are accurate diagnosis, prognosis and assessment of response to therapy. Reliable clinical biomarkers that facilitate these processes can greatly enhance patient care. As it is usually not possible to detect early intrapulmonary changes that occur during the establishment of chronic inflammatory lung diseases, the search for surrogate markers of lung disease has been ongoing for many years. Circulating RNA, and in particular non-coding RNA, represents a relatively recently discovered biomarker class that has potential in this field. This article will review the field of non-coding RNA in chronic inflammatory lung disease and present the current high-throughput methodologies that can be used for the discovery of non-coding RNA biomarkers of lung disease.

2. Non-coding RNA – an introduction

The human genome encodes a vast amount of non-coding RNA; only ~1.2% of the genome is protein-coding, yet much of it is transcribed¹. The Encyclopaedia of DNA Elements (ENCODE) consortium has recently reported that up to 75% of the human genome is transcribed². Thus many of these transcripts are non-coding but do have important roles nonetheless. Non-coding RNA (ncRNA) is a class of endogenous, non-protein coding RNA transcripts with very diverse, yet essential roles in almost all cellular processes. Several types of ncRNA transcripts exist, and are broadly characterised based on size; long ncRNA (lncRNA) over 200 nucleotides (nt) in length; and short ncRNA, less than 200 nt.

2.1 microRNA

Of the small ncRNA, microRNAs (miRNA) are the most notable and most extensively studied. miRNAs are 21-24 nucleotide duplex RNAs involved in the translational regulation of gene expression³. Although the term ‘microRNA’ was first coined in 2001, the first miRNA, *lin-4*, was discovered eight years earlier in the nematode *Caenorhabditis elegans*⁴. Having been initially discovered to have importance in developmental biology, interest in these small RNAs has dramatically increased since this time as they have been found to have significant roles in almost all biological processes. Expression levels of miRNAs vary greatly between cells and tissues, and indeed aberrant levels of miRNA are associated with many diseases in humans. miRNA generally regulate gene expression post-transcriptionally by

binding to the 3'UTR of a target messenger RNA (mRNA). Although most miRNA studies have largely focused on miRNA-mRNA interactions in the 3'UTR of target mRNA, these interactions can occur in the 5'UTR and coding sequence (CDS)^{5,6}. An interesting aspect of miRNA regulation of mRNA translation lies in the fact that full complementarity between miRNA and target mRNA is not required. In fact, only partial complementarity is required and a 2-8 nucleotide (nt) 'seed region' is thought to be crucial in the selection of targets by miRNA^{7,8}. Binding to miRNA responsive elements (MREs) in target mRNA appears to occur through this seed region. Messenger RNAs typically have many different MREs and can therefore be regulated by more than one miRNA.

2.2 Long non-coding RNA

Long non-coding RNAs (lncRNAs) are an emerging class of ncRNA. Unlike miRNA whose functions are generally mediated through binding with target mRNAs and affecting translation; lncRNAs are implicated in the regulation of gene expression at almost every level. Although their precise mechanisms of action are relatively unknown, common mechanistic themes are emerging. They have the ability to affect gene expression through a spectrum of interactions with proteins, RNA and DNA. With their capacity to act both *in cis* and *in trans*, they can guide epigenetic modifier complexes or transcription factors to particular genomic sites. These lncRNAs have been implicated in a variety of cellular processes, including genomic imprinting, chromatin modification and the regulation of cellular cycle.

3. Non-coding RNA in respiratory disease

Much recent evidence demonstrates that both miRNA and lncRNA are required to regulate lung development and maintain homeostasis^{9,10}. During both murine and human lung development, widespread changes in miRNA expression have been observed¹¹. For example, DICER, a key enzyme required for miRNA biogenesis, has been shown to be essential for proper lung epithelial morphogenesis; *Dicer* knockout mice, who have disrupted miRNA processing, display a lethal phenotype as a result of impaired lung growth¹². It is also known that lncRNA expression patterns are altered in pathological states, such as lung cancer, lung inflammation and various lung diseases.

3.1 Altered miRNA expression associated with lung disease

Many ncRNA expression and function studies in the lung have been performed in the context of lung cancer and the current understanding of ncRNA as lung cancer biomarkers were recently reviewed ¹³. Here we will provide a brief overview of miRNA studies that have been performed in a selection of chronic inflammatory lung diseases.

The World Health Organisation predicts Chronic obstructive pulmonary disease (COPD) to become the third leading cause of death worldwide by 2030 ¹⁴. COPD is a debilitating lung disorder which is characterised by progressive airflow obstruction that is not fully reversible. It is associated with an enhanced inflammatory response to gasses and noxious particles, in particular from cigarette smoke, and is a leading cause of morbidity ¹⁵. COPD is a heterogeneous disease with the main clinical manifestations being various combinations of chronic bronchitis and emphysema. It is currently incurable, although smoking cessation and symptomatic treatment can delay disease progression. Akbas and colleagues ¹⁶ were the first to screen COPD serum for miRNA-based biomarkers and found five miRNAs (miR-20a, miR-28-3p, miR-34c-5p, miR-100 and miR-7) to be significantly dysregulated. Another study which compared miRNA expression in plasma of COPD patients compared with healthy smokers, found nine differentially expressed miRNAs in COPD (miR-29b, miR-483-5p, miR-152, miR-629, miR-26b, miR-101, miR-106b, miR-532-5p and miR-133b) ¹⁷. The data on plasma levels of miR-106b in COPD ex-smokers led the authors to suggest that plasma miR-106b levels may reflect persistent and systemic changes even after smoking cessation. Another recent study has suggested serum miR-21 and miR-181a and their ratio as potential biomarkers for predicting COPD development in heavy asymptomatic smokers ¹⁸.

Unlike COPD, asthma is a chronic inflammatory airways disease that is characterised by reversible airflow obstruction. However in a significant minority of asthmatics the airflow obstruction is not reversible and in severe cases poor responses to steroid and beta agonist therapies can result in potentially fatal attacks ^{19,20}. Hence there is an unmet clinical requirement for biomarkers of asthma, and in particular the more severe phenotypes. Very few studies have investigated cell-free miRNA expression in serum and plasma of asthmatics, and most of the focus has been on their expression within circulating T cells ²¹. One of these studies has identified differential expression of miR-1248, miR-26a, Let-7a, and Let-7d in serum of asthmatic patients compared to controls ²². Some interesting work has begun on the use of exhaled breath condensate (EBC) as a noninvasive convenient representation of the

airway lining fluid. Interestingly miRNAs are detectable in these samples²³ and miRNAs such as miR-1248, miR-155 and let-7a have been found to be differentially expressed in EBC from asthmatics versus healthy controls²⁴. miR-570-3p has been shown to be induced in TNF α -treated airway epithelial cells and its levels in EBC inversely correlate with FEV1% in asthmatics²⁵.

Cystic fibrosis (CF) is a lethal multi-system autosomal recessive disease affecting many organs. It is characterised by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. A deletion of a phenylalanine residue at position 508 of the *CFTR* protein (F508del) is the most frequent mutation, accounting for approximately 70% of the alleles in CF world-wide, although over 1,900 mutations have been identified to-date²⁶. We have been the first to examine miRNA expression in people with CF²⁷. The expression levels of several miRNA were found to be altered in the bronchial epithelium of people with versus without CF. For example, miR-126 was significantly down-regulated in the CF lung. Down-regulation of this miRNA was associated with an up-regulation of target of Myb1 (TOM1), a negative regulator of IL-1 β , TLR2 and TLR4 induced signaling, and this mRNA was shown to be directly targeted by miR-126. Various groups have examined the role of miRNAs in controlling *CFTR* expression with various miRNAs being demonstrated to regulate *CFTR*²⁸⁻³³. These studies have found some differences in the miRNAs reported to regulate *CFTR* but this is most likely due to different experimental situations, such as different cell types and responses to different stimuli such as cigarette smoke. Such miRNAs include miRs -138³², -223³³, -224³⁴, -509-3p²⁸, -4921²⁸, however miRs -101, -145 and -494 have been replicated in more than one of these studies^{29-31,33,35}. For example, we have demonstrated that miR-145, miR-223 and miR-494 are up-regulated in CF bronchial epithelial cells and cell lines, are inversely correlated with *CFTR* levels and directly target *CFTR* mRNA³³. Their expression also correlated with p.Phe508del mutation and colonisation with *Ps. aeruginosa*. Other recent studies have postulated that miRNA may be useful in the therapy of CF^{36,37}. For example, Kumar *et al.* have identified that a miR-16 mimic can rescue F508del-*CFTR* protein function in primary cultures and airway cell lines³⁷. Several studies have also examined the roles of miRNA in inflammation in CF. *Ps. aeruginosa* infection induces the production of pro-inflammatory cytokines such as IL-8 in the CF airway epithelium. Fabbri and colleagues³⁸ have found a decrease in miR-93 levels in CF bronchial epithelial cells during infection with this major CF pathogen and that this miRNA directly targets IL-8 mRNA. We reported similar

findings for miR-17³⁹ whilst Zhang *et al.*⁴⁰ have shown that high levels of miR-199a-5p in human and murine CF macrophages and murine CF lungs leads to increased TLR4 signaling.

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive disease of the lung interstitium, usually occurring in middle-aged and elderly adults. It is characterised by epithelial cell damage, fibroblast proliferation and exaggerated accumulation of extracellular matrix within the lung parenchyma. Let-7d was the first miRNA demonstrated to be involved in the pathogenesis of IPF, reported in 2010⁴¹. Since then a number of studies have examined roles of miRNA in the pathogenesis of this condition^{42–45}. Yang *et al.*⁴⁶ reported 47 differentially expressed miRNAs in the serum of IPF patients compared to healthy controls. They validated six of these (miR-21, miR-199a-5p and miR-200c are increased whereas miR-31, let-7a, and let-7d are under-expressed) in an independent cohort, thus demonstrating the potential of serum miRNA as biomarkers of IPF. Another recent study demonstrated that altered serum levels of miR-21, miR-155 and miR-101-3p were associated with forced vital capacity (FVC) and radiological features in IPF⁴⁷.

3.2 Lung disease lncRNA expression patterns

lncRNAs have been less well characterised than miRNA in COPD, asthma and CF and there is just one report to-date on the expression and function of lncRNA in IPF⁴⁵. In COPD the most exciting development in this area relates to a recently discovered novel lncRNA, termed the smoke and cancer-associated lncRNA-1 (SCAL1). As its name hints, this lncRNA is elevated in numerous lung cancer cells lines, upregulated *in vitro* by cigarette smoke extract⁴⁸ and is also elevated in the airway epithelia of cigarette smokers versus non-smokers⁴⁹. In another report genome-wide expression analysis of lncRNA in the lung tissue of non-smokers without COPD, and smokers with and without COPD led to the discovery that various lncRNAs are altered in COPD patients compared to smokers without COPD⁵⁰. As in other chronic inflammatory lung diseases, lncRNA have also been under-explored in the pathophysiology of asthma. Perry *et al.*⁵¹ have detected altered expression of numerous lncRNAs in corticosteroid treated airway smooth muscle cells; key cells in the pathogenesis of asthma. Interestingly, they propose that four of these (RP11-46A10.4, LINC00883, BCYRN1, and LINC00882) act as miRNA ‘sponges’ for miR-150, miR-371-5p, miR-940, and miR-1207-5p, respectively. Widespread changes in the expression of lncRNAs have also been observed in circulating CD8+ T cells from patients with severe asthma²¹. Whilst very

little is known about the role of lncRNA in cystic fibrosis a recent study of ours has demonstrated an altered lncRNA expression profile in the bronchial epithelium of adults with CF when compared to non-CF controls ⁵².

4. Platforms for detection and quantification of non-coding RNA expression

Traditional methods for detection of non-coding RNA include *in-situ* hybridisation and northern blotting, and these are still very much applicable to the detection of ncRNA, however this review will focus on newer, higher throughput approaches (Figure 1. Steps in high-throughput profiling of ncRNA expression and some important considerations). Profiling of non-coding RNA, like mRNA, can essentially be categorised into three approaches, depending on the technology used; hybridisation, qPCR and RNA sequencing (RNA-seq). Although each method has its own benefits they also have significant limitations. Such limitations include the number of RNAs that can be analysed simultaneously; the source, quality and amount of RNA required; sensitivity and specificity of detection; and the requirement for prior knowledge of targets ^{53,54}. In comparison to mRNA species, primer and probe design for quantification of ncRNA abundance is difficult. The small size of miRNA, presence of isomiRs and sequence similarity between families, as well as lack of polyadenylation contribute to the difficulty of expression profiling of these RNAs. Previously inadequate annotation was a limiting factor in the design of primers and probes for miRNA detection but this has been significantly improved with the growth of miRBase; the comprehensive public repository of miRNA sequences ⁵⁵. Therefore, there has been a considerable increase in the choice of products available for detection of these small miRNAs. On the other hand, inadequate annotation is probably the greatest current limitation of lncRNA assay design, although this is rapidly improving. Detection of lncRNAs can be quite difficult, but this varies from transcript-to-transcript. lncRNAs are often found in intergenic regions or ‘gene deserts’ with high GC content, many are not polyadenylated, and some are antisense to protein coding genes with lower abundance, meaning strand specific PCR amplification is required ⁵⁶.

4.1 Hybridisation

Microarrays are the classical example of a hybridisation based approach to ncRNA expression analysis. Microarrays utilise the hybridisation of fluorescently tagged ncRNAs (be they miRNA or lncRNA) within a sample to glass printed DNA-based capture probes, followed by laser scanning of the slide and quantification of fluorescence associated with

each spot. A limitation of this method when examining miRNAs is that their small size means widely varying melting temperatures (T_m) of each probe can reduce specificity and/or sensitivity, especially for low GC content miRNAs⁵⁷. This issue may be overcome through the use of Locked Nucleic Acid (LNA) probes, which contain a synthetic DNA/RNA analogue which significantly increases the T_m ⁵⁸. A large number of microarray platforms exist, with differences ranging from probe design and labelling methods, to surface chemistry and printing technology⁵⁹. A major advantage of microarrays is their reproducibility, flexibility and ability to simultaneously detect a large number of ncRNAs. However, they do suffer from a limited linear range of quantification (typically over 2-3 logs), imperfect sensitivity; particularly for highly similar miRNA sequences, and they cannot be used for absolute quantification. The *NanoString nCounter* technology has been a recent innovation in RNA profiling based on hybridisation⁶⁰. This technology involves the use of colour-coded tags coupled to the reporter probe, and through permutations in the order of fluorochromes a molecular barcode is created for each ncRNA. This technology has been adapted for miRNAs and, more recently lncRNAs. However, this system has limitations when low-abundance RNA samples are used, such as serum/plasma. There is a requirement for 100 ng of concentrated RNA (33 ng/ μ l) and thus an additional concentration step may be required, and/or larger volumes of serum are recommended (400 μ l or more).

4.2 qRT-PCR

qRT-PCR is probably the easiest method for detection of ncRNA, and most laboratories have the equipment/hardware for performing PCR, especially those already examining mRNA expression. Analysis of ncRNA expression can be performed using either SYBR® Green or TaqMan® approaches, and various vendors offer pre-designed assays for miRNA and lncRNA.

Due to their small nature, miRNAs are difficult to detect by PCR, however various approaches have overcome this difficulty. TaqMan® miRNA assays contain a unique stem-loop RT primer to allow for reverse transcription of only mature miRNAs with high specificity. Reverse transcription using these primers results in a mature miRNA/primer chimera which extends the 3' end of the miRNA, creating a longer product for specific probes and primers to anneal during PCR⁶¹. Life Technologies have been the leader in the field of miRNA qPCR profiling, with microfluidic based *TaqMan® Low Density Array* (TLDA) cards, and more recently the adaptation of the nanofluidic based *OpenArray®* (OA)

system for miRNA profiling. Advances in nano- and micro-fluidic engineering has enabled a larger number of reactions to be performed in smaller reaction volumes⁶². Essentially, these fluidic platforms allow for multiplexing of TaqMan® miRNA assays. TLDA enables the profiling of 384 miRNAs per card and two cards (Cards A & B) cover 754 of the Sanger miRBase v14 annotated miRNAs. Card A allows for the profiling of well characterised miRNAs, with Card B containing assays for more recently discovered miRNAs. Therefore, this technology has been medium-throughput at best, but Life Technologies' OA system is a more recently devised high-throughput approach. OA allows for profiling of 754 miRNAs per sample, with the ability to profile up to 12 samples in a single run.

Other manufacturers have exploited the SYBR® Green route for miRNA profiling. The Qiagen SYBR® Green based *miScript PCR* system focuses on polyadenylation of small RNAs (including pre-miRNAs and other RNAs), and subsequent reverse transcription of mature miRNAs using an oligo-dT primer with a common tag sequence. The specificity is controlled by the proprietary dual buffer system used by the RT step⁶¹. PCR amplification proceeds with a universal reverse primer (utilising the tag sequence site) and a miRNA-specific forward primer, followed by SYBR® Green chemistry based detection and quantification⁵⁶. The Exiqon based miRCURY LNA™ PCR is also a SYBR® Green based approach, which uses polyadenylation of small RNAs followed by RT. However, specificity for miRNA amplification is achieved using LNA-enhanced forward and reverse primers. Both of these SYBR-based approaches are in essence, medium-throughput, being 384-well formats or lower.

As already discussed, there are some challenges posed to the expression profiling of lncRNA, when compared to mRNA. However, many manufacturers are allocating resources to improve the large scale quantification of lncRNA. For example, Life Technologies now have pre-designed TaqMan® assays for over 15,000 human ncRNAs, however these are currently only provided in individual assay format, and there is no option for multiplexing of yet. WaferGen have introduced a PCR array for lncRNA based on its SmartChip panel technology which contain over 1,700 triplicate, predisposed PCR assays that have been rigorously validated. Qiagen have extrapolated their RT² PCR array platform to include 100-well ring disc or 96 and 384 well plate formats for the quantification of multiple lncRNAs within a sample, but with a maximum of 84 lncRNAs detected per sample. Similarly, Systems Biosciences have developed 96-well plate PCR array formats for quantification of various well characterised lncRNAs. Clearly, high-throughput analysis of lncRNA expression

by qRT-PCR is lagging behind that of mRNA or miRNA, but is likely to improve with developments in annotation and characterisation of lncRNAs.

4.3 Next-generation sequencing/RNA-seq

RNA-seq is a relatively recent approach to transcriptomics and relies on Next-Generation Sequencing (NGS) technologies, which are excellently reviewed elsewhere^{63,64}. The major advantage of NGS is that prior knowledge of target ncRNA is not required, nor are specific primers or probes needed. Several manufacturers offer high-throughput NGS platforms, with varying technologies. RNA-seq offers many advantages over microarray approaches and is currently the best platform for both miRNA and lncRNA discovery. With single nucleotide resolution, RNA-seq works on a genome-wide scale, allowing for the precise identification of and quantification of RNAs. Compared with probe-based microarrays, RNA-seq has a much larger dynamic range (potentially infinite, depending on sufficient sequencing depth) and offers the ability to detect transcripts with very low expression⁶⁵. However, this approach does suffer from some limitations. It is relatively expensive and the resources required for downstream data analysis can be quite significant. Nonetheless, RNA-seq permits unbiased global ncRNA detection and expression quantification. It is important to decide whether to use total or poly-adenylated RNA before beginning RNA-seq, as the presence of ribosomal and transfer RNA can severely reduce cDNA library diversity during amplification⁶⁵. Some studies^{66,67} have suggested that depletion of ribosomal RNA may be superior in producing reliable expression data, as opposed to using polyadenylated RNA, given the high percentage of RNA transcripts that are not poly-adenylated (estimated at 40%⁶⁸). The sequencing approach also depends on the nature of the RNA species examined; i.e. whether miRNA or lncRNA are profiled. Generally, small RNAs such as miRNAs can be directly sequenced by adapter ligation, but larger RNAs require fragmentation to become compatible with most RNA-seq technologies⁶³.

4.4 Comparisons between platforms

Generally speaking, microarray technology is a very cost effective means of high throughput profiling of ncRNA expression. In addition, when comparing to low sequencing coverage RNA-seq, arrays may be superior in terms of having lower technical variation and better detection sensitivity for low-abundance transcripts which is a prominent feature of lncRNAs⁶⁹. RNA-seq with sufficient coverage and depth has the advantage of *de novo* ncRNA discovery, however as some ncRNA are in low abundance, higher costs of RNA-seq may be

envisaged due to higher depth of sequencing required to accurately detect these transcripts. Nevertheless, RNA-seq costs have reduced significantly in recent years and further reductions in cost and improvements in technology will facilitate more detailed RNA-seq studies. RNA-seq and microarray have the advantage over RT-PCR in terms of throughput, especially when examining lncRNA expression, however RT-PCR is still seen as a ‘gold-standard’ in expression studies, and is particularly useful in the validation of high-throughput data. Studies comparing the performance of these different platforms with respect to analysis of non-coding RNA expression are lacking, particularly for the more under-studied lncRNA. One recent study that comprehensively compared several platforms for miRNA expression profiling has concluded that each method has its strengths and weaknesses, which should help guide informed selection of the correct platform for particular study goals ⁷⁰, and this may also be applicable to lncRNA profiling. In addition, the use of in situ hybridisation (ISH) techniques must also be mentioned. This highly sensitive approach is based upon the detection of specific RNAs *via* hybridisation of a complementary (probe) strand to the sequence of interest in tissue or cell preparations. The ISH method, usually coupled with a fluorescent (FISH) or chromogenic (CISH) readout allows for the analysis at a single cell or subcellular level. Although this currently is a semi-quantitative and low throughput method, it is powerful in the fact that it allows for localisation of signals to specific cell-types. Interestingly, this situation is improving, particularly since the introduction of LNAs. For example, significant progress has been made quite recently, allowing for higher throughput when Battich and colleagues combined branched DNA technology with automated liquid handling, high-content imaging and quantitative image analysis and showed that this allows highly reproducible quantification of transcript abundance in thousands of single cells at single-molecule resolution ⁷¹.

4.5 Data interpretation and normalisation

Whatever profiling technology is utilised, particular attention must be placed on interpretation of biological meaning of the large amounts of data that ncRNA profiling generates. Raw data must be processed and a quality assessment performed. Special consideration must be given to data normalisation, and this can be difficult for samples with low RNA concentrations such as serum/plasma. This step is crucial in order to remove variation between samples which is not due to biological conditions studied ⁷² in order to improve detection of true biological differences. For qRT-PCR data, the most common approach is the use of endogenous control RNAs that are found to be relatively invariant

between samples in a given experiment⁷³. A second strategy involves a global measure of expression and has been used in large-scale qRT-PCR and, particularly microarray profiling datasets, where large numbers of RNAs are profiled simultaneously. Using this approach one assumes that although individual RNA levels may vary, the overall pattern of expression is expected to be invariable. Variations such as global geometric mean expression, Loess, quantile normalisation and others have been utilised in this approach, and this approach appears to perform better than when using endogenous controls^{74,75}. Another approach uses synthetic control RNAs, generally from a different species, which are spiked in at a known concentration into each RNA sample. This provides excellent quality control for various aspects of technical variation, for example reverse transcription for RT-PCR profiling, but it requires that assays for the spiked in controls be available on the platform used, which is not always the case. However, for samples with low RNA concentrations where variable amounts of RNA are used for profiling the spike-in method is unreliable, and therefore better used as an extraction efficiency tool⁶¹. Normalisation methods for RNA-seq data are less well developed. Many existing normalisation methods are insufficient at removing RNA composition bias and other unwanted complex technical effects, and efforts are underway to achieve an improvement in RNA-seq normalisation^{76,77}.

5. Properties of biomarkers

Biomarkers are important diagnostic tools reflecting health and disease that can be objectively measured, preferably in a minimally-invasive manner. There is an unmet clinical need for biomarkers of chronic inflammatory lung diseases that can provide earlier detection, ensure optimum treatment and refine prognosis. A common problem is that affected tissue is not easily accessible. Therefore specific, inexpensive biomarkers must be identified that are accessible using non-invasive protocols. Extracellular ncRNA in biofluids may represent suitable biomarkers for many diseases. In blood, over 100 miRNAs are normally detectable. Their presence is thought due to both passive and active mechanisms, including release from damaged cells and secretion. miRNAs are particularly attractive as biomarkers because of their stability in biofluids, resistance against freeze-thaw and pH, tissue-specific expression, conserved sequences, and roles in many human diseases. In cancer and other diseases, biofluid miRNAs display particular specificity and sensitivity as biomarkers. Many lncRNAs also have the characteristics required for a good biomarker.

6. Pre-analytical considerations for ncRNA profiling

6.1 Tissue versus circulating RNA

RNA can be easily extracted from lung tissue such as biopsy or brushings and also exists in relatively high concentrations in broncho-alveolar lavage fluid (BALF) ⁷⁸. BALF ncRNA analyses have proven valuable in the study of lung disease ^{79–81}. However as described earlier, biomarkers should be easily accessible using non-invasive methods. Therefore, there is a need for the discovery of biomarkers from fluids such as blood, saliva/sputum and urine. Although relatively high concentrations of RNA can be found in saliva/sputum, lower concentrations are found in blood and these are lower still in urine (Table 1). Also, before embarking upon a lung disease biomarker study one must consider the appropriate source, and whether cell-free RNA is required. For instance, although high relative concentrations of RNA exist in BALF and sputum, much of this is derived from the lysis of cells during the extraction procedure, and not cell-free RNA. This is also the case for urine, where cellular RNA contributes greatly to the total RNA found in urine. Given the difficulties in obtaining and preparing sputum for isolation of RNA, especially from children or those with low lung function ⁸², and the low RNA concentration in urine, circulating RNA is perhaps the most promising candidate for biomarker discovery. ncRNAs present in blood also have potential, given that these may be minimally invasive biomarkers of a range of lung diseases. When using blood one must first decide empirically whether serum or plasma should be used. Although the RNA profile of serum and plasma is generally similar, significant differences have been shown ⁸³. The primary difference between serum and plasma is the absence or presence, respectively, of fibrinogen and clotting factors ⁶¹. Plasma is the supernatant obtained after centrifugation of blood which has been collected in the presence of an anticoagulant. It is routinely collected in various tubes containing anticoagulants such as EDTA, sodium citrate and heparin. However, the selection of anticoagulant containing tubes must be carefully considered when examining ncRNA levels as some of these anticoagulants inhibit enzymes required for reverse transcription and PCR ^{84–86}. For example, citrate and heparin have inhibitory effects on RT-PCR, and therefore EDTA should be used. This also appears to be the case with RNA-seq, as heparin has been shown to inhibit reverse transcription and the effect of citrate is as of yet unknown.

Consideration must be given to other factors that can contribute to RNA quantification. The timing of food intake prior to sample collection should be recorded and consistency should be

maintained across sampling as regards time of day. For instance, high levels of lipoproteins can affect RNA extraction efficiency⁸⁷. Phlebotomy variables can also affect miRNA and lncRNA quantification, generally through affecting lysis of erythrocytes (and possibly leukocytes) and activation of platelets by using small gauge needles⁸⁸, as these cellular components can contribute to and skew the ncRNA profile of samples. Additionally, the time between sample collection and sample processing should be standardised. Ideally, protocols should be consistent in sample acquisition, storage, processing and quality control checks (particularly for haemolysis⁸⁹) and however these are currently under-reported in the literature.

Furthermore, if the aim of the study is the identification of circulating miRNA profiles specifically, one must bear in mind that miRNA may be transported in complex with proteins such as argonaute-2⁹⁰; lipoproteins⁹¹; or cellular fragments such as extracellular vesicles and exosomes^{92,93} and this must be considered when extracting RNA. The most common approach to RNA isolation from cell lines or tissue is the use of phenol:chloroform extraction and most current techniques are based on this approach. A difficulty in RNA isolation from fluids such as serum and plasma is the fact that these highly aqueous samples have very high concentrations of protein and extraction therefore involves the scaling up of the lysis reagent to specimen ratio several fold^{61,87}. Several other products are also available for RNA isolation from biofluids, with many combining phenol:chloroform extraction with an additional silica column-based RNA purification step, which may be particularly useful in removing PCR inhibitors. Alternatively, RNA may be separated based on size using polyacrylamide gel electrophoresis, and this step is particularly useful in the generation of appropriate cDNA libraries for RNA-seq. Finally, once the desired RNA has been obtained, its quality must be assessed. Spectrophotometric measurements are simple and cheap methods for determination of RNA concentration and quality control e.g. evaluation of phenol and protein contamination. However, these QC measures suffer from low sensitivity⁹⁴ and are not particularly useful with samples of low RNA concentration. Electrophoretic analysis of RNA integrity is useful and automated capillary electrophoresis instruments are widely available, such as the Agilent Bioanalyzer and the Bio-Rad Experion; but these are significantly more expensive as compared to spectrophotometric methods. Finally, as alluded to earlier, spike-in of synthetic RNAs can be useful in determining RNA extraction efficiency.

7. Expert commentary

With the current enthusiasm within the field of ncRNA, our understanding of the epigenetic arena is increasing. However, as we discover more detail about ncRNAs, how they function and their roles in health and disease, we realise that we are only at the very beginning of this process. Much more work is required to truly understand their roles, especially in the context of human disease. Nonetheless, we can benefit from the functional or epiphenomenal presence of these molecules in biofluids without necessarily knowing their exact functions. Certain characteristics of ncRNA, circulating or otherwise, underscore their potential use as diagnostic or prognostic non-invasive biomarkers of human disease, or response to therapy. With this in mind, their use as biomarkers deserves fuller interrogation, as current biomarkers of lung disease in particular are lacking. In order to do this, improvements in high throughput discovery and profiling technologies must be married with a simultaneous improvement in computational methodologies, to better identify true ncRNA biomarkers. Careful consideration must be taken with respect to the design of biomarker discovery and validation studies; from sampling to processing through to profiling, validation and clinical interpretation. Key aspects driving these studies must be accuracy, reliability and reproducibility. Great care must be taken in developing optimized methods for miRNA extraction, analysis and detection and it will be important that researchers in the field reach consensus on detection methods. International efforts should be undertaken via task forces or research networks to develop standard operating procedures to optimize biofluid processing, ensure quality control and include standardisation of assays. Beyond that extensive validation studies will need to be carried out in geographically distinct cohorts to truly determine the diagnostic/prognostic potential of any individual ncRNA panel.

8. Five-year view

Emphasis is ever-increasingly being placed on the study of non-coding RNAs by the scientific community at large, with an explosion in the number of studies and data produced pertaining to these intriguing biomolecules in the last few years. Their involvement in human disease has become increasingly apparent, and their use as biomarkers of disease is increasingly being investigated. Although up until now miRNAs have been extensively investigated, focus is increasing on other ncRNAs; and in particular long non-coding RNAs. The presence of miRNAs and lncRNAs in body fluids and in particular in blood means that

these have extremely high potential as clinical markers of lung disease, as diagnostic and/or prognostic tools, and also as predictors for assessing response to therapy. In fact, some diagnostic tests employing ncRNA are already commercially available^{13,95-97}, with many others expected to follow.

Most research to date on miRNA as biomarkers has focused on blood as a source (serum/plasma). This of course has advantages in terms of providing sufficient RNA. However, the risk of haemolysis and other cell-derived contaminants requires stringent quality controls. Additionally, the requirement for venepuncture can be limiting particular where repeated sampling is needed, where out-of-hospital care might require use of blood-based ncRNA biomarker profiling, and particularly with vulnerable groups such as children and infants. There is, therefore, significant interest and opportunity in exploring other biofluid sources for miRNA biomarkers.

More effort is required to evaluate novel sources of ncRNA biomarkers e.g. saliva, urine, that are more accessible from young patients and potentially more cost effective. Thus it is essential that we (1) develop optimized techniques for miRNA and lncRNA profiling of novel biofluids, (2) study normal variation and stability of ncRNA in these biofluids in relation to standards (e.g. plasma); and (3) assess the potential of these novel biofluid sources for future potential microfluidic diagnostic point-of-care/lab-on-a-chip devices.

Other challenges yet to be fully explored and discussed in a transdisciplinary manner are the most appropriate and robust biostatistical and bioinformatics methodologies to apply when analysing ncRNA data (e.g. normalisation, correlative statistics). Given the global interest in ncRNA biomarkers in lung disease and beyond, it is highly likely that the number of approved clinical biomarker tests and standardised methods for ncRNA biomarker quantification will be developed in the near future.

9. Key Issues

- miRNA and lncRNA expression is altered in COPD, asthma, CF and IPF. These non-coding RNAs are detectable in various biofluids and are potential clinical biomarkers for diagnosis, prognosis and assessment of response to therapy.
- The major high throughput approaches for profiling of non-coding RNA are hybridisation (microarrays), qRT-PCR and RNA sequencing (RNA-seq/NGS). Each method has its own benefits and limitations which can affect the choice of method when embarking on a biomarker study.
- qRT-PCR is the easiest and most widely used method for detection of non-coding RNA, with high-throughput analysis of lncRNA expression by qRT-PCR lagging behind that of miRNA. qRT-PCR remains the 'gold-standard' in expression studies and is useful in the validation of high-throughput data.
- Microarrays are reproducible, flexible and can simultaneously detect a large number of non-coding RNAs. Although cost effective, their limitations include a narrow linear range of quantification, imperfect sensitivity for highly similar miRNA sequences, and they cannot be used for absolute quantification.
- The major advantage of RNA-seq is that prior knowledge of target non-coding RNA is not required, nor are specific primers or probes needed. RNA-seq has a large dynamic range and can detect transcripts with very low expression, however the method is expensive.
- Data normalisation is essential to remove variation between samples not due to biological conditions and can be difficult for samples with low RNA concentrations such as serum/plasma. The two most common normalisation methods for large-scale qRT-PCR and microarray profiling are the use of reliable endogenous control RNAs or global normalisation utilising the entire dataset.
- Various biospecimens can be used for lung biomarker studies including whole blood, serum, plasma, peripheral blood cells, bronchoalveolar lavage fluid, bronchial brushings, sputum, exhaled breath condensates or urine. Ideally biomarkers should be easily accessible using non-invasive methods. Each biofluid has characteristics that can affect non-coding RNA detection.
- Accuracy, reliability and reproducibility of the methodologies are tantamount within this field. Current emphasis should focus on developing standard operating

procedures and standardisation of assays coupled with the most appropriate and robust data interpretation approaches.

Financial & competing interests disclosure

The authors were supported by the Royal College of Surgeons in Ireland Seed Fund B 2014. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

10. References

1. International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
2. Djebali, S., Davis, C. A., Merkel, A., *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

*** This paper from the ENCODE project reports evidence that three-quarters of the human genome is capable of being transcribed and that genes are highly interlaced with overlapping transcripts that are synthesized from both DNA strands.*
3. Rana, T. M. Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* **8**, 23–36 (2007).
4. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).

** This paper is the first report of microRNA function, which was *lin-4* and was initially discovered for its role in regulating developmental timing in nematodes.*
5. Lytle, J. R., Yario, T. A. & Steitz, J. A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104**, 9667–9672 (2007).
6. Hausser, J., Syed, A. P., Bilen, B. & Zavolan, M. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Res.* **23**, 604–615 (2013).
7. Liu, J. Control of protein synthesis and mRNA degradation by microRNAs. *Curr. Opin. Cell Biol.* **20**, 214–221 (2008).
8. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of MicroRNA–Target Recognition. *PLoS Biol.* **3**, e85 (2005).

*** An excellent article examining and establishing some requirements for microRNA::target binding and specificity, and provides an insight into the scope of miRNA control, providing evidence that an average miRNA has approximately 100 target sites.*
9. Booton, R. & Lindsay, M. A. Emerging role of microRNAs and long noncoding RNAs

in respiratory disease. *Chest* **146**, 193–204 (2014).

10. Szafranski, P., Dharmadhikari, A. V, Brosens, E., *et al.* Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder. *Genome Res.* **23**, 23–33 (2013).
11. Williams, A. E., Moschos, S. A., Perry, M. M., Barnes, P. J. & Lindsay, M. A. Maternally imprinted microRNAs are differentially expressed during mouse and human lung development. *Dev. Dyn.* **236**, 572–580 (2007).
12. Harris, K. S., Zhang, Z., McManus, M. T., Harfe, B. D. & Sun, X. Dicer function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci U S A* **103**, 2208–2213 (2006).
13. Gilad, S., Lithwick-Yanai, G., Barshack, I., *et al.* Classification of the four main types of lung cancer using a microRNA-based diagnostic assay. *J. Mol. Diagn.* **14**, 510–7 (2012).
14. Who. WHO | Burden of COPD. *Online 1* (2013), at <http://www.who.int/respiratory/copd/burden/en/index.html>
15. Vestbo, J., Hurd, S. S., Agustí, A. G., *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease GOLD executive summary. *Am. J. Respir. Crit. Care Med.* **187**, 347–365 (2013).
16. Akbas, F., Coskunpinar, E., Aynaci, E., Oltulu, Y. M. & Yildiz, P. Analysis of serum micro-RNAs as potential biomarker in chronic obstructive pulmonary disease. *Exp. Lung Res.* **38**, 286–94 (2012).
17. Soeda, S., Ohyashiki, J. H., Ohtsuki, K., *et al.* Clinical relevance of plasma miR-106b levels in patients with chronic obstructive pulmonary disease. *Int. J. Mol. Med.* **31**, 533–9 (2013).
18. Xie, L., Wu, M., Lin, H., *et al.* An increased ratio of serum miR-21 to miR-181a levels is associated with the early pathogenic process of chronic obstructive pulmonary disease in asymptomatic heavy smokers. *Mol. Biosyst.* **10**, 1072–81 (2014).
19. Comer, B. S., Ba, M., Singer, C. A. & Gerthoffer, W. T. Epigenetic targets for novel therapies of lung diseases. *Pharmacol. Ther.* **147**, 91–110 (2015).
20. Wenzel, S. Severe asthma in adults. *Am. J. Respir. Crit. Care Med.* **172**, 149–60 (2005).

21. Tsitsiou, E., Williams, A. E., Moschos, S. A., *et al.* Transcriptome analysis shows activation of circulating CD8⁺ T cells in patients with severe asthma. *J. Allergy Clin. Immunol.* **129**, 95–103 (2012).
22. Panganiban, R. P. L., Pinkerton, M. H., Maru, S. Y., *et al.* Differential microRNA expression in asthma and the role of miR-1248 in regulation of IL-5. *Am. J. Clin. Exp. Immunol.* **1**, 154–65 (2012).
23. Sinha, A., Yadav, A. K., Chakraborty, S., *et al.* Exosome-enclosed microRNAs in exhaled breath hold potential for biomarker discovery in patients with pulmonary diseases. *J. Allergy Clin. Immunol.* **132**, 219–22 (2013).
24. Pinkerton, M., Chinchilli, V., Banta, E., *et al.* Differential expression of microRNAs in exhaled breath condensates of patients with asthma, patients with chronic obstructive pulmonary disease, and healthy adults. *J. Allergy Clin. Immunol.* **132**, 217–9 (2013).
25. Roff, A. N., Craig, T. J., August, A., Stellato, C. & Ishmael, F. T. MicroRNA-570-3p regulates HuR and cytokine expression in airway epithelial cells. *Am. J. Clin. Exp. Immunol.* **3**, 68–83 (2014).
26. Sheridan, M. B., Hefferon, T. W., Wang, N., *et al.* CFTR transcription defects in pancreatic sufficient cystic fibrosis patients with only one mutation in the coding region of CFTR. *J. Med. Genet.* **48**, 235–41 (2011).
27. Oglesby, I. K., Bray, I. M., Chotirmall, S. H., *et al.* miR-126 is downregulated in cystic fibrosis airway epithelial cells and regulates TOM1 expression. *J. Immunol.* **184**, 1702–9 (2010).
28. Ramachandran, S., Karp, P. H., Osterhaus, S. R., *et al.* Post-transcriptional regulation of cystic fibrosis transmembrane conductance regulator expression and function by microRNAs. *Am J Respir Cell Mol Biol* **49**, 544–551 (2013).
29. Gillen, A. E., Gosalia, N., Leir, S. H. & Harris, A. MicroRNA regulation of expression of the cystic fibrosis transmembrane conductance regulator gene. *Biochem. J.* **438**, 25–32 (2011).
30. Megiorni, F., Cialfi, S., Dominici, C., Quattrucci, S. & Pizzuti, A. Synergistic Post-Transcriptional Regulation of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) by miR-101 and miR-494 Specific Binding. *PLoS One* **6**, e26601 (2011).

31. Hassan, F., Nuovo, G. J., Crawford, M., *et al.* MiR-101 and miR-144 Regulate the Expression of the CFTR Chloride Channel in the Lung. *PLoS One* **7**, e50837 (2012).
32. Ramachandran, S., Karp, P. H., Jiang, P., *et al.* A microRNA network regulates expression and biosynthesis of wild-type and DeltaF508 mutant cystic fibrosis transmembrane conductance regulator. *Proc Natl Acad Sci U S A* **109**, 13362–13367 (2012).
33. Oglesby, I. K., Chotirmall, S. H., McElvaney, N. G. & Greene, C. M. Regulation of cystic fibrosis transmembrane conductance regulator by microRNA-145, -223, and -494 is altered in DeltaF508 cystic fibrosis airway epithelium. *J. Immunol.* **190**, 3354–3362 (2013).
34. Fan, Y., Zhou, Y., Zhou, X., *et al.* MicroRNA 224 Regulates Ion Transporter Expression in Ameloblasts To Coordinate Enamel Mineralization. *Mol. Cell. Biol.* **35**, 2875–90 (2015).
35. Viart, V., Bergougnoux, A., Bonini, J., *et al.* Transcription factors and miRNAs that regulate fetal to adult CFTR expression change are new targets for cystic fibrosis. *Eur. Respir. J.* (2014). doi:10.1183/09031936.00113214
36. McKiernan, P. J., Cunningham, O., Greene, C. M. & Cryan, S.-A. A. Targeting miRNA-based medicines to cystic fibrosis airway epithelial cells using nanotechnology. *Int. J. Nanomedicine* **8**, 3907–3915 (2013).
37. Kumar, P., Bhattacharyya, S., Peters, K. W., *et al.* miR-16 rescues F508del-CFTR function in native cystic fibrosis epithelial cells. *Gene Ther.* 1–9 (2015). doi:10.1038/gt.2015.56
38. Fabbri, E., Borgatti, M., Montagner, G., *et al.* Expression of microRNA-93 and Interleukin-8 during *Pseudomonas aeruginosa*-mediated induction of proinflammatory responses. *Am J Respir Cell Mol Biol* **50**, 1144–1155 (2014).
39. Oglesby, I. K., Vencken, S. F., Agrawal, R., *et al.* miR-17 overexpression in cystic fibrosis airway epithelial cells decreases interleukin-8 production. *Eur. Respir. J.* (2015). doi:10.1183/09031936.00163414
40. Zhang, P.-X., Cheng, J., Zou, S., *et al.* Pharmacological modulation of the AKT/microRNA-199a-5p/CAV1 pathway ameliorates cystic fibrosis lung hyperinflammation. *Nat. Commun.* **6**, 6221 (2015).
41. Pandit, K. V., Corcoran, D., Yousef, H., *et al.* Inhibition and role of let-7d in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **182**, 220–9 (2010).

42. Liu, G., Friggeri, A., Yang, Y., *et al.* miR-21 mediates fibrogenic activation of pulmonary fibroblasts and lung fibrosis. *J. Exp. Med.* **207**, 1589–97 (2010).
43. Cushing, L., Kuang, P. P., Qian, J., *et al.* miR-29 is a major regulator of genes associated with pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **45**, 287–94 (2011).
44. Lino Cardenas, C. L., Henaoui, I. S., Courcot, E., *et al.* miR-199a-5p Is upregulated during fibrogenic response to tissue injury and mediates TGFbeta-induced lung fibroblast activation by targeting caveolin-1. *PLoS Genet.* **9**, e1003291 (2013).
45. Huang, C., Yang, Y. & Liu, L. Interaction of long non-coding RNAs and microRNAs in the pathogenesis of idiopathic pulmonary fibrosis. *Physiol. Genomics* *physiolgenomics.00064.2015* (2015). doi:10.1152/physiolgenomics.00064.2015
46. Yang, G., Yang, L., Wang, W., *et al.* Discovery and validation of extracellular/circulating microRNAs during idiopathic pulmonary fibrosis disease progression. *Gene* **562**, 138–44 (2015).
47. Li, P., Li, J., Chen, T., *et al.* Expression analysis of serum microRNAs in idiopathic pulmonary fibrosis. *Int. J. Mol. Med.* **33**, 1554–62 (2014).
48. Thai, P., Statt, S., Chen, C. H., *et al.* Characterization of a Novel Long Noncoding RNA, SCAL1, Induced by Cigarette Smoke and Elevated in Lung Cancer Cell Lines. *Am. J. Respir. Cell Mol. Biol.* (2013). at <<http://www.atsjournals.org/doi/abs/10.1165/rcmb.2013-0159RC#.VcpMMfIVjM0>>
49. De Smet, E. G., Mestdagh, P., Vandesompele, J., Brusselle, G. G. & Bracke, K. R. Non-coding RNAs in the pathogenesis of COPD. *Thorax* **70**, 782–791 (2015).
50. Bi, H., Zhou, J., Wu, D., *et al.* Microarray analysis of long non-coding RNAs in COPD lung tissue. *Inflamm. Res.* **64**, 119–26 (2015).
51. Perry, M. M., Tsitsiou, E., Austin, P. J., *et al.* Role of non-coding RNAs in maintaining primary airway smooth muscle cells. *Respir Res* **15**, 58 (2014).
52. McKiernan, P. J., Molloy, K., Cryan, S. A., McElvaney, N. G. & Greene, C. M. Long noncoding RNA are aberrantly expressed in vivo in the cystic fibrosis bronchial epithelium. *Int J Biochem Cell Biol* **52**, 184–191 (2014).
53. van Rooij, E. The art of microRNA research. *Circ. Res.* **108**, 219–34 (2011).

54. Lopez, J. P., Diallo, A., Cruceanu, C., *et al.* Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med. Genomics* **8**, 35 (2015).
 55. Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res.* **32**, D109–11 (2004).
 56. Perdomo, C., Campbell, J. & Schembri, F. in *Non-coding RNAs and Cancer* (ed. Fabbri, M.) 25–44 (Springer New York). doi:10.1007/978-1-4614-8444-8_3 (2015)
 57. Guerau-de-Arellano, M., Alder, H., Ozer, H. G., Lovett-Racke, A. & Racke, M. K. miRNA profiling for biomarker discovery in multiple sclerosis: from microarray to deep sequencing. *J. Neuroimmunol.* **248**, 32–9 (2012).
 58. Castoldi, M., Schmidt, S., Benes, V., *et al.* A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). *RNA* **12**, 913–20 (2006).
 59. Git, A., Dvinge, H., Salmon-Divon, M., *et al.* Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* **16**, 991–1006 (2010).
 60. Geiss, G. K., Bumgarner, R. E., Birditt, B., *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–25 (2008).
 61. Moldovan, L., Batte, K. E., Trgovcich, J., *et al.* Methodological challenges in utilizing miRNAs as circulating biomarkers. *J. Cell. Mol. Med.* **18**, 371–90 (2014).
- * This article reviews methodological considerations for miRNA biomarker studies, but these considerations can also be applied to other ncRNA types.*
62. Farr, R. J., Januszewski, A. S., Joglekar, M. V, *et al.* A comparative analysis of high-throughput platforms for validation of a circulating microRNA signature in diabetic retinopathy. *Sci. Rep.* **5**, 10375 (2015).
 63. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
 64. Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
 65. McHale, C. M., Zhang, L., Thomas, R. & Smith, M. T. Analysis of the transcriptome in molecular epidemiology studies. *Environ. Mol. Mutagen.* **54**, 500–517 (2013).

66. Cui, P., Lin, Q., Ding, F., *et al.* A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**, 259–265 (2010).
67. Huang, R., Jaritz, M., Guenzl, P., *et al.* An RNA-Seq Strategy to Detect the Complete Coding and Non-Coding Transcriptome Including Full-Length Imprinted Macro ncRNAs. *PLoS One* **6**, e27288 (2011).
68. Cheng, J., Kapranov, P., Drenkow, J., *et al.* Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution. *Science* **308**, 1149–1154 (2005).
69. Du, Z., Fei, T., Verhaak, R. G. W., *et al.* Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.* **20**, 908–13 (2013).
70. Mestdagh, P., Hartmann, N., Baeriswyl, L., *et al.* Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* **11**, 809–815 (2014).
71. Battich, N., Stoeger, T. & Pelkmans, L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* **10**, 1127–33 (2013).
72. Pritchard, C. C., Cheng, H. H. & Tewari, M. MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.* **13**, 358–369 (2012).

*** An excellent thorough review of factors to be considered when profiling the expression of microRNA across various platforms.*

73. Peltier, H. J. & Latham, G. J. Normalization of microRNA expression levels in quantitative RT-PCR assays: Identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA* **14**, 844–852 (2008).
74. Mestdagh, P., Van Vlierberghe, P., De Weer, A., *et al.* A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* **10**, R64 (2009).
75. Wylie, D., Shelton, J., Choudhary, A. & Adai, A. T. A novel mean-centering method for normalizing microRNA expression from high-throughput RT-qPCR data. *BMC Res. Notes* **4**, 555 (2011).
76. Meyer, S. U., Pfaffl, M. W. & Ulbrich, S. E. Normalization strategies for microRNA profiling experiments: a ‘normal’ way to a hidden layer of complexity? *Biotechnol. Lett.* **32**, 1777–88 (2010).

77. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
78. Weber, J. A., Baxter, D. H., Zhang, S., *et al.* The microRNA spectrum in 12 body fluids. *Clin. Chem.* **56**, 1733–41 (2010).
79. Huang, W., Hu, J., Yang, D., *et al.* Two microRNA panels to discriminate three subtypes of lung carcinoma in bronchial brushing specimens. *Am. J. Respir. Crit. Care Med.* **186**, 1160–7 (2012).
80. Levänen, B., Bhakta, N. R., Torregrosa Paredes, P., *et al.* Altered microRNA profiles in bronchoalveolar lavage fluid exosomes in asthmatic patients. *J. Allergy Clin. Immunol.* **131**, 894–903 (2013).
81. Tufman, A., Gamarra, F., Kiefl, R., *et al.* Feasibility of microRNA Measurements in Routine Bronchoalveolar Lavage (BAL) and Bronchial Washings in Patients With Lung Cancer and Nonmalignant Pulmonary Diseases. *Chest* **145**, 342A (2014).
82. Sagel, S. D., Chmiel, J. F. & Konstan, M. W. Sputum biomarkers of inflammation in cystic fibrosis lung disease. *Proc. Am. Thorac. Soc.* **4**, 406–17 (2007).
83. Wang, K., Yuan, Y., Cho, J.-H., *et al.* Comparing the MicroRNA spectrum between serum and plasma. *PLoS One* **7**, e41561 (2012).
84. Willems, M., Moshage, H., Nevens, F., Fevery, J. & Yap, S. H. Plasma collected from heparinized blood is not suitable for HCV-RNA detection by conventional RT-PCR assay. *J. Virol. Methods* **42**, 127–30 (1993).
85. Al-Soud, W. A. & Rådström, P. Purification and characterization of PCR-inhibitory components in blood cells. *J. Clin. Microbiol.* **39**, 485–93 (2001).
86. Garcia, M. E., Blanco, J. L., Caballero, J. & Gargallo-Viola, D. Anticoagulants Interfere with PCR Used To Diagnose Invasive Aspergillosis. *J. Clin. Microbiol.* **40**, 1567–1568 (2002).
87. Kroh, E. M., Parkin, R. K., Mitchell, P. S. & Tewari, M. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods* **50**, 298–301 (2010).
88. Witwer, K. W., Buzás, E. I., Bemis, L. T., *et al.* Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *Journal of Extracellular Vesicles* **2**, (2013).

89. Blondal, T., Jensby Nielsen, S., Baker, A., *et al.* Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods* **59**, S1–S6 (2013).
90. Arroyo, J. D., Chevillet, J. R., Kroh, E. M., *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5003–8 (2011).
91. Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D. & Remaley, A. T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* **13**, 423–433 (2011).
92. Mitchell, P. S., Parkin, R. K., Kroh, E. M., *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA* **105**, 10513–8 (2008).

** This article provides key evidence for the stability of endogenous microRNAs in blood and its fraction and therefore their usefulness as circulating, non-invasive biomarkers of disease.*

93. Ismail, N., Wang, Y., Dakhllallah, D., *et al.* Macrophage microvesicles induce macrophage differentiation and miR-223 transfer. *Blood* **121**, 984–95 (2013).
94. Glasel, J. A. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques* **18**, 62–3 (1995).
95. Meiri, E., Mueller, W. C., Rosenwald, S., *et al.* A second-generation microRNA-based assay for diagnosing tumor tissue origin. *Oncologist* **17**, 801–12 (2012).
96. Benjamin, H., Lebanony, D., Rosenwald, S., *et al.* A diagnostic assay based on microRNA expression accurately identifies malignant pleural mesothelioma. *J. Mol. Diagn.* **12**, 771–9 (2010).
97. van Gils, M. P. M. Q., Hessels, D., van Hooij, O., *et al.* The time-resolved fluorescence-based PCA3 test on urinary sediments after digital rectal examination; a Dutch multicenter validation of the diagnostic performance. *Clin. Cancer Res.* **13**, 939–43 (2007).

** A very interesting study of the diagnostic performance of the FDA-approved urinary test for the lncRNA PCA3 for the diagnosis of prostate cancer.*