# Genomic Considerations When Selecting a Model Organism

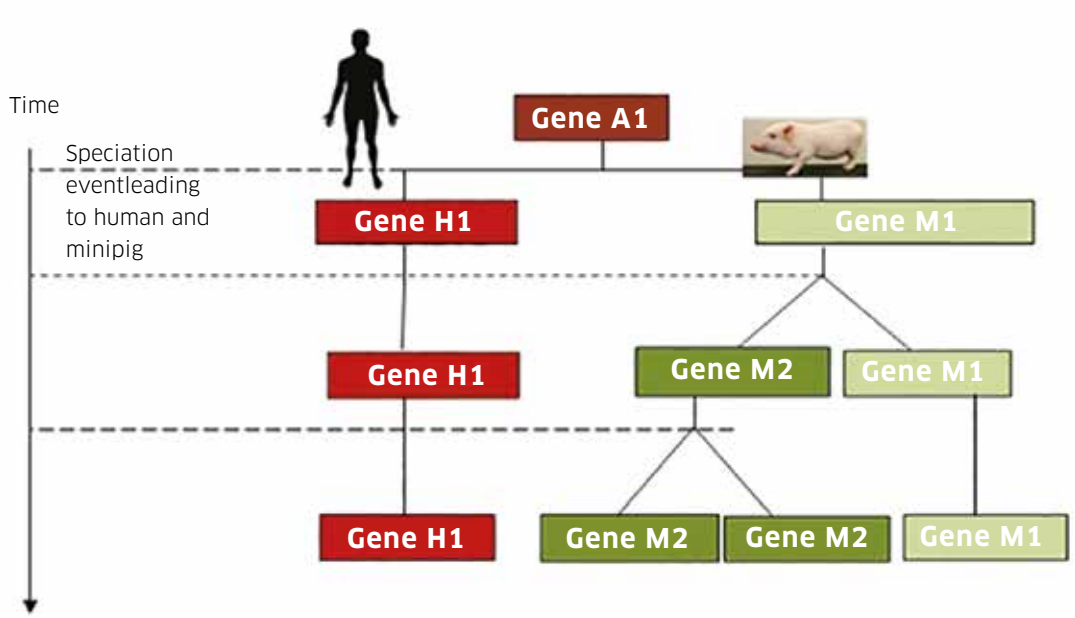
Peter Woollard,

*Computational Biology, GlaxoSmithKline.*

## Introduction

Safety concerns are a major factor in the high attrition rate in drug discovery and development. Regulatory authorities mandate pre-clinical safety assessment in at least one non-rodent animal model. Ideally we want to choose appropriate animal models for a particular drug that are likely to demonstrate similar safety signals to what would be expected in humans. This will help us reduce the unnecessary use of animals http://www.nc3rs.org.uk/, as well as help to provide society with better drugs to treat unmet medical needs. For example, if a safety signal for a drug is seen in a beagle for a reason that is specific to beagles, but that was not known, further animal trials may be needed to understand this before terminating a potentially useful human drug unnecessarily. Knowing the genomic sequence of organisms is a key piece of knowledge along with biochemical, phenotypic and other evidence in selecting potentially appropriate species. Even a basic molecular knowledge of the genes in the genomes of model organisms helps *(Figure 1)*.

Figure 1:
Many changes may have occurred to individual genes since the last ancestor that minipigs and humans had in common. Genes in either species may now be non-functional (pseudogenisation), duplicated, undergone a change of function, etc.




## The pig sequence exists, so why sequence minipig too?

A plethora of academically and agriculturally important animals have been sequenced in recent years, e.g. human, mouse, rat, fugu, pig [1], cow, domestic dog (boxer), etc. Many breeds/strains of domesticated animals are surprisingly different [2,3,4]. In 2011, we realised that we needed to know more about the key breeds of non-rodent species used in our studies, including the Ellegaard Minipig breed, to make a more informed species choice.

## Accessing Minipig Sequence Data

Beijing Genomics Institute(BGI) sequenced and assembled the minipig sequence into contigs and then analysed the genome [4]. The minipig contig data is now publicly available (e.g. GenBank and EMBL), so it can be searched for and freely downloaded. Internally, gene predictions and other annotation were performed to allow easier analysis. At least one other company has sequenced the minipig and also has useful transcriptomics (RNASeq) information, but this is not publicly available at present. In the long term, it would be fantastic if integrated minipig genomic data was maintained in the public domain, in the same way as rat data in Ensembl.

## Choosing the Right Species

The genes of particular interest to pharmaceutical companies are those to which drugs can typically be designed (e.g. by small molecules and biologicals), genes that metabolise drugs(ADMET) and other genes which have historically been implicated in adverse events. If a project has a promising therapeutic modulator, then there are some questions about the protein target in helping to determine which species is best for use in tests *(Table 1)*. Obviously, many factors are involved in determining the model's organism choice.

| Question | Comment |
|---|---|
| Is my gene present in the minipig? | Can be found by a sequence search. |
| What is the sequence of my gene in the minipig? Is the sequence well conserved? | Best to look at peptide level. Are the key active and structural domains conserved? |
| Is my human gene a pseudogene in the minipig? | E.g. is the protein-coding sequence full-length compared to human? |
| How many copies of my gene are there in the minipig? 1:1 or duplicated? | A phylogenetic sequence tree with other placental mammals helps. |
| Has my gene undergone positive selection/ functional divergence in the minipig? | N.B. Quite complex phylogenetics and other analysis are needed to answer that. |
| Is the ADMET profile for a compound likely to be similar in minipigs and humans? | ADME Sarfari may help here. |

Table 1: Sample target-validation questions that can be answered now that the minipig genome is known.

## ADME Sarfari integrates genomic, pharmacokinetics (PK) and transcriptomics data

Abundant pharmacokinetics (PK) and genomic sequence data exist in the public domain for human and model organisms. Tissue-specific expression data are increasingly available too. The need to answer certain fundamental ADME/translational science questions motivated the development of ADME Sarfari https://www.ebi.ac.uk/chembl/admesarfari at the European Bioinformatics Institute(EBI). Questions such as this are becoming easier to answer: a lead molecule has been identified for a certain disease indication; is it possible to predict from available molecular data, if there is a specific animal model that might be best suited to model human ADME?

## Summary

Having the minipig genome is already informing species choice for safety tests, this will improve as more information is integrated. Logically this helps to reveal obvious genomic reasons for choosing certain non-rodent species over others, reduces the unnecessary use of animals, helps to correctly fail drugs earlier and reduces the wrongful attrition of drugs.

**Acknowledgements: Samiul Hasan and Jessica Vamathevan for material and advice.**

In GSK all animal studies need to be ethically reviewed and carried out in accordance with the 1986 Animals (Scientific Procedures) Act and the GSK Policy on the Care, Welfare and Treatment of Animals.

## References

1. Humphray et al. (2007), A high utility integrated map of the pig genome, Genome Biol., 8 (2007), p. R139.
2. Groenen et al. Analyses of pig genomes provide insight into porcine demography and evolution, Nature, 491 (2012), pp. 393–398.
3. R. Rimbault and E A Ostrandar, So many doggone traits: mapping genetics of multiple phenotypes in the domestic dog, Hum Mol Genet. 2012 October 15; 21(R1): R52–R57.
4. Vamathevan et al. 2013, Minipig and beagle animal model genomes aid species selection in pharmaceutical discovery and development. Toxicol Appl Pharmacol. 2013 Jul 15;270(2):149-57. doi: 10.1016/j.taap.2013.04.007. Epub 2013 Apr 19.