NC3Rs Maths Study Group - Applying mathematics to 3Rs problems

Using mathematical modelling to optimise work flow in the Sanger Mouse Genetics Project

Chris Lelliott, Jennifer Salisbury, Hannah Wardle-Jones, Brendan Doe, James Bussell, Jacqui White

Sanger Mouse Genetics Project, Wellcome Trust Sanger Institute

Background to the problem

In recent years, the generation and phenotypic analysis of genetically modified mice was been increasing at breakneck pace around the world. Allied to this, a number of consortia operating under the banner of the International Mouse Phenotyping Consortium (IMPC) have the goal to produce and phenotype mice with null alleles for all coding genes (Ayadi *et al*. 2012; Ramirez-Solis *et al*. 2012). The Sanger Mouse Genetics Project (MGP) is one of the world leaders in this field and operates well established large-scale and high-throughput screening (HTS) pipeline for the generation, expansion, phenotyping and export of colonies of genetically-modified mice. This entails the coordination of several steps and teams to enable the delivery of mice for downstream applications. So far the MGP has phenotyped over 500 lines of mice and generated and exported many more, making it one of the largest non-commercial mouse breeding facilities in the UK. Due to the complexity and scale of the project, inefficiencies exist in our processes which will increase the number of surplus mice, extend the transit time of colonies through the pipeline and result in the suboptimal use of space within the facility.

We have already refined data generation and analysis to maximize the value of results from the animals used in our pipeline (Karp *et al*. 2012). To our knowledge however, mathematical modelling has not been applied to the workflows of biological in vivo HTS projects, presumably due to the intrinsic variability of model organisms. Optimization of the pipeline will generate significant benefits from the 3Rs perspective, as well as improving resource usage and distribution. Given the global scale of phenotyping efforts and the increasing use of high-throughput animal screens, lessons learned from the optimisation process for the MGP will have wide-ranging benefits for animal usage and welfare around the world.

Details of the problem

In a standard high-throughput pipeline, such as in a manufacturing plant, we know the absolute values for the transit time and capacity for each of these steps to a high degree of certainty. Linear programming and related methods can be used to identify the optimal outcome for a given set of values, such as those derived from a pipeline. In our pipeline (presented in Figure 1.), for each line of mice with a specific mutation (known as a colony) we require a specific number and type of mice for each of the end points (text boxes marked in red). The colony is only complete once the required delivery of mice to all end points has occurred and the results from the end point have been obtained. Our current target is to complete at least 160 colonies in our main phenotyping pipelines.

The biological nature of our pipeline means that for a given colony, each step is intrinsically variable, despite specific goals for the pipeline that need to be achieved with finite resources. The transit time of a specific colony through the pipeline cannot be determined prior to the entry of the colony into the pipeline. Some colonies display homozygous subfertility (around 5% of all tested) or are homozygous lethal or subviable (HV; approximately 40%), extending the time needed for generating mice for downstream phenotyping. Also, a variable proportion of mice are lost at each step due either to the nature of the genetic modification or due to the background mortality rate of the mice. These factors mean that overdelivery of mice can be required to meet the demand for each end point, leading to increased use of mice and an extended colony pipeline transit time.

One solution to overcome the variability in the pipeline is to generate and fill the pipeline with as many mice and colonies as possible, despite not knowing which colonies are subfertile, subviable or prone to loss. However, we do not have space available for such a strategy that would meet our current targets, as well as resources in terms of human resource for animal husbandry and the financial budget to deliver this kind of approach. In addition, during periods where loss is minimal and cohort generation is optimal, we risk overwhelming our resources and animal wastage would be markedly increased. This is detrimental for parts of the pipeline such as MGP Select, where we have a defined amount of time and people available for a particular technique, limiting the maximum capacity of the pipeline at this end point. An additional problem is that mouse fertility decreases with age, which means that the longer a colony is needed for different end points, the more breedings are required to generate new mice for cohort generation. Therefore, we need a strategy that will optimise the generation and expansion of breeding cohorts of mice to allow a consistent flow into the pipelines, which is aligned with our resources and with our overall goals.

Available data for informing possible mathematical models

Figure 1 shows the workflow of the pipeline and the numbers of mice that are expected to be delivered to each end point of the pipeline. Figure 2 gives further details of the phenotyping pipeline requirements and introduces the limitations on how and when the mice are delivered. Figure 3 illustrates the key questions to be answered to help optimize the pipeline. Some key limitations to be considered are: The requirement for single batches of mice in some end points; the use of specific genotypes i.e. homozygous or heterozygous; dependencies between the endpoints, so that Recessive Lethality is only used in colonies that fail the HV test. In addition, the genotype of mice delivered to an end point is also dependent on HV, since while homozygous may be used by default, colonies that fail HV will use heterozygous mice instead.

We have a large database with historical raw data for all mice, colonies and end points that have been performed over the 5+ year lifespan of the project. This database may be mined for specific questions before or during the Study Group Sessions. Data that may be derived includes statistics on colony breeding performance, cage utilization and occupancy, rate of colonies through the pipelines and specific end points and background rate of animal loss during breeding and after end point delivery.

Questions you would like to see answered

The three key questions that we would like answered are:

1. What is the optimal number of breeding mice in F1 expansion and the cohort generation steps to be able to deliver efficiently to all pipelines?
2. A. What are the best breeding strategies in order to deliver mice for the various end points?
   B. How do we adapt the workflow for subviable/subfertile and colonies with other problems which impact on delivery to the end points?
3. In which order should the end points/phenotyping pipelines be filled?

Ideally, the answers to these questions would fulfil the following criteria:

1. Reduce the number of animals generated that are subsequently not used.
2. No significant increase in the amount of resource (in terms of space, cost and workload).

The potential impact on animal use

According to Home Office data for 2011, over 3.7 million animals were bred and used of scientific procedures in Great Britain. Each year, the Wellcome Trust Sanger Institute Mouse Genetics Project generates, phenotypes and exports more than 160 novel lines of mice, each requiring multiple breeding and expansion steps. A 1% improvement in our breeding and issuing processes should lead to significant reductions in the numbers of animal generated and used. Should the methods derived from this project for our high throughput pipeline be successful, the biological pipeline optimisation approach could be rolled out to other breeding programmes, further reducing animal use.

References

1. Ayadi A., *et al*. Mamm Genome. 2012 23: 600-10 Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project.
2. Karp N., *et al*. PLoS One. 2012 12:e52410. Robust and sensitive analysis of mouse knockout phenotypes.
3. Ramirez-Solis R., *et al*. Wiley Interdiscip Rev Syst Biol Med. 2012 4:547-63 Large-scale mouse knockouts and phenotypes.

Figure 1. MGP Workflow for a single colony

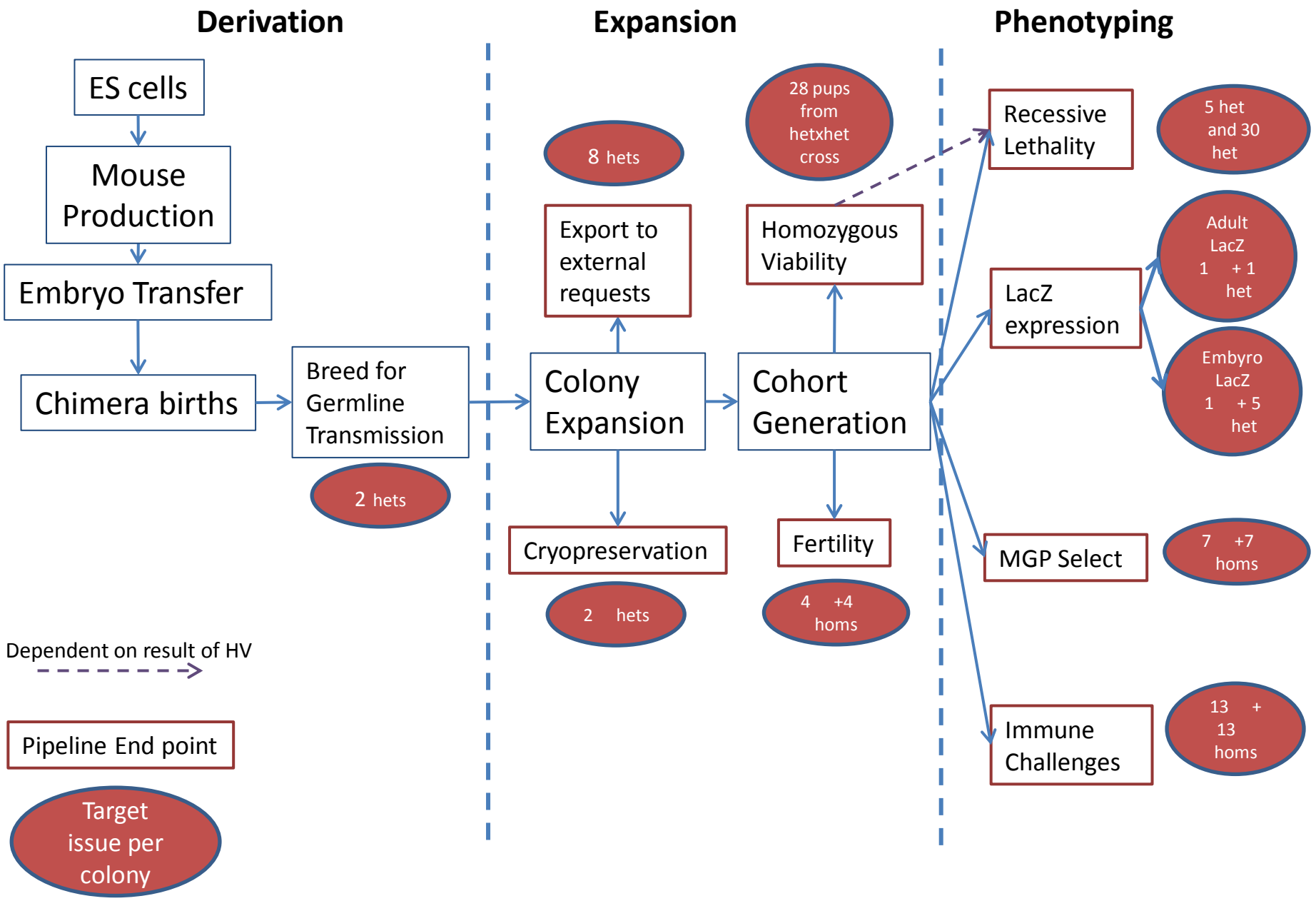

Figure 2. Specific Deliverables to Phenotyping Pipelines

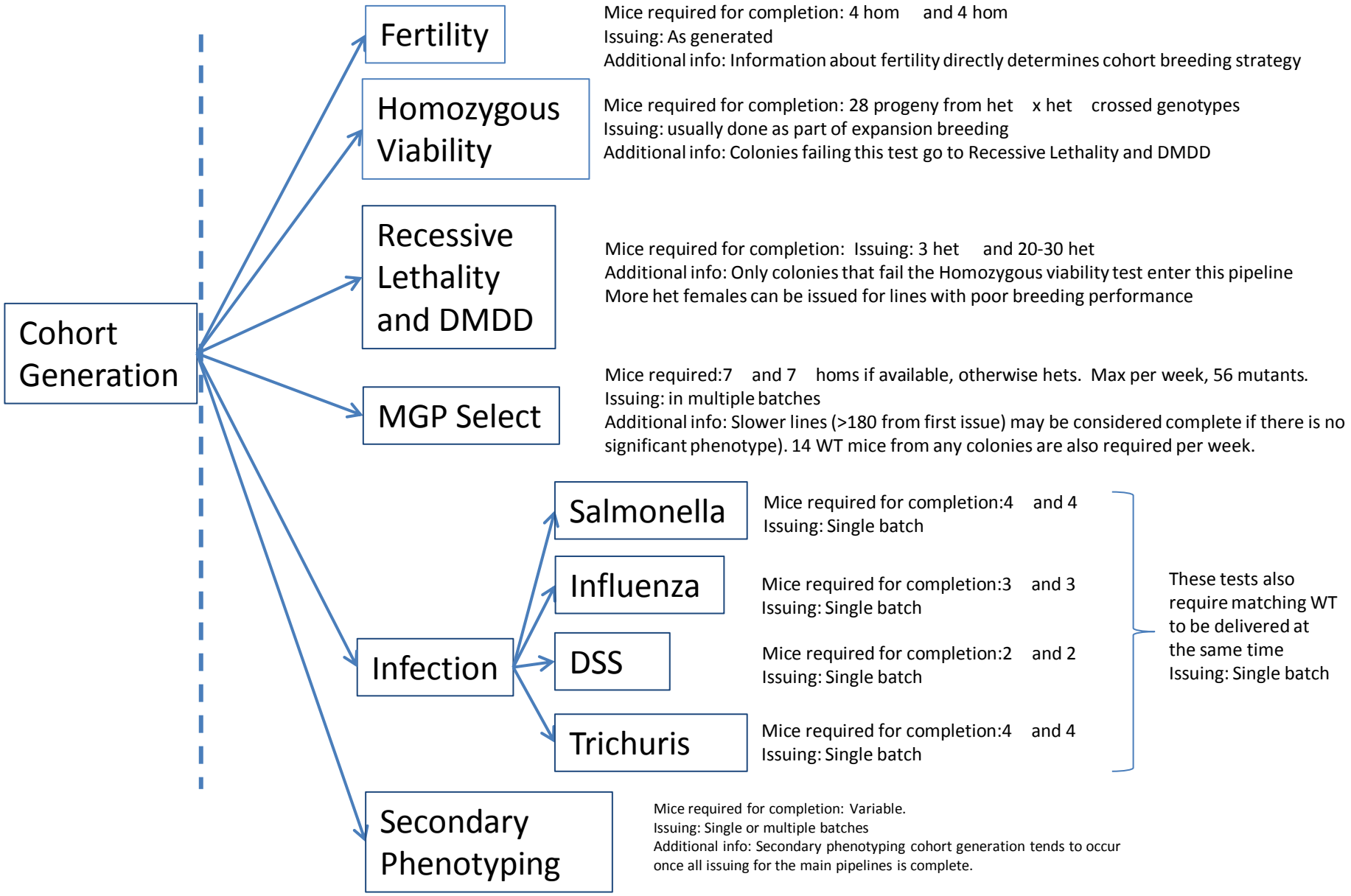

Figure 3. MGP Workflow – Expansion & Phenotyping