

MANUAL ANNOTATION OF THE MOUSE GENOME: HAVANA AND EUCOMM
Mark Thomas, Clara Amid, Denise Carvalho-Silva, Vivek Iyer, Manousos Koutsourakis,
Jane Loveland, Alejandro Mujica, Jeena Rajan, Catherine Snow, Marie-Marthe Suer,
Jennifer Harrow, William Skarnes, Tim Hubbard and Alan Bradley
Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

The European Conditional Mouse Mutagenesis (EUCOMM) project aims to establish a mutant resource containing up to 13,000 conditional mouse mutations in C57BL/6N embryonic stem cells. As part of this resource, up to 8,000 targeted conditional mutations will be generated for genes that can not be readily trapped by random gene trapping methods. Non-symmetrical critical exons are identified as part of the targeting approach, which when deleted produce a knockout (KO) transcript that is susceptible to nonsense-mediated decay (NMD). The success of this approach is dependent on detailed and accurate genome annotation, which is provided by the Human and Vertebrate Analysis and Annotation (HAVANA) group at the Wellcome Trust Sanger Institute. Based on a combination of EST, cDNA and protein evidence, HAVANA produces high quality manual annotation of gene structures, that includes alternative splice variants, polyadenylation features, pseudogenes and gene family clusters. Using this annotation we are able to identify suitable exons for targeting, whilst minimising potential disruption to adjacent genomic features. Annotation of both gene and KO transcripts can be viewed through the Vertebrate Genome Annotation (VEGA) browser (<http://vega.sanger.ac.uk>).

More information on the EUCOMM resource and the targeted genes is available at <http://www.eucomm.org>. Targeting vector design information can also be viewed through the Ensembl genome browser (<http://www.ensembl.org>) as DAS resources (http://das.sanger.ac.uk/das/KO_designs; http://das.sanger.ac.uk/das/KO_vectors). In addition to the EUCOMM project, HAVANA also contributes towards the mouse Consensus Coding DNA Sequence (CCDS) project and is involved in collaborations with RefSeq and MGI.