Institution: University of Manchester

Unit of Assessment: UoA05

Title of case study: The miRBase database – an essential resource for tool development and pharmaceutical research

1. Summary of the impact

MicroRNAs are a class of non-protein-coding RNA genes that regulate the expression of protein-coding mRNAs in animals and plants. Researchers at the University of Manchester (UoM) have developed a microRNA database (miRBase) which has become an essential resource for researchers both in academia and the pharmaceutical industry. The database is the central global repository for all published microRNA sequences and annotation.

miRBase data enables production of novel experimental kits and resources (including microRNA qPCR assays and microarrays) by companies including ABI, Invitrogen, Sigma-Aldrich and Exiqon. Tools produced by these companies underpin experimental microRNA research across academic and industrial settings, which benefit product development, drug discovery and clinical research. Exiqon reported more than 110 million DKK in revenue from their life science business in 2012, with main product lines involving microRNAs.

2. Underpinning research

Griffiths-Jones originally created miRBase at the Wellcome Trust Sanger Institute, with the first major publication in 2004. The impact is based on the continued development and improvement of the database carried out in the Griffiths-Jones research group at the UoM from 2007 onwards.

The key researchers involved in the development of the miRBase database at the UoM are:
Dr Sam Griffiths-Jones (Senior Lecturer, 2007 to date)
Dr Antonio Marco, (Post-Doctoral Research Associate, 2008 to 2013)
Dr Ana Kozomara (Post-Doctoral Research Associate, 2009 to date)

The key research that underpins the development of the miRBase database at the UoM is as follows:

- The Griffiths-Jones group developed methods and pipelines to determine evolutionary and functional relationships between microRNA gene sequences submitted to the database by microRNA researchers. miRBase is responsible for all microRNA gene nomenclature in published literature. The group have assigned gene names for over 8000 microRNA sequences provided by 300 researchers since 2008.
- They developed web tools and interfaces to facilitate access to microRNA sequences and annotation. These include methods for discovering microRNAs from deep sequencing data [1], novel views of the data [2], and procedures to allow the user community to contribute annotation, for example through a collaboration with the Wikipedia online encyclopaedia which began in 2012.
- The Griffiths-Jones group provided access to the primary underlying evidence for every microRNA gene annotation. They obtained and mapped over 300 third party deep sequencing datasets and developed methods to query the microRNA dataset based on domain of expression, experiment, author etc. [2].
- The group also curate and provide links to other microRNA tools and resources, including those that predict functional targets of microRNAs and roles in disease.

miRBase is the sole authoritative source of all published microRNA sequences. All microRNA resources, including other microRNA databases and companies that produce experimental resources source microRNA sequence data directly or indirectly from miRBase. The primary impact derives from the development and improvement of the database. The Griffiths-Jones research group also contributes content to the database from in-house novel microRNA discovery [3, 4].

3. References to the research

The research that led to the development of the miRBase database has been published in high impact journals and is highly cited. The Nucleic Acids Research papers describing the database [including 2, 4] have over 4800 Scopus citations including over 4200 since 2008.

1. **Marco, A., Griffiths-Jones, S.** (2012). Detection of microRNAs in color-space. *Bioinformatics*. 28 (3). p.318-323. DOI:10.1093/bioinformatics/btr686.
2. **Kozomara, A., Griffiths-Jones, S.** (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 39. p. D152-D157. DOI:10.1093/nar/gkq1027.
3. **Marco, A.,** Hui, J.H., Ronshaugen, M., **Griffiths-Jones, S.** (2010). Functional shifts in insect microRNA evolution. *Genome Biol Evol*. 2. p. 686-696. DOI:10.1093/gbe/evq053.
4. **Griffiths-Jones, S.,** Saini, H.K., van Dongen, S., Enright, A.J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 36. p. D154-D158. DOI:10.1093/nar/gkm952.

4. Details of the impact

Context

MicroRNAs are ~22 nucleotide non-protein-coding genes found in all animals and plants. They were discovered as a class in 2001. MicroRNAs act by binding to mRNA transcripts to trigger their down-regulation. They represent arguably the most widespread and important mechanism of post-transcriptional gene regulation. The human genome has over 1500 annotated microRNA genes and it is predicted that over half of all human genes are regulated by microRNAs. The microRNA field is a rapidly expanding one; more than 75% of the data has been deposited and curated since 2008. Commercial organisations, including large pharmaceutical companies, have been quick to adopt microRNA research programmes.

miRBase is the primary source of microRNA sequence data. All microRNA research globally is therefore impacted by the production, development and availability of miRBase.

Pathways to impact

miRBase was created by Griffiths-Jones and was initially hosted and supported by the Wellcome Trust Sanger Institute. UoM hosts and maintains the miRBase from 2007 to date.

Reach and significance of the impact

The primary use of miRBase data is via the website and by bulk data download from the FTP site. Since 2008 the miRBase website has consistently received 40,000-50,000 visits per month from 20,000-30,000 unique users viewing 120,000-150,000 pages. Empirical data shows that ~10% of email queries are from commercial users. The miRBase database development and use is described in a series of publications from 2004 to date. In total, these manuscripts have been cited over 4200 times since 2008. Of these citing journal articles, 10% (424) list commercial organisations as author affiliations (according to Scopus).

Derived commercial and monetary benefits:

The impact on the pharmaceutical industry of miRBase has been considerable. It has allowed new products and significant research gains to be made every year since 2008.

A number of companies, such as Exiqon [A], LC Sciences [B], Comprehensive Biomarker Centre (formerly Febit) [C], Agilent [D], Sigma-Aldrich [E], ABI, Exiqon, GenoSensor, Invitrogen and many others, make microRNA microarrays and other experimental resources from sequence data sourced from miRBase. For example, Exiqon sell 20 different products that rely on data from miRBase, including qPCR assays and LNA microarray technologies for identifying microRNAs. These products are built using miRBase sequence data, and are the most significant contributor to Exiqon's life science business which reported more than 110 million DKK in revenue in 2012 [F].

The Vice President for Research and Development at Exiqon stated: *"From interaction with our customers we understand that everyone in the miRNA research community consider the miRNA content of miRBase to represent the golden standard miRNA repository. Over the years, Exiqon has used miRBase as a reference for further development of new products and updating of existing products. Coverage of miRNAs in a given miRBase version has always been a central part*

of the product information accompanying products in the different portfolios like the miRCURY microarrays" [F].

ABI manufacture and sell Taqman microRNA PCR assays at an off-the-shelf cost of around £250 for sequences that are in miRBase, and custom-designed for around £400 for those which are not. Other companies follow a similar model. These commercial tools produced by third parties from miRBase data underpin experimental studies of microRNA genes in academic and industry labs, and curation of microRNA sequences lowers the cost of experimental assays for the user.

Companies aim to update their products as quickly as possible after a release of a new miRBase version; the schedule of production and sales of these companies is therefore partly governed by miRBase updates. For example, Exiqon *"plan timely product updates and new product developments taking the update and release of new miRBase versions into account"* [F].

Clinical applications:

MicroRNAs have been implicated in a wide range of disease processes and have been shown to act as biomarkers for cancer types, stages, disease prognosis and drug performance. Pharmaceutical companies, including AstraZeneca and GlaxoSmithKline, have active research streams investigating the use of microRNAs as biomarkers. For example, AstraZeneca has a number of lines of research into the roles and pharmaceutical relevance of microRNAs in oncology. A senior genomics scientist at AstraZeneca states that: "*we use miRBase microRNA sequences to design primers for PCR in order to validate specific microRNAs as potential biomarkers. We also explore links from miRBase to up-to-date predicted and validated microRNA target sites, in order to identify pathways and functions*" [G].

Drugs that target microRNAs are also showing clinical promise. For example, miRagen Therapeutics Inc. is a biopharmaceutical company dedicated to developing therapies that target microRNA in disease areas of high unmet medical need. Drugs for chronic heart failure (mir-208), post-myocardial infarction remodelling (mir-15/195) and cardiometabolic disease (mir-378) are in the pre-clinical phase [H].

Santaris Pharma also has a phase IIa clinical trial drug (miravirsen) that acts by inhibiting miR-122. The Associate Director states that: "*Santaris has so far conducted drug discovery projects to find drug candidates against > 30 specific microRNAs. In all cases these projects started from the sequence of the target microRNA as annotated in miRBase*" [I].

Miravirsen is the first microRNA therapy for hepatitis C. At the 2012 International Liver Congress, a renowned professor from the Department of Gastroenterology and Hepatology, University of Amsterdam, stated in a press release issued by Santaris: *"Due to its ability in targeting miR-122, miravirsen has the potential to change the way hepatitis C is treated."*

Further trials are planned to test miravirsen in people with all HCV genotypes, using longer durations, and in combination with direct-acting antivirals. The success of these and newly emerging programmes and the downstream benefits to health are dependent on the availability of miRBase.

Influencing policy:

miRBase is held as the official arbiter of microRNA gene status and annotation by the following bodies: the human (HUGO/HGNC) and mouse (MGI) gene nomenclature committees, RefSeq at the National Center for Biotechnology Information (NCBI) and the International Union of Basic and Clinical Pharmacology (NC-IUPHAR). miRBase datasets are also distributed by the NCBI, the University of California, Santa Cruz and Ensembl genome browsers. These resources and bodies have their own substantial user bases which are also significantly impacted by miRBase work.

5. Sources to corroborate the impact

Example press releases and products from companies that make microRNA resources from miRBase data:

A. http://www.exiqon.com/microrna-microarray-analysis-microrna-array *Exiqon use miRBase v16 data to produce microRNA microarrays*

B. http://www.lcsciences.com/news/lc-sciences-launches-v17-microrna-microarrays *LC Sciences*

use miRBase v17 data to produce microRNA microarrays

C. http://www.cbioc.com/en/services/microarray-services/microrna-profiling/#range-arrays *Comprehensive Biomarker Centre (formerly Febit) use miRBase v16 and V17 to produce microRNA microarrays*

D. http://www.genomics.agilent.com/CollectionSubpage.aspx?PageType=Product&SubPageType=ProductDetail&PageID=1531 *Agilent use miRBase v16 data to produce human microRNA microarray*

E. http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/mirna/microrna-mimics.html *Sigma Life Sciences use miRBase V17 data to produce microRNA mimics*

Corroborating letters and sources:

F. Letter from Vice President for Research and Development, Exiqon, Copenhagen. *Describes range of products that rely on miRBase data.*

G. Letter from Oncology Bioinformatics group, AstraZeneca. *Describes use of miRBase sequence data in screens for cancer biomarkers.*

H. http://www.miragentherapeutics.com/pipeline/ *miragen website showing pipeline of drug discovery using microRNA sequences.*

I. Letter from Santaris Pharma. *Corroborates how the miRBase has played a role in developing the miR-122 inhibitor, miravirsen.*