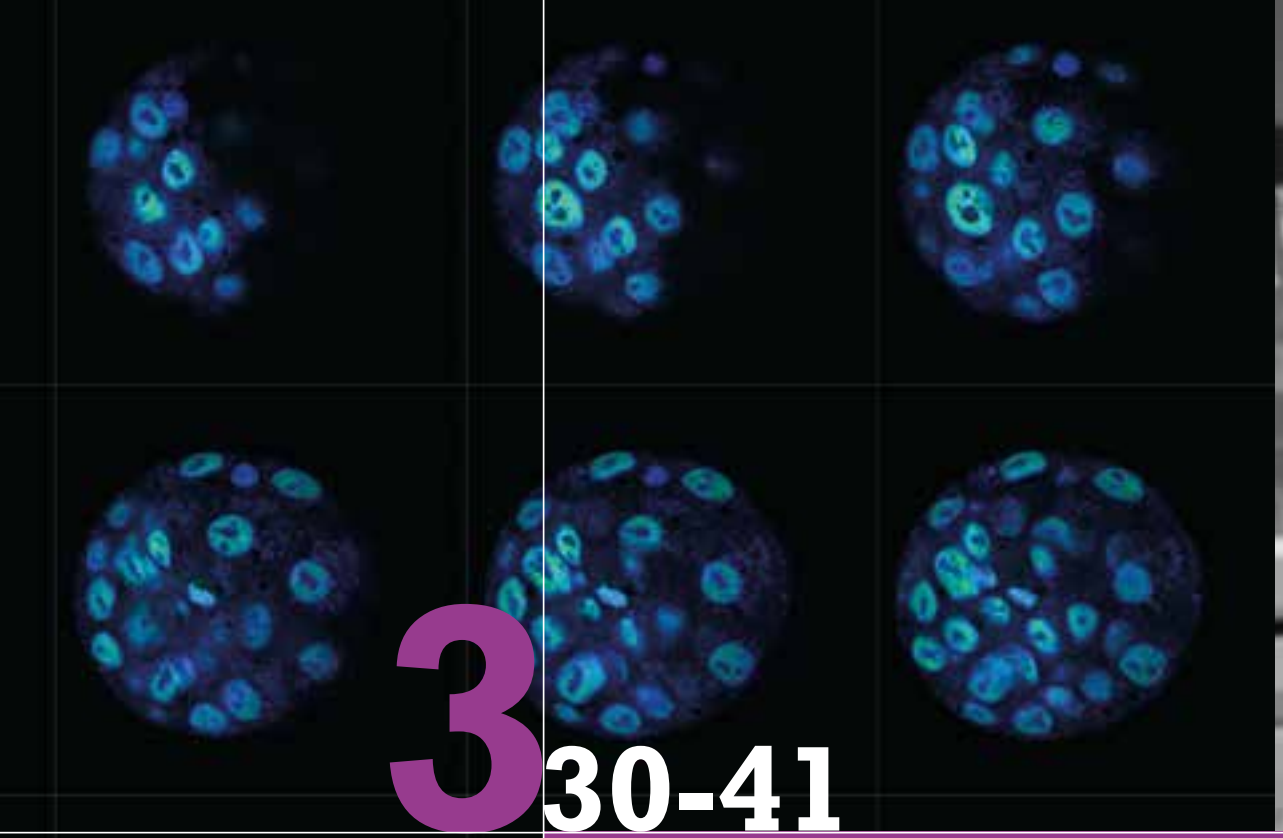
Babraham Institute

2018

Babraham Institute
Annual Research Report
Epigenetics



3 30-41

Epigenetics

Inside cells, genetic information stored in DNA is packaged by proteins into a structure called chromatin. Epigenetics is the study of chemical modifications to DNA and to chromatin and the effects that these modifications have on genome function. Epigenetic marks are involved in the creation of different types of cells from stem cells and epigenetic changes over time are associated with ageing. Epigenetic marks also provide a form of cellular memory, recording certain information about past events and potentially carrying it between parent and child.

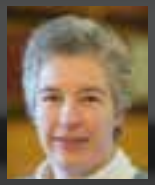
Our work in this area aims to enhance our understanding of how epigenetics shapes human development and affects healthy ageing by examining:

- How stem cells develop into different types of cells
- How subtle epigenetic differences influence cell diversity
- The impacts of diet on epigenetics, health and ageing
- The inheritance of epigenetic memory between generations
- How life events affect biological ageing through the epigenetic clock
- New approaches and technologies to drive further progress

Group leaders

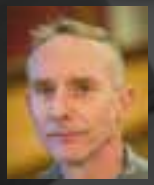
Wolf
Reik

Olivia
Casanueva



Myriam
Hemberger

Jon
Houseley



Gavin
Kelsey

Peter
Rugg-Gunn

Stefan
Schoenfelder

Wolf Reik
Programme leader

Group members

Senior research scientists:
Stephen Clark
Wendy Dean
Melanie Eckersley-Maslin
Fatima Santos
Ferdinand Von Meyenn
(Left in 2018)

Research fellow:
Carine Stapel
(Started 2018)

Postdoctoral researchers:
Irina Abnizova
Rebecca Berrens (Left in 2018)
Poppy Gould (Left in 2018)
Irene Hernando Herraez
Nelly Olova
Aled Parry (Started in 2018)
Solenn Patalano

PhD students:
Celia Alda
Diljeet Gill
Oana Kubinyecz
(Started in 2018)
Georgia Lea
Tim Lohoff
Juliette Pearce
Julia Spindel

Research assistant:
Laura Benson
(Started in 2018)

Visiting scientists:
Romina Durigon
(Started in 2018)

Single-cell epigenome landscape of development and ageing

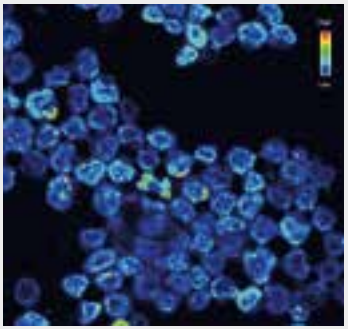
We are interested in epigenetic mechanisms in mammalian development and ageing. Epigenetic marks are able to regulate gene expression and can behave as a form of memory that records the history of a cell. These marks include chemical changes to DNA or DNA-associated proteins. We are particularly interested in the epigenetic rules that govern cell fate decisions in early development, and how cell fate degrades during ageing. Our research uses single-cell sequencing methods to investigate cell fate decisions at the level of individual cells.

Current Aims

The group's research focuses on understanding how cell fate decisions are first programmed or primed, and which epigenetic layers this involves. We would then like to know which cell signalling-induced epigenetic rearrangements occur during fate change. Finally, we are exploring whether epigenetic memory keeps cell identity of fully differentiated cells intact for the rest of our lives (or at least until we start to age). We are also working on identifying DNA binding proteins which are involved in epigenetic priming of enhancers or promoters for future lineage-specific gene expression.

Progress in 2018

In order to address these questions we are developing single-cell multi-omics sequencing approaches which can reveal molecular hierarchies involved in fate decision-making. Our most advanced method combines sequencing of the transcriptome, the methylome, and chromatin accessibility from the same single cell. Our collaborators have developed computational methods by which biologically meaningful relationships between the regulatory layers can be identified. In cells that exit from pluripotency and prepare for differentiation, we made the surprising observation that cells become hugely heterogeneous in their methylation patterns especially in enhancers, and this may be associated with transcriptional heterogeneity (or noise) which may help with cell fate decisions. We are also aiming to connect epigenetic marks in enhancers with histone dynamics which may be important for dynamic gene regulation.



Mouse embryonic stem cells (ESCs) stained for DNA methylation, and pseudo-coloured according to signal intensity (higher signal - red; lowest - blue), revealing the heterogeneity of this epigenetic mark.

Selected Impact Activities

- Members of the group formed a core part in developing and delivering the Institute's exhibit: Race Against the Ageing Clock at the 2018 Royal Society Summer Science Exhibition.
- Wolf Reik was featured in an article on big data 'How big data is changing science' for the Wellcome Trust's Mosaic platform, which was subsequently featured in the Independent online on 11th November, reaching over 22M people.
- A visiting employee from Shift Biosciences has been based in our lab over the past year to apply the lab's epigenetic ageing clock model to drug discovery.

Publications

www.babraham.ac.uk/our-research/epigenetics/wolf-reik

@ReikLab

- Rulands, S. *et al.* (2018) Genome-scale oscillations in DNA methylation during exit from pluripotency. ***Cell Syst.*** 7(1):63-76
- Clark, S.J. *et al.* (2018) scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. ***Nat. Commun.*** 9(1):781
- Raiber, E.A. *et al.* (2018) 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. ***Nat. Chem.*** 10:1258-1266

Olivia Casanueva

Group members

Senior research assistant:
Sharlene Murdoch

Postdoctoral research scientists:
Laetitia Chauve
Cheryl Li (Left 2018)
Celia Raimondi
Boo Virk (Left 2018)

PhD students:
Janna Hastings
Abraham Mains (Left 2018)
Manusnan Suriyalaksh

Research assistants:
Francesca Hodge
Sheikh Mukhtar

Visiting students:
Fatemeh Masoudzadeh
Pia Todtenhaupt (Left in 2018)

Other visitors:
Rebecca Aldunate (Left in 2018)
Rob Jelier (Left in 2018)

Understanding the interplay between stress and metabolism during early stages of ageing

The discovery that genes control longevity has been quite significant for the understanding of ageing because it changed the view from a gradual stochastic process, to a genetically controlled process that we can interfere with and potentially slow down. Since then, thousands of genes and conditions have been found to influence lifespan, with many of them controlling the way in which organisms deal with external challenges brought by stress and nutrition. Such findings underscore the key interplay between genes and the environment and may explain the high degree of discordance among identical twins. Our lab's interest is to understand the non-genetic influences on lifespan and stress related phenotypes using genetically identical lab strains of *C.elegans* as a model organism.

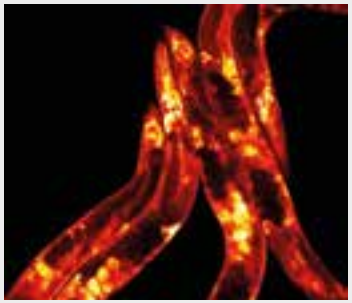
Current Aims

Our overarching aim is to understand the molecular details of ageing and to discover new ways to slow or even reverse the ageing process. With that goal in mind, we use *C. elegans* to understand:

1. The significance of non-genetically encoded variability in the expression of genes that respond to external cues such as temperature and nutrients. We are interesting in finding how early molecular differences in the way worms respond to stress can influence and be predictive of lifespan.
2. How lipid and energy metabolism interplay with signalling pathways that mediate healthy ageing. Metabolism is a key mediator of longevity, however its complexity makes it difficult to study within the particular context of ageing. With this goal in mind, we have developed computational tools to study metabolic fluxes during ageing.

Progress in 2018

- We launched WormJam, a community-driven platform that improved the status of the existing model of *C. elegans* by reconciling and manually curating its metabolic pathways into a single consensus model (refs 1 & 3).
- We also discussed the relevance of these approaches in *Current Opinions in Systems Biology* (Hastings, in print), pointing at one of the key challenges that we face when studying ageing with these tools, namely that the modelling tools available are optimised for animals or cells that are in the process of growing, which is not happening in aged animals.
- Confronted with this challenge, we re-optimised the modelling tool by driving information from multi-omic sources (both transcriptomics and metabolomics) and were able to optimise this tool to study metabolic fluxes during ageing (2). This re-optimisation represents a significant technical advance for the field and will allow more accurate predictions of metabolic fluxes during the course of ageing.



C. elegans worms labelled with a fluorescent protein and imaged using a confocal microscope. C. elegans is a really useful system for our research because we can monitor fluorescent reporters of gene expression in vivo and study inter-individual variability in stress response gene expression in isogenic individuals.

Selected Impact Activities

- The group was involved in several public engagement events throughout 2018, including being part of the team that developed and presented the Institute's 'Race Against the Ageing Clock' exhibit at the Royal Society Summer Exhibition, and sharing the Institute's science at the Cambridge Science Festival and events held as part of the LifeLab project for European Researchers' Night.
- Lab members were involved in organising the second EU-LIFE postdoc retreat.
- The lab coordinates the organisation of local area Cambridge worm meetings.

Publications

www.babraham.ac.uk/our-research/epigenetics/olivia-casanueva

- Witting, M. *et al.* (2018) Modeling meets metabolomics - The WormJam consensus model as basis for metabolic studies in the model organism *Caenorhabditis elegans* ***Front. Mol. Biosci.*** 5: 96
- Hastings, J. *et al.* Multi-omics and genome-scale modeling reveal a metabolic shift during *C. elegans* aging ***Front. Mol. Biosci.*** (in print)
- Hastings, J. *et al.* (2017) WormJam: A consensus *C. elegans* metabolic reconstruction and metabolomics community and workshop series ***Worm*** 6(2): e1373939

Myriam Hemberger

Group members

Senior researcher:
Claire Senner

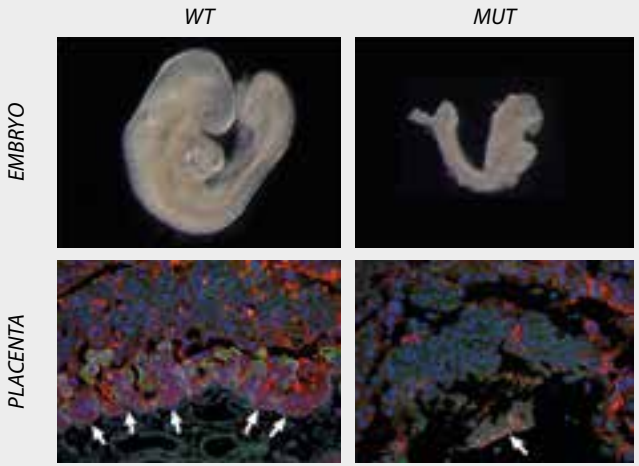
Postdoctoral researchers:
Sarah Burge
Ruslan Strogantsev

PhD student:
Natasha Morgan

Visiting scientists:
Courtney Hanna
Vicente Perez Garcia
Laura Woods

Visiting student:
Dominika Dudzinska

The placenta at the heart of development



Comparison of a normal (WT) and mutant (MUT) mouse embryo that lacks a functional Nubpl gene. The corresponding placentas are depicted underneath. The placenta associated with the mutant embryo is severely malformed and contributes significantly to the developmental defects of the embryo. The placentas are stained for the nutrient exchange surface (green) and for an essential placental progenitor cell type (red, highlighted by arrows).

A functional placenta is critical for normal embryonic development and lifelong health. The placenta is an integrated unit that develops from cells derived both from the embryo and from the mother. Our aim is to gain a comprehensive understanding of the collection of genes that contribute to normal placentation, as well as the maternal factors that affect this process such as advancing maternal age.

Current Aims

Our work focuses on gaining a better molecular understanding of placental development. We use genetic and stem cell models to identify critical factors and pathways involved in the early placentation process. This critical time window is when most defects occur that lead to common pregnancy disorders and developmental defects. We use state-of-the art technologies to manipulate gene function and correlate this with differences in placental cell differentiation and function. We are also integrating the impact of the maternal environment on placental development, and are specifically interested in the changes induced by advancing maternal age and how they affect placental development.

Progress in 2018

It has long been appreciated that development relies on a functional placenta. However, the extent to which the placenta potentially contributes to embryonic defects had remained vastly under-estimated. In 2018, we reported that the placenta is abnormal in about two-thirds of gene mutations that cause embryonic lethality. We also found that placental defects often correlate with abnormal heart and brain development. Thus, the placenta may be a frequent contributor to developmental and birth defects.

In parallel, we contributed to the establishment of a human placental stem cell model, a breakthrough advance that will greatly facilitate research into placenta-based pregnancy complications.

Selected Impact Activities

- Conference organiser: Reproduction and Development Meeting, Cambridge, UK (10-12 March 2018).
- Speaker at the British Society of Developmental Biology (BSDB) Meeting, Oxford, UK (10-13 September 2018).
- Received the Institute's Athena SWAN Best Practice Award 2018 for pioneering initiatives and commitment to women's careers in science.

Publications

www.babraham.ac.uk/our-research/affiliated-scientists/myriam-hemberger @HembergerLab

- Perez-Garcia, V. *et al.* (2018). Placentation defects are highly prevalent in embryonic lethal mouse mutants. ***Nature*** 555: 463-468
- Turco, M.Y. *et al.* (2018). Trophoblast organoids as a model for maternal-fetal interactions during human placentation. ***Nature*** 564: 263-267
- Chrysanthou, S. *et al.* (2018). A critical role of TET1/2 proteins in cell-cycle progression of trophoblast stem cells. ***Stem Cell Rep.*** 10: 1355-1368

Jon Houseley

Group members

Senior research associate:
Cristina Cruz

Postdoctoral researchers:
Anna Channathodiyil
Ryan Hull
Alex Whale

PhD students:
Dorottya Horkai
Andre Zylstra

Research assistant:
Michelle King

Visiting students:
Sebastian Parker (Left in 2018)
Fabiola Vacca (Started in 2018)

How cells interact with their environment

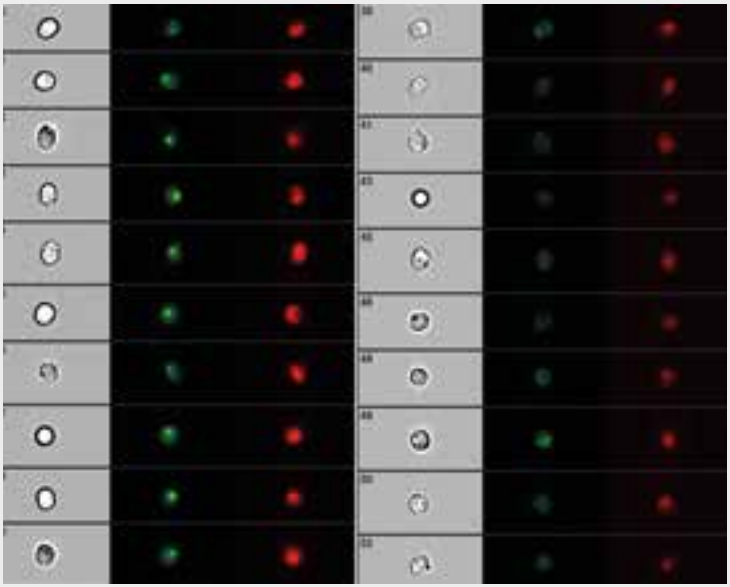
We study how cells adapt to their environment at the genetic and epigenetic level, particularly how they adjust to challenging and toxic environments. This contributes to our understanding of how our cells change in response to environmental pressures and as a consequence of ageing. Our work aims to discover ways of improving health throughout life and to find better approaches to chemotherapy.

Current Aims

- To determine how novel mutations occur and whether these can be stimulated by the environment.
- To establish when and how drug resistance mutations occur in cancer cells.
- To understand the mechanistic link between nutrient environment and the ageing process.

Progress in 2018

Our optimisation of yeast ageing methods has finally borne fruit, allowing us to show that certain epigenetic marks become important as cells age to facilitate appropriate gene expression patterns. We are now able to functionally profile multiple epigenetic marks during the chromatin upheavals that are thought to accompany ageing in all eukaryotes, allowing new insights into the relationships between environment, diet and ageing. Work on non-random mutation is also progressing; we are elucidating new pathways of extrachromosomal DNA formation in response to environmental change and setting up industrial collaborations to address mechanisms of drug resistance acquisition.



Images of very old yeast cells that have been aged under unhealthy (left) or healthy diets (right). Red and green channels show two fluorescent markers of ageing that allow us to quantify ageing pathology. Images acquired using an ImageStream imaging flow cytometer (within the Institute's Flow Cytometry facility) by Dori Horkai.

Selected Impact Activities

- Members of the lab participated in the 2018 Royal Society Summer Exhibition 'Race Against the Ageing Clock' exhibit.
- Jon Houseley presented on extrachromosomal DNA formation at the EMBL Molecular Evolution and Ecology meeting.
- The lab opened a Twitter account - @HouseleyLab.

Publications

www.babraham.ac.uk/our-research/epigenetics/jon-houseley

@HouseleyLab

- Cruz, C., *et al.* (2018). Tri-methylation of histone H3 lysine 4 facilitates gene expression in ageing cells. ***eLife*** 7, e34081
- Frenk, S. & Houseley, J. (2018). Gene expression hallmarks of cellular ageing. ***Biogerontology*** 19(6): 547-566
- Cruz, C. & Houseley, J. (2018). Protocols for northern analysis of exosome substrates and other non-coding RNAs. ***Methods Mol Bio.***, in press

Gavin Kelsey

Group members

Research fellow:
Antonio Galvao

Postdoctoral researchers:
Hannah Desmond
Elena Ivanova

PhD student:
Gintare Sendzikaite

Visiting scientists:
Zahra Anvar (Left in 2018)
Salah Azzi (Left in 2018)
Joomyeong Kim (Left in 2018)
Evelyne Oller (Left in 2018)
Laura Saucedo-Cuevas (Left in 2018)

Visiting student:
Anna Townley (Started in 2018)

Epigenetic legacies from eggs

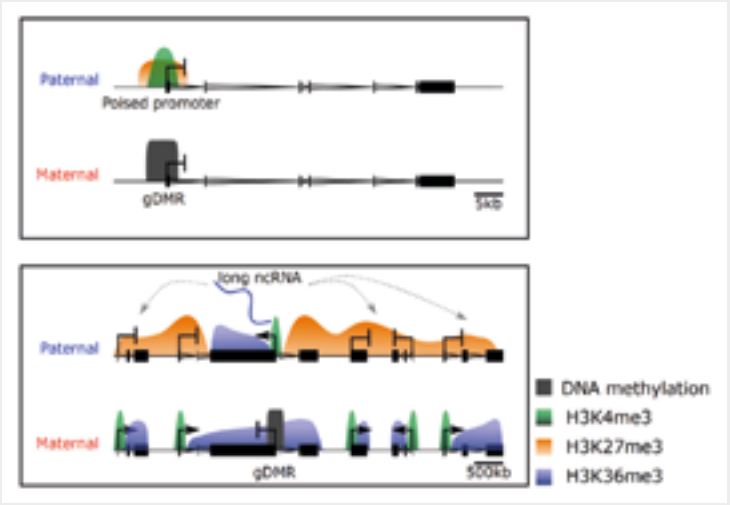
As well as genetic information, the egg and sperm also contribute epigenetic annotations that may influence gene activity both before and after fertilisation. We examine epigenetics during egg development and the effects of epigenetic marks on gene activity in the embryo. Our goal is to understand whether, through epigenetics, factors such as a mother's age or diet have consequences on the health of a child.

Current Aims

We investigate how epigenetic states are set up during oocyte development and influence gene expression in the embryo; for example, how repressive chromatin marks in oocytes lead to long-term silencing of maternal alleles particularly in cells that will form the placenta. We are also interested in how variations in DNA methylation can come about in oocytes and whether we can use methylation variation as a marker for oocyte quality and embryo potential. To investigate these questions, we develop methods to profile epigenetic information in small numbers of cells or even single cells.

Progress in 2018

Using these sensitive methods, we have developed high-resolution epigenetic maps of mouse embryos shortly after implantation, in which we can separate the epigenetic information obtained on maternal and paternal chromosomes (alleles). This has allowed us to distinguish imprinted genes – genes that express a single allele – that depend on DNA methylation conferred in the oocyte from those that depend on repressive chromatin in oocytes. Importantly, we are beginning to see a pattern to explain how chromatin-dependent imprinting is controlled and why it persists selectively in extra-embryonic tissues.



Alternative roles for the repressive chromatin mark H3K27me3 in controlling genomic imprinting. In the upper panel, a paternal allele is silenced by H3K27me3 at its promoter but poised for later expression. In the lower panel, extensive domains of H3K27me3 dependent on monoallelic expression of a long non-coding RNA repress multiple genes.

Selected Impact Activities

- Participation of multiple lab members in the 2018 Royal Society Summer Exhibition 'Race Against the Ageing Clock'.
- Participation in the Cambridge LaunchPad partnership for STEM engagement (Hannah Demond).
- Co-organisation of 2018 EU-LIFE postdoc retreat, Wellcome Sanger Institute, November 2018 (Hannah Demond).

Publications

www.babraham.ac.uk/our-research/epigenetics/gavin-kelsey

- Hanna, C. W., Demond, H. & Kelsey, G. (2018) Epigenetic regulation in development: is the mouse a good model for the human? ***Human Reproduction Update*** 24: 556-576
- Clark, S. J., *et al.* (2018) Joint profiling of chromatin accessibility, DNA methylation and transcription in single cells. ***Nat. Commun.*** 9:781
- Hanna, C. W. *et al.* (2018) MLL2 conveys transcription-independent H3K4me3 in the oocyte. ***Nat. Struct. Mol. Bio.*** 25: 73-82

Peter Rugg-Gunn

Group members

Senior researcher:
Clara Novo

Postdoctoral researchers:
Mandy Collier
Claudia Semprich
(Started in 2018)

PhD students:
Adam Bendall
(Started in 2018)
Charlene Fabian
Andrew Malcolm
(Started in 2018)

Visiting scientist:
Paola Serena Nisi
(Left in 2018)

Epigenetic regulation of human development

How DNA is packaged in cells and the use of biochemical switches in the genome are key aspects of the epigenetic control of gene activity. We are interested in understanding how epigenetic processes are established during human development and during the differentiation of stem cells to form various cell types. This is important for understanding health and for finding ways to use stem cells in regenerative biology.

Current Aims

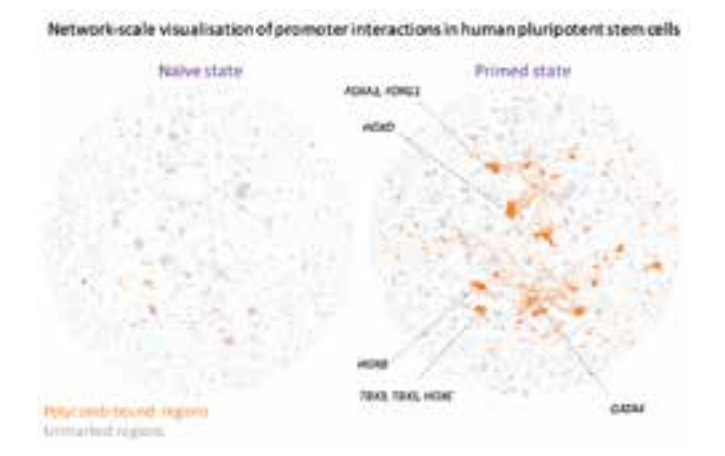
We seek to define the epigenetic and gene regulatory mechanisms that operate in unspecialised pluripotent stem cells and in cells transitioning towards more specialist cell types. We examine how these mechanisms are established in development, how they control cell state changes, and how their alteration can be helpful to reprogramme mature cell types back into an unspecialised form. Our current work also investigates how changes to epigenetic marks can alter the organisation of DNA interactions and associated gene activity. Applying this information will allow us to more precisely control cell fate decisions and to better understand the processes that shape human development.

Progress in 2018

One highlight was discovering that very active regulatory regions, called super-enhancers, are brought into close physical proximity to different gene promoters in different cell types through long-range DNA interactions (1). We further showed that the regulatory interactions connecting super-enhancers with their target gene promoters are controlled by cell type-specific transcription factors. Working with the Corcoran and Schoenfelder groups, we have now developed computational methods to study genome interactions at a network scale (see figure). Investigating chromatin topology and activity in pluripotent cells offers new insights into features of gene regulatory control during development and stem cell differentiation.

Selected Impact Activities

- We participated in several public events including Pint of Science and the Cambridge Science Festival, and we contributed to a podcast by the Naked Scientists.
- We all enjoyed being part of the Royal Society Summer Science Exhibition 'Race Against the Ageing Clock' – a very successful and rewarding team effort.
- Peter took over the co-organisation of the Cambridge Epigenetics Club (@EpigeneticsClub).



Visualisation of promoter-capture Hi-C data at a network-scale uncovers global changes in gene regulatory interactions as human pluripotent stem cells transition from a 'naïve' state to a 'primed' state. The example shown highlights the acquisition of long-range, Polycomb-associated DNA interaction networks that contain the majority of genes encoding key developmental regulators. The figure was produced by Peter Chovanec and Amanda Collier, as part of a collaborative project with Anne Corcoran and Stefan Schoenfelder.

Publications

www.babraham.ac.uk/our-research/epigenetics/peter-rugg-gunn

@RuggGunnLab

- Novo, C. *et al.* (2018) Long-range enhancer interactions are prevalent in mouse embryonic stem cells and are reorganized upon pluripotent state transition. ***Cell Rep*** 22: 2615-2627
- Lupo, G. *et al.* (2018) Molecular profiling of aged neural progenitors identifies Dbx2 as a candidate regulator of age-associated neurogenic decline. ***Aging Cell*** 17: e12745
- Collier, A. and Rugg-Gunn, P. (2018) Identifying human naïve pluripotent stem cells - evaluating state-specific reporter lines and cell-surface markers. ***Bioessays*** 40: e1700239

Stefan Schoenfelder

3D genome organisation in stem cells

The three-dimensional organisation of our genome is tightly linked to its function. Dispersed throughout the non-coding part of our genome, regulatory elements (such as enhancers and promoters) function as 'molecular switches' to turn genes on and off. These regulatory elements are brought into spatial proximity through cell-type specific folding of chromatin, which represents a key regulatory mechanism to control gene expression programmes during cell lineage specification.

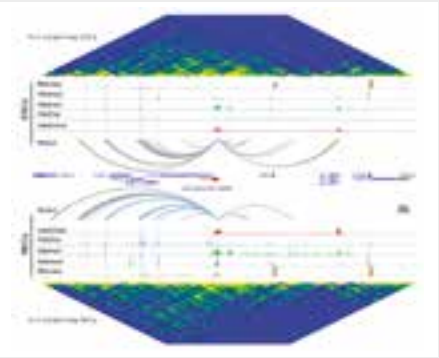
Current Aims

Our aim is to dissect 3D gene regulatory networks and enhancer–promoter interactions during normal development, and to understand how their function is perturbed in disease and ageing. Our research focuses on pluripotent stem cells, as they are a unique model to study early mammalian development, and hold great promise for applications in regenerative medicine, disease modelling and compound screening. We address three main questions:

1. How is the folding of the genome rewired during early mammalian development?
2. Which regulatory elements control cell fate decisions during lineage specification?
3. How do sequence variants in regulatory elements affect gene expression levels and thereby cell function?

Progress in 2018

We have mapped promoter–regulatory element interactions in mouse embryonic stem cells and trophoblast stem cells, demonstrating that profound differences in 3D genome organisation are already established after the first cell lineage differentiation process during early mammalian development. We have also shown that haematopoietic stem cells respond to growth factor signalling through widespread epigenome and transcriptome remodelling, but with very limited changes to promoter–regulatory element interactions. This indicates that priming of promoter–regulatory element contacts may contribute to efficient signalling pathway responses at the transcriptional level. We have further analysed how the 3D genome organisation changes during ageing in mouse B lymphocytes.



Developmental control of enhancer–promoter contacts during early human cell lineage specification. Transcriptional upregulation of PAX6 during the conversion of human embryonic stem cells (ESCs) into neuroectodermal cells (NECs) is accompanied by the formation of enhancer–promoter contacts in NECs (mapped by Promoter Capture Hi-C; PCHi-C), which involves both novel enhancer–promoter contacts and the emergence of the enhancer associated mark acetylation at lysine 27 of histone H3 (H3K27ac) at genomic regions that already interacted with PAX6 in ESCs (modified from Freire-Pritchett et al., eLife 2017).

Selected Impact Activities

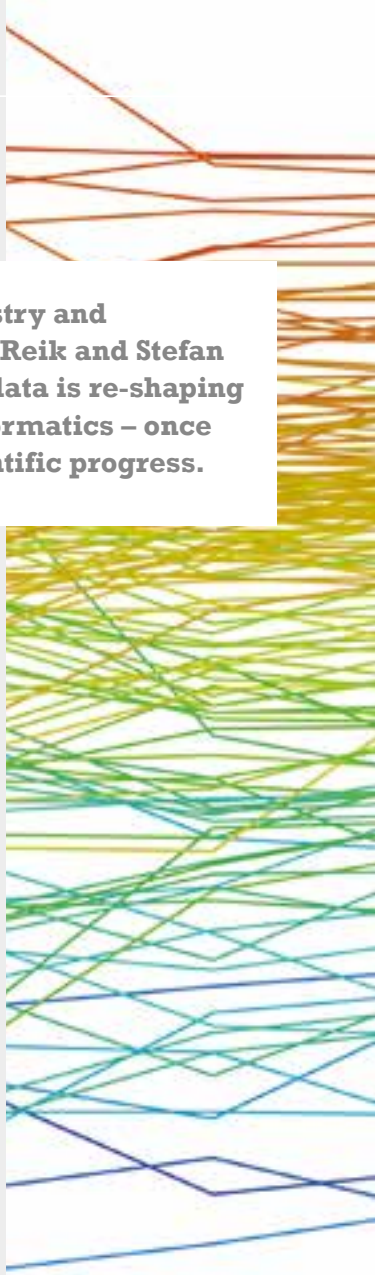
- Presentation at EMBL-EBI Industry Programme workshop 'Epigenetics for ageing and disease' (November 2018).
- Participation in a STEM Insights teacher event hosted at the Babraham Institute (October 2018).
- Interview with Tom Chivers that contributed to an online article 'How big data is changing science' for the Wellcome Trust's Mosaic platform, which was subsequently featured in the Independent online on 11th November, reaching over 22M people.

Publications

www.babraham.ac.uk/our-research/epigenetics/stefan-schoenfelder

- Schoenfelder S. *et al.* (2018) Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. ***Nat. Commun*** 9: 4189
- Schoenfelder S. *et al.* (2018) Promoter Capture Hi-C: High resolution, genome-wide profiling of promoter interactions. ***J. Vis. Exp.*** 136. doi: 10.3791/5732
- Comoglio F. *et al.* (2018) Thrombopoietin signalling to chromatin elicits rapid and pervasive epigenome remodelling within poised chromatin architectures. ***Genome Res.*** 28: 295-309

Riding the data wave



Big data is revolutionising science. But as well as changing physics, chemistry and biology, it's changing the nature of science itself. Institute researchers Wolf Reik and Stefan Schoenfelder and bioinformatics expert Simon Andrews reflect on how big data is re-shaping not only the way they work, but how they think. And we discover how bioinformatics – once considered a geeky corner of biology by some – has become central to scientific progress.

When Professor Wolf Reik, head of the Epigenetics programme, thinks about how big data has revolutionised his field he remembers the Swiss anatomist Karl Theiler. Theiler spent a lifetime creating 'The House Mouse: Atlas of Embryonic Development', painstakingly taking embryos at different stages of development and using staining and microscopy to identify each tissue type.

"Now, we take the same embryos and put them in a big machine which sequences up to 100,000 cells at a time. Through gene expression, it gives us an equally detailed atlas of development, but because we can now use multi-omics methods that link together different layers in a single cell – the epigenome and the transcriptome – we can ask much deeper questions about how these patterns arise mechanistically. That's what we really want to know," says Reik.

Despite being a relic of an earlier scientific age, Theiler's atlas remains on Reik's bookshelves: an illustration of how big data has transformed the scientific questions he can ask and an embodiment of how it's reshaped the way his younger colleagues think.

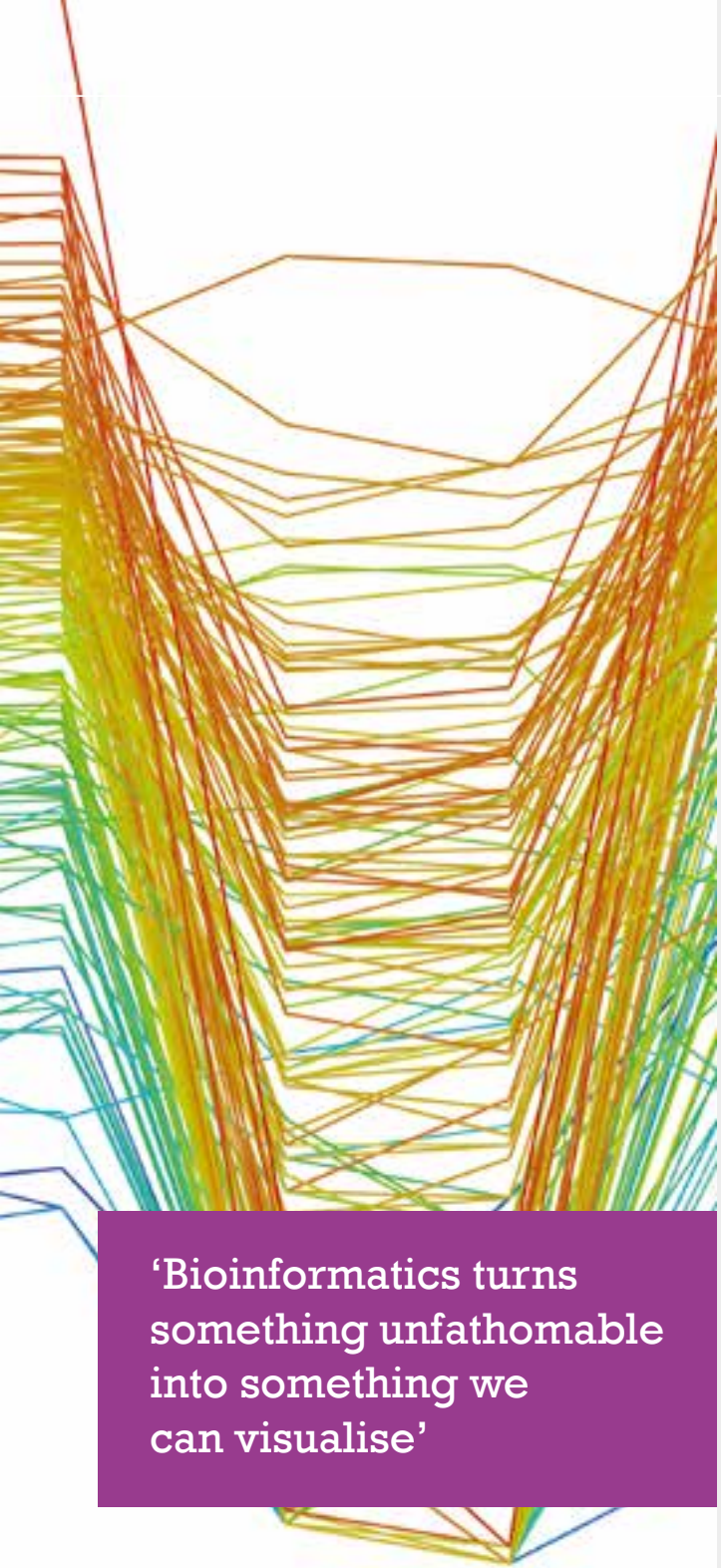
Before big data, researchers thought and worked on single genes – how they were regulated and their role in development, health and ageing. Now, thanks to the recent developments in next-generation sequencing, the focus is firmly on the genome as a whole. "We can now look at 20,000 genes or 20,000 promoters and get huge amounts of information. The younger members of my group get excited about the whole genome and what it's doing, whereas I was brought up in an era of asking what single genes do; it's a fundamental difference in thinking," says Reik.

Big data brings huge opportunities, but using techniques that generate massive amounts of increasingly complex data also presents huge computational challenges. So how do Reik and other researchers extract meaning from this deluge of data? The answer lies in bioinformatics, the science that has emerged at the intersection of biology, computer science and statistics.

Dr Simon Andrews, head of Bioinformatics, belongs to this new breed of experts. Since joining the Institute in 2001, he's seen the group expand from two to 10 staff, many of whom have their roots in biology.

"Lots of people in my group were once biologists who happened to play with computers for fun. My mother was a primary school teacher. Sometimes she'd turn up with a computer that had been donated to the school, point me towards it and say 'make this do something that I can take back into the classroom!'" Andrews recalls. "At university I built my own computers because we couldn't afford to buy them, and when I started research we were beginning to get electronically-generated data."

His PhD generated a respectable 1,000 bases of DNA sequence. Today, a single sample at the Institute yields 40 billion. "The fundamental change is that many experiments generate amounts of data that are



‘Bioinformatics turns something unfathomable into something we can visualise’

‘Big data changes how we think – and how we work’

impossible to understand without a computer. Before, computers were a nice add-on; now they are fundamental,” he says.

One of the Institute’s core facilities, the Bioinformatics group provides computational power and data analysis plus expert advice and bespoke development work. “What fires me up are computational problems that spring from biology,” says Andrews, and what researchers often need most are ways of making their data more accessible. Over several years, Andrews’ group has developed packages capable of visualising sequencing data sets with billions of data points. “These are unfathomable on their own, but we can turn them into billions of positions in a genome, and visualise what they look like,” he explains.

Like Reik and Andrews, Dr Stefan Schoenfelder has lived through the revolution wrought by next-generation sequencing and big data. “It changes the way you think and changes the way we work,” he says. “When I did my PhD 15 years ago I spent all my time doing experiments in the lab. Now it’s the analysis that takes the time.”

Schoenfelder is interested in how gene function and gene expression are controlled by non-coding bits of DNA known as regulatory sequences. In linear terms, genes and their regulatory elements may be some distance apart, so how the genome is organised in three dimensions is one of his key questions. “Whereas we used to look at individual examples, now it’s possible to address those questions genome wide. We can get a complete picture of all the interacting sequences in a cell,” he says. “When I came here after my PhD, it was something I thought might happen at the end of my career. That it’s happened so quickly is incredible.”

It also means that researchers need to learn how to interpret data, so the Institute’s Bioinformatics group makes a major difference. “The skills I was equipped with in my PhD are not enough anymore. It’s normal to keep learning in science, but this is a quantum leap,” says Schoenfelder. “In a competitive field you need to work rapidly. I often work with dedicated bioinformaticians because it’s almost impossible to be an expert in both.”

The next scientific revolution is anyone’s guess, but Schoenfelder is sure it will only underscore how much more we need to understand. “Sequencing and its impact on personalised medicine will continue to grow. High-resolution microscopy, observing live cells and even individual molecules, will be another game changer,” he concludes. “We make contributions all the time, but we know so little. That’s humbling – but it’s also very exciting to be a part of.”

Babraham Institute
Babraham Research Campus
Cambridge
CB22 3AT
UK

www.babraham.ac.uk
Tel: +44 (0)1223 496000
babraham.contact@babraham.ac.uk
@BabrahamInst

The Babraham Institute

Registered office:
Babraham Hall, Babraham,
Cambridge, CB22 3AT
Registered in England and Wales
No. 3011737 as a company
limited by guarantee
Registered Charity No. 1053902

Athena SWAN Silver Award

BBSRC | UK Research and Innovation

The Babraham Institute receives core funding in strategic programme grants from the BBSRC.

Designed & Produced by www.pickeringhutchins.com