Bench philosophy (34): Tools for miRNA analysis

Getting to know your favourite microRNA

Detailed information about protein coding genes is readily available from online resources. However, it is not always as easy to find out about a microRNA.

Many times I've been asked by a colleague, "What can you tell me about this microRNA I've found?" Fortunately, we now have access to many online resources with useful information about microRNAs (miRNAs). The aim of this short overview is to highlight some of these tools and how you can get the most out of them. But first things first: what is a microRNA?

MicroRNAs are single-stranded RNA sequences, of about 22 nucleotides in length, processed from longer transcripts. MicroRNAs are important regulators that repress the translation of messenger RNA transcripts. First discovered in 1993 in the round worm *Caenorhabditis elegans*, they were regarded as an oddity of this atypical species. Today, we know of more than 15,000 microRNAs in over 150 species. They are crucial regulators of gene expression, in particular during development. Importantly, microRNAs are also associated with human diseases.

First things first

When working with protein-coding genes, one must understand the nomenclature to distinguish between gene loci, transcripts and protein products. Likewise, there is a specific nomenclature for microRNAs we should be familiar with. Typically, any microRNA is named with the word "mir" followed by a number, for instance: mir-31 or mir-4991. There are a few exceptions such as let-7 or bantam (discovered before the naming system described became a standard). Additionally, the microRNA name is preceded by three letters that are specific for each species. For instance, *hsa-mir-31* is a human (*Homo sapiens*) microRNA locus, whilst *dme-mir-1* is from fly (*Drosophila melanogaster*). Very often, multiple microRNAs are evolutionarily related. We use a letter after the number to differentiate among multiple members of the same family, for example: *hsa-mir-33a* and *hsa-mir-33b*. Sometimes, two different

loci produce identical mature products. In this case, there is an additional number after the full name. For instance, both *hsa-mir-24-1* and *hsa-mir-24-2* produce the same final microRNA product: hsa-miR-24.

When the names detailed above refer to genomic loci, they should be written in italics. These loci encode transcripts or primary microRNAs. The precursor microRNA is formed by the processing of a hairpin-like structure within the primary transcript. The precursor is processed again to produce a double-stranded RNA molecule, the microRNA duplex. From this molecule, one of the strands is usually the functional microRNA and the other is often degraded. Sometimes, both strands become functional products. Mature sequences are tagged with the 'miR' label (note the capital 'R'). That means, *dme-mir-1* will produce the mature microRNA dme-miR-1. Also, it is good practice to add a tag to the name indicating, which strand the mature sequence comes from (either the 5' or 3' arm of the precursor). In our example, *dme-mir-1* locus will ultimately produce two microR-

NAs: dme-miR-1-5p and dme-miR-1-3p. Although this may seem a bit confusing at first, you should become familiar with this nomenclature to navigate throughout the different microRNA resources on the Internet.

Searching for miRNA informations

The first thing you should do is to search for your microRNA in miRBase (*www.mirbase.org*). miRBase is the official repository of microRNAs, where known microRNAs are named, stored and lovingly cared for. Enter the name of your microRNA and you'll get a bunch of useful information. You'll see the hairpin sequence of the precursor and the sequences of the mature microRNAs. Instead of the endless list of references you may get from a PubMed search, you will see the most relevant papers like those, in which the microRNA was described for the first time. There are also many links to external databases. A relatively recent feature of miRBase is the "deep sequencing" field, in which you can see for yourself the range of expressed short RNA sequences

Experts estimate that nearly one-third of the encoded genes in mammalian cells are regulated by miRNAs. Various bioinformatics databases, tools and algorithms may help to predict the sequences of miRNAs and their target genes.

from high-throughput experiments. Here you can find, among other things, the relative contribution of both arms of the precursor to the mature microRNA pool. The functional mature sequence will be represented by many more copies than the nonfunctional one, which is rapidly degraded by the cell. As we mentioned before, there are many instances, in which both strands are highly abundant (i.e., functional).

Another piece of useful information is the "gene family" field. If you click on the gene family link you will get a list of evolutionarily-related microRNAs. (If your *Drosophila* microRNA has a highly conserved copy in the human genome, you may boost your next grant application by emphasising that your fly microRNA is of clinical importance). From this option you can also retrieve a full sequence alignment and run your favourite phylogenetic analysis software. Please take into account that the evolutionary models used for protein-coding sequences may not always apply to microRNAs – but that's another topic.

Detailed miRNA expression patterns

Now that you're familiar with your microRNA, you probably want to know where (and when) it is expressed. You can get this information from the "deep sequencing" data available from miRBase. Another extremely useful database is the smirnaDB resource (*www.mirz.unibas.ch/cloningprofiles*), where you can find detailed expression patterns of microRNAs. SmirnaDB compiles information from different sources including diseased tissues. Therefore, you should carefully select the samples you want to use to avoid any bias. But that's what you also do when working with protein-coding gene expression databases, don't you?

Mature microRNAs physically bind to target transcripts by sequence complementarity (often in their 3' untranslated regions). MicroRNA target prediction algorithms exploit this characteristic. As you may guess, however, since microRNAs are short, target prediction is problematic. Indeed, different programmes give different predictions and no one can agree, which is the best approach. Two popular algorithms are DianaT-microT (*http://diana.cslab.ece.ntua.gr/DianaTools*) and TargetScan (*www.targetscan.org*). miRBase entries host links to these and other pre-calculated target predictions. You may find it useful to scan your microRNA using two or even more of these prediction programmes. It is recommended that you work with each list of predict-

ed targets separately, since a recent study shows that the combination of multiple algorithms may not lead to higher quality predictions.

Experimentally validated targets are also available for some microRNAs, although this information is limited and generally biased towards disease-associated microRNAs. Three good resources for validated targets are TarBase (*http://diana.cslab.ece.ntua.gr/DianaTools*), miRTarBase (*http://mirtarbase.mbc.nctu.edu.tw*) and miRecords (*http://mirecords.biolead.org*). These databases are extensive compilations collected from experimental research papers.

MicroRNAs have been associated with multiple human diseases, mainly to different types of cancer. If you're working with human microRNAs (or evolutionarily related sequences in other species), you may want to check miR2Disease (*www.mir2disease.org*). Here, you will find any disease, with which your microRNA has been associated and the reference papers that support this association. Needless to say, you should scrutinise the original papers to be sure what type of relationship exists between your microRNA and the given disease.

Now you have all the basic information you need to prepare your next experiment. For designing probes and primers you'll find the deep-sequencing information from miRBase useful. If you're planning to knock-down a microRNA, take into account that members of the same family could create cross-hybridisation problems. If you want to detect targets, restrict your experimental validation to transcripts with predicted target sites so you save time (and resources). Tissue expression information may also help to narrow down the list of potential targets. And make sure you use the appropriate nomenclature to discern between loci, precursors and the mature microRNAs. As you become more familiar with microRNAs, you will realise that these sequences have extraordinary features that are uncommon in the protein-coding realm. It is then only a matter of time until you join the increasingly large community of small RNA biologists.   Antonio Marco

---

Fancy composing an installment of "Bench Philosophy"?

Contact Lab Times
E-mail: *editors@lab-times.org*