

The model organism programme generates and distributes a wide range of community resources, accelerating research globally. Its principal focus is on mouse and zebrafish genetics, information and biological resources. In-house research focuses on the biological function of genes identified in other Sanger Institute programmes, such as human genetics and the Cancer Genome Project, or of potential importance to human development and health.

- **A mouse ES cell mutant resource**
- **The Mouse Genetics Programme**
- **Zebrafish Mutation Resource**
- **Experimental cancer genetics**
- **Recessive genetic screens**
- **Genes to Cognition Programme**
- **Analysing cancer genes in the mouse**
- **Mouse developmental genetics**
- **Genetics of deafness**
- **Vertebrate development and genetics**
- **Cell surface signalling laboratory**
- **Model Organism Genome Projects**

➤ **Precision excision**

Adult cells can be reprogrammed into stem cells without being permanently altered genetically.

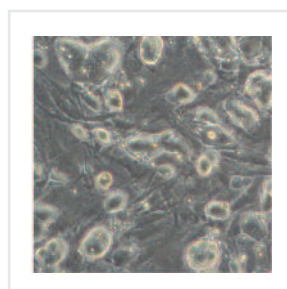
Induced pluripotent stem cells (iPS cells) can be generated from differentiated cells by transgenic expression of just four stem cell factors (Oct4, Sox2, Klf4 and Myc). Up till now, the huge medical potential of iPS cells has been undermined by the use of retroviral vectors to introduce the stem cell factors, which increases the risk of tumourigenic transformation.

Transposon-mediated approaches offer a way round this problem. Allan Bradley and Pentao Liu have developed methods by which genes for the four stem cell factors are introduced via *piggyBac* transposons. Transient expression of a transposase neatly excises the four genes, leaving the original cell reprogrammed but genetically unaltered.

Mouse iPS cells produced in this manner fulfill all the standard criteria of pluripotency (gene expression profiles, teratoma formation and contribution to chimaeras). Furthermore, the reprogramming technology has also been successfully applied to human cells, generating human iPS cells from skin cells. As well as their value in a medical context, such cells offer a valuable 'human model' in which to understand gene function and disease processes.

Woltjen K et al. *piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature*. 2009 Apr 9;458(7239):766-70. PMID: 19252478

Yusa K et al. Generation of transgene-free induced pluripotent mouse stem cells by the *piggyBac* transposon. *Nat Methods*. 2009 May;6(5):363-9. PMID: 19337237



Mouse iPS cells.

The discovery of disease-associated genetic variants in the human genome rarely provides conclusive evidence of causality or a mechanistic explanation of how a specific variant causes disease. Anticipating the need to follow up discoveries in human and cancer genetics, and to provide a platform for addressing vertebrate gene function generally, we established programmes in model organism genetics – principally mice and zebrafish.

Our investment in model organism genetics provides the global community of researchers with both data and genetic resources. To ensure our resource and data generation is technologically leading, we support individual Institute faculty groups who validate their usefulness internally. We also support faculty groups whose innovations provide the foundation for continued scientific advances.

A major focus over the last year has been the zebrafish genome, eagerly awaited by the community. The high degree of inter-individual variation between fish has challenged our efforts to complete this genome to the necessary standard for several years. Major improvements to the zebrafish genome have been achieved over the last year and we anticipate publishing the finished genome early in 2010.

Faculty members

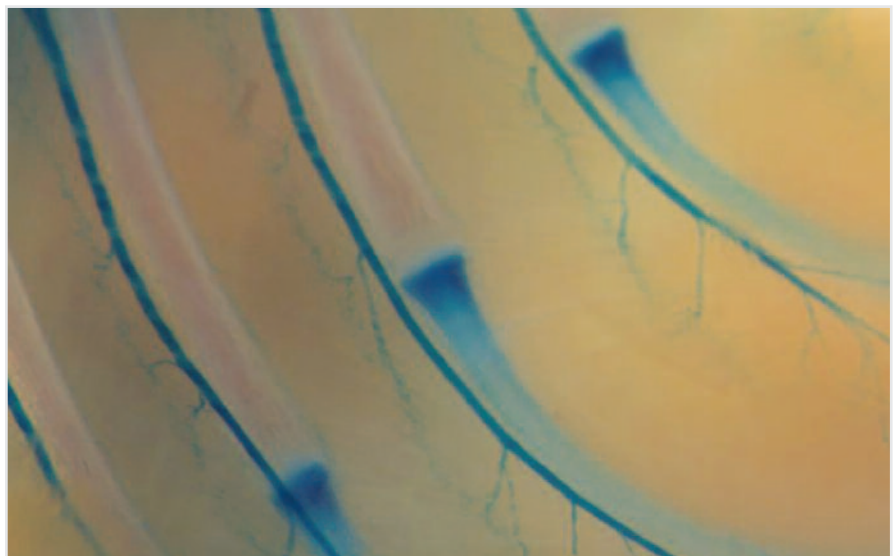
Allan Bradley, Head
David Adams
Seth Grant
Pentao Liu
Bill Skarnes
Karen Steel
Derek Stemple
Gavin Wright

Although a high-quality mouse reference genome has been widely used for several years, new sequencing technologies are enabling multiple additional sequences to be generated at a fraction of the original cost. With support from the UK Medical Research Council, David Adams is sequencing the genomes of 17 strains of laboratory mice, which will support the interpretation of many large-scale genetic studies by documenting the differences between widely used mouse strains.

Our zebrafish genetics platform is heavily oriented towards support of the external community, with mutant alleles available 'on request'. This project, directed by Derek Stemple, is based on a library of chemically mutagenised fish, which are sequenced to identify individuals with a mutation in genes of interest. To date more than 200 different mutants have been identified, most of which have been distributed to scientists from all over the globe. Next-generation sequencing methods promise to increase significantly the output of mutant alleles.

Our internal use of zebrafish mutants is a relatively small scale effort. Gavin Wright's laboratory is defining protein-protein interactions at the surface of the cell. Large-scale screens of zebrafish genes have begun to define cell surface receptor-ligand pairs of molecules that mediate intercellular communication, the significance of which can then be explored *in vivo* using the zebrafish mutation resource. This approach revealed a pair of genes involved in the fusion of myoblasts to form multinucleate muscle fibres.

A similar principle has been used to select genes for analysis in the Genes to Cognition programme managed by Seth Grant. In this case the approach is heavily biochemical and is based on purification of protein complexes from the synapse, which are analysed by mass spectrometry. Protein interactions identified by this approach are the basis of genetic studies, with the animals produced being analysed for behavioural as well as neurophysiological changes.



Thoracic wall of an adult mouse showing strong lacZ expression in the rib cartilage and intercostals nerves. The lacZ gene is under the control of the *Cttnl1* regulatory sequences.

In contrast to the germline mutagenesis approach used in zebrafish, mouse genetics studies generally start with ES cell lines with targeted mutations. ES cells can be turned into mice locally, but more importantly these cells are now the accepted currency for moving targeted alleles throughout the world.

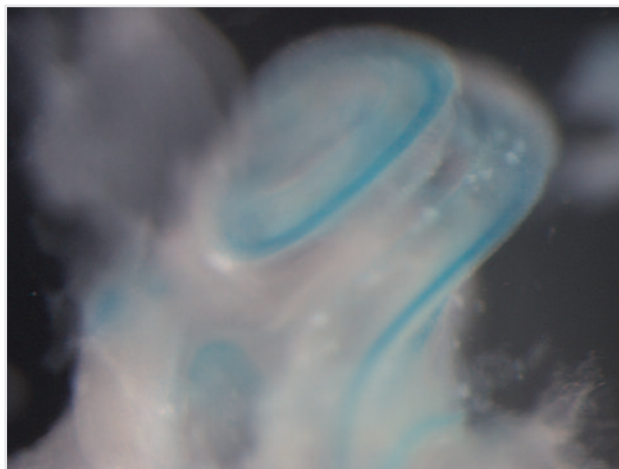
Our productivity in mouse genetics has been built on our ability to manipulate the genome of an ES cell, supported in part by major awards from the National Institutes of Health (NIH) and the EU under the KOMP and EUComm programmes respectively. These awards are part of a coordinated global effort to knockout or mutate every mouse gene. The Sanger Institute is responsible for the largest part of this effort, some 13 000 genes, in work directed by Bill Skarnes.

This year progress has accelerated, with more than 3000 knockouts available. These knockouts are available in two repositories, which have already distributed several hundred cell lines. This resource is being generated on a pure inbred black (C57BL/6N) genetic background. We have recently corrected the mutation that makes these mice black back to its wild-type configuration, so both the production of chimaeras and germline transmission of the ES cell genome can be monitored on the basis of a wild-type (brownish) coat colour.

The growing genetic resource of mutations in ES cells fuels our effort to convert 250 of these alleles into mice annually and describe their phenotype in an unbiased way. As the mice are produced, their embryos are frozen and passed onto the European Mutant Mouse Repository (EMMA) for distribution. Over the coming year the Mouse Genetics Programme will be reshaped to support major external interactions, in the light of progress towards a large-scale phenotyping consortium.

While the Mouse Genetics Programme is configured to generate large numbers of mutants, faculty groups can study mutants in enough depth to reach mechanistic conclusions. Following on from our first description of the function of a mammalian microRNA last year, Karen Steel's laboratory has described a mutation in a microRNA that causes deafness in mice. The point mutation identified in mice has been found in the corresponding human microRNA, and also causes deafness. As well as being mutually confirmatory, these observations also illustrate the value of mouse mutants to human genetics. Unlike the human, the mouse can be studied further to elucidate how this genetic variant causes disease.

Technology development is central to our work on model organisms. Particularly noteworthy has been the rapid utilisation of DNA transposons, which are being employed for a variety of purposes, including somatic mutagenesis *in vivo* and in ES cells. The Liu laboratory is using *Sleeping Beauty* transposons to express mutant genes in mice, directly testing the oncogenic potential of sequence variants detected in the Cancer Genome Project. As the potential oncogene is transposed to different positions in the genome throughout development, it will inevitably come under the control of different regulatory elements, thereby testing its capacity to cause cancer in different cell types and developmental stages. The system has been validated using a range of known oncogenes.



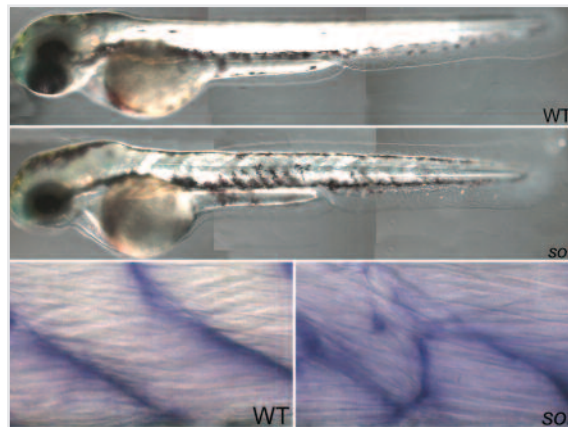
Side view of a mouse inner ear with sensory hair cells labelled in blue, forming a blue stripe coiling from the base to the apex of the coiled cochlea.

The Adams and Bradley laboratories are performing insertional mutagenesis using *Sleeping Beauty* and *PiggyBac* transposons, respectively, in gain- and loss-of-function screens for cancer-causing genes. David Adams has identified genes that cooperate with known oncogenes such as KRAS, thereby identifying genetic networks important in the development of tumours.

Through the Mouse Genetics Programme, individual genes can be studied in the context of the whole organism, throughout development and into adult life. Even so, full phenotypic analysis is labour-intensive and inevitably limited to a subset of genes. Greater productivity can be achieved through genetic screens in mammalian cell lines. Until recently, however, it has not been possible to generate and screen large numbers of homozygous mutations. We have solved this problem by insertional mutagenesis in Blm-deficient ES cell lines, which segregate homozygous cells at high frequency. In combination with *PiggyBac* transposon mutagenesis, we have generated excellent mutation libraries, screens of which have revealed host genes required for retroviral infection and for ricin toxicity.

Mutant cell lines isolated from these stem cell libraries promise to provide insights into many aspects of mammalian biology, from basic cellular functions to differentiation and susceptibility to infection. However, the best models for human systems are human cell lines – an area greatly facilitated by the reprogramming of fibroblasts to iPS cells. The Liu and Bradley laboratories have significantly improved reprogramming, using *PiggyBac* vectors to introduce and then excise the genes encoding stem cell factors, thereby creating transgene-free iPS cells (see box: Precision excision). Human iPS cells are likely to be a greater focus of our future work.

Elisabeth Busch-Nentwich



Under polarised light, the highly organised muscle fibres appear white in a wild-type (WT) zebrafish embryo at two days post-fertilisation while the disrupted muscle fibres of the *softy* (*sof*) mutant embryos are dark.

The myotendinous junctions are visualised with an antibody and ectopic myotendinous junctions are apparent in the *softy* mutant.



A fish called *softy*

A zebrafish mutant with severe muscle degeneration can survive to adulthood – a finding of potential relevance to human muscular dystrophies.

Several 'dystrophic' zebrafish mutants show severe muscle degeneration during embryogenesis, and have become useful models of human muscular dystrophies. Research on the fish version of the Duchenne muscular dystrophy gene, for example, has shed light on the molecular and cellular mechanisms of muscle degeneration. Working with colleagues in Australia, Derek Stemple has identified an unusual dystrophic mutant, *softy*, which can survive to adulthood despite severe muscle degeneration.

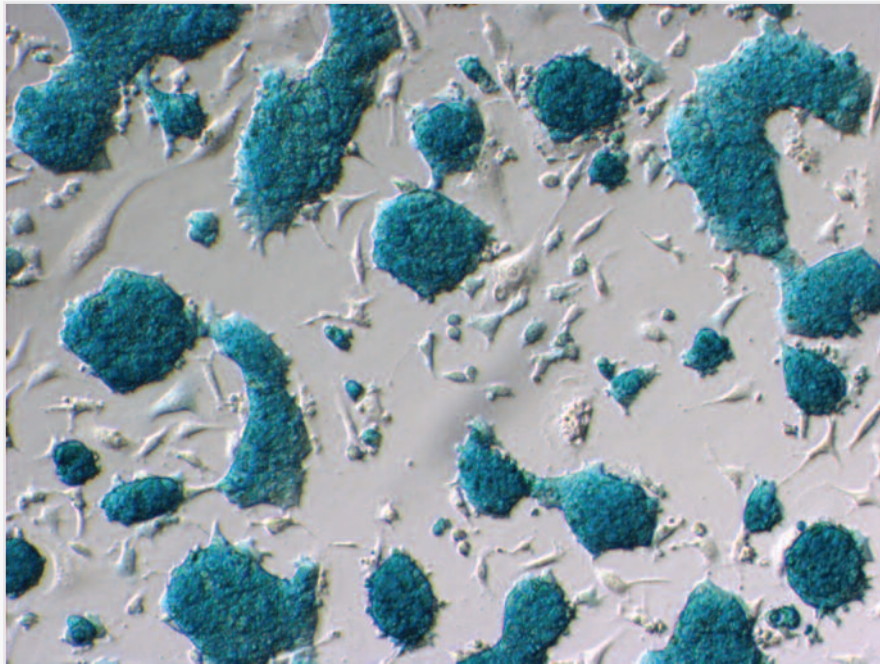
The *softy* phenotype results from a mutation in the *lamb2* gene, which codes for laminin $\beta 2$. This protein is one of the components of the basement membrane, which acts as a tough yet flexible meshwork encasing muscle fibres and is essential for muscle development and muscle function in adulthood.

The critical point of failure in dystrophic mutants is generally where the ends of muscle fibres meet tendons. If the attachment is not strong, muscle fibres tend to detach and die once muscles begin to contract. While this detachment is seen in *softy* fish, a significant proportion of detached fibres go on to develop secondary points of contact to the basement membrane outside the muscle fibre. These secondary attachments seem to enable the muscle to function and the fish to survive.

This unexpected discovery suggests a novel therapeutic strategy for human muscular dystrophies. If the processes responsible for the development of secondary attachments could be identified, it might be possible to develop therapies that generate similar attachments in human muscles affected by dystrophic mutations.

Jacoby AS et al. The zebrafish dystrophic mutant, *softy*, maintains muscle fibre viability despite basement membrane rupture and muscle detachment. *Development in press*

We are creating a high quality resource of conditional mutations in C57BL/6 mouse embryonic stem (ES) cells.



LacZ expression in C57BL/6 mouse embryonic stem cells in which the Cbx7 gene has been conditionally targeted.

The aim of the ES cell mutant resource is to accelerate the functional analysis of genes in the mammalian genome. With substantial external funding from the EU and the NIH, we are developing new technologies for targeting genes in ES cells on an unprecedented scale. Mutant ES cell lines are made available to the scientific community without restriction and are being used by the Sanger Mouse Genetics Programme.

We have derived robust, highly stable, germline-competent ES cells from the C57BL/6N inbred genetic background that are suitable for high-throughput genetic manipulation. These cells can be propagated using standard ES cell culture conditions, obviating the need for expensive specialty medium. Based on blastocyst injections of hundreds of targeted ES cell lines, at least 65% of clones contributed to the germline of chimaeric mice.

We have established a pipeline for high-throughput targeting of genes, from vector design to archiving targeted ES cell clones. We are achieving a throughput of 200 targeted genes per month. To date, we have constructed more than 7000 targeting vectors and have successfully knocked out 3000 genes in C57BL/6 ES cells. For further information please see www.sanger.ac.uk/htgt

To simplify the generation of pure inbred mutant mice, we recently produced C57BL/6 cells engineered to re-express the wild-type agouti gene. Use of these cells simplifies the breeding strategy required to generate mutant mice in a pure C57BL/6N genetic background.

Based on our current rate of production, we expect to generate targeted mutations in 10 000 genes by the end of the programme in 2011. These targeted ES cell resources are used by the Sanger Mouse Genetics Programme to generate hundreds of mutant mouse strains per year for high-throughput phenotyping. A list of available mutant ES cells, mouse strains and phenotype data is accessible via the Sanger mouse portal.

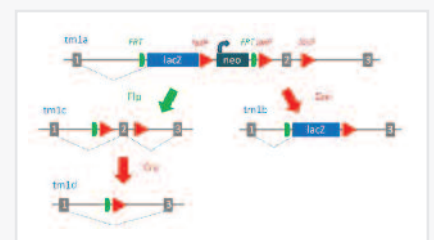


Wellcome Library, London

➤ **A total of 3000 targeted mutations have been generated in C57BL/6 ES cells.**

➤ **Validation of C57BL/6 ES cells for high-throughput targeting.**

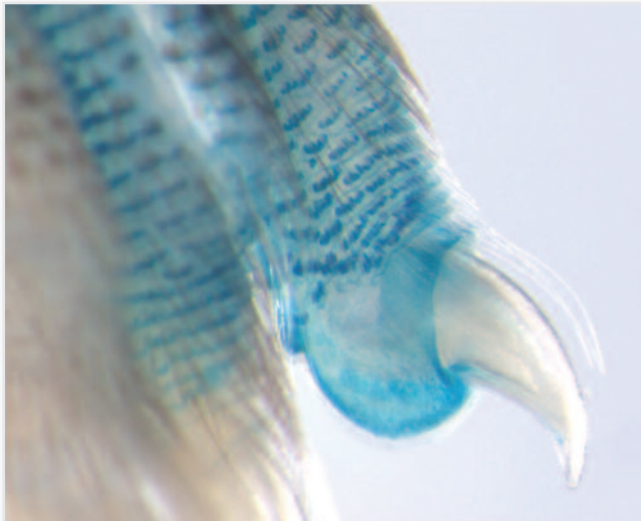
Pettitt S et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat Methods* 6: 493 (2009)



Knock-out first conditional allele.

The Mouse Genetics Programme generates large numbers of mutant mouse lines and carries out primary phenotypic screens.

Adult finger and claw, showing strong lacZ expression in the hair follicles. The lacZ marker is controlled by the *Mta3* gene.



The Mouse Genetics Programme (MGP) uses mutant mouse ES cells generated by the KOMP (Knock-Out Mouse Project) and by the European Union Conditional Mouse Mutagenesis project (EUComm) to produce and analyse a large number of mutant mouse lines. During the last year, the MGP reached its target production rate of 250 lines per year.

The mutant lines were chosen with input from the wider scientific community or nominated by Sanger Institute investigators based on biological themes, such as chromatin regulators, synaptic proteins and homologues of known or candidate human disease genes. Also included were suspected modulators of the innate immune response and genes potentially involved in hearing.

A few lines targeted genes of unknown function, potentially opening completely novel research areas. This approach revealed that the *Spns2* gene has an important role in hearing. To support this thematic approach, a new gene targeting team focuses on high-priority lines that may require customised mutation designs.

In parallel we have streamlined the phenotypic analysis pipeline, following input from internal and external experts. The MGP test battery includes key metabolic assessments as well as multiple tests exploring behaviour and motor and sensory functions. Immune responses are tested in naïve animals and mice challenged with bacteria. Embryonic lethality and fertility are assessed, and adult and embryonic expression patterns are determined.

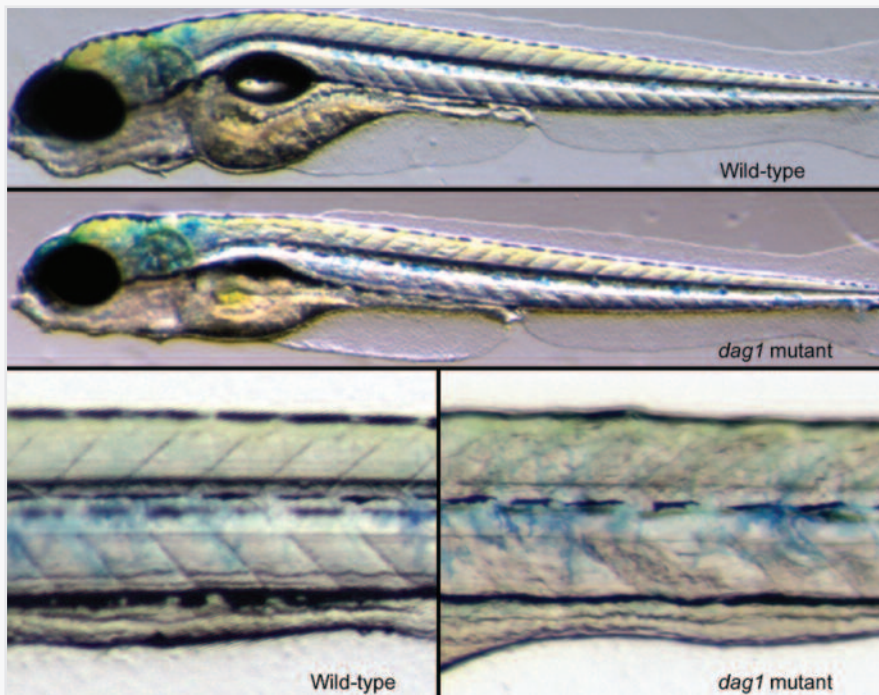
Mutant lines and phenotypic data are freely available to the scientific community. Mouse strains are advertised on the International Knock-Out Mouse Consortium website and, if available, are sent directly from the Sanger Institute. For long-term sustainable distribution, frozen embryos are deposited at the European Mutant Mouse Archive. Over the last year, the MGP received more than 100 external requests.

The successful output of the MGP has placed the Sanger Institute in a leadership position in several worldwide large-scale mouse mutagenesis and phenotypic analysis programmes, including EUComm, KOMP and EUMODIC. It provides leadership on operational, scientific and strategic discussions, and is collaborating with the MRC Harwell, the Wellcome Trust, and Medical Research Council to define a vision and plan for future large-scale phenotypic analysis.



Side view of a mouse embryo showing widespread x-gal staining. The lacZ marker is under the control of the *Kctd10* regulatory sequences.

We aim to generate zebrafish lines carrying mutations in at least 1000 genes, creating a permanent freely available resource for the fish research community.



Elisabeth Buch-Nerwich

Dystroglycan is a protein important for muscle integrity. We model muscular dystrophy using *dystroglycan* mutant zebrafish. Lateral views of wild-type and *dystroglycan* (*dag1*) mutant embryos are shown at four days post fertilisation (top panels). Higher magnification views (lower panels) show disrupted muscle in *dag1* mutant embryos.

We identify chemically induced nonsense mutations in specific zebrafish genes, then characterise the phenotype of homozygous mutants. We have identified new alleles in over 200 genes to date, but aim to identify mutations in around 1000 genes by October 2011. We have recently applied new sequencing technology to the mutation discovery problem, and using this strategy we expect to increase our mutation recovery rate four to 10-fold.

We are also exploring more comprehensive sequencing strategies for individual fish. While complete sequencing of individual fish is achievable, it is not yet cost-efficient. In collaboration with Edwin Cuppen in Utrecht, we are testing a more limited exon-based sequencing strategy covering around 50% of the most 5' exons.

In addition to mutation discovery, we are carrying out an initial phenotypic analysis for each gene, including morphological screening over the first five days after fertilisation and transcriptional profiling. Thus far we have phenotyped 76 alleles representing 66 genes; around 20% of genes produce a homozygous mutant phenotype scorable by morphology during the first five days after fertilisation.

These data will be published on public databases such as ZFIN, the zebrafish model organism database, and ArrayExpress. We supply mutant alleles directly to the research community and to the Zebrafish International Resource Center for future, wider distribution. Thus far we have distributed more than 140 alleles to researchers around the world.



Wellcome Library, London

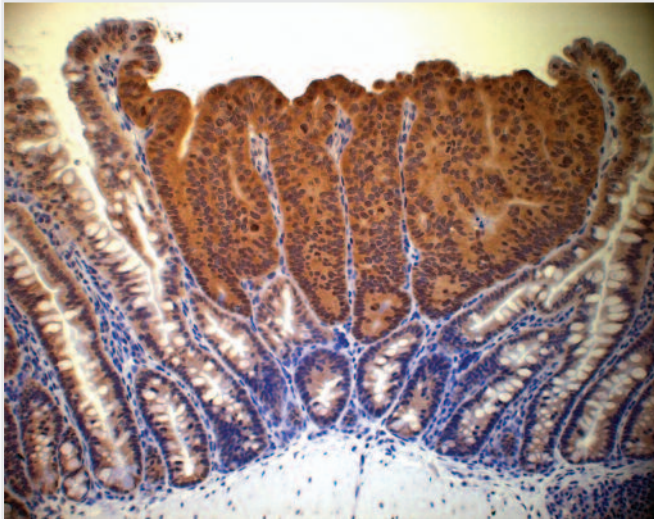
➤ Mutations identified by the Zebrafish Mutation Resource have been used in studies resulting in at least 14 publications.

➤ More than 140 alleles have been distributed to researchers around the world.



An adult zebrafish awaits genotype identification in a single fish holding tank.

We aim to identify cancer genes in the mouse and explore their relevance to human cancer. In addition, as part of the Mouse Genomes Project, we are delivering high-quality sequence from 17 key mouse strains.



A spontaneous intestinal adenoma/tumour from a histone methylase *Mll3* knockout mouse stained for expression of β -catenin which promotes tumour cell growth.

We aim to identify genes that can participate in tumour formation, using various strategies, including genetic screens in mice using retroviruses and transposons, *in vitro* screens in culture, and screens in other model organisms such as the worm. These data are being used in conjunction with human tumour datasets, such as from the Cancer Genome Project, to identify new genes relevant to tumour formation in humans. This cross-species approach has identified several cancer genes that we are currently studying in detail.

We have performed genetic screens to identify genes that co-operate with loss of tumour suppressor genes (for example in *p53*, *Rassf1a* and *Tslc1* knockout mice) to promote cancer. For each of these backgrounds we have collected more than 150 tumours, which we are sequencing on the 454 platform to determine insertion sites and to discover new cancer pathways. We are also screening for genes that specifically drive cancer in the intestine, pancreas and haematopoietic systems. Importantly we have shown that loss of the gene *Mll3* and activating mutations in the *Gnas* gene are important driver mutations of intestinal tumorigenesis.

The second activity of the group is to deliver high quality sequence from 17 key mouse strains as part of the Mouse Genomes Project. Access to complete sequence of multiple inbred strains will add to a wealth of existing genetic resources, providing a foundation for a systems biology approach to phenotypic variation in the mouse. We have generated 20x coverage of the mouse strain *NOD/LtJ* and around 15x coverage of 13 other mouse strains. The first phase of data release took place in July 2009.

In the worm, we have identified a new gene, *CWR* (co-operates with *Ras*). Intriguingly, in the mouse this gene is critical for the viability of cells of the haematopoietic system and intestine, and we have data to suggest that somatic inactivation of *CWR* may lead to cancer formation.



As part of the Mouse Genomes Project we are sequencing all of the commonly used mouse strains including all of the strains of the collaborative cross.

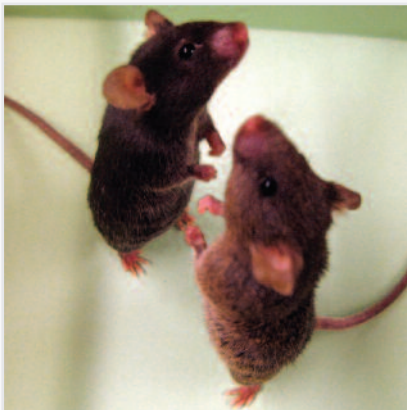


Wellcome Library, London

Megaesophagus in aged *Rassf1a* knockout mice is a good model for human megaesophagus/achalasia, an oesophageal abnormality of unknown genetic cause.

van der Weyden L et al. Megaesophagus in *Rassf1a*-null mice. *Int J Exp Pathol* 2: 101 (2009)

We develop technology and resources to support genetic analysis of the mouse, and use these technologies to identify key genetic determinants of important biological processes.



Stephen Pettitt

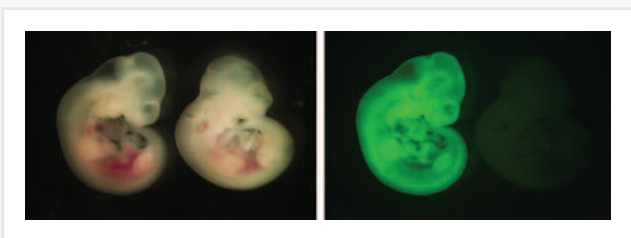
Mice carrying the dominant agouti (brownish) coat colour, (right), have been generated from a modified C57BL/6 embryonic stem cell line. Use of this cell line in large scale knockout programmes can simplify further breeding of the resulting mice.

Several projects in the laboratory use *Blm*-deficient ES cell lines as a genetic background in which to conduct genetic screens. Such lines segregate cells with homozygous mutations at a high frequency, making it easier to isolate recessive mutations. Over the last year we have used the *PiggyBac* transposon to generate mutation libraries in *Blm*-deficient ES cells covering much of the mouse genome. Using these libraries we have identified new host genes required for a variety of processes, including retrovirus infection and ricin-mediated toxicity.

We also conduct genetic screens in somatic cells *in vivo*, to identify tumour suppressor genes. We again use the *Blm*-deficient genetic background and *PiggyBac* mutagenesis. By regulating when and in which tissues transposase is active, we can knock out (or turn on) gene activity in somatic cells at very precise times and places.

The *PiggyBac* transposon is a versatile vector with a large cargo capacity, and can be excised from the mouse genome without leaving a footprint. These features have enabled us to reprogramme embryonic fibroblasts to generate transgene-free iPS cells.

We are contributing to the Knock Out Mouse Programme (KOMP) and EU Conditional Mouse Mutagenesis (EUCOMM) genetic resources, and have collaborated with Bill Skarnes and the Mouse Genetics Programme to establish and characterise a C57BL/6N ES cell line. To simplify breeding schemes, we have corrected the agouti mutation, and this modified genetic background will be used in the KOMP and EUCOMM programmes. We are also producing a microRNA knockout resource in ES cells. To date, we have assembled a knockout vector resource for more than 300 genes and derived ES cells with mutations in more than 100 microRNA genes.



Contribution of integration-free induced iPS cells to somatic tissue in chimeric embryos. iPS cells expressing GFP were microinjected into blastocysts. Chimeric embryos were taken out at embryonic day 10.5 and observed under the fluorescence microscope. Right embryo is non-GFP control. Yusa K et al. *Nature Methods* 2009 6(5): 363-69 doi:10.1038/nmeth.1323



Wellcome Library, London

➤ **Analysis of clonal isolates of mouse embryonic stem cells reveals a high degree of copy number variation, suggesting that adult somatic cells may vary significantly within an individual.**

Liang Q et al. Extensive genomic copy number variation in embryonic stem cells. *P Natl Acad Sci USA* 105: 17453 (2008)

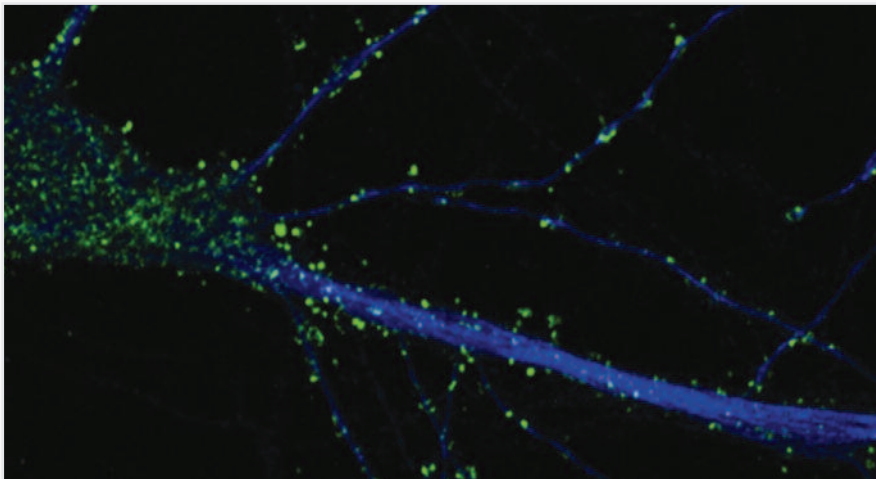
➤ **The *PiggyBac* transposon can be used to reprogramme embryonic fibroblasts to iPS cells without introducing permanent genetic changes.**

Yusa K et al. Generation of transgene-free induced pluripotent mouse stem cells by the *piggyBac* transposon. *Nat Methods* 6: 363 (2009)

➤ **Production of embryonic stem cells from C57BL/6N mice with a wild-type agouti coat colour gene will simplify breeding schemes and provide a foundation for mouse genetic resources.**

Pettitt SJ et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat Methods* 6: 493 (2009)

The Genes to Cognition Programme integrates genome biology with neuroscience, addressing problems of fundamental biological and medical importance.



Esperanza Fernandez

Branches (blue) of a brain neuron expressing a genetically engineered protein (green dots) for isolating synapses and disease relevant proteins. Fernández E et al. *Molecular Systems Biology* 2009 5:269 doi:10.1038/msb.2009.27

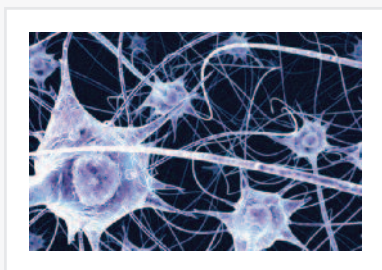
The unifying focus of the Genes to Cognition Programme is the synapse. Our recent human work indicates that synapse proteins are involved in over 200 diseases, including common diseases of great economic and social significance.

We have defined the protein composition of synapses in multiple species and developed a new model that traces the origin of the brain to an ancestral set of proteins that predate multicellular animals and affect the behaviour of unicellular animals.

The mammalian synapse contains more than 1000 proteins, physically linked in protein networks. We have mapped the building blocks of the network's connections and defined molecular circuitry of extraordinary complexity and computational capacity. The symptoms of genomic disorders such as schizophrenia may reflect perturbation of this molecular circuitry, potentially explaining why complex genetic diseases involve so many genes.

In the mouse, we have performed the largest genetic study of the vertebrate synapse, analysing the consequences of mutations on neurophysiology and behaviour. By studying the most basic adaptive and learned responses at the cellular and whole animal level, we gain unique insight into the genomic organisation of behaviour. We have defined sets of genes that encode interacting protein networks in the synapse that underlie behavioural traits. This work should help us to explain how genetic complexity contributed to the assembly of the behavioural repertoire of animals and its relationship to human diseases.

For more information on the Genes to Cognition Programme please visit <http://www.genes2cognition.org>



Benedict Campbell, Wellcome Images

Artist's impression of a network of nerve cells in the brain.



Wellcome Library, London



Comparative proteomic and genomic analyses of synapse multiprotein complexes shed light on the evolution of the synapse.

Emes RD et al. Evolution and anatomical specialization of synapse proteome complexity. *Nat Neurosci* 11: 799 (2008)



High-throughput analysis identifies distinct patterns of protein modification in post-synaptic neurons after stimulation by different neurotransmitters.

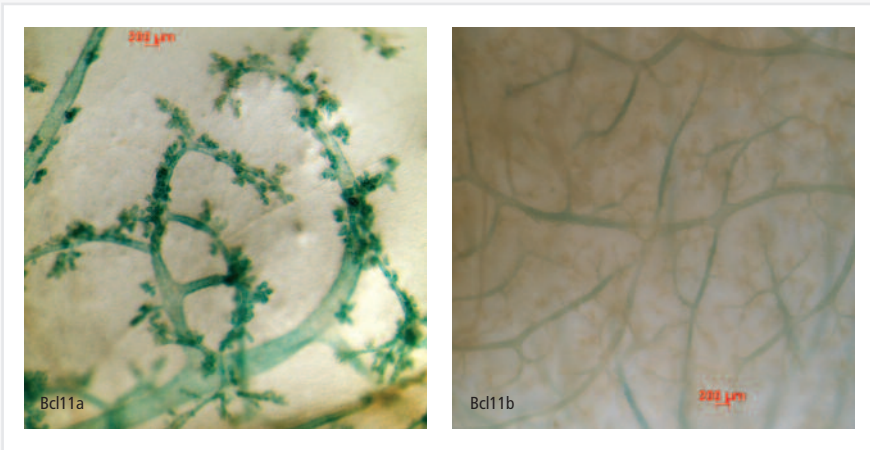
Coba MP et al. Neurotransmitters drive combinatorial multistate postsynaptic density networks. *Sci Signal* 2: ra19 (2009)



Tagging of genes encoding synapse proteins provides a way to identify clusters of interacting proteins, including those binding to proteins associated with increased risk of schizophrenia.

Fernández E et al. Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Mol Syst Biol* 5: 269 (2009)

We study cancer genes that play important roles in development.



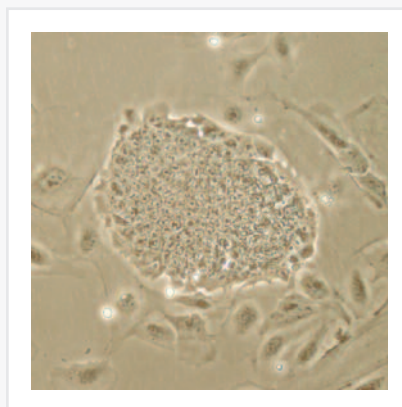
Expression of Bcl11 genes in the mouse mammary gland during early gestation. Note: *Bcl11a* is highly expressed in the alveolars, which produce milk. In contrast, *Bcl11b* expression is excluded from the alveolars.

Understanding the molecular and cellular mechanisms of action of cancer genes provides fundamental knowledge and identifies possible new therapeutic approaches. We have focused in particular on *Bcl11a* and *Bcl11b*, which encode C2H2 zinc finger transcription factors and are mutated in both human and mouse tumours. We have previously shown that they have critical roles in B and T lymphocyte development.

We have now further characterised *Bcl11a* and *Bcl11b* in haematopoietic cells and have identified novel functions of both genes. As well as haematopoietic cells, *Bcl11a* and *Bcl11b* are expressed in other cell types. We have recently discovered that *Bcl11a* and *Bcl11b* are critical regulators in the mouse mammary gland. They are active in the mammary epithelial stem cell compartment, where they direct subsequent lineage specification and commitment, and maintain the differentiation status of mammary epithelial cells.

We have also developed new transposon-based genetic tools. For validating genetic mutations identified from cancer genome sequencing, we developed a platform based on *Sleeping Beauty* DNA transposition to study the oncogenic potential of newly discovered mutations. Transposition of *Sleeping Beauty* transposons carrying mutant genes can reveal in which cell types and which developmental stages activation of the gene leads to cancer.

We have characterised the *PiggyBac* DNA transposition system in mouse ES cells, which can be used to introduce genetic elements into mammalian cells. We have cloned into the *PiggyBac* transposon genetic cassettes for mutagenesis, putative cancer genes, microRNAs, and genes for stem cell factors, which can be used to produce iPS cells free of exogenous genes. These iPS cells are likely to be a valuable tool in functional genomics studies.



A colony of undifferentiated mouse embryonic stem cells growing in culture in serum-free conditions without feeder cells.



Wellcome Library, London

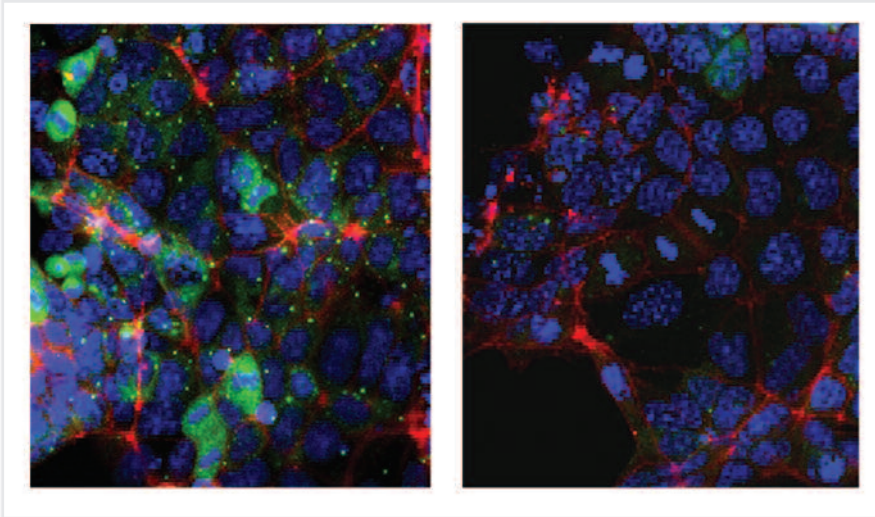
➤ **An engineered DNA transposon can reprogramme fibroblasts into induced pluripotent stem cells then be excised, leaving the cells genetically unchanged.**

Woltjen K et al. *piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458: 766 (2009)

➤ **Transposons containing putative oncogenes can be used to test their oncogenic potential in all cell types and all developmental stages.**

Su Q et al. A DNA transposon-based approach to validate oncogenic mutations in the mouse. *P Natl Acad Sci USA* 105: 19904 (2008)

We are studying the underlying genetic basis of early cell fate decisions in the mammalian embryo.



Arid1a protein is enriched at mitosis in uninduced ES cells (green staining in cells on left). Activation of a conditional mutation (cells on right) affects self-renewal, cytoskeletal (red staining) and nuclear organisation (blue staining).

We carry out detailed genetic, transcriptomic and proteomic studies to elucidate the signalling pathways and regulatory networks that delineate specific developmental programmes. We are using mouse ES cells as a model system to understand lineage decisions and confirm our findings in the mouse.

We are developing a high-throughput process to create a resource of homozygous ES cells for cellular phenotyping. We either use an insertional targeting vector to disrupt the second allele of heterozygous mutant cells or, for genes that are essential in ES cells, we generate inducible homozygous mutations dependent on the activation of Cre recombinase.

By applying these approaches to chromatin proteins expressed in ES cells, we have identified several genes required for the growth and maintenance of stem cells. The transcriptional targets of these chromatin factors can be identified by expression profiling in cells after ablation of gene function.

Genes that exhibit mutant phenotypes in ES cells are epitope-tagged using a highly efficient 'knock-in' strategy. We then use tandem affinity purification mass spectrometry (TAP-MS) to catalogue protein-protein interactions. Epitope-tagging can also be used to identify protein-DNA interactions in cells using chromatin immunoprecipitation and mass parallel sequencing (ChIP-seq). Combining these approaches, we are defining protein interaction and transcriptional networks that operate in stem cells.

We have identified several chromatin-associated protein complexes required for ES cell self-renewal, and detailed studies are ongoing to elucidate their molecular function. Our genetic characterisation of Arid1a, for example, has identified a critical role for the BAF chromatin-remodelling complex in maintaining ES cell self-renewal. Our data suggest that the BAF complex physically interacts with many regulators of pluripotency and acts as a transcriptional co-factor of one in particular, helping to maintain the stem cell transcriptional programme.



Wellcome Library, London

Embryonic stem cell pluripotency depends on the chromatin-remodelling complex BAF.

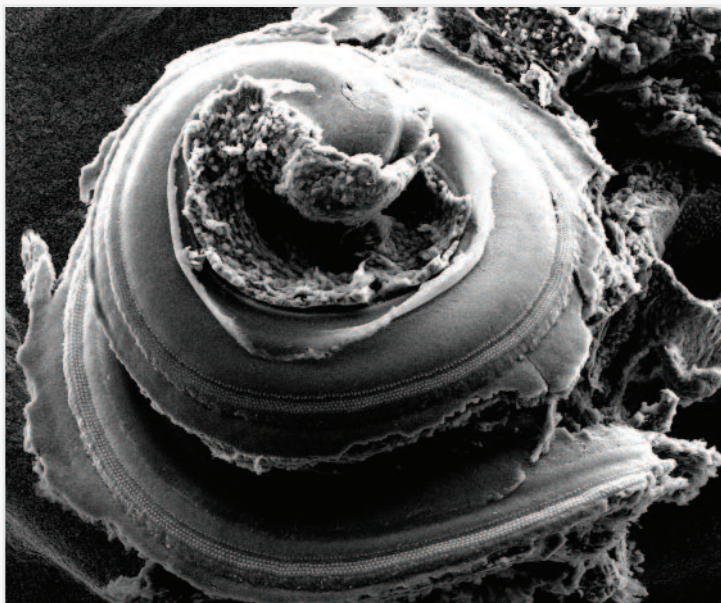
Gao X et al. ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *P Natl Acad Sci USA* 105: 6656 (2008)



The chromatin remodelling factor Arid1a is expressed (blue staining) in cells of the early mouse embryo (3.5 days after fertilisation).

We aim to identify genes involved in deafness and establish their roles in normal hearing, using the mouse as a model system.

A mouse cochlea dissected open to reveal the surface of the sensory region. The four rows of spots coiling up the cochlea from base to apex show the hair bundles on top of sensory hair cells.



Graham Froggatt

To find new genes underlying hearing impairment, we screen new mouse mutants generated by the Sanger Institute's Mouse Genetics Programme. We have developed auditory brainstem response measurements as a screening tool to look for mutants with raised thresholds for a physiological response to sound stimuli or other signs of abnormal auditory processing.

So far, we have screened over 70 new mutants and detected three genes that influence auditory function, causing mild (*Matn1*), moderate (*McpH1*), and severe (*Spns2*) hearing impairment. None of these genes was suspected of involvement in deafness. Unlike earlier tests, our screen can detect a wide range of impairments, suggesting that many new genes involved in deafness are likely to be found with this approach, particularly as the screen builds up to five new genes each week. We are investigating the underlying pathology in these three mouse mutants.

Deaf mouse mutants from an earlier screen for ENU-induced hearing and balance disorders continue to yield insight into the molecular basis of deafness. We have found that a point mutation in the critical seed region of a microRNA, miR-96 (*Mirn96*), causes progressive hearing loss in mice carrying one mutant allele, and severe early-onset deafness in mice with two copies of the mutant gene. The mutation affects the expression of hundreds of target genes. We found five genes that were greatly downregulated in the inner ear, any one of which could explain the hearing deficit.

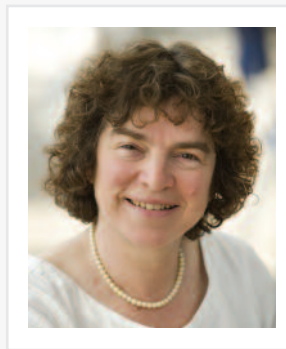
Our collaborators have found two further mutations in the seed region of miR-96 in two human families with progressive hearing loss, emphasising the value of studying mouse mutants as models for human deafness.

Left: An outer hair cell from a 4-week-old normal mouse showing the stereocilia are arranged neatly on the surface of the cell in a typical W shape pattern.

Right: An outer hair cell from a 4-week-old heterozygous mutant: stereocilia are disordered and the organisation lost. *Lewis M et al. Nature Genetics 2009 41(5): 614-18 doi:10.1038/ng.369*



Agnieszka Rzażewska



Wellcome Library, London

➤ **A mutation in a microRNA causes progressive hearing loss in the diminuendo mouse, and mutations in the same microRNA in humans also lead to hearing loss.**

Lewis M et al. An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat Genet* 41: 614 (2009); Mencía A et al. Mutations in the seed region of the human miR-96 are responsible for non-syndromic progressive hearing loss. *Nat Genet* 41: 609 (2009)

➤ **The chemically induced catweasel mutation disrupts the *Six1* gene, creating a model for a rare human disorder affecting hearing, the kidney and urinary system.**

Bosman EA et al. Catweasel mice: a novel role for *Six1* in sensory patch development and a model for Branchio-Oto-Renal syndrome. *Dev Biol* 328: 285 (2009)

➤ **Identification of three new genes underlying hearing impairment by screening new mouse mutants generated by the Mouse Genetics Programme.**

We apply vertebrate genetics and genomics to understand early development and to model human genetic disease, principally using zebrafish and *Xenopus tropicalis*.

Our research addresses the development and maintenance of muscle, and genes controlling early vertebrate development. The Sanger Institute Zebrafish Mutation Resource and the NIH-funded *Xenopus tropicalis* Mutation Resource provide mutant strains to complement our genetic studies.

Our work on muscle integrity has provided insight into muscle assembly and shed light on several inherited diseases. In a collaborative study we identified *lamb2* as a candidate gene for some forms of human muscular dystrophy. The condition can also be caused by mutations in glycosyltransferases such as fukutin-related protein (FKRP). We have found that loss of FKRP affects laminin 1, as well as dystroglycan, previously thought to be the only target of FKRP. In addition, loss of glycogenin produces an essentially identical phenotype, suggesting that mutations in the *glycogenin* gene might underlie some forms of muscular dystrophy.

We identified another candidate disease gene following positional cloning of the zebrafish *buzz off* locus. We mapped *buzz off* to the zebrafish *nebulin* gene, mutation of which leads to a condition called nemaline myopathy in humans. This led us to knock out the *tropomodulin 4* gene, which also produces a nemaline myopathy phenotype.

We have identified a mutation in the *capza1a* gene by positional cloning of the *schnecke* locus. This mutation leads to a contraction-dependent weakening of the sarcomere structure, ultimately causing breakage at Z-disks. By examining sarcomere assembly in combined loss-of-function studies, we found that the basic sarcomere structure will form in the absence of either thick filaments or thin filaments, but not when both types are disrupted.

In collaboration with the laboratory of Inês Barroso and Phil Beales (University College London), we have found that, although *FTO* encodes a nucleic acid demethylase, it is required for the function of primary cilia, which have many important roles during development and in adult tissues.



Wellcome Library, London

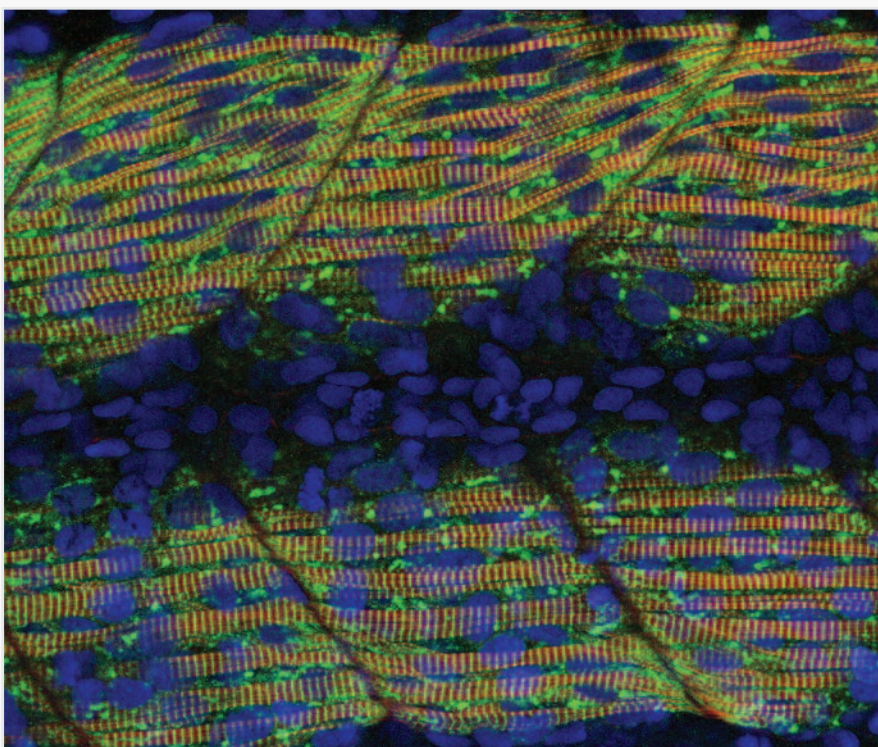
The zebrafish version of the human oral-facial-digital type 1 syndrome gene is required for cilia function and cell intercalation during development.

Ferrante MI et al. Convergent extension movements and ciliary function are mediated by *ofd1*, a zebrafish orthologue of the human oral-facial-digital type 1 syndrome gene. *Hum Mol Genet* 18: 289 (2009)

Loss of laminin $\beta 2$ disrupts muscle cell detachment, pointing to *LAMB2* as a candidate gene for some forms of human muscular dystrophy.

Jacoby AS et al. The zebrafish dystrophic mutant, *softy*, maintains muscle fibre viability despite basement membrane rupture and muscle detachment. *Development in press* (2009)

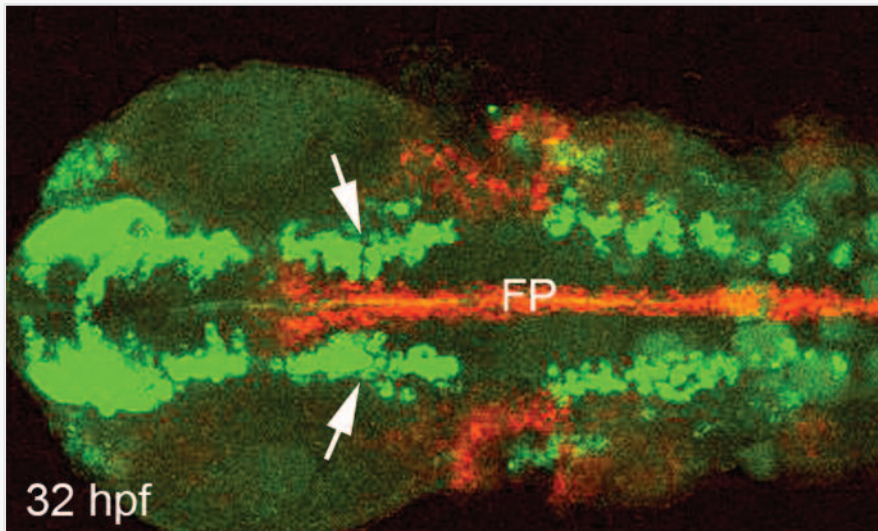
Zebrafish *fto* is required for normal function of primary cilia. Loss of function leads to a randomisation of left-right asymmetry, kidney cysts, and other abnormalities.



Maniella Ferrante

A lateral view of a 30-hour-old zebrafish embryo, highlighting newly formed muscle. This confocal fluorescence micrograph shows the muscle proteins Titin (green) and Actin (red) as well as the nuclei (blue). Two types of muscle fibres are seen. Most are slow twitch fibres (horizontal) but a few fast twitch fibres (diagonal) are seen in the upper part of the image.

Our goal is to identify novel cell surface receptor–ligand pairs that initiate intercellular communication in vertebrates.



Detailed expression pattern of two interacting cell surface receptor proteins from our network. One gene (shown in red) is expressed in the cells that constitute the floorplate (FP) of a 32-hour-old zebrafish and its interacting partner is expressed in neurons (shown in green).

Extracellular protein–protein interactions between secreted and membrane-tethered proteins are critical for initiating intercellular recognition events. However, although 15% of our genes encode proteins predicted to form extracellular interactions, most have no documented binding partner.

We have developed mammalian protein expression technologies and an interaction assay, AVEXIS (avidity-based extracellular interaction screen), for undertaking systematic large-scale screens for such interactions. We are using this assay to study basic biological processes and clinically relevant problems, such as how pathogens recognise and invade host cells.

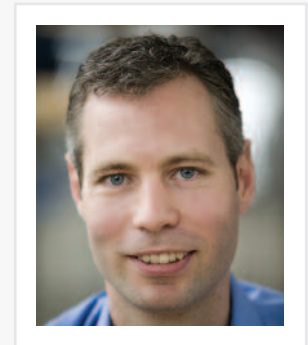
This year, we have analysed a large network of interactions between 250 zebrafish receptor proteins belonging to the immunoglobulin and leucine-rich repeat superfamilies – the first large-scale view of extracellular glycoprotein interaction networks. Data are publicly accessible through the online database ARNIE (<http://www.sanger.ac.uk/arnie>).

We have functionally analysed a subset of these interactions using mutant zebrafish provided by the Zebrafish Mutation Resource. We have also investigated a receptor–ligand pair involved in the cellular fusion of myoblasts to form multi-nucleate syncytial muscle fibres. The zebrafish process differs fundamentally from that seen in *Drosophila*, the standard model for muscle formation.

We have also used our unique library of zebrafish recombinant proteins to raise high-quality mouse monoclonal antibodies – a much-needed resource for the zebrafish community. We are currently optimising the technique and aim to generate antibodies to 200 proteins over the next five years.

Finally, we are searching for novel ligands used by the blood stage of the malaria parasite to bind and invade host erythrocytes. Candidate receptor–ligand pairs will be assessed by *in vitro* invasion assays in collaboration with Julian Rayner’s malaria group.

Although we can screen for interactions between hundreds of proteins, vertebrate genomes contain several thousand. We therefore plan to improve the throughput and sensitivity of these technologies, miniaturising the assay by spotting the proteins in a microarray format.



Wellcome Library, London



Our extracellular glycoprotein interaction network determined by AVEXIS. Further details and expression patterns can be seen at <http://www.sanger.ac.uk/arnie>.

The model organism genome projects are generating high-quality reference genome sequences for zebrafish (Derek Stemple) and mouse (Julian Parkhill).

Zebrafish

The zebrafish (*Danio rerio*) genome comprises 25 pairs of chromosomes and around 1.4Gb of haploid DNA sequence. The main goal of the project is to completely cover the genome with high-quality, finished, large-insert (>100kb) BAC clones.

Over the past year we have substantially improved the genome assembly by more closely tying the physical DNA sequence to meiotic genetic maps. Anchored on an existing low-resolution meiotic map, we published a new annotated assembly, Zv8, in Ensembl 54, released in May 2009.

To further refine genome assemblies, we have been developing a high-density and high-resolution meiotic map with a physical resolution of around 65kb, containing more than 100 000 SNP markers. The new map will allow us to accurately place nearly every finished large-insert clone. It will be published as a new assembly, Zv9. Up to 2400 more clones need to be finished to provide complete genome coverage. Although some will be left for genome maintenance, most will be submitted for sequencing by the end of 2009. To realise significant savings, we have piloted the finishing of clones using Illumina sequencing.

As only about half of all zebrafish genes are represented by complete cDNA sequence in public databases, we have undertaken a comprehensive RNA sequencing project, generating more than 50Gb of cDNA sequence data from 10 different developmental stages and adult tissues. In collaboration with the Ensembl team, we have devised algorithms to generate gene models directly from these data. Thus far we have rough models for more than 25 000 poly-A+ transcripts, providing a wealth of quantitative information on tissue- and/or stage-specific transcription.

By the end of 2009, we aim to submit a genome sequencing paper and to turn the project over to the Genome Reference Consortium (see Bioinformatics section, page 67) for maintenance.

Mouse

Ongoing efforts to close the remaining gaps in the mouse genome project are coordinated through the Genome Reference Consortium. The current essentially finished genome sequence was assembled at the National Center for Biotechnology Information in April 2007 (Build 37). We currently have around 20 clones to finish, to close the remaining gaps on chromosomes 2, 4, 11 and X.



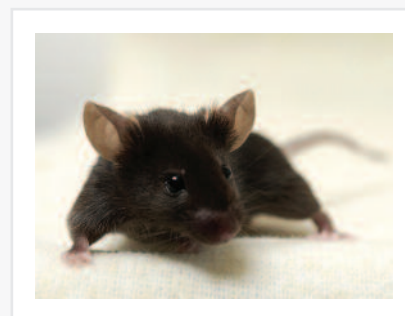
Wellcome Library, London



Wellcome Library, London



Wellcome Library, London



Wellcome Library, London

Zebrafish and mice are used as model organisms by geneticists to test their theories of inheritance and gene function.