

The model organism programmes generate and make freely available a range of resources to accelerate research worldwide. These include a large collection of zebrafish carrying specific mutations, genetically modified mouse embryonic stem cells and other resources to support genetic analysis of these pivotal model organisms. In-house research focuses on the biological function of genes identified in other Sanger Institute programmes or those of potential importance to human health and development.

- ↘ Experimental cancer genetics
- ↘ Recessive genetic screens
- ↘ The Mouse Genetics Programme
- ↘ Genes to Cognition Programme
- ↘ Cancer genes in the mouse
- ↘ Genetics of instinctive behaviour
- ↘ Mouse developmental genetics and ES cell mutagenesis
- ↘ Genetics of deafness
- ↘ Haematological cancers
- ↘ Vertebrate development and genetics
- ↘ Cell surface signalling laboratory
- ↘ The Zebrafish Genome Project
- ↘ Zebrafish Mutation Resource



Hunting cancer genes

Combining human data with experimental studies in mice has revealed a multitude of new cancer-causing genes.

Although a mass of information has been gathered about human cancer genomes, interpreting genomic data can be tricky – cancer cells are a genetic disaster area, and it is rarely clear which mutated genes are contributing to cancer and which simply represent collateral damage. But collaborative efforts between Sanger Institute teams have shown that combining human analyses with studies on mice can focus attention on the truly important genes.

The collaboration looked for genes that were mutated in both naturally occurring human tumours and experimentally induced mouse cancers. Human genes were identified by screening nearly 600 human cancer cell lines for regions of the genome that had been deleted or duplicated. These aberrations typically span a group of genes, any one of which could be a critical cancer-causing ‘driver’ mutation.

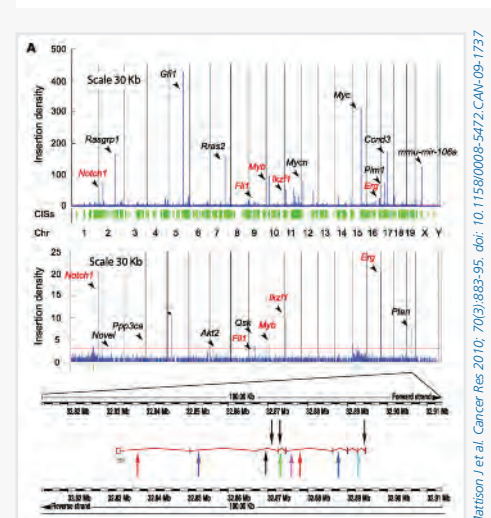
In the mouse work, a virus or a mobile genetic element (known as *Sleeping Beauty*) was used to disrupt genes at random points in the genome. The advantage of this approach is that once a cancer has been identified, the inserted element acts as a ‘tag’, and the disrupted gene can be rapidly pulled out of the genome.

More than 1000 induced tumours were analysed. All told, some 80 genes were mutated in both human and mice tumours, making them highly likely to be *bona fide* cancer-causing genes.

As well as known genes, the study identified several new potential cancer-causing genes. Interestingly, a number of insertions in mice were close to genes controlled by stem cell factors such as Oct4 or Nanog, suggesting that alterations to the transcriptional networks controlled by these factors are particularly likely to generate cancerous cells.

Mattison J et al. Novel candidate cancer genes identified by a large-scale cross-species comparative oncogenomics approach. *Cancer Res* 2010; 70(3):883-95

While human genetics studies continue to generate a wealth of information on genetic contributions to disease, further functional characterisation generally requires experimentally tractable model organisms such as the mouse and zebrafish. An important activity of the Sanger Institute is to generate and distribute resources to facilitate genetic research in these organisms, often as part of systematic coordinated efforts being undertaken by international consortia. The Institute’s impact thus stretches well beyond the research carried out in house, which is itself making important contributions in several biological areas.



Insertions in the mouse genome generated by using murine leukaemia virus and the *Sleeping Beauty* transposon. Around 80 genes were found to be mutated both in naturally occurring human and induced mouse tumours, suggesting that they are likely to be cancer-causing variants.

Faculty members

*Derek Stemple, Acting Head
David Adams
Allan Bradley
Seth Grant
Pentao Liu
Darren Logan
Bill Skarnes
Karen Steel
George Vassiliou
Gavin Wright*

A major achievement has been the completion of the zebrafish reference genome sequence, with publication of the Zv9 assembly in Ensembl release 60. An extensive cDNA sequencing exercise, generating more than 50Gb of data, significantly increased the number of identified protein-coding genes in the genome, which now stands at around 24 000. The Zv9 assembly is a significant improvement, correcting most of the misalignments of previous releases, and is likely to be a much-used community resource. The project has now been passed onto the Genome Reference Consortium for long-term curation.

Next-generation sequencing technologies are enabling more rapid progress to be made on the Zebrafish Mutation Resource project. Technical challenges have limited the speed at which knock-outs can be created, but next-generation techniques should accelerate production during 2011, enabling us to identify more than 1000 new disruptive mutant alleles. Even so, more than 400 alleles have been distributed to researchers in more than 16 countries, and the Resource has been acknowledged in a growing number of papers.

Mutant fish supplied by the Sanger Institute have been used to study a wide range of biological processes, from sleep regulation to cancer development. In-house research has focused on early development and muscle formation, which is shedding light on the mechanisms of vertebrate muscle development as well as genes potentially involved in human muscle-wasting conditions.

Insight into zebrafish muscle development has also come from high-throughput analysis of extracellular protein interactions, using innovative techniques developed within the Institute. As well as revealing significant differences between muscle development in vertebrates and the fly, the most commonly used model, the techniques are also being used to examine interactions between the malaria parasite and the red blood cell.



Dave Sayer, Wellcome Trust

The Zv9 assembly of the zebrafish reference genome sequence is a significant improvement for the research community, by correcting most of the misalignments of previous releases.

Mouse genetics

The Mouse Genetics Programme (MGP) has developed a high-throughput platform for production of mutant mouse lines, and plays a leading role in several large-scale international collaborations. MGP is one of the world's most productive facilities for mouse genetics, and to date has generated more than 450 mutant strains (2 per cent of protein-coding genes). Around half have been phenotyped, with phenotypic data being made available via the web. Phenotyping has revealed numerous interesting features in mutant animals, including in one case altered lipid metabolism and adiposity. There is strong demand for mutant lines from the research community, and some 400 or so lines have been distributed for more detailed analysis.

Targeting of specific genes is possible in mouse embryonic stem (ES) cells. The Sanger Institute is playing a significant role in international consortia systematically knocking out genes in the mouse genome. To date, more than 8000 genes have been targeted to date, with ES cell resources being made freely available to researchers worldwide. Also, a microRNA knockout vector resource has been developed that has been used to create ES cells with mutations in more than 400 microRNAs – more than 75 per cent of known microRNA genes.

A significant advance in recent years has been the development of vectors based on transposable elements, such as *Sleeping Beauty* and *PiggyBac*. Transposons can be engineered into mouse lines and transposition initiated by crosses with mice expressing the appropriate transposase. The payloads of transposons can be modified and the time and place of transposition controlled by use of suitable promoters to drive transposase expression.



Dave Syer, Wellcome Trust

Our Mouse Genetics Programme is one of the most productive facilities in the world, creating mutant strains for 2 per cent of protein-coding genes.

One application has been in the identification of new cancer genes. The transposition of *Sleeping Beauty* during mouse development has been used to generate tumours by disrupting (or activating) genes. Analysis of genes affected in particular tumour types, and comparisons with human data collected by the Cancer Genome Project, can identify likely cancer genes. Such studies revealed more than 900 genes involved in cancer of the colon, as well as genes linked to pancreatic cancer, melanoma and acute lymphoblastic leukemia.

Next-generation sequencing technologies are also providing a way to identify cancer genes directly, often through side-by-side sequencing of mouse and human tumours – emphasising the value of working on both mice and humans.

More recently, the *PiggyBac* transposon has been used in combination with *Sleeping Beauty* to identify cancer genes. Compared with *Sleeping Beauty*, *PiggyBac* can carry more DNA, tends to integrate further away from its point of origin and leaves no 'footprint' when it excises. By controlling expression of its transposase to certain cell types, several novel and interesting cancer genes contributing to haematopoietic cancers and solid tumours have been identified.

From gene to function

The Sanger Institute's expertise in mouse genetics has enabled it to make significant in-house contributions to research. Techniques pioneered at the Institute are being used to engineer ES cells and study the factors controlling stem cell pluripotency. While work into the genetic control of immune cell development has identified a key role for the *Bcl11b* gene in T-cell development. Elimination of *Bcl11b* generates natural killer-like cells with potent anti-tumour properties, suggesting a possible use in cell-based therapies for cancer.

A new area of investigation for the programme is into the genetic basis of olfaction. Mice rely heavily on their sense of smell, which influences a wide range of social behaviours, from suckling to predator evasion. It is thus an ideal system in which to study the links between genes, neural circuits and behaviour.

Hearing is an important sense for both mice and people. Our continuing work to screen mutant mice for defects in hearing, has identified several new genes that affect hearing functions in a variety of ways. More than 300 mutant lines have now been screened, and many of the genes identified would not have been predicted to be involved in hearing, illustrating the advantage of using an unbiased approach to detect relevant mutants.

The Genes to Cognition Programme has continued to probe the origins and function of multiprotein complexes at the synapse. Comparison of synapse proteins across species suggests that the synapse is a very ancient structure and may even pre-date the origins of multicellular animals. Mutations in components of synapse protein complexes have been implicated in a wide range of neurological and psychiatric conditions, illustrating their central role in brain function and behaviour.

Looking forward, there is likely to be a growing emphasis on human induced pluripotent stem (iPS) cells. We have developed methods to reprogramme human somatic cells into iPS cells that closely resemble mouse ES cells. Human iPS cells are likely to become an important model in which to study cellular mechanisms of disease and the impact of genetic manipulations.



A turn for the better

Knocking out a key gene in immune cell development generates cells with potent anti-tumour activity.

'Natural killer' (NK) cells are an important part of early immune responses to invading organisms and also dispose of cells that are in the process of turning cancerous. By eliminating a 'master gene' in immune cell development, an abundant supply of NK-like cells can be created that are long-lasting, seem to have no ill-effects and are highly effective at eliminating cancer cells *in vitro* and *in vivo*.

Cells of the immune system come in myriad forms, each with a different role in defence. As they develop, the fate of individual cells depends on which genes are active within them.

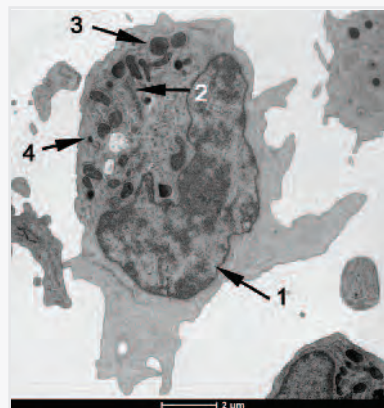
Work is ongoing to identify genes that play critical roles in generating different types of immune cell. Of particular interest is the *Bcl11b* gene, already known to be involved in immune cell function.

Gene knockout experiments revealed that *Bcl11b* was absolutely required for formation of T cells, one of the main categories of immune cell. Without *Bcl11b*, mice produced no T cells.

Instead, the animals produced a type of cell closely resembling an NK cell. Christened 'induced T to natural killer' or ITNK cells, these cells turned out to be have potent effects on cultured lymphoma and melanoma cells and were more effective than 'ordinary' NK cells at preventing the development of cancers in animal models. Moreover, as well as surviving for several months, ITNK showed no signs of attacking healthy cells.

The research sheds important light on the genetic mechanisms controlling the development of T cells. Furthermore, in terms of practical application, ITNK cells have great potential as new cell-based therapies for cancer.

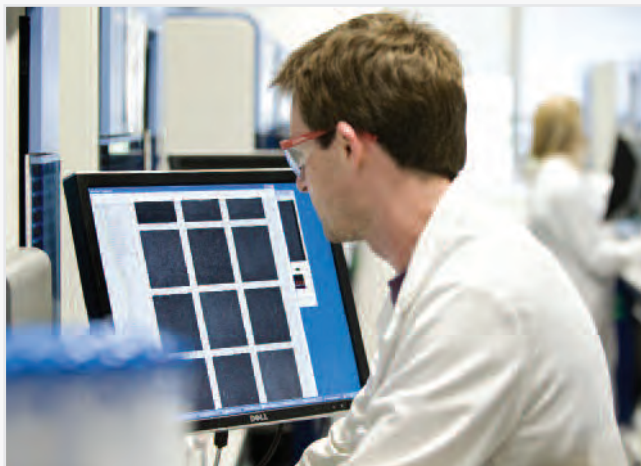
Li P et al. Reprogramming of T cells to natural killer-like cells upon *Bcl11b* deletion. *Science* 2010; 329(5987):85-9.



Li P et al. *Science* 2010; 329(5987):85-9.
doi:10.1126/science.1188063

Transmission electron micrograph of an induced T to natural killer (ITNK) cell. These cells survive for several months and show no signs of attacking healthy cells.

We aim to identify cancer genes by analysing mouse and human cancer genomes, and to understand how these genes participate in tumour formation. In addition, as part of the Mouse Genomes Project, we are delivering high-quality sequence from 17 key mouse strains.



Wellcome Library, London

By applying next-generation sequencing technologies to mouse cancer genomes, we have been able to analyse mice and human cancers side by side.

We are using a combination of mouse and human genetics to identify and characterise new cancer genes. This work involves the use of high-throughput sequencing technologies and genetically engineered mice to understand how the genome changes as a cancer forms.

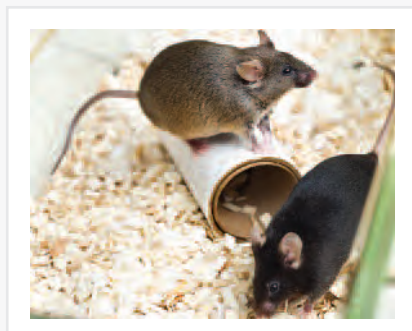
Over the past year we have completed the largest ever genetic screen for cancer genes in the mouse using a mobile genetic element called a transposon. This screen allowed us to identify over 900 genes involved in cancer of the colon. In addition, we have also performed experiments to identify genes involved in pancreatic cancer, melanoma and several haematopoietic malignancies, including acute lymphoblastic leukaemia. We have also been the first to apply next-generation sequencing to the analysis of mouse cancers, profiling the genomes of mouse and human cancers side by side. Using this approach we have identified genes that are mutated in humans and in the mouse which makes them strong candidate cancer genes, but also determined that different mutagenic processes are operative suggesting that there are fundamental differences in how cancers evolve in these in mouse and man.

In collaboration with Great Ormond Street Hospital in London, the University of Leeds, Cardiff University and the University of

Oxford, we are also exome sequencing DNA from families who have unknown high-penetrance cancer predisposition syndromes. Using this approach, we have identified candidate mutations for several conditions.

In addition, we have sequenced the genomes of 17 key mouse strains as part of the Mouse Genomes Project (<http://www.sanger.ac.uk/resources/mouse/genomes/>) and also produced draft sequences for a further 13 mouse strains.

More recently we have generated the first collection of transcriptomes for these mice, data which we have released to the community. We continue to develop resources for researchers world-wide, including indexed collections of bacterial artificial chromosomes for mouse strain genomes and software tools for analysis of high-throughput sequencing data.



Dave Sayer, Wellcome Trust

Using mice as our model, we have identified more than 900 genes involved in colon cancer.



Wellcome Library, London

➤ **PARK2, a gene previously associated with Parkinson's disease, is somatically mutated in colorectal cancer.**

Poulogiannis G et al. *PARK2* deletions occur frequently in sporadic colorectal cancer and accelerate adenoma development in *Apc* mutant mice. *Proc Natl Acad Sci USA* 2010; 107(34);15145-50

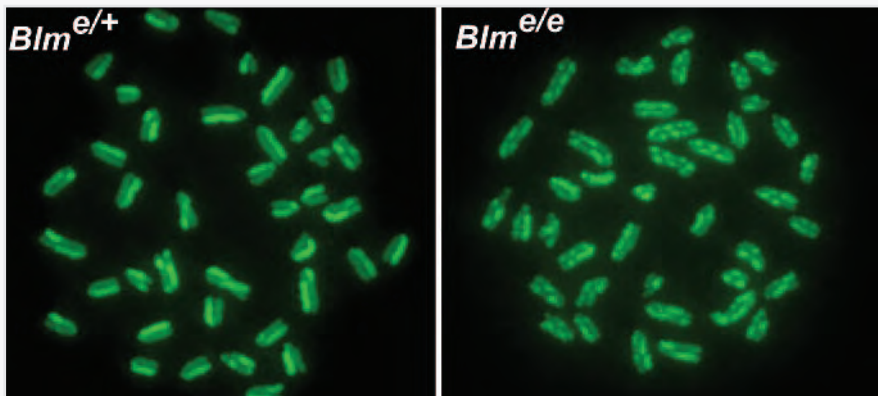
➤ **Deletion of the *IFITM* gene cluster in mice increases susceptibility to influenza A H1N1, West Nile and dengue viruses.**

Brass AL et al. The *IFITM* proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* 2009; 139(7);1243-54

➤ **Pilot for the Mouse Genomes Project confirms that deep sequencing of mouse chromosomes can be used to profile genomic variation and identify candidate genes in quantitative trait loci.**

Sudbery I et al. Deep short-read sequencing of chromosome 17 from the mouse strains AJ and CAST/Ei identifies significant germline variation and candidate genes that regulate liver triglyceride levels. *Genome Biol* 2009; 10(10);R112

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



Meng Li, Genome Research Limited

Increased chromatid exchanges in *Blm*-deficient embryonic stem cell lines elevates the rate of generating homozygous mutants by 20-fold.

Several projects in the laboratory are based on genetic screens in *Blm*-deficient embryonic stem (ES) cell lines. Such lines segregate cells with homozygous mutations at a high frequency, so recessive mutations are relatively easy to recover.

Over the past year we have developed this technology further so that homozygous mutant cells can be isolated by drug selection before genetic screens are carried out. We have also developed more active versions of the *PiggyBac* transposase, which will improve our ability to generate complex libraries of mutants. We have found that *PiggyBac* can mobilise fragments of DNA of more than 100 000 bases in ES cells, which will open up wider use of this mobile element in mammalian genetics and gene therapy. Large genomic fragments have significant advantages in these contexts. A complete genomic locus supports transcription of alternative isoforms. Moreover, regulatory elements which are often fragmented and dispersed across a genomic locus will be captured in large contiguous genomic fragments and will provide normal physiologically regulated gene expression.

We also conduct genetic screens in somatic cells *in vivo*, again using *PiggyBac* mutagenesis, to identify cancer genes. By regulating when and where the transposase is active, we can control mutagenesis, enabling us to initiate cancer in particular organs or even specific cell types within a tissue. We

have developed a series of mice with several different types of *PiggyBac* transposons which can knock out (or turn on) gene activity. When mice with these transposons are bred to mice expressing transposase, their progeny develop a variety of different cancer types. These experiments have already uncovered several new cancer genes.

We continue to contribute to the Knock Out Mouse Programme (KOMP) and EU Conditional Mouse Mutagenesis (EUCOMM) programme. We have assembled a microRNA knockout vector resource and derived ES cells with mutations in more than 400 microRNAs, representing more than 75 per cent of known microRNA genes.

Over the coming year we will begin to develop the *PiggyBac* transposon for use in human induced pluripotent stem cells for genetic screens and for repairing inherited genetic lesions.



Roland Rad, Genome Research Limited

This mouse line was engineered to possess the *PiggyBac* transposon system (red) on chromosome 8. From there transposons start jumping across the genome once activated by the transposase. In this way several new cancer genes in mice have been identified.



Wellcome Library, London

A mouse model of autism with an artificial chromosomal duplication shows behavioural differences similar to those seen in the human condition.

Nakatani J et al. Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. *Cell* 2009; 137(7):1235-46.

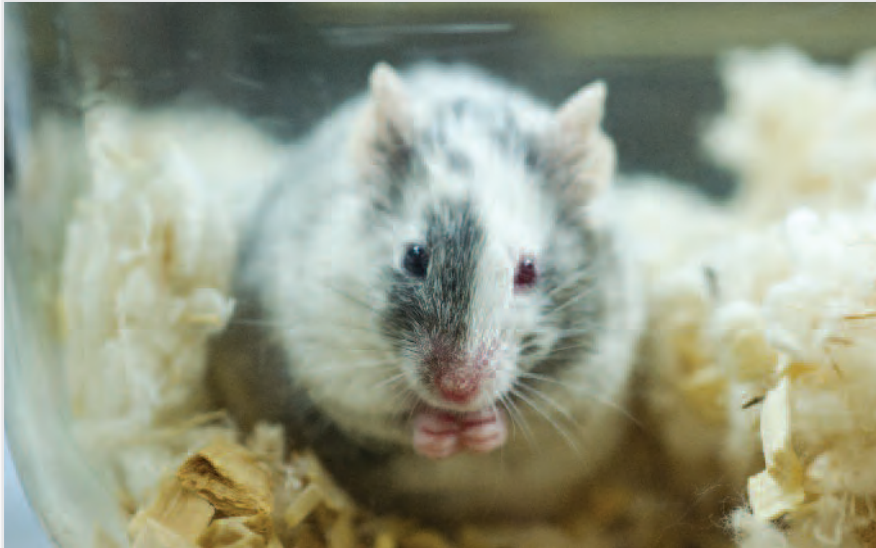
Ten years of work using chromosome engineering has culminated in a faithful genetic model of Down syndrome in which the role of individual genes can be investigated.

Yu T et al. A mouse model of Down syndrome trisomy for all human chromosome 21 syntenic regions. *Hum Mol Genet* 2010; 19(14):2780-91.

The *PiggyBac* transposon can be used as an insertional mutagen in mice, facilitating the identification of new cancer genes.

Rad R et al. *PiggyBac* Transposon Mutagenesis: A Tool for Cancer Gene Discovery in Mice. *Science* 2010; 330(6007):1104-7.

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



Dave Sayer, Wellcome Trust

Chimaeric mouse produced as part of the Mouse Genetics Programme.

Over the past year, the Mouse Genetics Programme has consolidated its international leadership position in the generation, distribution and primary phenotypic characterisation of mutant mice. We have produced more than 450 mutant strains (representing more than 2 per cent of protein-coding genes). The effects of mutation have been analysed in nearly half these strains, providing a rich source of information that is made available to the scientific community via weekly updates at the Sanger Institute's Mouse Resources Portal (<http://www.sanger.ac.uk/mouseportal>).

The scientific community can also obtain the strains themselves, and the past year has seen continued high demand. To date, more than 400 shipments have been made to individual scientists or to repositories for long-term sustainable distribution.

Phenotypic analysis has uncovered numerous novel genotype–phenotype relationships in all major biomedical areas, potentially opening up several interesting new areas of research. Male mice carrying a mutation of the *Tbc1d10a* gene, for example, show a reduction in fat percentage, accompanied by significant reductions in cholesterol and low-density lipoprotein, among other interesting phenotypes.

We have continued to interact with the wider scientific community, seeking input at the gene selection stage and into the range of phenotypic tests undertaken. We are also directly involved in research in several areas, including eye histology, bone, magnetic resonance imaging of embryos, brain morphology, and hair and skin structure. This year, references to the Mouse Genetics Programme began to appear in the scientific literature, and the programme is receiving an increasing number of collaboration requests.

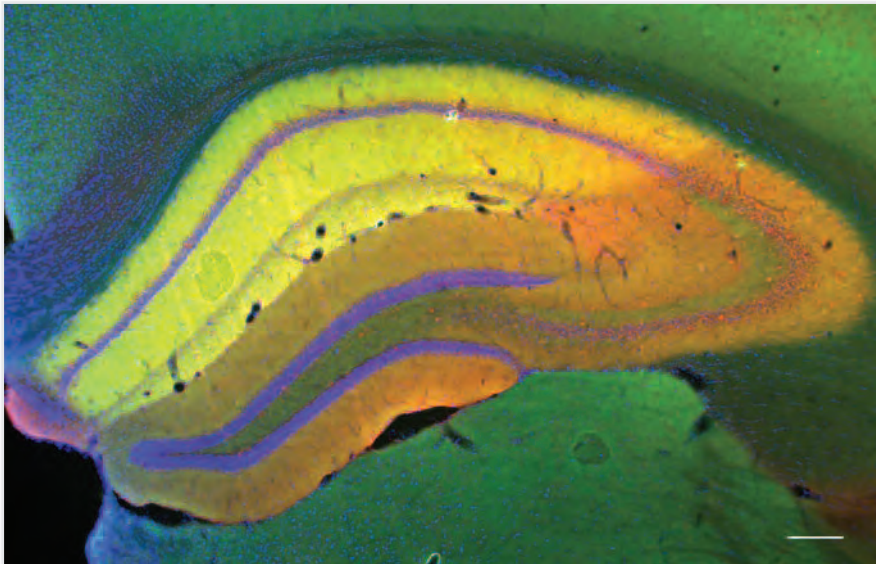
The Mouse Genetics Programme plays a significant role in several large-scale international collaborations, including the Knock-Out Mouse Project (KOMP), the European Union Conditional Mouse Mutagenesis project (EUCOMM) and the European Mouse Disease Clinic (EUMODIC).



Dave Sayer, Wellcome Trust

Mouse tracker in action. The research support facility curates and manages the large numbers of mutant mice lines generated by the Mouse Genetics Programme.

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



Fei Zhu, Genome Research Limited

Immunofluorescent staining of PSD95-EGFP mouse brain. Brain slices from mouse hippocampus were fixed and immunostained for GluR1 (red) and DAPI (blue), combined with direct fluorescence of eGFP(green). In contrast to PSD95-EGFP, GluR1 displays strong staining throughout all subregions of hippocampus, suggesting the differential expression patterns of AMPA receptors and PSD95.

The unifying focus of the Genes to Cognition Programme is the synapse, the microscopic connection between neurons. Synapses are the fundamental functional unit of the nervous system, underlying the plasticity fundamental to learning and memory, emotional responses and many forms of behaviour, including maladaptive behaviour such as drug addiction.

More than 1000 proteins are present at the human synapse, physically linked together in large protein complexes. Our work has focused on the composition of these complexes and the networks of interactions between individual components. We have also examined the functions of individual proteins, particularly those of the postsynaptic density (PSD), a large signalling complex present in many synapses in the brain. These studies are revealing more about how signalling proteins associate in the cell and how intracellular signalling is regulated. They provide a way of linking molecular interactions at the synapse, with changes in neuron behaviour and a corresponding impact on higher-level processes – as seen in work linking the PSD95 protein to inflammatory pain, learning and memory and other forms of plasticity.

Mutations can disrupt the function of signalling complexes at the synapse, resulting in behavioural changes and a range of neurological and psychiatric conditions. Our recent work indicates PSD proteins are involved in more than 100 brain diseases, including common conditions of great economic and social significance such as schizophrenia, autism and Alzheimer's disease. For example, with Nigel Carter's group, we have identified rare copy number variation in individuals with schizophrenia, some affecting genes not previously implicated in the condition.

Our current research is focused on understanding the genomic basis of synapse complexity and diversity and how different diseases arise from mutations affecting synapse proteins. We have created a set of models and computational approaches that will aid in understanding the origins of conditions linked to problems with brain function and in the identification of new therapeutics for such conditions. For more information on the Genes to Cognition Programme please visit <http://www.genes2cognition.org>.



Wellcome Library, London

The human postsynaptic density is involved with over 100 brain diseases.

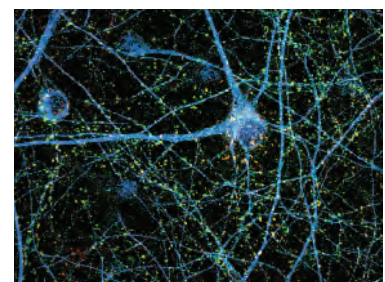
Bayés A et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci.* 2011; 14(1):19-21.

PSD95 protein's role in inflammatory pain is linked to its binding to a key enzyme in intracellular signalling.

Arbuckle MI et al. The SH3 domain of postsynaptic density 95 mediates inflammatory pain through phosphatidylinositol-3-kinase recruitment. *EMBO Rep* 2010; 11(6):473-8.

A review of the evolutionary relationships among synapse proteins and how this influenced the organization of brain complexity.

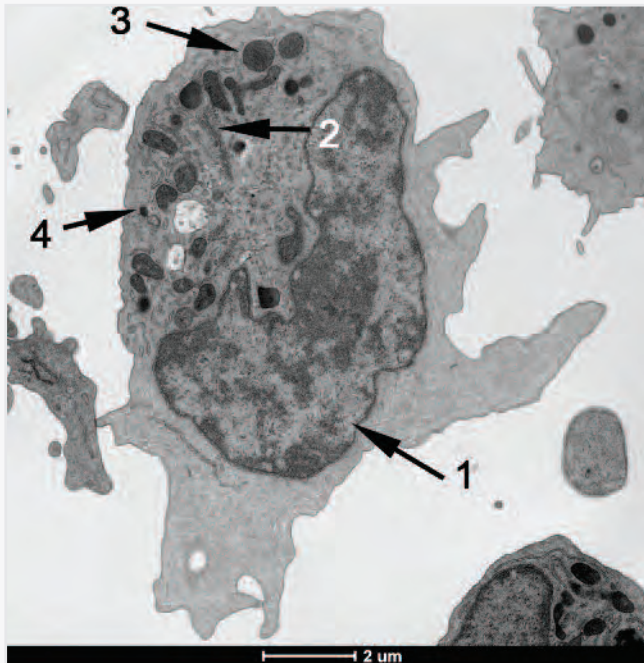
Ryan T, Grant SGN. The origin and evolution of synapses. *Nature Rev Neurosci* 2009; 10(10):701-12.



Fei Zhu, Genome Research Limited

Primary neuronal culture from dissociated PSD95-EGFP mouse cortex and hippocampus. Mutations affecting human postsynaptic density proteins cause more than 130 neurological and psychiatric conditions.

We study cancer genes that play important roles in development.



Li P et al. *Science* 2010; 329(5987):85-9. doi: 10.1126/science.1188063

Transmission electron micrograph of an 'induced T to NK' (ITNK) cell. ITNK cells can be produced in large numbers and have potent effects on tumour cells. 1. Nucleus; 2. Golgi body; 3. Mitochondrion; 4. Granule.

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 transcription factors and are mutated in both human and mouse tumours.

To gain insight into the function of *Bcl11b*, we have deleted the *Bcl11b* gene in cultured T cells. Such *Bcl11b*-deficient T lymphocytes are reprogrammed into cells that resemble natural killer (NK) cells. Loss of *Bcl11b* has a profound impact on T-cell differentiation: T-cell progenitors lose their ability to develop into T cells, and the identities of mature T cells are dramatically altered. *Bcl11b* is thus an essential transcription factor in T-cell development and for maintenance of T-cell lineage identity.

We have called the reprogrammed NK-like cells 'induced-T-to-NK (ITNK)' cells. ITNK cells have potent effects on tumour cells *in vitro* and in the mouse. Unexpectedly, they also kill tumour cells expressing major histocompatibility complex type I molecules, which normally inhibit NK cell killing.

The potent tumour-killing ability of ITNK cells suggests that they could provide a source for cell-based therapies in cancer and infectious disease. We are testing whether loss of *BCL11B* in human T cell would have the similar phenotype.

Over the past few years, we have been developing a new technology to reprogramme somatic cells to induced pluripotent stem cells (iPS cells). We have identified a signalling pathway that is required for reprogramming. Modulating this pathway enables us to rapidly reprogramme mouse fibroblast cells to ground-state iPS cells, and to produce human iPS cells that are similar to mouse embryonic stem cells in morphology, culture behaviour, signalling dependency and receptiveness to genetic manipulation.

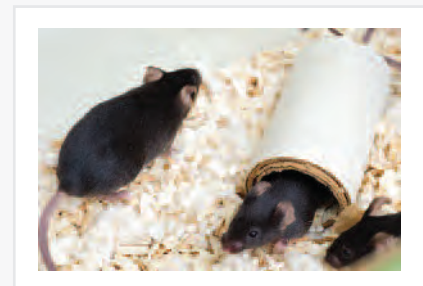
These human iPS cells will enable us to assess the effects of genetic manipulations, in particular those relevant to human disease, on cellular differentiation and differentiated cell function.



Wellcome Library, London

➤ **Deletion of the *Bcl11b* gene disrupts T-cell development and creates cells with potent tumour-killing capabilities.**

Li P et al. Reprogramming of T Cells to Natural Killer-Like cells upon *Bcl11b* deletion. *Science* 2010; 329(5987):85-9.



ITNK cells have been shown to kill tumour cells in mice as well as in the test tube.

Dave Sajer, Wellcome Trust

Using the sense of smell in mice as our model system, we aim to identify the genes and neural circuits that drive social behaviour in mammals.



Dave Sajer, Wellcome Trust

Olfaction is critical for suckling: 97 per cent of mice that cannot smell their mother at birth die within a day of being born.

Understanding the neurobiology of human behaviour is a daunting challenge, due to both the complexity and diversity of our social interactions and the plasticity of our brains. The stereotyped nature of instinctive behaviours such as fear, sex, aggression and feeding suggests they are 'hardwired' and under a strong genetic influence. Our goal is to identify neural and genetic networks that underpin instinct by studying mice that have genetic defects in these behaviours.

In collaboration with colleagues in the USA and Brazil, this year we have completed a project exploring why animals are fearful of predators, even without any prior experience of them. We found that mice lacking *Trpc2*, a gene important for some specialised neurons in their nose, were no longer instinctively afraid of predators. We further isolated the predator-derived signals that activate these fear circuits – proteins found in the urine and saliva of predators such as cats and rats.

Ongoing work is examining how mothers promote suckling in their newborns. There is a widely held theory that suckling is initiated in response to a pheromone present in maternal milk. However, our studies suggest that mice lack a mammary pheromone and newborns instead rapidly learn their mother's signature smell as they are born. Mice that cannot smell their mother die soon after birth.

We have recently begun an ambitious multiyear project to engineer mice that are missing genes for more than 100 specialised olfactory receptors. These receptor genes are grouped together, in clusters, across the genome. The precise function of the receptors is unknown, but they probably detect different types of pheromone. In the coming year, we will produce the first of these engineered mice and begin to test them for abnormal instinctive responses to other animals and purified pheromones.



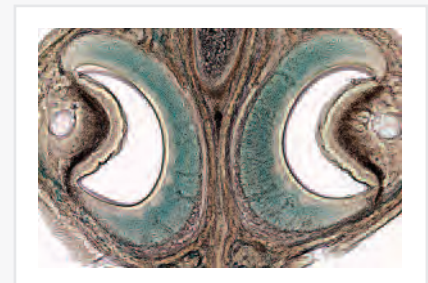
Genome Research Limited

➤ **Mice are instinctively fearful of conserved olfactory signals encoded within the genome of their predators.**

Papes F et al. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* 2010; 141:692-703.

➤ **A review of recent evidence supporting the existence of sensory neurons dedicated to detecting social odours, such as pheromones.**

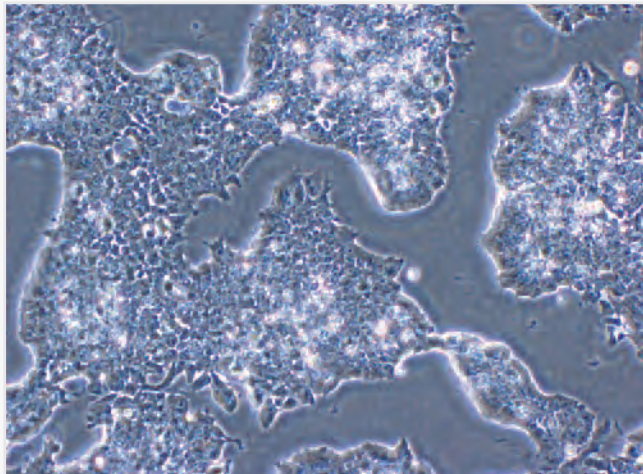
Stowers L, Logan DW. Olfactory mechanisms of stereotyped behavior: on the scent of specialized circuits. *Curr Opin Neurobiol* 2010; 20:274-280.



A coronal section of the vomeronasal organ in the nose of a mouse. Predator cues diffuse through the fluid-filled lumen (white crescents) and are detected by *Trpc2*+ olfactory sensory neurons (blue crescents).

Darren Logan, Genome Research Limited

Our aim is to produce mutant mouse embryonic stem (ES) cells for the scientific community and to use this resource to investigate early cell fate decisions in the mammalian embryo.



Genome Research Limited

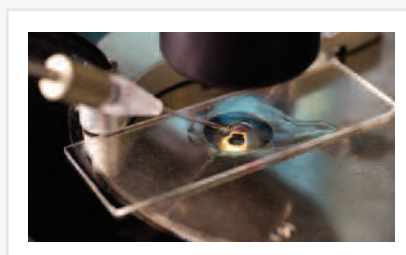
Embryonic stem (ES) cells. A total of 8000 targeted mutations have been generated in C57BL/6 ES cells.

Our large-scale mutant embryonic stem (ES) cell resources are designed to accelerate the functional analysis of genes in the mammalian genome. With substantial funding from the European Union and the US National Institutes of Health, we are developing new technologies for targeting genes on an unprecedented scale. We have established a pipeline for high-throughput targeting of genes, from vector design to archiving of targeted ES cell clones. To date, we have constructed more than 13 000 targeting vectors and have successfully knocked out more than 8000 genes in C57BL/6 ES cells (see <http://www.knockoutmouse.org>).

ES cells represent a genetically tractable model system in which to study basic cell biological and developmental processes on a genome-wide scale. We are particularly interested in the genetic pathways required for ES cell self-renewal and pluripotency. A better understanding of these pathways may suggest ways to direct mouse ES cell differentiation along specific cell lineages. Furthermore, it may ultimately be possible to engineer and reprogramme human stem cells to provide specific cell types for somatic cell replacement therapies.

Genetic screens in ES cells are hampered by the difficulty in generating homozygous mutant cells. Moreover, we need to use conditional strategies to study loss-of-function mutations in genes essential for the growth of stem cells. We have developed highly efficient strategies to generate inducible homozygous mutations in ES cells, with gene activity eliminated by activation of Cre recombinase. This approach is ideal for the study of genes required for stem cell pluripotency and self-renewal.

We have identified several chromatin-associated protein complexes required for ES cell self-renewal, and are now attempting 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 cofactor of one in particular, helping to maintain the stem cell transcriptional programme.



Dave Sayer, Wellcome Trust

Microinjection to produce embryonic stem (ES) cells. An understanding of ES cell self-renewal and pluripotency could suggest ways to reprogramme human cells for therapeutic use.



Wellcome Library, London

An efficient method for generating other useful alleles from targeted ES cell resources.

Osterwalder M et al. Dual RMCE for efficient re-engineering of mouse mutant alleles. *Nat Methods* 2010; 7(11):893-5.

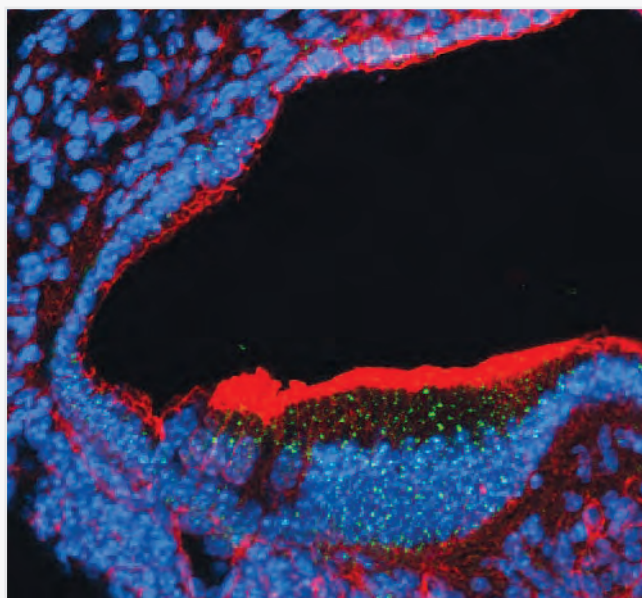
A web portal providing access to detailed information about genetic resources generated by the international knockout programmes.

Ringwald M et al. The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res* 2011; 39(Database issue):D849-55.

Knockout of *Jarid2* reveals an essential role for this chromatin protein in ES cell pluripotency.

Landeira D et al. *Jarid2* is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nature Cell Biol* 2010; 12(6):618-24.

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



Jing Chen, Genome Research Limited

Immunolabelling shows Spns2 protein (green spots) is present in and around the sensory hair cells of the cochlea and also in the stria vascularis at the top of the image.

We are continuing our screen of auditory function in new mutant mice generated by the Sanger Institute's Mouse Genetics Programme, in order to find new genes underlying deafness. So far, we have screened more than 300 new lines using the auditory brainstem response, a physiological measure of responses to sounds.

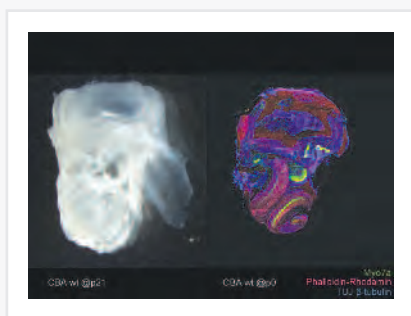
We have found several new genes associated with varying levels of hearing impairment. Affected mice show a range of impairments, including one with severe deafness (*Spns2*), two with poor hearing at high frequencies only (*Acs14* and *Wbp2*), several with milder hearing impairment across all frequencies, two with highly variable hearing thresholds, plus a number with normal detection thresholds but reduced amplitudes of response. None of these genes was suspected of involvement in deafness beforehand, illustrating the value of a broad screening approach. Mutations in any of these genes could potentially be a cause of deafness in humans.

The severe deafness of *Spns2* mutants results from their failure to maintain a high endocochlear potential, the resting potential in the fluid that bathes the top of sensory hair cells. This potential acts as a battery, increasing the flow of ions into hair cells in response to sound stimuli, which in turn triggers synaptic activity at the other end of

the hair cell. We are currently investigating the structure that generates the potential, the stria vascularis, to discover how its function is affected by the *Spns2* mutation.

Deafness can be caused by changes in many different genes required for normal function of different parts of the auditory system, and our work on new mouse mutants continues to reveal new pathological mechanisms underlying deafness. For example, two other mutant lines show inflammation of the middle ear, which can impair conduction of sound to the inner ear.

Phenotypic data on mouse mutants can be found at: <http://www.sanger.ac.uk/mouseportal/>.



Georg Steffes, Genome Research Limited

On the left, a normal mouse inner ear with coiled cochlea at the bottom and balance organs at the top. On the right, the patches of sensory hair cells within the inner ear are labelled in green (*Myo7a*).



Wellcome Library, London

➤ **The homeodomain protein Emx2 is required for establishing the normal arrangement and orientation of cells in the inner ear.**

Holley M et al. *Emx2* and early hair cell development in the mouse inner ear. *Dev Biol* 2010; 340(2):547–56.

➤ **The human connexin 30 T5M mutation leads to mild hearing impairment in the mouse.**

Schütz M et al. The human deafness-associated connexin 30 T5M mutation causes mild hearing loss and reduces biochemical coupling among cochlear non-sensory cells in *Hum Mol Genet* 2010; 19:4759–73.

➤ **Olfactory and reproductive abnormalities are less common in a mouse model of CHARGE syndrome than in patients.**

Bergman JE et al. Study of smell and reproductive organs in a mouse model for CHARGE syndrome. *Eur J Hum Genet* 2010; 18(2):171–7.

We study the pathogenesis of haematological cancers using both *in vitro* approaches and *in vivo* models designed to closely mimic the human diseases, in order to develop targeted treatments.

George Vassiliou's group formed in April 2010, before this time he worked in Allan Bradley's team

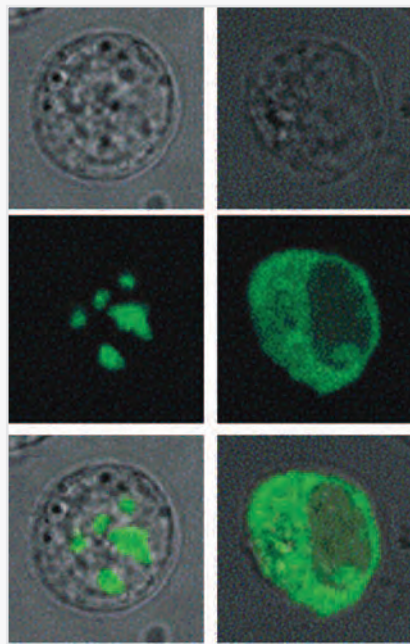
Acute myeloid leukaemia (AML) is a devastating disease with a poor prognosis. The most common type of AML is characterised by somatic mutations in *NPM1*, the gene for the nucleolar shuttle protein Nucleophosmin. These mutations, termed *NPM1c* and uniquely found in AML, lead to aberrant localisation of nucleophosmin to the cytoplasm; however, it is not clear how this promotes the development of leukaemia.

We have generated 'knock-in' mice carrying a conditional humanised form of *NPM1c* mutations. Activation of the mutation in blood stem cells leads to an expansion of myeloid cells and a reduction in mature B cell numbers, with one third of mice going on to develop AML.

Using a novel *Sleeping Beauty* transposon, we have found that genes signalling through the Ras and Stat pathways, namely *Csf2*, *Fli3*, *Rasgrp1* and *Nras*, have a striking ability to enhance the oncogenicity of *NPM1c*. This could explain why such mutations are common in human *NPM1c+* AML and identifies this combination of mutations as a crucial requirement for leukaemogenesis.

We are studying the nature of this interdependency and ways in which it might be targeted therapeutically: using data from whole genome sequencing studies, we are developing mouse models that enable simultaneous activation of all three putative 'AML-driver' mutations in the same cell. This approach aspires to recapitulate the complement of genetic events seen in human AMLs and define the nature of their leukaemogenic interactions, a critical step in designing targeted therapies against this disease. By testing these mutations in three combinations of two as well as all three together, we hope to decipher their collaborative roles.

To identify other cancer-causing genes, we are using *Sleeping Beauty* and *PiggyBac* transposons designed to activate or disrupt genes. Genes recurrently disrupted by transposons in several independent cancers are likely to be important cancer genes.



Nucleophosmin is normally a nucleolar protein (left panels). *NPM1c* mutations displace the protein to the cytoplasm (right panels) by generating a *de novo* nuclear export signal.

Vassiliou G et al. Mutant Nucleophosmin and cooperating pathways drive leukaemia initiation and progression in mice. Nature Genetics; in press.



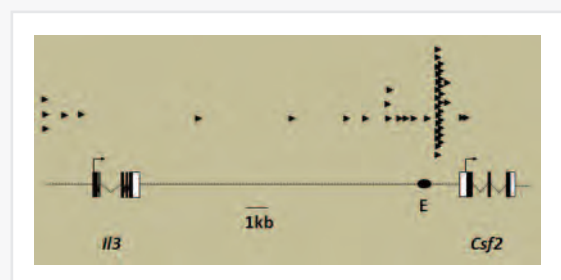
Genome Research Limited

A guide to the practical use of transposons to identify cancer genes in mice.

Vassiliou G et al. The use of DNA transposons for cancer gene discovery in mice. *Methods Enzymol* 2010; 477:91-106.

We have used this method to identify cancer genes that collaborate with *NPM1c* to cause AML and are now using it to identify genes involved in the pathogenesis of myeloma, another common blood-derived cancer.

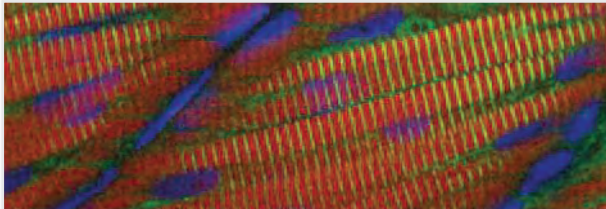
We have also developed a leukaemia transplant method which is enabling us to distinguish 'driver' from 'passenger' mutations in transposon-induced cancers. This is based on our observation that 'passengers' disperse whilst 'drivers' persist through the transplantation process, therefore identifying themselves as such.



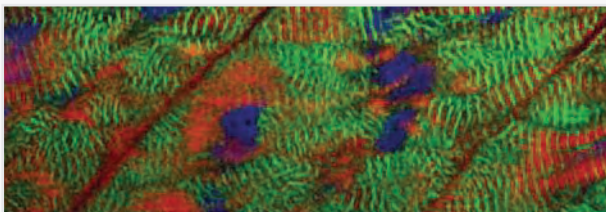
Directional activating *GrOnc* transposon insertions upstream of the *Csf2* gene in 42 of 70 *Npm1c+* mouse AMLs (arrowheads). Most insertions lie between the gene's enhancer and its first exon (E).

Vassiliou G et al. Nature Genetics; in press.

We use genetics and genomics to understand early vertebrate development and to model human genetic disease.



Ferrante MI et al. *J Cell Sci* 2011; 124 (Pt 4):565-77. doi: 10.1242/jcs.071274



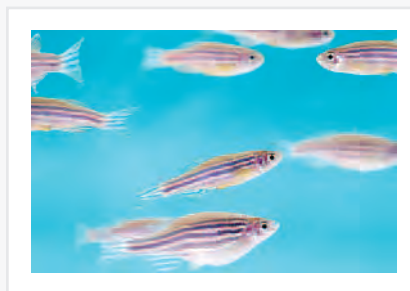
Muscle proteins actin (red) and myomesin (green) (yellow where the two proteins overlap) are shown in normal zebrafish muscle (bottom) and muscle where expression of troponin T3a has been disrupted. Unregulated myosin contraction destroys the integrity of sarcomere organisation in muscle that lacks functional troponin.

Our work focuses on two major areas: how the dorsal organiser and its derived tissues are specified and which genes control their development; and the development and maintenance of muscle. For these two areas, the Sanger Institute Zebrafish Mutation Resource and the *Xenopus tropicalis* Mutation Resource (funded by the US National Institutes of Health) provide mutant strains to complement our genetic studies.

Our studies of the dorsal organiser have revealed an essential role for Nodal signalling in the regeneration and positioning of the organiser during early zebrafish development. We have also shown that the normal up-regulation of coatmer (proteins involved in trafficking of intracellular vesicles) seen in developing embryos is dependent on the unfolded protein response.

In collaboration with Inês Barroso, we have been studying the role of the obesity susceptibility gene *fto* in zebrafish and *Xenopus* embryos. As well as its contribution to weight control, *fto* also has important roles in development. We have found that *fto* is required for normal function of primary cilia, and loss-of-function mutants show a variety of abnormalities, including randomisation of left-right asymmetry, kidney cysts, a down-curved body shape and abnormal numbers of otoliths.

Our work on muscle has shed important light on normal development and potential causes of human disease. A comparison of muscle development in various zebrafish mutants and in human muscular dystrophies has identified several candidate genes for inherited human conditions. Such studies have also provided insight into the basic biology of muscle development, including a surprising role for troponin T in the formation of thin filaments, apparent only after inhibition of myosin contractility.



Dave Sayer, Wellcome Trust

Zebrafish are an ideal model organism for studying muscle development in humans, helping to identify candidate genes for human muscular dystrophies.



Wellcome Library, London

Troponin T is required for the formation of thin filaments in zebrafish muscle.

Ferrante MI et al. Troponin T is essential for sarcomere assembly in zebrafish skeletal muscle. *J Cell Sci* 2011; 124 (Pt 4):565-77.

Knockdown in zebrafish confirms that LRRFIP1, identified in an association study, affects platelet function.

Goodall AH et al. Transcription profiling in human platelets reveals LRRFIP1 as a novel protein regulating platelet function. *Blood* 2010; 116(22):4646-56.

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



Dave Syer, Wellcome Trust

Studies in zebrafish have revealed that muscle development in vertebrates is fundamentally different to that seen in *Drosophila*, which had been the standard model.

We are interested in understanding how cells recognise and communicate with each other. Frequently, this involves interactions between the extracellular regions of membrane-embedded receptor proteins. Despite their importance in both genetic and infectious diseases, such extracellular interactions have been little studied as they are difficult to detect with current large-scale protein interaction techniques. To address this issue, we have developed an assay that enables us to screen for thousands of extracellular protein interactions in a systematic way.

Functional analysis of interactions has focused this year on a receptor–ligand pair expressed in developing vertebrate muscle. Using mutant zebrafish provided by the Sanger Institute Zebrafish Mutation Resource, we have shown that the interaction between these proteins is an essential part of the recognition process that precedes cellular fusion of myoblasts to form multinuclear muscle fibres.

Analysis of mutant zebrafish has revealed that regulation of muscle formation in vertebrates is fundamentally different from that seen in *Drosophila*, which has been the standard model for more than 15 years. As well as revealing a key aspect of vertebrate muscle formation, these findings could also have important implications for the understanding and treatment of degenerative muscle diseases.

We have also made important advances in understanding how the merozoite – the blood stage of the malaria parasite – recognises and invades human red blood cells. A key step in this research has been the development of a method to systematically produce the extracellular regions of merozoite proteins in a functionally active form.

We are now using these proteins and our interaction assay to screen a library of human red blood cell receptors to identify novel host–pathogen receptor–ligand interactions. These proteins will be a valuable resource for the malaria research community, and may suggest new targets for vaccine development.



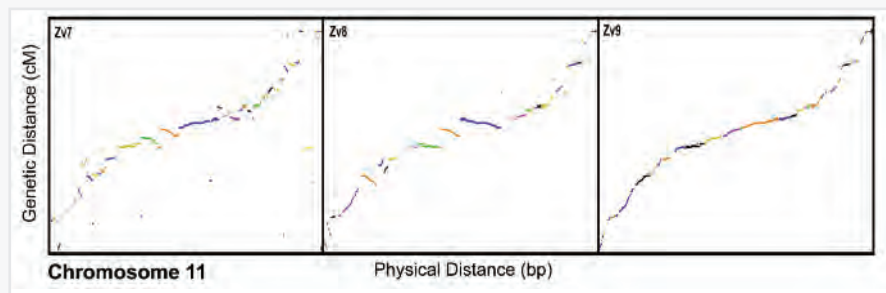
Wellcome Library, London



High-throughput screen for interactions between extracellular zebrafish protein domains reveals extensive network of interactions.

Martin S et al. Construction of a large extracellular protein interaction network and its resolution by spatiotemporal expression profiling. *Mol Cell Proteomics* 2010; 9(12): 2654-65.

We are producing a high-quality zebrafish reference genome sequence to support research in this vertebrate model organism.



Genome Research Limited

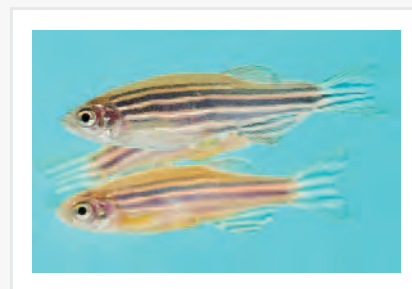
Three views of zebrafish chromosome 11, comparing genetic distance in centimorgans (cM) with physical distance in base pairs (bp) for three assemblies (Zv7, Zv8 and Zv9). Each coloured segment represents a different fingerprint contig. The new high-resolution meiotic map has greatly improved the assembly, with Zv9 showing the expected smooth sigmoidal curve with minimal gaps and inversions.

The zebrafish (*Danio rerio*) genome comprises 25 pairs of chromosomes and around 1.4 Gb of haploid DNA sequence. In 2001 the Sanger Institute took on the task of providing the research community with a high-quality finished genome sequence. Progress has been slower than anticipated because of three major difficulties. Firstly, initial libraries were generated from a group of individuals that, although related, carried a remarkably high degree of polymorphism. Secondly, more than half the genome is made up of repetitive DNA. Finally, the initial maps used to assemble the genome sequence contained many errors.

To provide a reliable framework for reference genome assembly, we have generated a new high-resolution and high-density meiotic map based on two groups of doubled haploid zebrafish (AB-strain and Tübingen-strain). The 0.1 cM resolution of this new map corresponds to a physical distance of approximately 65 kb (smaller than the average clone size). The map was used to position existing clones and contigs in the latest assembly of the Zebrafish Reference Genome, Zv9.

Since only about half of all zebrafish genes are represented by complete cDNA sequence in public databases, we undertook a comprehensive RNA sequencing project to obtain more complete gene coverage. We sequenced more than 50 Gb of cDNA from 10 different developmental stages and adult tissues. In collaboration with the Ensembl team, we have devised algorithms to generate gene models from these data. These data and algorithms have been incorporated into the Ensembl gene-building pipeline. Gene models have been built for the zebrafish Zv9 assembly and we now have evidence for more than 24 000 zebrafish protein-coding genes, including 8000 previously unidentified genes. These models and the fully annotated Zv9 assembly were released in Ensembl version 60.

Our plan is to publish a description of the zebrafish reference genome based on the new assembly, Zv9. As of October 2010, the Genome Reference Consortium, which is responsible for maintenance of the human and mouse reference genomes, has also taken over responsibility for the zebrafish reference genome.

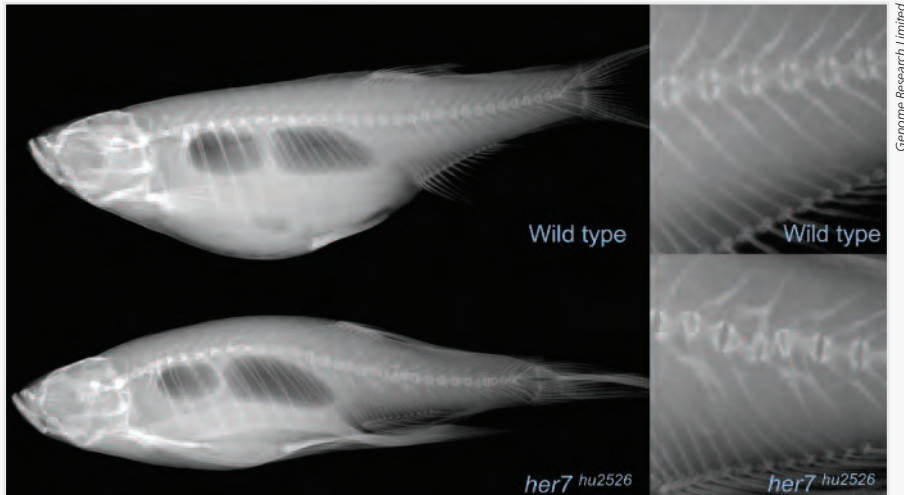


Dave Sayer, Wellcome Trust

We have identified 8000 previously unknown zebrafish genes.



We aim ultimately to generate zebrafish lines carrying mutations in each protein-coding gene, carry out a baseline phenotypic analysis of mutant lines, and make mutant alleles available to the research community.



Genome Research Limited

Comparison of *her7* mutant with wild type. The mutation is homozygous viable but affected adults have a range of bone and vertebral malformations. The Zebrafish Mutation Resource is aiming to generate new nonsense alleles in at least 1000 genes by October 2011.

Over the past three years, we have used capillary sequencing to identify new alleles in around 200 genes, following chemical mutagenesis. By applying next-generation sequencing technology we have dramatically accelerated our discovery rate such that we expect to generate new nonsense alleles in at least 1000 genes by October 2011.

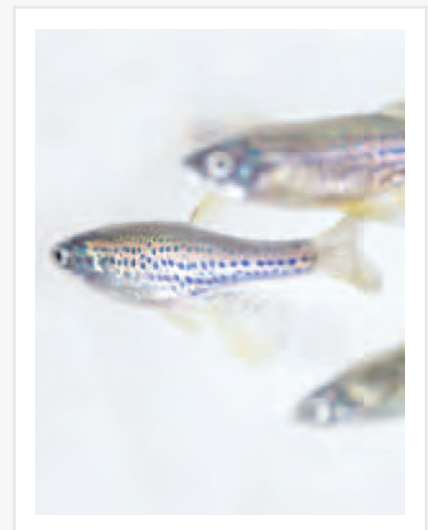
As individual fish are likely to carry multiple mutations, we are exploring comprehensive sequencing strategies to identify newly induced mutations. Our calculations indicate that complete sequencing of individual fish is achievable, but is not currently cost-efficient. We have recently switched to whole-exome sequencing and analysis, which has further enhanced our mutation discovery capability.

In addition to mutation discovery, we also carry out a first-pass phenotypic analysis for each gene, including morphological screening and transcriptional profiling. Once genotyped, carriers are mated and progeny are scored for defects during the first five days of development. We initially planned to use a custom microarray of about 3000 genes whose expression changes during the first five days after fertilisation, but have recently switched to direct transcript

counting using next-generation sequencing technology. This approach will produce a genome-wide and more accurate measurement of the effect of mutations on gene expression early in development.

Data are made public through databases such as ZFIN, the zebrafish model organism database, and ArrayExpress. Mutant alleles are made available to the research community for the cost of shipping. Thus far we have made more than 400 shipments to researchers around the world and we are providing the Zebrafish International Resource Center with alleles for wider distribution.

Fish supplied by the Zebrafish Mutation Resource have been used to study a wide range of biological processes, including cancer development, small RNA function, heart development and sleep regulation.



Dave Sayer, Wellcome Trust

We have shipped more than 400 mutant alleles to researchers worldwide.