

systems biology goes multicellular

Caenorhabditis elegans is one of the most elegant model organisms

by Marlon Stoeckius

Caenorhabditis elegans is a transparent nematode about 1 millimetre long that lives in the soil (Fig. 1). In the 1960s, biologist Sydney Brenner first used it as a model for researching nerve development. In the years that followed, *C. elegans* became established as a model system for cell division and embryonic development. Many fundamental mechanisms in this tiny worm function in a way that is similar to the way they function in humans. As a result, it is now one of the most popular model organisms for developmental biologists and is an equally widely used model system for researching the molecular bases of many diseases, including cancer, diabetes, Parkinson's and Alzheimer's. Another, fairly recent field of research is research into the function of small RNAs, such as microRNAs, which were first discovered in *C. elegans*. Molecular biological methods of examining the nematode are very well established and offer many approaches for analysing fundamental phenomena and mechanisms. This has already resulted in the awarding of three Nobel prizes for discoveries in *C. elegans*.

A toolbox filled to the brim for systems biologists

In recent times, *C. elegans* has also become a model organism of systems biology – mainly because its molecular biological and biochemical toolbox is so lavishly equipped. A consortium of scientists laid the foundations in 1998 when they fully decoded the worm's genome. *C. elegans* was the first multicellular organism to have its genome sequenced (Sequencing Consortium, 1998). Even now, ten years after the human genome was sequenced, the nematode's comparatively small genome (roughly 100 mega base pairs) is still the best-mapped genome of a multicellular organism. Using modern methods, the worm's genome can be manipulated relatively easily. This means that foreign or modified genes can be inserted accurately into the genome (MosSCI; Frøkjær-Jensen *et al.*, 2008). To modify genes, researchers using the worm can make use of "gene banks" where almost all genes (ORFeome; Lamesch *et al.*, 2004) and regulators (Promoterome; Dupuy *et al.*, 2004, 3'UTRome; F. Piano laboratory) from the nematode's genome have been isolated. They can then be freely combined in the manner of a construction kit. Genes in *C. elegans* can thus be comparatively easily disabled by feeding the worm special bacteria (RNAi feeding; Boutros and Ahringer, 2008). Likewise, specific sections can be deleted from the genome (MosDEL;

Fig. 1: *Caenorhabditis elegans*



The one millimetre long nematode *Caenorhabditis elegans* is a popular model organism (Image: Marlon Stoeckius).



Marlon Stoeckius is using the nematode for research into the molecular bases of early embryonic development (Photo: Alexander Baltz @ BIMSB).

Frøkjær-Jensen *et al.*, 2010). The worm can be cultivated on agar plates with up to 384 chambers, as if in separate habitats. This makes it possible to carry out parallel, automated, high-throughput screening whereby approximately every gene of the nematode is disabled individually in just one week. The *C. elegans* embryo and its fully-grown worm are both transparent, and it always develops stereotypically. That makes it possible to observe individual cells in the living embryo and nematode under a microscope. In this way, the origin and the fate of every cell has been described, from the single-cell stage via the hatching larva with 558 cells to the adult worm with 959 somatic cells (Sulston *et al.*, 1983). Nowadays, embryonic development can be monitored automatically with the help of imaging methods. When specific genes in the embryo are deactivated, changes in the characteristic pattern of cell division can thus be recognised automatically. This makes it possible to construct hypotheses about the gene's function. Scientists have generated more than 13,000 genetically manipulated nematodes (CGC, March 2011) in recent decades, with individual genes deleted, mutated or marked with a fluorescent marker. This has produced a huge resource that is shared among scientists in line with the principle of open source access. Unlike many other model organisms, these worms can simply be stored in a freezer until further use. In addition to laboratory method resources, hundreds of datasets from high-throughput experiments are available in databases. A recently completed project for the systematic mapping of regulators in the genome, in addition to establishing new methods, produced several terabytes of data that is now freely available (ModENCODE; Gerstein *et al.*, 2010). Altogether, these resources make *C. elegans* an ideal multicellular model for systems biology.

What is (was) lacking

Modern DNA sequencing technologies permit quantitative analysis of all activated genes (RNA expression) in all organisms. So far, in most model organisms it has not been possible to quantitatively measure proteins as has been done for several years in cell culture (SILAC; Ong *et al.*, 2002). Until now,

therefore, hypotheses about the function of certain genes could only be constructed on the basis of microscopic changes in the worm, of changes in RNA expression, and by measuring a small number of selected proteins.

Ultimately, in order to understand biological processes, the quantity, identity and dynamics of the metabolic products in an organism must be determined quantitatively. Metabolism both in single-celled and in multicellular model systems is an insufficiently developed area of research (Metabolomics; Oliver, 2006). High-throughput research projects involving *C. elegans* require both the optimisation of methods and an innovative approach to data analysis. After the worm has hatched, it can be cultivated in great numbers synchronously throughout its life cycle. As a result, sufficient experimental data (e.g. RNA and proteins) is available for analyses of worms of the same age. Nonetheless, previous quantitative analyses have been based mainly on measurements of the whole organism, which has led to mixing of the different tissues. This makes it difficult, for example, to distinguish which genes are specifically activated in muscles or in nerves. In future, methods must be established that permit selective high-throughput measurements in individual tissues and cells.

Similar hurdles must be overcome for research into embryonic development. The adult nematode carries embryos of different ages inside its body, from the single-celled zygote to the hatching larva with 558 cells. More material-intensive measurements have always been based on heterogeneous populations of embryos at different cell-development stages. Until now, in order to examine the sequence of chronological events during embryonic development, embryos had to be sorted by age under a microscope. When sorting manually in this time-consuming manner it is impossible to isolate the quantities of embryos required for analysing RNA expression or protein expression. With a steady hand, it is possible to collect around 500 embryos a day. However, modern high-throughput analyses require more than 100,000 embryos.

Single-celled *C. elegans* embryos (Image: M. Stoeckius).

Automatic sorting of embryos of the same age

We have developed a method of isolating tens of thousands of embryos at the same stage of development within a short time with the help of a conventional cell sorter (FACS) (Fig. 2). The fundamental prerequisite for this method, which we christened eFACS, is a strain of worm that specifically produces a fluorescent protein only at one unique cell stage. The fluorescence serves as a marker for the desired stage and is specifically recognised by the cell sorter. We use this method to enable us to sort large quantities of single-celled and two-celled embryos. In principle, however, it can be used to isolate embryos at any stage of development. eFACS thus opens the door to systematic research on embryonic development in *C. elegans* using modern high-throughput methods (Stoeckius *et al.*, 2009).

Quantifying thousands of proteins

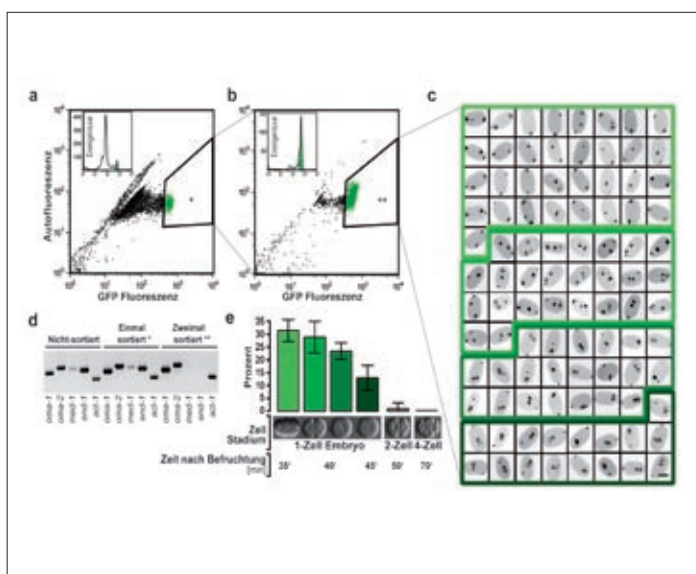
In collaboration with the laboratory of Matthias Selbach (MDC Berlin) we have developed in the laboratory of Nikolaus Rajewsky (BIMSB@MDC) a new method that enables us to quantify the activity and quantity of thousands of proteins by comparing two samples of *C. elegans* (Thierfelder *et al.*, in preparation).

For the first time, this provides insights into changes in the quantities of thousands of proteins, thereby permitting one to draw conclusions about their function in the living organism. It is now possible to examine how protein quantities change throughout the worm's development and how certain proteins behave when a specific gene is deactivated. This method also enables the quantitative measurement of proteins activated by chains of signals (phosphoproteome) and makes it possible to improve the sensitivity of other methods used with the nematode (e.g. Co-IP).

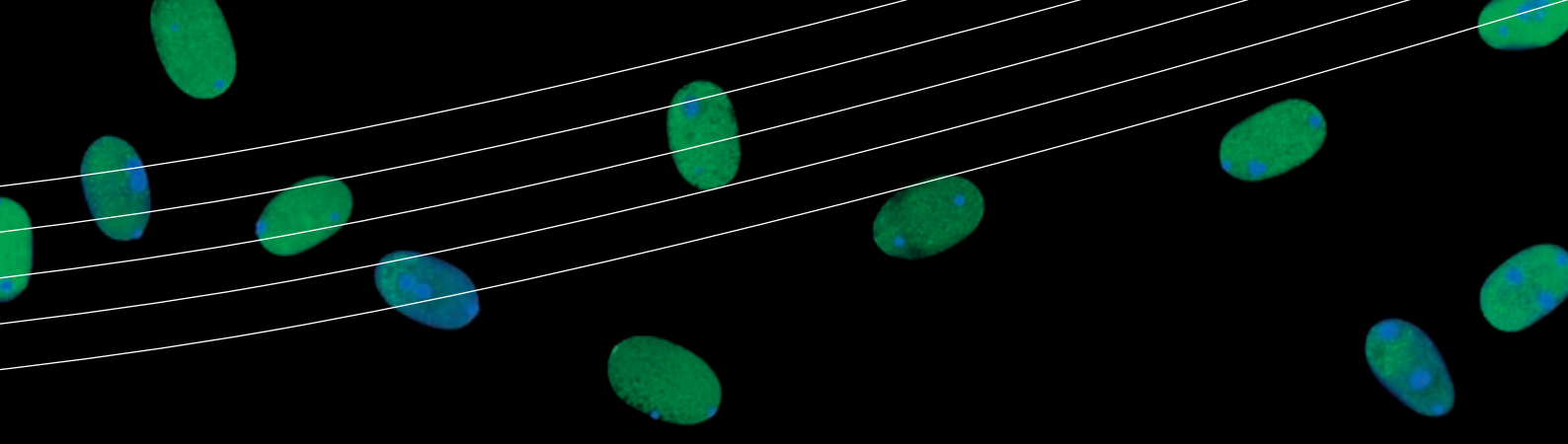
How do a sperm and an egg become an embryo?

The fusion of two highly specialised cells, a spermatozoon and an oocyte, leads within a short time to the emergence of a self-dividing cell, the zygote, which then develops into a complete organism. How this fundamental process in the development of every living creature functions is to a large extent unknown. True, many fundamental factors are known. For example, around 600 genes are essential for the embryo of *C. elegans* to develop (Sönnichsen *et al.*, 2005). However, the global processes are yet to be illuminated. Interestingly, in all creatures stud-

Fig. 2: eFACS



We use eFACS to sort large quantities of single-celled and two-celled embryos. For this we use a strain of nematode that expresses a gene that is specific to the single-celled embryo and has been linked to a fluorescent marker (OMA-1-GFP). This method routinely results in a recovery rate of several tens of thousands of embryos with purity in excess of 98% (a). The scatter plot shows the first stage of sorting of single-celled OMA-1-GFP embryos. The GFP-positive population (~3–5% single-celled embryos) is selected and sorted (b). After the first sorting, the GFP-positive population has a purity of roughly 70% and must be sorted again. The purity of the GFP-positive population that has been sorted twice is then checked by examination under a microscope (c and e) and by RT-PCRs (d). These confirm a purity of more than 98% of single-celled embryos. (Chart from Stoeckius *et al.*, 2009. Copyright Nature Methods).



ied, these “early” processes are controlled by factors that are present in oocyte and spermatooon prior to fertilisation. That leads one to ask how big and how essential is the part played by gene products of oocyte and sperm.

In the laboratory of Nikolaus Rajewsky (BIMSB@MDC) we use *C. elegans* and the methods established here to learn about these fundamental reorganisation and reprogramming processes, taking a genome-wide approach. For this project, we measure both coding and non-coding RNAs and proteins in oocytes, spermatozoa, single-celled and two-celled embryos. This data provides us with the first-ever glimpse into a time frame covering the first minutes in early embryonic development. We are observing dynamic changes in the quantities of RNAs and nearly 5,000 proteins. RNAs and proteins behave very differently, which leads one to conclude that there is a high degree of post-transcriptional gene regulation. We are also observing dynamic changes within and between all known short, non-coding classes of RNA (miRNAs, 21U-RNAs, 22G-RNAs, 26G-RNAs, and other endo-siRNAs). The data permits the hypothesis that a large number of RNAs, mRNAs and proteins in the embryo likely originate in the sperm. This points to an unexplored paternal contribution to early embryonic development. The data also suggests that the single-celled *C. elegans* embryo may already be transcriptionally active. These enormous datasets are currently being analysed and validated in independent tests.

The research project in brief:

“High-throughput research into the proteome and transcriptome during the reprogramming of an oocyte into a zygote and the early embryonic development of *C. elegans*” in the group of Nikolaus Rajewsky, BIMSB Berlin. Funded since 2009 by the BIMSB/NYU PhD exchange programme (www.mdc-berlin.de/en/bimbs/index.html). This work is part of a longstanding collaboration with Fabio Piano (NYU, USA).

Thanks:

This project is part of my doctoral research in the laboratory of Nikolaus Rajewsky. I would like to thank my colleagues Jonas

Maaskola (co-lead author of the Nature Methods publication) and Dominic Grün (co-lead author of the publication in preparation) for all their bioinformatics and statistical analyses.

References:

- Boutros M, and Ahringer J (2008) The art and design of genetic screens: RNA interference. *Nature Reviews Genetics*, 9, 554-66.
- Dupuy D, *et al.* (2004) A first version of the *Caenorhabditis elegans* Promoterome. *Genome Research*, 1421, 69-75.
- Frøkjær-Jensen C, *et al.* (2010) Targeted gene deletions in *C. elegans* using transposon excision. *Nature Methods*, 7, 451-53.
- Frøkjær-Jensen C, *et al.* (2008) Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature Genetics*, 40, 1375-83.
- Gerstein MB *et al.* (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science*, 6012, 1775-87.
- Lamesch P *et al.* (2004) *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Research* 14, 2064-9.
- Oliver SG (2006) From genomes to systems: the path with yeast. *Philosophical Transactions of the Royal Society*, 361, 477-82.
- Ong SE *et al.* (2002) Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular Cellular Proteomics*, 1, 376-86.
- Sönnichsen B *et al.* (2005) Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature*, 434, 462-69.
- Stoeckius M *et al.* (2009) Large-scale sorting of *C. elegans* embryos reveals the dynamics of smallRNA expression. *Nature Methods*, 6, 745-51.
- Sulston JE *et al.* (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology*, 100, 64-119.

Contact:

Marlon Stoeckius

Berlin Institute for Medical Systems Biology

MDC-Berlin

marlon.stoeckius@mdc-berlin.de

www.mdc-berlin.de/en/bimbs

ein fadenwurm bereichert die systembiologie

Caenorhabditis elegans ist einer der elegantesten Modellorganismen

von Marlon Stoeckius

Caenorhabditis elegans ist ein etwa 1 Millimeter langer, transparenter, im Boden vorkommender Fadenwurm (Abb.1). Er wurde in den 60er Jahren vom Biologen Sydney Brenner anfangs als Modell zur Erforschung der Nervenentwicklung genutzt. In den darauffolgenden Jahren etablierte sich *C. elegans* als ein Modellsystem für die Zellteilung und die Embryonalentwicklung. Viele grundlegende Mechanismen funktionieren im Wurm sehr ähnlich wie im Menschen. So ist er heute einer der beliebtesten Modellorganismen für Entwicklungsbiologen und ebenso ein weit verbreitetes Modellsystem für die Erforschung der molekularen Grundlagen von vielen Krankheiten wie zum Beispiel Krebs, Diabetes, Parkinson und Alzheimer. Ein weiteres und eher junges Forschungsfeld ist die Funktion von kleinen RNAs, wie zum Beispiel microRNAs, die in *C. elegans* erstmals entdeckt wurden. Molekularbiologische Methoden zur Untersuchung des Wurms sind sehr gut etabliert und bieten den Forschern viele Ansätze, grundlegende Phänomene und Mechanismen zu analysieren. Dies führte nicht zuletzt dazu, dass für Entdeckungen in *C. elegans* bereits drei Nobelpreise vergeben wurden.

Randvoller Werkzeugkasten für Systembiologen

Auch aus der Systembiologie ist *C. elegans* seit jüngster Zeit nicht mehr wegzudenken. Der Hauptgrund dafür ist, dass sein molekularbiologischer und biochemischer Werkzeugkasten üppig ausgestattet ist. Die Grundsteine legte ein Konsortium von Wissenschaftlern, das 1998 das Erbgut (Genom) des Wurms als ersten vielzelligen Organismus vollständig entschlüsselte (Sequencing Consortium, 1998). Noch heute, auch 10 Jahre nach der Entschlüsselung des menschlichen Genoms, ist das vergleichsweise kleine Genom des Wurms (etwa 100 Megabasen) das am besten kartierte Genom eines Vielzelllers. Das Erbgut des Wurms kann mit modernen Methoden relativ einfach manipuliert werden. So können fremde oder veränderte Gene zielgenau in das Genom eingebaut werden (MosSCI; Frøkjær-Jensen *et al.*, 2008). Für die Veränderung von Genen können sich Wurmforscher 'Genbanken' zunutze machen, in denen fast alle Gene (ORFeome; Lamesch *et al.*, 2004) und regulatorischer Elemente (Promoterome; Dupuy *et al.*, 2004. 3'UTRome; F. Piano Labor) aus dem Genom des Wurms isoliert wurden. Diese können dann wie in einem Baukasten frei kombiniert werden. Gene können in *C. elegans* durch Fütterung mit speziellen Bakterien vergleichsweise einfach ausgeschaltet werden (RNAi feeding; Boutros and Ahringer, 2008). Darüber hinaus können ebenfalls spezifische Regionen aus dem Genom gelöscht werden (MosDEL; Frøkjær-Jensen *et al.*, 2010). Der Wurm

Abbildung 1: *Caenorhabditis elegans*



Der ein Millimeter große Fadenwurm *Caenorhabditis elegans* ist ein beliebter Modellorganismus.



Marlon Stoeckius nutzt den Fadenwurm für die Erforschung der molekularen Grundlagen der frühen Embryonalentwicklung (Photo: Alexander Baltz @ BIMSB).

kann auf Agarplatten mit bis zu 384 Kammern, wie in separaten Lebensräumen, kultiviert werden. Das erlaubt die Durchführung von parallelisierten automatisierten Hochdurchsatzscreens, in denen etwa jedes Gen des Wurms im Zeitraum von nur einer Woche einzeln ausgeschaltet wird. Sowohl der Embryo als auch der erwachsene Wurm ist bei *C. elegans* transparent, und seine Entwicklung verläuft immer stereotyp. Das erlaubt die Beobachtung einzelner Zellen im lebenden Embryo und Wurm unter dem Mikroskop. So ist der Ursprung und das Schicksal einer jeden Zelle, vom Einzell-Stadium zur schlüpfenden Larve mit 558 Zellen bis zum erwachsenen Wurm mit 959 somatischen Zellen, beschrieben (Sulston *et al.*, 1983). Heute kann die Embryonalentwicklung mithilfe bildgebender Verfahren automatisiert überwacht werden. In Kombination mit dem Ausschalten von bestimmten Genen im Embryo können Veränderungen der charakteristischen Zellteilungsmuster automatisch erkannt werden. Damit können Hypothesen zur Funktion des Gens aufgestellt werden. Über 13.000 genetisch manipulierte Würmer (CGC, März 2011) wurden in den vergangenen Jahrzehnten von Wissenschaftlern generiert, in denen einzelne Gene gelöscht, mutiert oder mit einem Fluoreszenzmarker markiert wurden. Dadurch ist eine enorme Ressource entstanden, die nach dem 'open-access' Prinzip unter den Wissenschaftlern geteilt wird. Im Gegensatz zu vielen anderen Modellorganismen können diese Würmer einfach im Tiefkühlschrank bis zur weiteren Verwendung gelagert werden. Neben den Ressourcen an Labor-Methoden stehen mittlerweile auch hunderte von Datensätzen aus Hochdurchsatzexperimenten in Datenbanken zur Verfügung. Ein kürzlich beendetes Projekt für die systematische Kartierung von regulatorischen Elementen im Genom hat neben der Etablierung neuer Methoden auch mehrere Terabytes an Daten produziert, die nun frei zur Verfügung stehen (ModENCODE; Gerstein *et al.*, 2010). Zusammen genommen machen diese Ressourcen *C. elegans* zu einem idealen vielzelligen Modell für die Systembiologie.

Woran es fehlt(e)

Moderne DNA-Sequenzierstechnologien erlauben quantitative Analysen von allen eingeschalteten Genen (RNA-Expression) in sämtlichen Organismen. Das quantitative Erfassen von Proteinen, wie es seit einigen Jahren in der Zellkultur genutzt wird (SILAC; Ong *et al.*, 2002),

kann bislang in den meisten Modellorganismen noch nicht angewendet werden. So wurden bislang Hypothesen über die Funktion von bestimmten Genen nur durch mikroskopische Veränderungen am Wurm, durch RNA-Expressionsänderung und durch die Messung von einzelnen wenigen ausgewählten Proteinen aufgestellt.

Letztendlich müssen für das Verständnis von biologischen Prozessen auch die Menge, die Identität und die Dynamik der Stoffwechselprodukte in einem Organismus quantitativ bestimmt werden. Der Stoffwechsel ist sowohl in einzelligen als auch in vielzelligen Modellsystemen ein unzureichend erschlossenes Forschungsgebiet. (Metabolomics; Oliver, 2006). Hochdurchsatz-Forschungsprojekte in *C. elegans* erfordern hier sowohl die Optimierung der Methoden als auch eine neuartige Herangehensweise an die Datenanalyse. Nachdem der Wurm geschlüpft ist, kann er über seinen Lebenszyklus hinweg synchron in großen Mengen kultiviert werden, wodurch genug Untersuchungsmaterial (z.B. RNA und Proteine) für Analysen von gleichaltrigen Würmern zur Verfügung steht. Quantitative Analysen beruhen bisher allerdings überwiegend auf Messungen des kompletten Wurms, was zu einem Vermischen der verschiedenen Gewebe führt. So ist es beispielsweise schwer, auseinanderzuhalten, welche Gene spezifisch in Muskeln oder Nerven eingeschaltet sind. In Zukunft werden Methoden etabliert werden müssen, die Hochdurchsatzmessungen gezielt in den einzelnen Geweben und Zellen erlauben.

Ähnliche Hürden müssen für die Erforschung der Embryonalentwicklung überwunden werden. Der erwachsene Wurm trägt Embryonen unterschiedlichen Alters in sich – von der einzelligen Zygote bis zur schlüpfenden Larve mit 558 Zellen. Materialintensive Messungen beruhen stets auf heterogenen Populationen von Embryonen unterschiedlicher Zellstadien. Für genauere Untersuchungen der Abfolge von zeitlichen Ereignissen während der Embryonalentwicklung mussten bisher Embryonen nach Alter unter dem Mikroskop sortiert werden. Bei diesem zeitaufwendigen manuellen Sortieren können nicht die Mengen an Embryonen isoliert werden, die für Analysen der RNA-Expression oder Proteinexpression benötigt werden. Mit einer ruhigen Hand können etwa 500 Embryonen am Tag gesammelt werden, moderne Hochdurchsatzanalysen benötigen jedoch über 100.000 Embryonen.

Einzellige *C. elegans* Embryonen.

Automatisiertes Sortieren von Embryonen gleichen Alters

Wir haben eine Methode entwickelt, mit der zehntausende Embryonen gleichen Entwicklungsstadiums innerhalb kurzer Zeit mithilfe eines klassischen Zellsortierers (FACS) isoliert werden können (Abb. 2). Grundvoraussetzung für das Verfahren, welches wir eFACS taufen, ist ein Wurmstamm, der ein fluoreszierendes Protein in nur einem bestimmten Zellstadium spezifisch produziert. Die Fluoreszenz dient dann als Marker für das gewünschte Stadium und wird vom Zellsortierer spezifisch erkannt. Wir nutzen diese Methode, um große Mengen von einzelligen und zweizelligen Embryonen sortieren zu können. Grundsätzlich kann die Methode aber für die Isolierung von Embryonen jedes beliebigen Entwicklungsstadiums eingesetzt werden. Dadurch eröffnet eFACS das Tor zur systematischen Erforschung der Embryonalentwicklung mit modernen Hochdurchsatzmethoden in *C. elegans* (Stoeckius *et al.*, 2009).

Quantifizieren von tausenden Proteinen

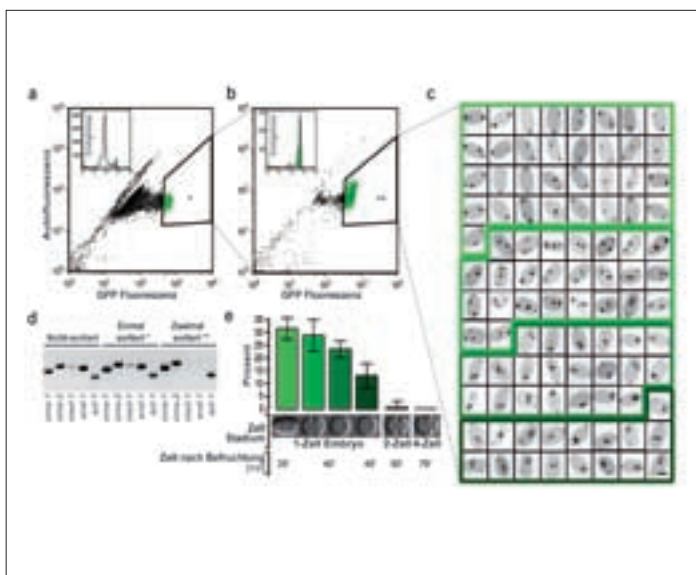
In Zusammenarbeit mit dem Labor von Matthias Selbach (MDC Berlin) haben wir im Labor von Nikolaus Rajewsky (BIMSB@MDC) eine neue Methode entwickelt, die es erlaubt, die Aktivität und

Menge tausender Proteine im Vergleich zweier *C. elegans*-Proben zu quantifizieren (Thierfelder *et al.*, in Vorbereitung). Dies eröffnet erstmalig Einblicke in Mengenveränderungen von Tausenden von Proteinen und erlaubt dadurch Rückschlüsse auf deren Funktion im lebenden Organismus. So kann nun untersucht werden, wie sich die Proteinmengen während der Entwicklung des Wurms verändern und wie sich gewisse Proteine verhalten, wenn ein bestimmtes Gen ausgeschaltet wird. Die Methode ermöglicht außerdem, die durch Signalketten aktivierten Proteine quantitativ zu messen (Phosphoproteom) und die Sensitivität von anderen im Wurm verwendeten Methoden zu verbessern (z.B. Co-IP).

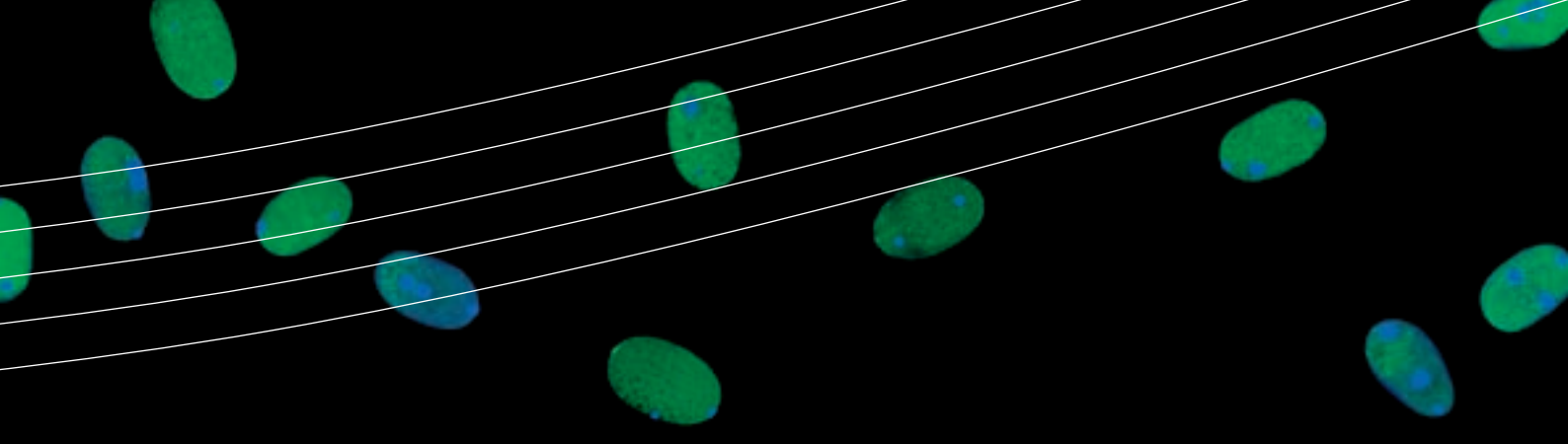
Wie wird aus Spermium und Eizelle ein Embryo?

Die Fusion von zwei hochspezialisierten Zellen - Spermium und Eizelle - führt in kurzer Zeit zur Entstehung einer sich teilenden Zelle - der Zygote - aus der ein vollständiger Organismus hervorgeht. Es ist weitgehend unverstanden, wie dieser fundamentale Prozess in der Entwicklung jedes Lebewesens funktioniert. Zwar sind viele essentielle Faktoren bekannt, beispielsweise sind etwa 600 Gene für die Embryonalentwicklung von *C. elegans* essentiell (Sönnichsen *et al.*, 2005), die globalen Vorgänge bleiben hingegen

Abbildung 2: eFACS



Wir nutzen eFACS, um große Mengen von ein- und zweizelligen Embryonen zu sortieren. Dafür nutzen wir einen Wurmstamm, der ein Gen exprimiert, welches spezifisch für den einzelligen Embryo ist und an einen Fluoreszenzmarker (OMA-1-GFP) gekoppelt wurde. Die Methode hat routinemäßig eine Ausbeute von mehreren zehntausend Embryonen mit einer Reinheit größer als 98% (a). Der Scatterplot zeigt den ersten Sortierungsschritt von einzelligen OMA-1-GFP-Embryonen. Die GFP-positive Population (~3-5% einzellige Embryonen) wird selektiert und sortiert (b). Die GFP-positive Population hat nach der ersten Sortierung eine Reinheit von etwa 70% und muss erneut sortiert werden. Die Reinheit der zweifach sortierten GFP-positiven Population wird dann durch mikroskopische Untersuchung (c und e) und RT-PCRs (d) untersucht. Diese bestätigen eine Reinheit von mehr als 98% einzelliger Embryonen. (Bild aus Stoeckius *et al.*, 2009. Copyright Nature Methods).



ungeklärt. Interessanterweise werden diese 'frühen' Prozesse in allen studierten Lebewesen von Faktoren kontrolliert, die bereits vor der Befruchtung in Eizelle und Spermium vorhanden sind. Das wirft die Frage auf, wie groß bzw. wie essentiell der Anteil der Genprodukte von Eizelle und Spermium ist.

Im Labor von Nikolaus Rajewsky (BIMSB@MDC) nutzen wir *C. elegans* und die hier etablierten Methoden, um im genomweiten Ansatz Aufschluss über diese fundamentalen Reorganisations- und Reprogrammierungsprozesse zu bekommen. Für dieses Projekt messen wir sowohl kodierende- als auch nicht-kodierende RNAs und Proteine in Eizellen, Spermien, einzelligen und zweizelligen Embryonen. Die Daten geben uns nie dagewesene Einblicke in ein Zeitfenster der ersten Minuten in der frühen Embryonalentwicklung. Wir beobachten dynamische Mengenänderungen von RNAs und knapp 5.000 Proteinen. RNAs und Proteine verhalten sich sehr unterschiedlich, was auf ein großes Maß an posttranskriptioneller Genregulation schließen lässt. Darüber hinaus beobachten wir dynamische Veränderungen innerhalb und zwischen allen bekannten kleinen nicht-kodierenden RNA-Klassen (miRNAs, 21U-RNAs, 22G-RNAs, 26G-RNAs, und andere endo-siRNAs). Die Daten lassen die Hypothese zu, dass eine große Anzahl von kleinen RNAs, mRNAs und Proteinen im Embryo vermeintlich aus dem Spermium stammen. Dies weist auf einen unerforschten väterlichen Beitrag zur frühen Embryonalentwicklung hin. Außerdem legen die Daten den Schluss nahe, dass der einzellige *C. elegans*-Embryo möglicherweise schon transkriptionell aktiv ist. Die riesigen Datensätze werden zurzeit ausgewertet und in unabhängigen Versuchen validiert.

Steckbrief Forschungsprojekt:

„Hochdurchsatzuntersuchungen des Proteoms und Transkriptoms während der Reprogrammierung von Eizelle zur Zygote und der frühen Embryonalentwicklung von *C. elegans*“ in der Gruppe von Nikolaus Rajewsky, BIMSB Berlin. Förderung seit 2009 durch das BIMSB/NYU-PhD Austauschprogramm (www.mdc-berlin.de/en/bimbsb/index.html). Diese Arbeit ist Teil einer langjährigen Kollaboration mit Fabio Piano (NYU, USA).

Danksagung:

Dieses Projekt ist Teil meiner Promotion im Labor von Nikolaus

Rajewsky. Ich danke meinen Kollegen Jonas Maaskola (Co-Erstautor der Nature Methods Publikation) und Dominic Grün (Co-Erstautor der in Vorbereitung befindlichen Publikation) für alle bioinformatischen und statistischen Analysen.

Referenzen:

- Boutros M, and Ahringer J (2008) The art and design of genetic screens: RNA interference. *Nature Reviews Genetics*, 9, 554-66.
- Dupuy D, *et al.* (2004) A first version of the *Caenorhabditis elegans* Promoterome. *Genome Research*, 14, 69-75.
- Frøkjær-Jensen C, *et al.* (2010) Targeted gene deletions in *C. elegans* using transposon excision. *Nature Methods*, 7, 451-53.
- Frøkjær-Jensen C, *et al.* (2008) Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature Genetics*, 40, 1375-83.
- Gerstein MB *et al.* (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science*, 307, 1775-87.
- Lamesch P *et al.* (2004) *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Research* 14, 2064-9.
- Oliver SG (2006) From genomes to systems: the path with yeast. *Philosophical Transactions of the Royal Society*, 361, 477-82.
- Ong SE *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular Cellular Proteomics*, 1, 376-86.
- Sönnichsen B *et al.* (2005) Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature*, 434, 462-69.
- Stoeckius M *et al.* (2009) Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. *Nature Methods*, 6, 745-51.
- Sulston JE *et al.* (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology*, 100, 64-119.

Kontakt:

Marlon Stoeckius

Berlin Institute for Medical Systems Biology

MDC-Berlin

marlon.stoeckius@mdc-berlin.de

www.mdc-berlin.de/en/bimbsb