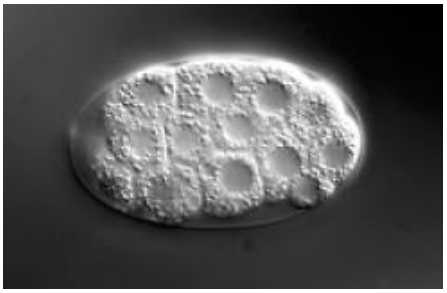
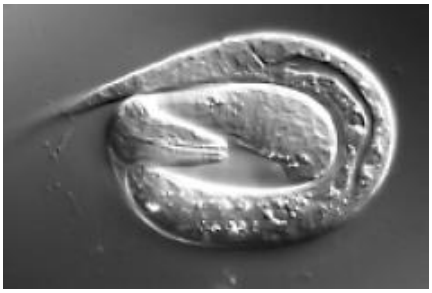
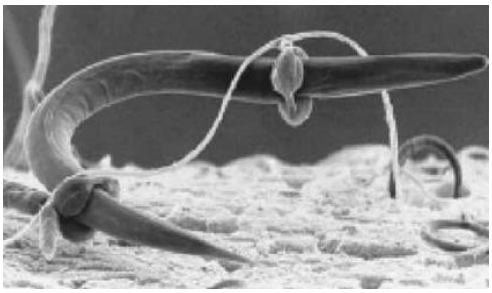
Caenorhabditis elegans: an overview of milestones, genome manipulation techniques and perspectives for the future.

Master thesis, November 2011

Dylan Bergen

Supervised by Dr. Rik Korswagen, Hubrecht Institute, Utrecht

Universiteit Utrecht

cancer genomics & developmental biology

Hubrecht Institute

This master thesis with a value of 7.5 ECTs was written as an assignment for the Master program Cancer Genomics and Developmental Biology, a program held by the Graduate School of Life Sciences of the Utrecht University, the Netherlands.

Student:	D.J.M. Bergen
Student number:	3446549
E-mail address:	d.j.m.bergen@students.uu.nl

Front page figures: clock wise; upper left: 16 cell stage of a *C. elegans* embryo, seen through a standard Normanski microscope (adapted from the Nobel lecture of J.E. Sulston, 2002); Upper right: Hatching of a *C. elegans* larvae, observed through a standard Normanski microscope (adapted from the Nobel Lecture of J.E. Sulston, 2002); An adult *C. elegans* captured by its natural predator, a fungi (adapted from the Nobel lecture of C.C. Mello, 2006).

Caenorhabditis elegans: an overview of milestones, genome manipulation techniques and perspectives for the future.

Summary

C. elegans is a well studied animal model and allows researchers to study basic and complex biological processes in a relatively simple multi-cellular environment. However, until recently, researchers did not have the possibilities to create gene targeted knock outs with an efficient protocol. Here we review these recent genome modifying technical developments, as Mos1 random integration, EMS mutagenesis combined with Whole Genome Sequencing, Mos1-single copy integration, and Mos1-deletion. These developments are put in the perspective of the already achieved Nobel prize milestones, as RNAi, exogenous GFP expression, and cell lineage tracing. Moreover, we propose that completion of the two mutant databases, which are generated by both random Mos1 integrations and EMS random mutagenesis will open a valuable source of potential data. We state that combining these loss of function mutant libraries will enable performing research in a more conclusive manner in the upcoming decade.

Table of contents

Introduction

The need of a model system to study certain biological processes is a never ending search. The new era of genetics led to an enormous expansion of available animal model systems. The past decades animal model systems such as the fruit fly (*Drosophila melanogaster*), yeast (*Saccharomyces cerevisiae*), and the zebrafish (*Danio rerio*) arose next to the existing rodent models. In biomedical sciences and developmental biology, the nematode *Caenorhabditis elegans* (*C. elegans*) has become a major animal model to study a variety of fundamental and applied biological processes. For studying fundamental questions in cell biology, cell behavior, cell fate and migration this multi-cellular organism has multiple advantages. One of the major advantages lies within the reproducibility of each sexually dimorphic individual, as every individual hermaphrodite's development is invariant with ~959 somatic nuclei. Until now *C. elegans* is the only multi-cellular organism of which the complete cell lineage is traced[1]. Furthermore, the worms are transparent, allowing cell tracking experiments on a single cell scale in an *in vivo* setting. In addition to physical advantages for a researcher, practically the nematode is advantageous as well. The hermaphrodites are self-fertilizing, have a large progeny (about 300 eggs), require room temperature storage, occupy little lab space, are kept on agar plates, have a short life cycle of 3 and a half days, and are fed with *Escherichia coli* (*E. coli*) [for more details about maintenance:[2]]. In rare situations, the hermaphrodites produce males, which can be used for crossing loci of interest into other strains.

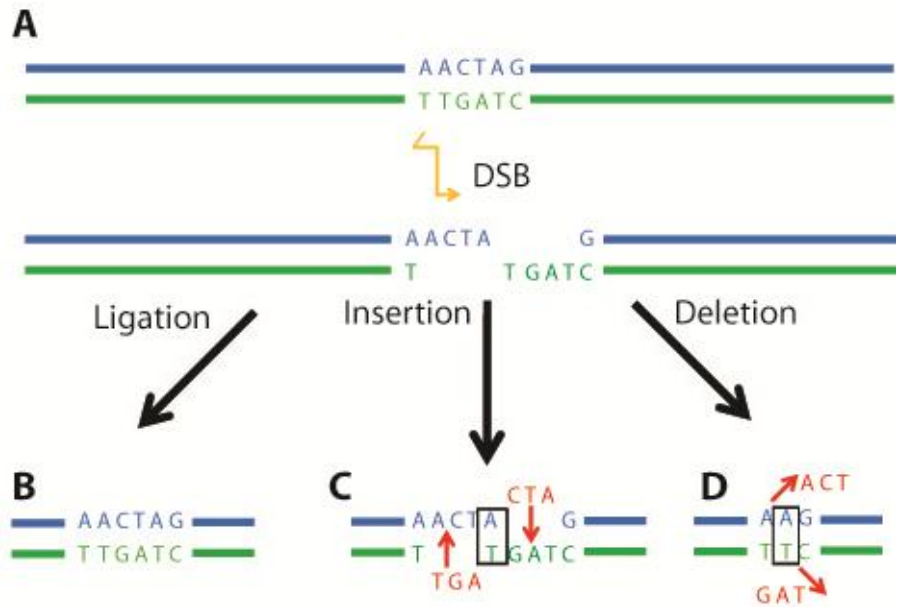


Figure 1: Non-homologues end joining is a mechanism to repair DSBs quickly. A DSB occurs (**A**) and the overhangs can be repaired directly via simple mutation-less ligation (**B**), the gaps can be filled in (red nucleotides) when impaired alignment (black box) occurs (**C**), or the sequence can be removed (**D**). These processes are initiated with the *Ku* and Artemis bridging factors, and DNA-protein kinases (DNA-PKs) which eventually recruit ligase IV and other mediating factors[7].

The genome of a hermaphrodite *C. elegans* is 100 mega bases (Mb) long and consists of six pairs of chromosomes, of which five are autosomes (I-V), and one pair is sex determining (X). Males are haploid for the sex determining chromosome, which could spontaneously occur during meiosis. All the six chromosomes are approximately equal in length (ranging from ~13 Mb to ~20 Mb) and the genetic map length of each chromosome consists out of roughly 50 centimorgans (cM) [3]. At the chromosome arms (~100 kb per cM) the physical distance of a cM is on average shorter compared to

defined gene cluster areas. These gene cluster are located around the core of the autosomes (~1500 kb per cM), resulting that the chance of meiotic recombination is higher in less gene concentrated areas during meiosis [4]. One of the genetic characteristics of *C. elegans* is the high gene density of coding sequences (DNA to mRNA: 40%, DNA to protein: 25.5%, untranslated RNA; mRNA - protein: 14.5%), which makes the genome spectacularly information-rich[5]. In contrast to higher vertebrates, the chromosomes are holocentric and do not contain discrete centromeres, which means that chromosome segregation and spindle microtubule attachment occurs at multiple sites. This characteristic is an advantage as well as a disadvantage for experiments performed with extrachromosomal arrays (e.g. plasmids, chunks of DNA, or other types of exogenous DNA). An advantage is that centromeric sequences are not necessary to maintain the exogenous DNA for multiple generations, and dilution of the construct is limited. On the other hand, due to multiple spindle attachments the arrays could be segregated and maintained in smaller pieces, or duplicated, resulting in an unequal distribution of the extrachromosomal array per individual (mosaic formation) [6]. Preventing mosaics in individuals, homogenously distributed knock-outs, knock-ins, or transgenics could be obtained by applying homologues recombination techniques. These recent developments are based on repair of a double stranded break (DSB) caused by transposon excision of Mos1, which normally is repaired via the non-homologues end joining (NHEJ) pathway as schematically depicted in **figure 1**. Agents causing lesions in the DNA structure, such as DSBs are repaired by the cell to maintain genome stability. NHEJ is a simple mechanism leading to a successful ligation without mutation, insertion due to impaired annealing and DNA polymerase repair, or deletion of the non-homologues sequence (see **figure 1**). All of the three endings of the NHEJ pathway are mediated by ligase IV and bridging factors (mammalian orthologues: *Ku70* and Ku86). These bridging factors have a positively charged asymmetric ring allowing unspecific sequence binding to the negatively charged bases of the DNA structure[7]. The broken ends of the strands are ligated to each other by the protein-protein interaction of the bridging factors. DSBs caused by transposon 'jumping' are in most cases repaired through this pathway (see **figure 1**) in a non-mutative manner at the excision site[8]. However, in rare cases the DSBs are repaired via the homologues recombination pathway. Homologues recombination Single Strand Dependent Annealing (SSDA) mainly occurs in gametes during meiosis[9], and normally results in non-crossover products. However, to exchange genetic information from two chromosomes to retain genetic diversity, crossover of chromosomal sequences between sister chromatids of maternal and paternal origin should happen to enhance this diversity. Canonical homologues recombination uses the homologues strands to synthesize DNA between two flanks of the DSB (see **figure 2**). The sister chromatids are in a non-experimental setting the templates for genetic recombination during meiosis or DSB repair in the nucleus. The annealing of the 3' overhang to the other homologues strand allows DNA replication by the DNA polymerase. The newly synthesized strands detach, are processed to complete genetic crossover between chromosomes[10]. In principle this process has a low mutation rate, since only the nucleotide overhangs are removed by nuclease activity[11]. Thus, in theory supplying an extrachromosomal array with flanking sequences could lead to the integration of the intra homologues flanking sequence of the chromosomal array via this rare event of DSB repair. Due to years of research a way is found to manipulate *C. elegans* genome via canonical homologues recombination in a controlled experimental setting.

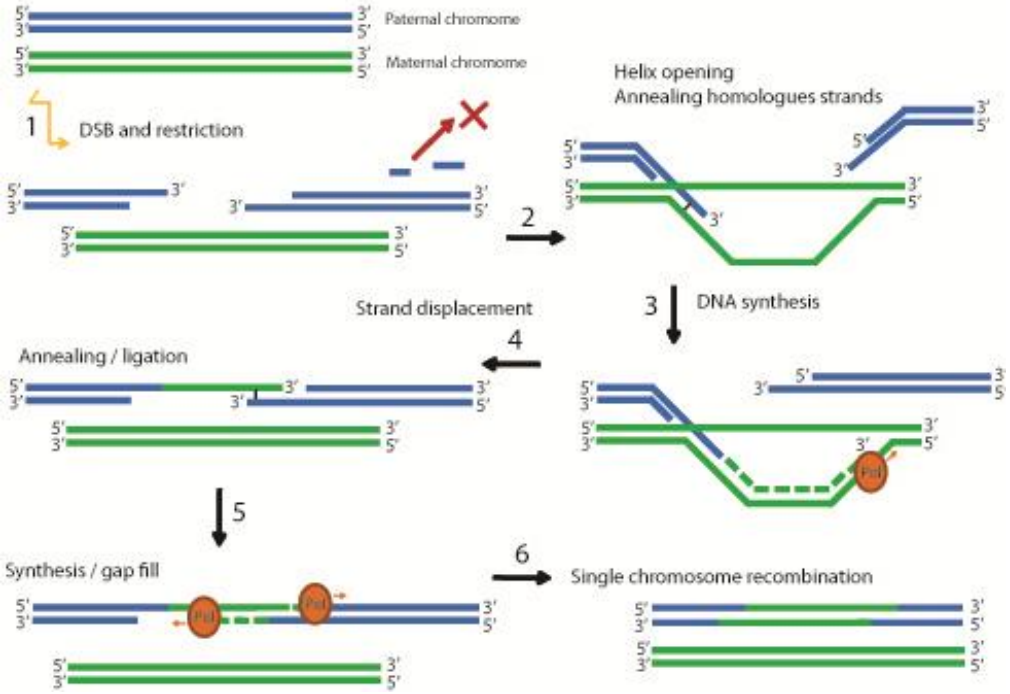


Figure 2: Schematic depiction of canonical homologues recombination process. A DSB initiates restriction and formation of 3' overhangs (**1**), which allows opening of the homologues sequence of the sister chromatid by invading and annealing of one 3' overhang (**2**). The DNA polymerase complex synthesizes the sequence downstream of the homologues region according to the complementary strand (**3**). The newly synthesized strand detaches and gets annealed and ligated with the other broken part (**4**), which results in completion of the gaps by the DNA polymerase complex (**5**). Now one chromosome contains the sequence of a complementary region located on a sister chromatid (**6**), green regions in blue regions). Please note that homologues recombination could have multiple outcomes as described in: M. Sasaki *et al.* 2010[10].

From the '80s on, *C. elegans* has grown from a minor model system to a mature model system. As stated in the Nobel lecture of Craig Mello in 2006[12], the 'lineage of researchers' started with the pioneering work of Sydney Brenner[13, 14]. From then on, many researchers joined the 'worm community' and increased knowledge for the whole biology community. The worm community is efficient, since it works in consortia and it established multiple databases, such as Wormbase and the freely available resource Wormbook. In this review, we will briefly show the three Nobel prize rewarded milestones that emphasized how complex life is build up, and which changed our way of thinking and therefore influenced many research projects. This will be followed up with an overview of the state of the art of the current exciting developments happening in genome manipulation techniques. We think that those developments make the worm a more attractive model system to study, since it is in the race to be the second organism carrying a loss of function mutation in every single gene.

Milestones for science

The first Nobel prize was given for describing *C. elegans* biology and for showing what the advantages of the nematode are to study fundamental biological processes. This prize rewarded this model system for establishing a model system that allows science to further understanding the complexity of life. Due to both its complexity and simplicity in one model system, *C. elegans* will increase knowledge about genes and mechanisms, which are responsible for the development of organs, cell lineages, and cell identity in the near future.

Organogenesis: how to approach the complexity of life

How several sets of genes can set up a body plan and give cells the information to perform a certain function at a certain spatio-temporal point is one of the most intriguing questions currently to be answered in biology[14]. Choosing and characterizing genetic players involved in the organogenesis of *C. elegans* in 1974[13] was a pioneering step in biomedical sciences. The nematode was by then poorly studied, however, it had several advantageous features (e.g. transparency, short life cycle, high amount of progeny, etc.) to study basic processes as cell division, tissue organization, cell fate and reproduction. This pioneering work in characterizing the nematode's biological composition and cell "behavior" led to the first Nobel prize for the worm as a model system in 2002. Sydney Brenner[14] shared the prize with his co-workers John Sulston, who traced the complete cell lineage and their decisions in the nematode[1, 15] (see **figure 3**). The other co-worker H. Robert Horvitz explained these features with genetic analysis, and therefore linking certain genes (read: elucidate their function) to certain fates[16, 17]. The nematode could be considered as a complex organism, since it has a fixed body plan, which is established by complex cell movements during gastrulation. Identifying the apoptotic cells, mother-daughter cell divisions, and exact knowledge about the occurring asymmetric cell divisions makes the worm an ideal model system for developmental biology.

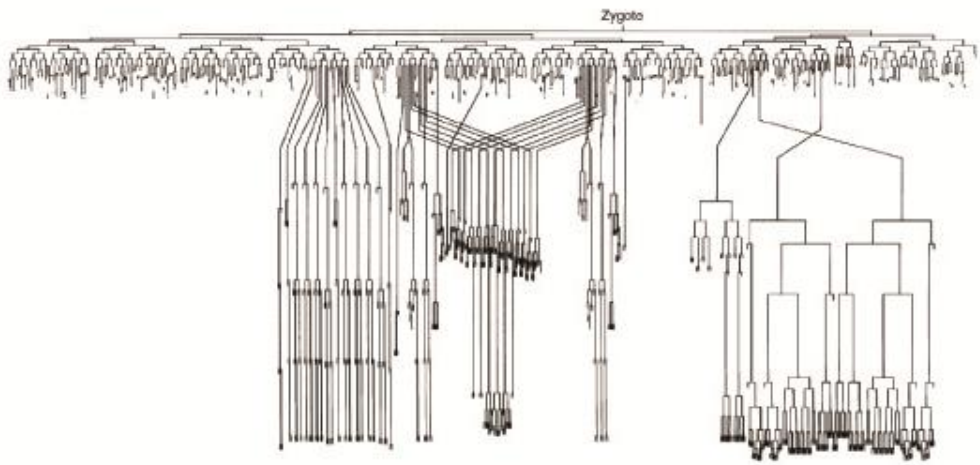


Figure 3: Snapshot of the work of Sulston *et al.* Adapted from the Nobel lecture of H. Horvitz, 2003[16]. Complexity of different cell lineages is shown and how these are formed during embryogenesis.

Robert Horvitz joined the lab of Sydney Brenner as a postdoc, and together with John Sulston they started investigating cell divisions of neuroblasts in mature neurons controlling movement (motor neurons). The rest of the somatic cell lineage and cells programmed to die were elucidated as well, and moreover, the gonad cell lineage was traced two years later[18]. Until now, *C. elegans* is the only multi-cellular organism where the whole cell lineage is described, due to its invariant cell lineage[16]. With laser microsurgery in John White's lab, specific cells were targeted with a laser[19]. It showed that cell-cell contact is a crucial factor in determining cell identity and cell fate. Hence, cell fate is strictly related to a relative position in the body plan, which explains the little variances in the cell lineages between individuals[19]. Isolation of mutants showing aberrancies in development was a second approach to elucidate the genetic cascades that are responsible for asymmetric cell divisions and regulation of the 'tight' cell lineages in the worm. Screenings in Sydney Brenner's lab revealed over 4,000 mutant lines showing impaired development in the cell lineage and these lines which were characterized were abbreviated with *lin* (lineage). Narrowing down the number of genes that have a direct effect on cell fate or cell division, the Sydney Brenner lab set up criteria to score mutants, the so called 'IDCGs', or Important Developmental Control Genes. The lab was looking for genes that could change cell fate in an opposite manner by using both loss of function and gain of function mutants[16]. The primary screening of the '*lin*' mutant databank had *lin-12*[20] and *lin-14*[21] as candidates fulfilling the criteria in changing cells their fate in direct opposite directions (e.g. from fate A-B to B-A). As known now due to intensive studies in the nematode, these "*lin*" genes are involved in several cell-cell signaling pathways such as *notch* signaling (*lin-12*)[22], or involved as a small RNA (*lin-4*)[23]. Interestingly, key functional studies in *C. elegans* allowed extrapolation to mammalian model systems. Many of these genes have mammalian orthologues with several important developmental functions, for example the *lin-12/Notch* signaling pathway is involved in stem cell niche determination[24] and several types of cancer, such as leukemia[25].

Another striking feature of organogenesis is programmed cell death, or apoptosis. H. Robert Horvitz's lab had a fundamental contribution to the apoptosis field by studying *C. elegans* apoptotic cells. Interestingly, of the 959 cells that a hermaphrodite generates during development, another 131 generated cells are apoptotic and die as a cell fate. One of the over represented tissues having apoptosis as cell fate is the nervous system, namely 105 of the in total 131 apoptotic cells[26]. Programmed cell death is considered to be a cell fate that contributes to the morphogenesis that occurs during development. The first cell death gene that was discovered that is crucial to initiate programmed cell death during early nematode nervous system development is the DNA endonuclease regulator *nuc-1* (named after *nuc*lease deficient)[15]. Several cell death (*ced*) genes were identified during screening for mutant worms showing abnormal apoptosis. A significant increase in knowledge has been established with the so called 'killer cascade' of enzymes regulating life or death. *ced-1* and *ced-2* mutants lack the ability of phagocytosis[27], and therefore proper removal of dead cells. Since these cells were not engulfed, other phenotypes related to impaired programmed cell death could be studied in mutagenized *ced-1* animal[16]. This led to the identification of the so called 'killer genes' like *ced-3* where all the 131 supposed apoptotic cells survived and escaped the programmed cell death program[28]. This strongly implied that *ced-3* is involved in the cascade inducing apoptosis. Many more interesting mutants were found such as, *ced-4*[29] with a lot of phenotypical similarities to *ced-3* in neurons, and *ced-9* with a protecting function for programmed cell death[30, 31]. This fundamental research in the worm led to a better understanding of apoptosis in vertebrates. As is true for the *lin* genes, the 'killer genes' have mammalian orthologues as well. The

pioneering work in the worm strongly contributed in understanding cellular and molecular cell death mechanisms in 'higher' organisms. One of the most classical examples of fundamental findings in the worm is *ced-3*, which appeared to be an orthologue of the *interleukin-1-β-converting enzyme* (*ICE*). This enzyme belongs to the caspase family[32] and activates the apoptosis inducing cytokine IL-1β. Later on by homology, *ced-4*'s orthologue *apoptotic protease activating factor-1* (*Apaf-1*) was identified and this enhanced understanding mechanisms of cell death in humans significantly[33]. These examples revealed that organogenesis is largely conserved from nematode to higher vertebrates until a certain extent. Furthermore, it shows that those fundamental findings in the worm can be directly extrapolated to other model systems and humans.

RNA interference: from a cellular defense mechanism to a screening tool

Another good example of a complex mechanism which was discovered and extensively investigated was the finding of RNA interference (RNAi). The laboratory of Craig Mello observed that double stranded RNA is activating a RNA transcript silencing mechanism in the cell. This RNAi mechanism appeared to work more efficiently with dsRNA than single stranded RNA[34]. This important finding which changed the view on regulatory mechanisms during gene expression was rewarded with the Nobel Prize for Medicine or Physiology for Andrew Fire and Craig Mello in 2006[12, 35]. Elucidation of this important mechanism gave opportunities to perform very efficient gene specific knock down assays. Moreover, it was implied that this repressing system on the RNA level is a cellular immune system, by screening for (in general exogenous) dsRNA and degrading these transcripts[34]. Before this observation, it was already known that ssRNA complementary to their targets which were directly injected into the cytoplasm were able to silence genes specifically. This so called 'antisense-mediated silencing' process was firstly shown with the *par-1* mRNA[36], and strikingly, also sense RNA could silence the target genes, both in somatic cells and germ cells[34, 36]. Single stranded RNA molecules show the tendency to form dsRNA molecules with complementary sequences (other molecules, hairpins in secondary structure, etc.). However, this secondary structure is also present in the test tube, since a 100% pure ssRNA solution does not exist. During experiments with sense ssRNA molecules the observed silencing response can be explained by this type of contamination in the supplied ssRNA molecule population.

The silencing signal as a response to foreign (ds)RNA could be sent to different cell types, uniformly distributed in the nematode, and even passed to the progeny[37]. The cellular defense system against exogenous RNA, present in all the cells forming the organism, was used to design experiments to test gene specific knockdowns. Two strategies could be used, either the animals could be soaked in artificially produced dsRNAs[38] or the animals consume bacteria transformed with a vector with the RNAi sequence [39]. Big RNAi screens resulted in an increase of genetic pathway knowledge (e.g. *Wnt* signaling[40]) and unraveled components of the evolutionary conserved[41] small RNA producing pathway (see **figure 4**)[42]. Now this pathway is often used to knock down genes of interest in several model systems, since the response is quick, it can be performed on several time points and the constructs are relatively cost effective.

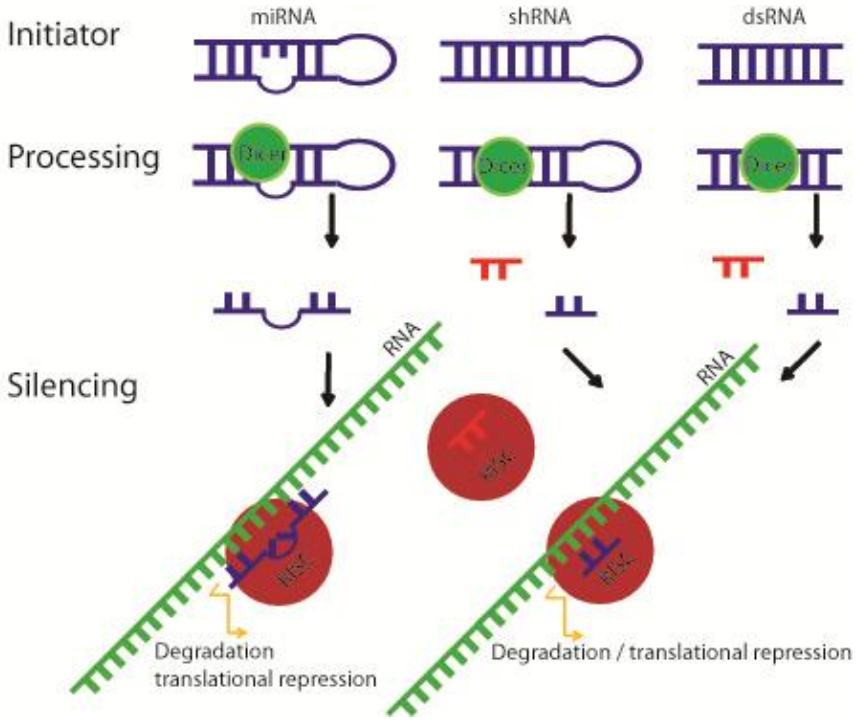


Figure 4: Simplified representation of the RNA silencing pathways. There are three types of small dsRNA molecules that can induce the RNAi pathway. Endogenous initiators as micro RNAs (miRNA) and short hairpin RNAs (shRNA) are recognized by the DICER protein and cleaved in fragments. Exogenous inducers as foreign dsRNAs are processed via this mechanism as well. These fragments are transferred to the RISC complex to scan for complementary sequences (note: miRNA appear to be less stringent). When a match is found the translation could be repressed or degradation of the RNA strand could occur.

The RNAi pathway, produces siRNA (short interference RNA) and some components process naturally produced miRNAs (micro RNA), which are small hairpin RNAs (so double stranded, 20-24 nt long)[43]. An important component in the production process is *Dicer*, which is the so called 'sensor' of the pathway[44]. It was shown that synthetically produced gene specific RNA structures similar to *Dicer* derived siRNAs could silence target genes[45]. In summary, the pathway is triggered by exogenous RNA as in the form of dsRNA (short hairpin RNAs or siRNAs) or by endogenous miRNA. These molecules are recognized by the Dicer complex and cleaved in smaller dsRNA fragments (from ~20 to ~30bp length) with 2 nucleotide long 3' overhang[46] and one of both strands of RNA fragments is incorporated (known as guide RNA) in the RNA-induced silencing complex (RISC) to scan for target mRNAs (see **figure 4**). Post transcriptional repression could occur via either translational repression by simply blocking the translational binding sites in the untranslated regions (UTRs) of the gene transcript, or the targeted transcript is degraded[41]. There is also strong evidence that dsRNAs can silence genes through chromatin modifications relying on a gene family called RNA-dependent RNA polymerases (RdRPs) and RNA-induced transcriptional silencing complex (RITS)[47]. Discovering this silencing pathway showed the important function that *C. elegans* has within biology to conduct experiments leading to new insights, due to its physiological advantages. Unraveling this pathway has led and will lead to (primary) data about gene function, redundant gene functions, and signaling pathways by knocking down the ubiquitously expressed genes in the preferred model system. In conclusion, the wide range of applications and, moreover, the strong contribution to fundamental knowledge of these findings will be a big contribution to understanding the regulation of life. Despite the many advantages it should be noted that RNAi is a knockdown system, and genes are not fully repressed by the Dicer-RISC pathway. Therefore data should be analyzed with care and critically reviewed, before being conclusive. The current opinion of many scientists is that RNAi provided experiments should contain supporting experiments such as gene knockouts from the Mos1 databank or EMS derived knockouts. This combination of gene knock outs and knock downs is a powerful and a more conclusive approach, since two independent means (on DNA level and RNA level) are disrupted. As Craig Mello stated during his Nobel lecture, every time after a big finding we

start to overestimate our confidence, and thus underestimating the complexity of life[12], referring to the complex regulatory mechanisms and great flexibility to adapt to circumstances. It would not be a big surprise whether there will be another milestone achieved in understanding regulation of life, which could be used to study this incredible complexity.

Green fluorescent protein coding sequence: cloning and expression of the visualization marker

Martin Chalfie, winner of the Nobel Prize for Chemistry in 2008 together with Osamu Shimomura[48] and Roger Tsien[49], for the discovery of GFP considers this Nobel Prize to be the third prize awarded to the nematode[50]. This third "worm prize" underlined that this model system is a powerful tool to develop new techniques, which could change the perception of biology in general. After a seminar of Paul Brehm about biological light production in sea organisms, he had an exciting idea to put the coding sequence (cds) of a green fluorescent protein (GFP) after a tissue specific promoter[51]. The GFP was discovered in a jellyfish (*Aequorea victoria*) as a by-product of aequorin, which is able transmitting blue light to green light [52]. The full protein is a stable, simple and short protein (238 AA), and moreover, has a relatively small absorption and excitation range. Its chromophore is a hexapeptide starting from amino acid 64, and the primary sequence requires a cyclization of serine-dehydrotyrosine-glycine to perform its illuminating function, which could occur in *E. coli* and *C. elegans*[51]. Thus, GFP does not require any additional jellyfish specific post translational modification (PTM) enzymes to get the same functional characteristics in other eukaryotes and could therefore be applied in different kinds of organisms without any toxic effects, including prokaryotes[51]. Since Martin Chalfie was interested in the development of sensory neurons in the nematode, he and his co-workers cloned the GFP cds downstream of the *mec-7* promoter, which normally allows specific expression of *β-tubulin* in neurons[53]. With expression from an extrachromosomal array, strong GFP expression was detected in the cell bodies of the touch receptor neurons during development[51].

The green fluorescent protein has several advantages over other previous used systems. In contrast to previously available techniques as *in situ* hybridization, *β*-galactosidase assay, and gene specific antibodies to visualize gene products, GFP could be used in living cells as a fusion protein or driven by a promoter of interest. Moreover, GFP is applicable in every cell type, GFP is non-invasive, does not disrupt the cell's metabolism, and due to its stability it is inheritable to next generations, which could be an advantage or a disadvantage. The usage of GFP allows studies involving cellular localization, timing, and other dynamics of gene expression. These studies have been shown to be very valuable for understanding the characteristics of genes[50]. The application of GFP as a bio-marker is therefore considered to be one of the biggest changes in fundamental biomedical research, and it led to key findings in understanding the complex organization of life.

A new era in C. elegans genome modification

The nematode already plays an important role as an animal model in biomedical science. With the new sequencing methods, new mapping algorithms, and screening approaches, screens for mutated genes are costing less time and less labor. Taking a look into the future, developments of these techniques and the advantages that the nematode has for research, this model system will teach us a lot about how life is organized. One of the ways will be via a complete library of mutants carrying a mutation in every single gene leading to a dysfunctional gene product. We give an overview of the techniques that will contribute in establishing this valuable source.

Recent developments in techniques for genome manipulation open possibilities to complete the mutant library in C. elegans

To gain knowledge about the function of a gene in an organism, genomes are screened for mutations with a phenotype of interest (forward genetics) or genes of interest are targeted directly and made dysfunctional to observe a phenotype (reverse genetics). There is a pallet of techniques available to achieve a dysfunctional gene, regardless of the purpose and animal model. Forward genetics is a laborious, though a random approach to link single genes to a specific phenotype, which can propose a biochemical, biological, and mechanistic function[54]. In this review we will briefly describe what impact 'high throughput sequencing' or 'whole genome sequencing' (WGS) methods are. *C. elegans* could become the second completed 'mutanome' (a loss of function mutation in every known gene), as *Saccharomyces cerevisiae* was the first eukaryote to have this database completed in 2002[55]. Furthermore, we will extensively describe the variety of newly available techniques based on transposon excision. Initially these transposon insertions were used as an alternative for EMS mutagenesis. However, these transposon sites are now also used as platforms to perform gene deletion, insertions and development of transgenes. These newly developed reverse genetic approaches allow studying gene function in the nematode in a more controlled way. Reverse genetics normally starts with a gene of interest and the search for an available mutant. When applicable in the preferred animal model, the gene is cloned (or for deletions, an alternative sequence) in a so called 'cassette'. This cassette contains flanking sequences and is recombined within the genome by the homologues recombination system to detect the impact of the gene of interest[56]. The first reverse genetic approaches were performed in *Drosophila melanogaster* by using P element based transposase mediated insertions[57]. In several model organisms it is shown that reverse genetics is a powerful method to focus on a single gene or a genetic cascade, and elucidate its function.

Here, we will focus on the methods that are available to manipulate the *C. elegans*'s genome, with emphasis on recently developed techniques such as homologues recombination, RNA interference (described with Nobel prize winners), and Mos-1 mediated insertions, replacements and deletions. We state that these techniques strongly contribute to our understanding of fundamental processes.

Endogenous transposon elements as a tool to study gene function

Endogenous DNA transposon elements allowed modifying the nematode's genome under controlled circumstances in the lab. These techniques were further improved during the 1980s and 1990s, and moreover, incredible progressions were reported during the last years[58]. Transposition of genomic DNA could be classified in two categories, replicative transposition and nonreplicative transposon activity. The best characterized endogenous transposon elements in *C. elegans* are the *Tc1* and *Tc3* elements, which were primarily applied to perform experimental procedures[59]. Tc1 transposase is a member of the nonreplicative *Tc1/mariner* superfamily and is a homologue to one of the first characterized transposons in biology, the bacterial Mu transposase[8]. Moreover, this family is unique in its kind, since it is widely represented in the evolutionary phylogenetic tree and until now the most common DNA transposon family found in nature, which implies that this family uses a host independent mechanism. The *Tc1/mariner* element is characterized by a 1,610 bp long DNA sequence coding for a transposase[60] in a single open reading frame[61], which is flanked by two 54 bp long non-conserved terminal inverted repeats (except for the 5'CAGT terminal sequence)[62, 63](**see figure 5A**). When cellular circumstances allow transcription and activation of the elements, the transposase binds to its corresponding flanking sequence with its active DDE domain (consisting out of two aspartic acid residues and a glutamic acid residue). Interestingly, this DDE domain was as well identified in a retroviral integrase[64] and is also present in bacterial transposon elements, suggesting a common ancestral origin[65]. The DDE domain allows phosphate transfers, resulting in a double stranded break (DSB) with a 3'TG overhang excised transposon sequence. This excised element integrates at the 3' end of a target TA sequence in the genome. Transposition of a single copy DNA occurs, and a 4bp gap is repaired by the cellular machinery[59] (**see figure 5B**). Via this way of activating the endogenous transposon elements, insertional mutagenesis screens have been performed.

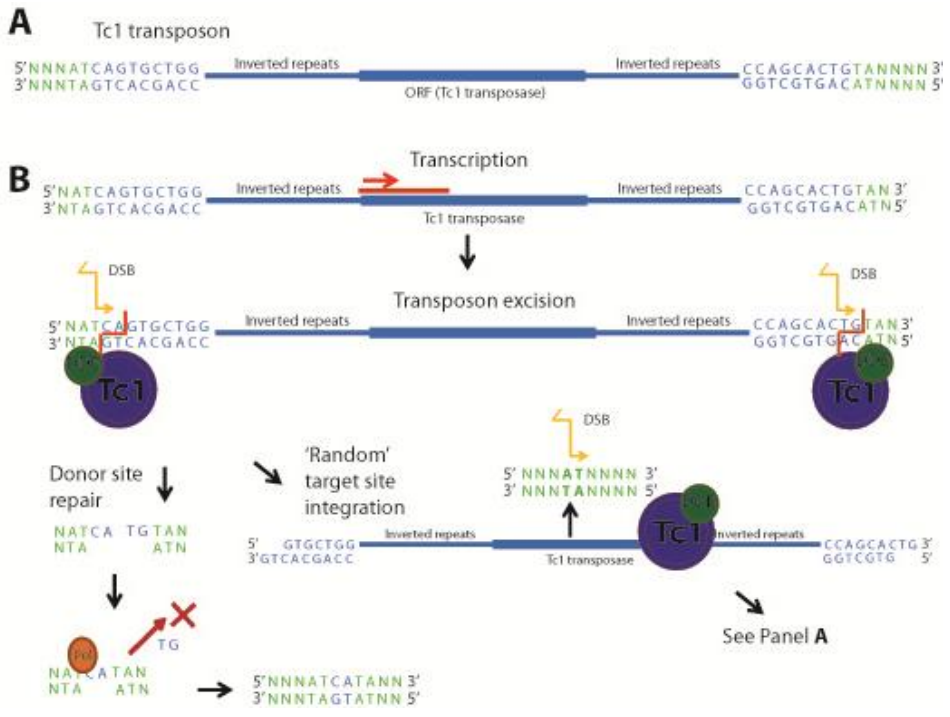


Figure 5: Excision of the Tc1 transposon superfamily. A Tc1 transposon element (**A**) consist of one open reading frame (ORF), flanked by non-conserved repeats with at the outside ends a 8 nt long conserved sequence (blue letters) which are integrated in the genome with a TA sequence (green). Cellular circumstances can allow transcription of this ORF (**B**) and this gene product (Blue with green circle) can cleave this element out of the genome mediated by the DDE domain with a DSB and a 3' TG overhang as consequence. The donor site got repaired and the 1.6 kb long element got integrated elsewhere in the genome containing a TA site.

However, the endogenous transposon excision method with Tc1 has a major disadvantage. There are approximately 31 endogenous copies of Tc1 active and 22 copies of the Tc3 transposon element in the standard Bristol N2 wild type laboratory strain[66]. These elements are strongly repressed in the germline[67] by the RNAi suppression complex, which actively degrades transposase RNA transcripts[68, 69, 70]. Strikingly, the *mutator* (*mut*) strains that were used to bypass the germline RNAi suppression while the transposon excision protocol was performed[71], seemed to be members of the RNAi complex[42, 58]. To obtain inheritable mutants out of insertional mutagenesis screens, the *mut* strains were used, and different backgrounds (*mut-7*, *mut-14* and *mut-16*) were available to reactivate Tc1 transposition in the germline[70, 72]. Nevertheless, these *mut* backgrounds were genomically instable, and a high rate of background mutations as random insertions, unrepaired double stranded breaks, or even chromosomal rearrangements occurred and were difficult to control in the lab[73, 74]. Furthermore, those studies were not performed in wild type backgrounds, and therefore more difficult to extrapolate to other studies, since linked mutations could not be fully excluded[75]. These problems can be partly circumvented with several outcrossings and rescue experiments to confirm the link between the observed phenotype and the mutated genomic locus. Recently, transposon excision based techniques were developed to avoid the RNAi defense mechanism, by using exogenous *Mos1*. This transposase was isolated from *Drosophila mauritiana*[76, 77], and activation of this 'foreign' transposon element in the nematode allowed performing screens under controlled circumstances in the standard wild type Bristol N2 background. Initially, *Mos1* was used for forward genetic approaches by *Mos1* based transposon mutagenesis, whereas recently, a variety of gene specific insertions and deletion techniques have been developed which are based on these identified random integration sites of *Mos1*. These techniques were called *Mos1* single copy insertion (MosSCI), *Mos1* deletion (MosDel), and *Mos1* excision-induced transgene-instructed gene conversion (MosTIC). Each of the available approaches is described in this overview.

Mos1: random insertional mutagenesis

Multiple copies of the Tc1 and Tc3 transposon elements in the genome made it difficult and laborious to identify the mutated genes due to background effects. In 2001 the *Mos1* transposase was isolated from the fruit fly and it was introduced in the nematode under control of the heatshock promoter[78]. This wild type transgenic background (strain: genotype *lin-15(n765); oxEx166[HSP ::MosTRANSPOSASE, lin-15(+), unc-122 ::GFP]*) carried a GFP cds under an unc122 promoter, resulting in marking coelomocytes (innate immune system). To check the presence of the extrachromosomal array containing the rescuing cds, the used strain carries a point mutation in *lin-15*. Since mutations in the *lin-15* lead to a multi vulva phenotype[79], absence of the array induces this phenotype. To perform random insertional mutagenesis in the wild type strain, a transgene carrying the *Mos1* transposase template sequence and a wild type strain expressing GFP under a pharynx associated promoter *myo-2* (strain: genotype *oxEx229 [Mos1 Substrate, myo-2::GFP]*) are combined. This crossing lead to double transgenic animals (F0), which were used to perform the heat-shock protocol (In general: 1 hour at 33°C, 1 hour at 15°C and 1 hour at 33°C). Transpositions that occurred in the germline, and therefore could be passed to the progeny, were selected and grown under normal conditions (F1). This generation was crossed out once, and non-fluorescent F2 animals were characterized by inverse PCR analysis with *Mos1* primers at the termini of the integration sites (mostly left side of the *Mos1* terminus) and random PCR primers allowing product formation[78, 80].

Individuals carrying successful insertions were analyzed by sequencing of the PCR purified band, resulting in a region where the insertion occurred, which allows to associate the region with a gene region with standard Blast analysis. The insertion is eventually confirmed by sequencing the whole integration site with flanking primers up and downstream of the putative *Mos1* site.

A common argument against transposition approaches is that the transposase does not 'cut and paste' randomly in the genome. Indeed a pilot study showed that *Mos1* has a strong preference for the ATATATAT consensus sequence[81], which is closely related to the endogenous Tc1 and Tc2 transposase consensus sequence[82]. Furthermore, the fruit fly transposase has a bias to insert its template into introns, even corrected for the overall presence of the TA availability in the nematode genome. A hotspot for the number of integration sites was detected in a 4 kb region on chromosome I (corresponding to the rDNA locus). Other preferences however, could not clearly be determined. There was a weak spot found as well, a relatively infrequent hit chromosome by *Mos1* is chromosome V[81]. Nevertheless, taking in account the advantages in time and therefore costs compared to classical random mutagenesis screens, transposase-based mutagenesis is a competing tool in mutagenizing the majority of *C. elegans* genes. Moreover, this approach becomes more competitive when placed in the current developments of newly derived high throughput techniques[54], and semi automatically identification of insertions[83]. Another striking advantage lies within the fact that the generated libraries of *Mos1* insertions can be used for gene replacement or gene deletions, as we will describe later on in this overview.

Mos1 initiated deletions through homologues recombination

Directly targeted deletion of genes in the nematode was lacking until recently, since the traditional homologues recombination assays were missing, beside the usage of the ultraviolet sensitive mutagen 4-5',8-trimethylpsoralen (UV/TMP)[84]. To perform reverse genetics, the desired gene deletion was identified in randomly UV/TMP treated worms by PCR and di-deoxy sequencing[85]. However, the availability of libraries containing *Mos1* insertions throughout the genome, allowed the usage of the transposition and recognition capacities of the *Mos1* transposase. The initiated DSB is repaired with the cellular repair mechanism via NHEJ or in rare events with homologues recombination. The latter mechanism allowed insertion of selection markers and deletion of targeted genomic DNA located closely to a *Mos1* insertion, which were generated during the random mutagenesis screens of the *C. elegans* consortium. The principle of this technique lies within the supplied extrachromosomal DNA, mainly in plasmid form. The targeting constructs were designed by choosing the closest available *Mos1* insertion site to the gene of interest. This *Mos1* insertion site's upstream and downstream sequences were used to perform the deletion. The right homology arm was designed approximately 2kb downstream of the insertion (normally ~2 kb). Subsequently, the left homology arm of ~3 kb is designed, which lies within or upstream of the gene of interest. Those elements were cloned into a targeting vector containing a selection marker for *unc-119* in between the homology arms. After the deletion procedure in *unc-119 (-)* worms[86], the plates are screened for the rescue of the *unc-119* mutation. Furthermore, successful integration could also be visualized as loss of the red fluorescent co-injection markers and thus the extrachromosomal array. With this procedure, multiple closely located genes could be deleted and basically allowing investigation of

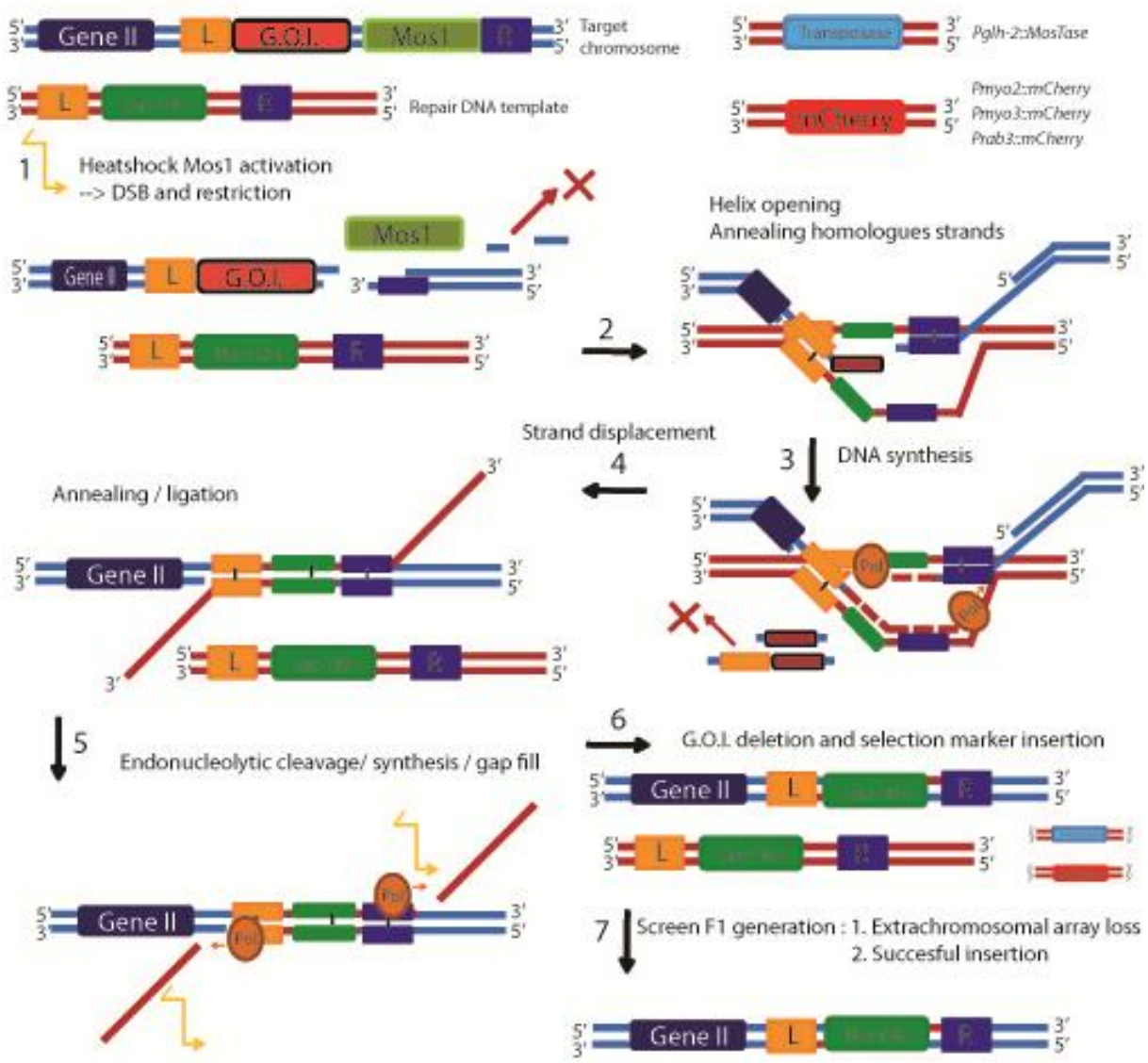


Figure 6: Achievements of gene deletions through homologues recombination caused by closely located Mos1 excision. Mos1 activation causes a DSB, and subsequent processing of the DSB with 3' overhangs (**1**). In rare occasions these overhangs are homologously recombined by the cellular repair mechanism, either with the homologues region on the sister chromatid or in this case a supplied repair template containing the replacing sequence containing the desired selection marker (**2**). During the repair, the homologues regions are the primers, and DNA polymerase synthesizes the complementary sequence of the repair template (**3**), and the remained chromosomal overhang containing the gene of interest (G.O.I.) is cleaved off (**3**). The strands are displaced and ligated (**4**), and DNA gaps are repaired after the excessive produced strands were removed (**5**). The selection marker replaced the g.o.i. (**6**), which allows screening of this selection marker in the next generation (**7**).

gene clusters, since approximately 25 kb regions could be deleted[86]. Moreover, adjacent genes or tandem duplicated genes having redundant function, could be deleted in one procedure [86]. It is proposed that 99.4% of the 10,154 genes lacking a mutant strain fall within the 25 kb range of the homologues recombination system and therefore could be deleted. However, in some of the cases multiple genes should be deleted when a gene lies in between the *Mos1* template and the target gene, which is a major disadvantage of this technique. Therefore more strains carrying random *Mos1* insertions in the genome should be generated. Nevertheless, this MosDel approach will be a key technique to support reverse genetic approaches and might be helpful in reaching the goal of knocking out all genes.

Mos1 excision induced transgene-instructed gene conversion and Mos1-mediated single copy insertions

Using the publically accessible *Mos1* insertion library[83], which was generated by the GENETAG consortium allowed taking advantage of the DSB that occurs during transposition. The essence of the approach is activating the *Mos1* transposase under certain conditions (e.g. the germline) and breaking the DNA structure. A repair template injected into the nematode's germline, which has homologues arms with a Mos1 region of interest closely located to the target gene, allows specific homologues recombination to occur. A proof of principle experiment was performed with two *Mos1* inserted elements, located in the *unc-5* and the *unc-63* sequence [73]. One of the two experiments showed that *unc-63* mutants could be rescued (2 out of 33,000 screened animals), since the associated impaired locomotion[87] was restored. Thus, the re-introduced gene was recombined and inserted through a neighboring Mos1 initiated DSB break. Those findings resulted in further optimization of the protocol, which resulted in two approaches to create gene specific transgenes or insert (or replace) genes carrying a mutation. The first technique based on homologues recombination is Mos1 excision induced transgene-instructed gene conversion (MosTIC) that was designed by the Bessereau lab around 2009 [75]. From this, some Mos1 insertions allow having single copy insertions (MosSCI) under certain circumstances[6].

The customized alleles could be inserted by activating homologues recombination machinery, which is available in the cell. DSB could be repaired by two cellular mechanisms, namely through a process that is called non-homologues end-joining (NHEJ) which involves ligation with the evolutionary conserved ligase IV of the DNA ends. The second mechanism to stabilize the broken DNA occurs via homologues recombination. In general it is rarely activated by the cell, however, when a template is supplied the frequency of these events increases[71]. Homologues recombination could be distinguished in two mechanisms. One is based on homology within the same chromosome, which is more deleterious since it uses 5' to 3' exonuclease activity extensively and non-homologues regions are therefore lost[75]. MosSCI and MosTIC are based on the fact that homologues recombination occurs with the homology of the sister chromatid to perform gene conversion (see **figure 2**). The initiated homologues recombination gene conversion pathway reacts on the DSB caused by the Mos1 excision. The 3' overhang regions are acting as primer sites, and the DNA synthesis complex synthesizes the DNA. The sister chromatid has not necessarily to be the repair template, as other homologues regions flanking the DSB break could also be the source of synthesis, as are other chromosomes or supplied homologues regions, which is the principle of this mutagenesis approach (see **figure 7**).

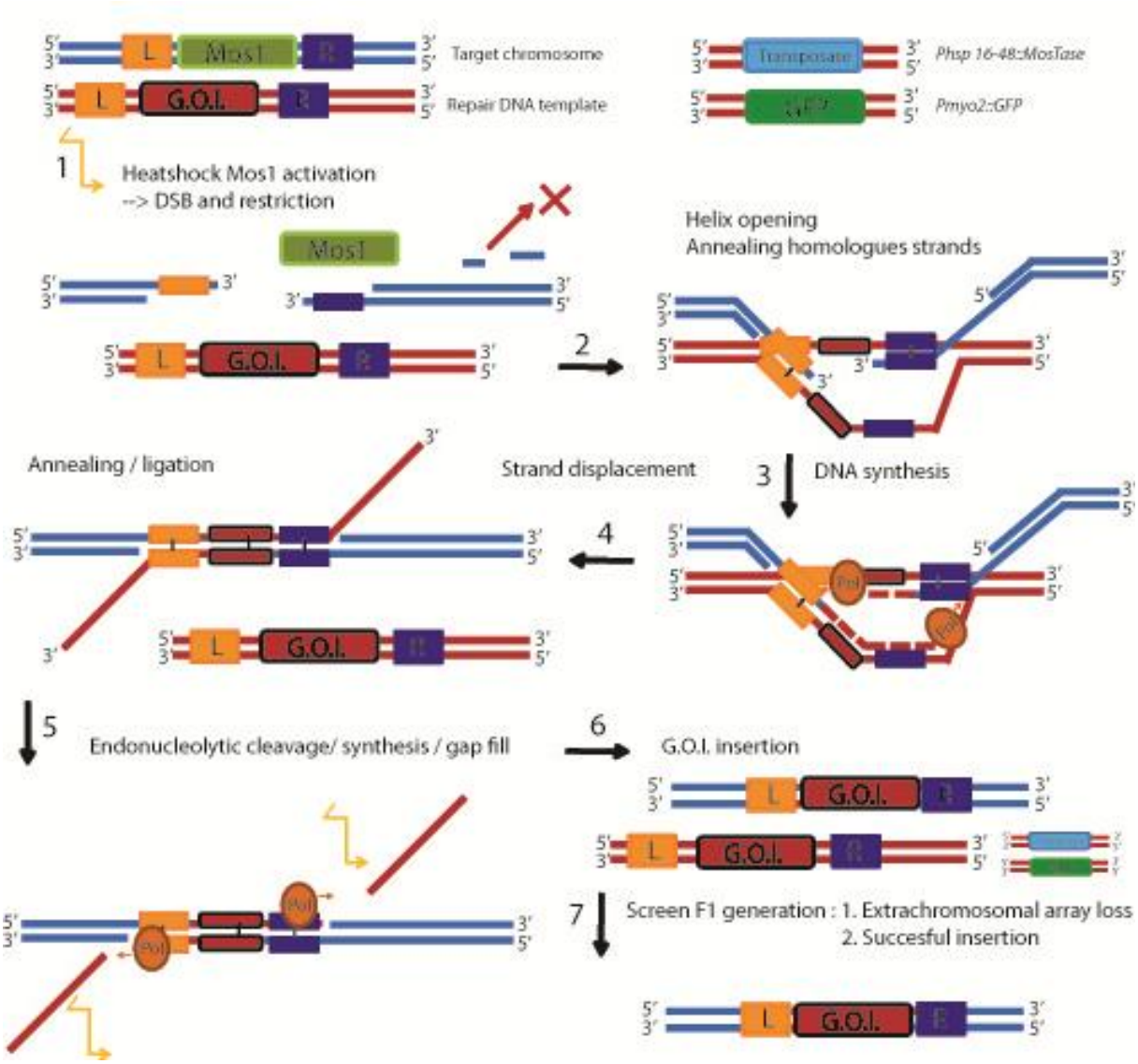


Figure 7: Representations of the events and supplied materials for gene insertions via Mos1 initiated homologues recombination. Mos1 activation causes a DSB, and subsequent processing of the DSB with 3' overhangs (**1**). In rare occasions these overhangs are homologously recombined by the cellular repair mechanism, either with the homologues region on the sister chromatid or in this case a supplied repair template containing the replacing sequence having the desired sequence of interest (G.O.I.) (**2**). During the repair, the homologues regions (L and R) are the landing sites, and DNA polymerase synthesizes the complementary sequence of the repair template (**3**), and the remained chromosomal overhangs are cleaved off (**3**). The strands are displaced and ligated (**4**), and DNA gaps are repaired after the excessive produced strands were removed (**5**). The sequence of interest replaced the targeted *Mos1* site (**6**), or as not depicted here with another genomic region of interest. The extrachromosomal arrays should be lost and confirmation of integration should be checked in the next generation (**7**).

MosTIC is dependent on a certain genomic position where the Mos insertion is localized, whereas MosSCI uses one position to introduce insertions. The MosTIC procedure is almost the same as the MosDel experimental steps. During MosTIC a certain sequence of interest (e.g. GFP insertion or a specific mutation) will be exchanged for the wild type sequence by homologues recombination with supplied homologues sequences for the Mos sequence and the genomic region of interest in the 'repair vector'. MosTIC gives opportunities to tag a gene of interest in an *in vivo* context in a time efficient manner, without having extrachromosomal arrays during expressional analysis studies[75]. Here we will describe the MosSCI, since this technique allows single insertions in the genome, and therefore comparing two genomic identical strains with a single variance (read inserted strain vs.

"WT" strain). Moreover, for structure and functional studies this technique had multiple major advantages like expression levels on an endogenous level, germline expression is possible, and between generations the expression levels are equal.

To achieve single copy insertions, the position of the Mos1 insertion should be at a neutral site, as it does not interfere with promoter, enhancer or operon activity. Additionally, the genomic region should not influence the transgene's expression. During the design of the technique[86], there was a focus on genomic regions with an insertion from 3' to coding sequences (tail-to-tail), and the key paper of the Jorgenson lab focused on the *ttTI5605* Mos1 insertion in the centric region of Chromosome II. The strain carrying this Mos insertion was crossed with the *unc-119(ed3)* mutant line. Starvation of these mutant animals prohibits reproduction, and therefore an extra selection criterion for the insertion procedure. Phenotypically these worms could be characterized by their reduction in size and decreased locomotion activity due to paralysis [88]. The cassette to perform homologues recombination to integrate the gene of interest with the *unc-119* selection marker, contains homologues arms of 1.4 kb upstream (left arm) and downstream (right arm) of the Mos insert in *ttTI5605* genomic region. Note that *C. briggsae* derived *unc-119* rescued worms either are carriers of the extra-chromosomal array or have had successful integration of the cassette after the procedure. To distinguish the injected worms with a successful transposase excision event with the false positive worms, co-injection of this cassette with extra-chromosomal non-homologues plasmids is necessary. Negative selection markers such as mCherry (cocktail: pGH8; *Prab-3::mCherry*, pCFJ104; *Pmyo-3::mCherry*, pCFJ90; *Pmyo-2::mCherry*) and a gain of function K^+ channel gene *TWK-18* are added to the array. The *TWK-18(gf)* gene is temperature sensitive and is driven under a *myo-3* promoter (*Pmyo-3::twk-18*), allowing muscle deactivation when the worms are put on 25°C. At 15°C this negative selection marker does not cause paralysis, however, the mCherry negative selection marker is visible to check extra-chromosomal array depletion.

This strategy could be performed in two ways, the first way is a heat shock promoter driven Mos1 transposase expression (pJL44; *Phsp-16-48::transposase*) or the second, with Mos1 transposase expressed in the germ line (pJL43.1; *Pglh-2::transposase*). Initially to show that this approach works, a 3 kb fragment containing GFP under control of a coelomocytes specific promoter (*punc-122::GFP*) was inserted in the worm's genome via the heat shock protocol (see **figure 8**). Screening of 1,000 animals from five independent *unc-119* rescued lines, ten individuals were not paralyzed and showed no mCherry expression, and had correct integration in the *ttTI5605* region on chromosome II. Interestingly, only seven out of ten lines had visible GFP expression in coelomocytes. Northern blot analysis showed no expression in these three lines, and further focusing on the PCR analysis showed aspecific product formation. This emphasizes that PCR screenings should be checked with sequence based approaches. Moreover, biased deletions could be observed more directly with such an approach, since *Mos* excisions are frequently accompanied with deletions[75, 86]. Some experiments do not allow usage of chromosome II, therefore the researchers introduced a strain carrying an integration site in the genomic region *cxTi10882* on chromosome IV accomplishing to the strict criteria as well. This genomic region showed the same integration ratio of approximately 1%. Both sites are able to integrate large templates with the same efficiency (up to 10 kb fragments), allowing to generate complex transgenic lines with this approach.

Since transgenic expression of the identified lines seems uniformly equal in every line, it is highly likely that single copies of the fragment are integrated. The experiment with Pglh2::GFP

fragment integration, only one line showed tandem integration of the cassette. Conclusive was the experiment with western blot showing equal expression of a similar experimental approach of 9 kb integration cassette containing *unc-18* designed with a C-terminal *mCherry* tag. *unc-18* loss of function mutants could be rescued with this integration through Mos excision induced homologues recombination (see **figure 8**). The insertion efficiency was compared with biolistic transformation[89], where gold coated DNA molecules induce random integration into the genome. The authors observed different levels of *mCherry* expression and different expression levels in the two isolated biolistic transformed strains[86]. In three out of the four MosSCI integrated worms however, *unc-18-mCherry* was expressed equally as the wild type strain, implying that single copy insertions have endogenous like expression levels.

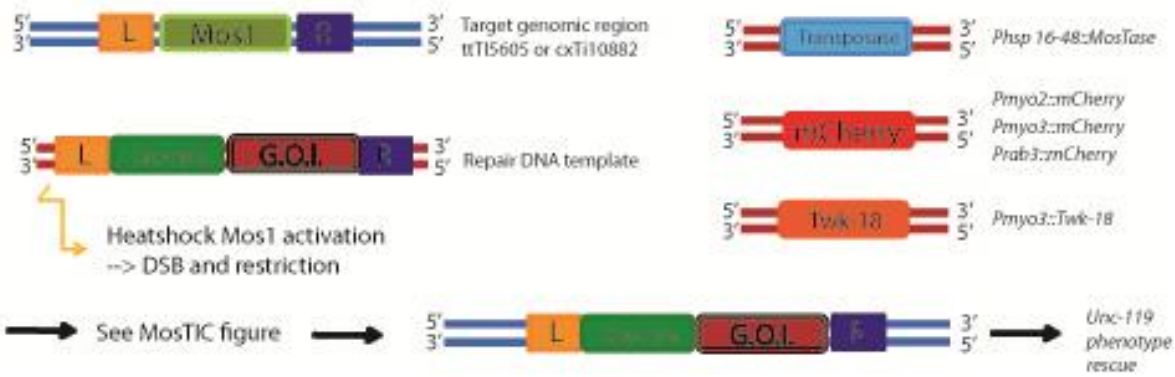


Figure 8: Required adaptations to the MosTIC protocol to insert a single gene copy into the nematode its genome and applying the MosSCI. Instead of designing homologues regions around the *Mos1* site close to the region of interest, the homologues sequences are used for *Mos1* sites in selected regions. With the integration sequence selection markers are enclosed to perform quick screens.

To achieve expression of the transgene in the germline is more complicated with extra-chromosomal arrays, since there are multiple replicates available in the gonad cells, activating the RNAi pathway. The approach with single copy transgenes may circumvent the activation of the RNAi pathway. The strain containing the Mos1 insertion in the ideal genomic region *ttTI5605* on chromosome II is permissive to express transgenes in the germline. With an integration efficiency of 12.5% of a cassette with Histon2b-GFP driven by a germline specific promoter (*pie-1*) in heatshocked P0 animals the protocol had a higher efficiency than integration in somatic cells. However, only 26% of the transgenic animals did show a specific GFP expression pattern. Strikingly, over multiple generations the expression showed an increase, suggesting that the endogenous RNAi defense system adapts to the single copied transgene. The other silenced strains could be explained by gene conversion caused mutations, for example, due to amplification of extra copies of the inverted repeats in the *pie-1* promoter region, which could occur during the annealing and synthesis process when homologues recombination is performed. Nonetheless, single copy transgenes could be generated, in both mitotic and meiotic cells, and moreover shows endogenous expression levels. To ensure that the transgene is well integrated in the genome, this tool requires controls to exclude errors with several selection steps. As recently was shown with the protein synthesis inhibitor G-418 antibiotic selection marker[90], adaptations to the protocol could enhance selection steps more efficiently [91].

Instead of co-injecting a Mos1 transposase cassette driven under a heatshock promoter in the gonads, the array could be supplied with a Mos1 transposase driven by the germline specific promoter *Pglh-2*. All the other selection constructs were co-injected together with the integration vector containing the *unc-119* and *Punc-122::GFP*. P0 *unc-119* mutant animals were grown on separate plates to get F2 worms. Those worms were screened, and *unc-119* rescued worms showed an integration ratio of 18.8%, which was significantly higher than the simultaneously performed heatshock protocol (8.6%). Expression of the integrated transgene in the germline was successful and similar to endogenous levels[86]. Moreover, this direct integration protocol has less handling and is more efficient, although the protocol takes slightly more time than the heatshock protocol. In addition, the other identified integration site on chromosome IV is also functional for direct integration. These methods are now widely used for analyzing gene function, structure and other genetic related questions.

Remarks on transposon based genome modification

Forward genetic screens with transposon excisions is a valuable way to perform screenings, since it is highly time efficient. Especially when there is an exogenous transposase injected to limit the amount of endogenously caused background mutations, as observed with Tc1 based screens. The advantage of the occurrence of DSB during the Tc1/marinier superfamily DNA fragment translocation has been fully used to develop techniques to modify the genome of *C. elegans* more easily and precisely than before. This gives opportunities to answer fundamental questions about gene function during several complex biological processes by applying MosSCI or MosDel.

An important objection to using transposon based screens is that the “jumping” of the elements does not occur randomly in such an extent that every coding sequence of a gene could be mutated, since these elements need a consensus sequence to unwire and break the DNA helix. Despite the fact that *Granger L. et al.*[81] showed that every chromosomal region is hit “equally”, except for one hot spot on chromosome I, it is likely to be difficult to get every gene mutated to complete the knock out database, and since it is shown that the mutational frequency of transposon excisions is significantly lower than EMS or UV/TMP.

However, in addition to mutating every single gene, this mutant bank could also be used to perform conditional gene expression experiments with the *mec-8* conditional expression system, which was developed recently[92]. In principle the technique is based on a member of the splicing machinery, the ubiquitously expressed *mec-8*[93]. This protein is able to alternatively splice transcripts when the *mec-2 intron-9* is introduced in the gene of interest (via the Mos inserts). Since the used *mec-8* mutated allele is temperature sensitive, a simple temperature shift induces re-expression of the gene of interest at a desired time point is established, within 15 minutes of temperature change[92]. This technique is a great addition to the already existing tools, which allows studies to investigating spatio-temporal effects of the gene of interest.

Single nucleotide polymorphisms generated by EMS mutagenesis

Another way to perform random mutagenesis is to generate point mutations with the efficient mutagen EMS (ethyl methanesulfonate)[94]. These single base mutations could lead to a premature stop codon in a cds, which is presumably to be accepted as a full null mutant (when a gene product is largely truncated). However, these random point mutations could also give rise to other types of mutations in genes such as hypermorphic, hypomorphic or antimorphic lesions, which are worth to study as well[95]. With the current rapid development and broader appliance of high throughput techniques as whole genome sequencing (WGS), big genetic screens could be performed in a more time and money efficient way[54]. Especially the fine mapping step used to be time consuming with the classical PCR mapping systems, which were applied during the previous decade[95]. In several animal models WGS was successfully applied to identify genomic lesions causing a specific phenotype (e.g. yeast[96, 97], plants, flies[98], bacteria[99] etc.). Moreover, the background mutations closely located to the artificially induced polymorphism of interest, also referred to as linked background mutations, could be more easily identified and removed by selected out crossing to obtain an as pure as possible WT population carrying a single mutation[95]. With the so called deep sequencing (based on short reads) appropriate software to align the reads to a reference genome was challenging to design. Developed alignment software, such as MAQGene, shows the EMS-induced polymorphisms and deletions, between the wild type reference genome and the mutagen treated genome[100]. Deletions induced by EMS treatment could be over a kilobase long[101], and therefore the short reads produced by deep sequencing make it difficult to identify. MAQGene software simply adds gene regions to the output list, which have a significantly lower coverage. Subsequently, if these deletions are a point of interest, confirmation could be performed with the traditional di-deoxy sequencing. Comparing different *C. elegans* mutant strains with the reference genome and the out crossing strain revealed that out crossing induces deletions as well, although in a lower extent than the highly mutagenic EMS[102]. Hence, it is recommended to apply classic rough mapping in parallel with deep sequencing of the mutant strain, to link the phenotype of interest to a certain genomic locus, since every chromosome contains approximately 300 genetic variants per chromosome (of which 30 in protein coding sequences)[95].

There are different second generation sequencing platforms available, e.g. ABI Solid (Applied Bioscience) or Illumina Genome Analyzer (Solexa) allowing sequencing of parts of genomes, exomes, transcriptomes (RNAseq) and complete genomes (for an extensive overview of all the technical details of next generation sequencing techniques[103, 104]). Different approaches could be chosen to identify SNPs or mutations induced by a mutagenesis protocol[102]. Briefly, sequencing with emulsion PCRs, as the Solexa or Applied Bioscience platform, amplifies isolated genomic DNA (or retro-transcribed RNA). For the Applied Bioscience platform these fragmented strands are ligated to an adaptor, and a single ligated DNA molecule is annealed to a bead to perform sequencing. Subsequently, random octameric primers with two terminal fluorescent bases are ligated to the adapted target strand and cleaved of to determine two bases. Each of the four bases has a different fluorescent label, allowing the camera to determine the 3' base and the 5' base. This step is repeated several times to get an eventual strand of ~35 bases[103] (see **figure 9**). Currently a run takes ~4.5 days to sequence a fully fragmented library, with a data output up to 10 gigabase. For the other important platform, the Illumina Genome Analyzer, the adaptor molecules are attached to a solid surface. Since there are adaptors on both the 5' end and the 3' end, these DNA molecules will form a bridge. Denaturation of the double stranded DNA results single stranded molecules that function as probes

on a DNA microarray. During the annealing phase the adaptor specific primers can anneal to the adaptor molecules at the 3' side of the template. During the synthesis phase the DNA polymerase recognizes the nucleotides with a 3' situated fluorescent dye. Only one nucleotide per synthesis cycle is incorporated, since the fluorescent dye functions as a 3' blocker[103, 104] (see **figure 9**). With the current up-scaling of the high throughput techniques and the upcoming single molecule sequencing the efficiency in a cost reducing manner is increased (e.g. labor, chemicals etc.), while subsequently the error rate of the techniques is reduced[104].

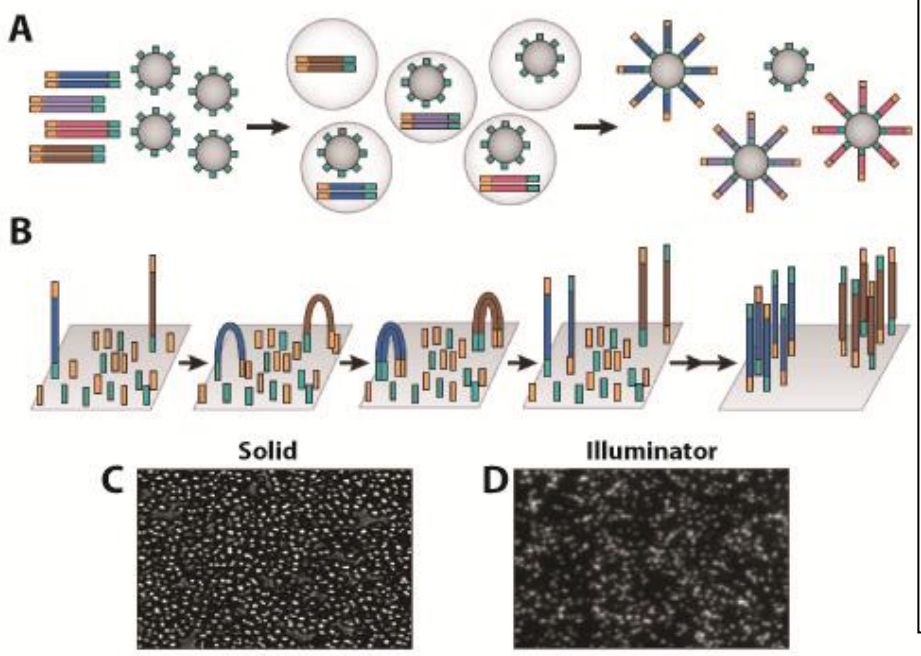


Figure 9: Schematic representation of the next generation techniques of the Solid and the Illuminator. Adapted from J. Schendure *et al.*, 2008[110]. The Solid platform (**A**) is based on beads (**C**) where the adaptor sequences can bind to. Fluorescent nucleotides excite, which allows to determining the order of the sequence during sequencing every single round (**A**). The Illuminator platform has an array where adaptor sequences are put on, and these strands could form bridges (**B**) to multiply the sequence. Every round there is a measurement of newly added nucleotides which are fluorescent (**D**).

Especially considering the fact that the genome is a bit smaller than higher multi-cellular organisms (~30 times smaller than humans), screening of the whole genome or exome would be even more cost effective. When F1 animals are screened for mutations in the genome, several outcrosses could be performed to obtain multiple phenotypes from one single F1 population (approximately 5 premature stop codons per screened genome[104]). This would save time consuming phenotype screenings to complete the mutant database of *C. elegans*. Furthermore, WGS allows the possibility to gain knowledge about gene clusters, regulatory sequences in the genome, and splicing variants. In summary, the further fine tuning of the WGS technique in combination with EMS or transposon excision based mutagenesis will give a complex resource of future data about gene functions, their regulation, their roles in nematode development, and extrapolation to mammalian development or disease.

Concluding remarks

C. elegans is a well determined and dynamic model system, which allows researchers to investigate intriguing fundamental questions from a biologically interesting and relatively simple perspective. Due to major scientific achievements during the past decades, intensive characterization of the model and Nobel prize winning findings made the nematode increase its scientific status and it is now a well accepted model system. With respect to other model systems, the nematode has all the necessary genetic tools to modify the genome, perform conditional experiments, reporter assays, and knock down assays. Although homologues recombination could not occur in embryonic stem cells as homologous recombination is performed in the mouse[105], the eventual result is equal. However, there are still a few wishes in the worm community, such as stable cell lines of *C. elegans* to perform experiments on a cell specific level *in vitro*[106, 107]. Additionally, the two-component Gal4 system [108], which is based on cell or tissue-specific expression of the yeast transcription factor Gal4, which activates the expression of a second transgene containing Gal4 binding sites in the promoter (Gal4::transgene) is still lacking. The recent developments of gene specific homologues recombination with e.g. the MosSCI protocol will lead to such an implementation of Gal4 like technique in the *C. elegans*'s genome in the future.

More exciting results with mutant lines are expected in the near future. The strongly improved protocols will allow quick determination of new mutant lines. In this review we gave an example of an algorithm that is able to recognize gaps in obtained sequences by next generation sequencers. These developments allow screening of mutagenized genomes in a cost and time effective manner. In an ideal scientific environment, a phenotype or gene function is confirmed with an independent biological replica. With the current upcoming mutant databases this could be established by taking an identified mutant of the Mos1 mutagenized databank and one of the EMS mutant libraries carrying a predicted loss of function mutant (either a premature stop codon, or a severe amino acid mutation). However, completing the 'mutanome' via both ways is one of the milestones to achieve, and therefore this biological replica could be filled up with RNAi knock down assays. Overall, as a consequence of quick technical developments, a boost for fundamental biological research is on the way with the upcoming generation of new potential data sources. It is not a question whether we will achieve the complete libraries, but the ultimate challenge lies within the question of how we are going to analyze all this potential data via a high throughput pipeline? A part of this pipeline will be lying within the development of worm sorter platforms like the Union Biometrica COPAS instruments. These platforms allow grouping of whole organisms on size[109] or presence of fluorescence[110], and these instruments will streamline the selection in a more objective manner and save laborious screening work.

On the other hand, fundamental questions still need to be answered, in the sense of the biochemical properties of the used mutagens. Although each mutagen has its advantages (e.g. transposon integrations allow homologues recombination, EMS a variety of mutations) there are disadvantages (e.g. transposon integrations do not occur randomly, EMS could cause severe background mutation) that should be considered when a conclusion is taken from a mutant. The solution may be lying within the biological replicas. Combining the mutant libraries (read: use two independent mutants from two independent techniques) enables studying gene function in a more

definite manner. Subsequently, these mutants of different origins can be rescued with MosSCI or reproduced with MosDel to confirm the observed phenotype with the linked genomic lesion. We state that the nematode is the ideal platform to initiate this 'objective' scientific approach, since *C. elegans* allows imaging due to its transparency, cell lineage tracking, studying of complex cell behavior and tissue formation, 'simple' animal behavior, and gene specific tests of drugs. In conclusion, we reviewed an exciting animal model, which already learnt us a lot and recent technical developments will give more interesting insights for the worm community, biology in general, and the recent established research area of systems biology.

Besides technical questions that remain, there is another fundamental question which needs to be answered: do we still need animal model systems when we are able to identify gene alterations with the new high throughput techniques?[111] In the past decade, several comparing methods have been developed, such as Genome Wide Association Studies (GWAS), where patients are grouped in a class of disease (e.g. severe obesity, obese, normal weight) and common genomic alterations are identified. Fundamental objections to couple an inheritable Single Nucleotide Polymorphism (SNP), insertions or deletions to gene function or a genomic region with a regulating function (e.g. enhancers, tandem repeats etc.) are that the used subjects live in different environments and carry a heterogenic gene pool. These features reduce the statistical power, neglect subtle variations which can be the key to an answer or treatment and in many cases GWAS studies only show a part of the complex underlying mechanism[112]. There are several arguments to continue with animal models in the new era of high throughput techniques, and among those even favoring animal models than being a danger. Animal models allow reducing the number of variables, the desired setting can be generated, and direct effects can be measured more easily. Ethically charged experiments can even be reduced in population size due to newly developed high throughput techniques, since one sacrifice creates a source with a huge amount of data (e.g. expression analysis, WGS etc.)[111]. For the nematode, the use of which creates less ethical concerns, these developments are even more an advantage than a disadvantage, as discussed previously. This new era will only strengthen this model system, since there is always a desire to extrapolate findings to other organisms. With studies in humans where it is difficult to perform follow up experiments, the nematode may function as a good model system to pursue findings. When the worm community is able to streamline and combine the several high throughput techniques in one platform, this simple multi-cellular organism will be an ideal model to widen the scope of biomedical research. Furthermore, completely focusing on one model system, will affect scientific objectivity and that will be a negative and unwanted development. Overall, the ongoing developments will only enhance the position of model systems such as *C. elegans*, and expand knowledge in multiple disciplines.

Literature

1. Sulston, J. E. C. elegans: the cell lineage and beyond. *Biosci. Rep.* **23**, 49-66 (2003).

2. Stiernagle, T. Maintenance of C. elegans. *WormBook*, 1-11 (2006).

3. Hodgkin, J. Karyotype, ploidy, and gene dosage. *WormBook*, 1-9 (2005).

4. Barnes, T. M., Kohara, Y., Coulson, A. & Hekimi, S. Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. *Genetics* **141**, 159-179 (1995).

5. Spieth, J. & Lawson, D. Overview of gene structure. *WormBook*, 1-10 (2006).

6. Frokjaer-Jensen, C. *et al*. Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat. Genet.* **40**, 1375-1383 (2008).

7. Hefferin, M. L. & Tomkinson, A. E. Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair (Amst)* **4**, 639-648 (2005).

8. Benjamin, H. W. & Kleckner, N. Excision of Tn10 from the donor site during transposition occurs by flush double-strand cleavages at the transposon termini. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4648-4652 (1992).

9. McMahill, M. S., Sham, C. W. & Bishop, D. K. Synthesis-dependent strand annealing in meiosis. *PLoS Biol.* **5**, e299 (2007).

10. Sasaki, M., Lange, J. & Keeney, S. Genome destabilization by homologous recombination in the germ line. *Nat. Rev. Mol. Cell Biol.* **11**, 182-195 (2010).

11. Paques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* **63**, 349-404 (1999).

12. Mello, C. C. Return to the RNAi world: rethinking gene expression and evolution. *Cell Death Differ.* **14**, 2013-2020 (2007).

13. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94 (1974).

14. Brenner, S. Nature's gift to science (Nobel lecture). *Chembiochem* **4**, 683-687 (2003).

15. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.* **56**, 110-156 (1977).

16. Horvitz, H. R. Worms, life, and death (Nobel lecture). *Chembiochem* **4**, 697-711 (2003).

17. Horvitz, H. R. & Sulston, J. E. Isolation and genetic characterization of cell-lineage mutants of the nematode Caenorhabditis elegans. *Genetics* **96**, 435-454 (1980).

18. Kimble, J. & Hirsh, D. The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. *Dev. Biol.* **70**, 396-417 (1979).

19. Sulston, J. E. & White, J. G. Regulation and cell autonomy during postembryonic development of Caenorhabditis elegans. *Dev. Biol.* **78**, 577-597 (1980).

20. Greenwald, I. S., Sternberg, P. W. & Horvitz, H. R. The lin-12 locus specifies cell fates in Caenorhabditis elegans. *Cell* **34**, 435-444 (1983).

21. Ambros, V. & Horvitz, H. R. Heterochronic mutants of the nematode Caenorhabditis elegans. *Science* **226**, 409-416 (1984).

22. Greenwald, I. Lin-12, a Nematode Homeotic Gene, is Homologous to a Set of Mammalian Proteins that Includes Epidermal Growth Factor. *Cell* **43**, 583-590 (1985).

23. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-854 (1993).

24. Watt, F. M. & Hogan, B. L. Out of Eden: stem cells and their niches. *Science* **287**, 1427-1430 (2000).

25. Ellisen, L. W. *et al*. TAN-1, the human homolog of the Drosophila notch gene, is broken by chromosomal translocations in T lymphoblastic neoplasms. *Cell* **66**, 649-661 (1991).

26. Sulston, J. E. Neuronal cell lineages in the nematode Caenorhabditis elegans. *Cold Spring Harb. Symp. Quant. Biol.* **48 Pt 2**, 443-452 (1983).

27. Hedgecock, E. M., Sulston, J. E. & Thomson, J. N. Mutations affecting programmed cell deaths in the nematode Caenorhabditis elegans. *Science* **220**, 1277-1279 (1983).

28. Ellis, H. M. & Horvitz, H. R. Genetic control of programmed cell death in the nematode C. elegans. *Cell* **44**, 817-829 (1986).

29. Yuan, J., Shaham, S., Ledoux, S., Ellis, H. M. & Horvitz, H. R. The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme. *Cell* **75**, 641-652 (1993).

30. Hengartner, M. O., Ellis, R. E. & Horvitz, H. R. Caenorhabditis elegans gene ced-9 protects cells from programmed cell death. *Nature* **356**, 494-499 (1992).

31. Ellis, R. E. & Horvitz, H. R. Two C. elegans genes control the programmed deaths of specific cells in the pharynx. *Development* **112**, 591-603 (1991).

32. Nicholson, D. W. Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death Differ.* **6**, 1028-1042 (1999).

33. Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to C. elegans CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405-413 (1997).

34. Fire, A. *et al*. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811 (1998).

35. Fire, A. Z. Gene silencing by double-stranded RNA. *Cell Death Differ.* **14**, 1998-2012 (2007).

36. Guo, S. & Kemphues, K. J. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611-620 (1995).

37. Grishok, A., Tabara, H. & Mello, C. C. Genetic requirements for inheritance of RNAi in C. elegans. *Science* **287**, 2494-2497 (2000).

38. Tabara, H., Grishok, A. & Mello, C. C. RNAi in C. elegans: soaking in the genome sequence. *Science* **282**, 430-431 (1998).

39. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).

40. Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**, 1017-1026 (1998).

41. Mello, C. C. & Conte, D.,Jr. Revealing the world of RNA interference. *Nature* **431**, 338-342 (2004).

42. Tabara, H. *et al*. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123-132 (1999).

43. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350-355 (2004).

44. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363-366 (2001).

45. Elbashir, S. M. *et al*. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494-498 (2001).

46. Nowotny, M. & Yang, W. Structural and functional modules in RNA interference. *Curr. Opin. Struct. Biol.* **19**, 286-293 (2009).

47. Baulcombe, D. RNA silencing in plants. *Nature* **431**, 356-363 (2004).

48. Shimomura, O. Discovery of green fluorescent protein (GFP) (Nobel Lecture). *Angew. Chem. Int. Ed Engl.* **48**, 5590-5602 (2009).

49. Tsien, R. Y. Nobel lecture: constructing and exploiting the fluorescent protein paintbox. *Integr. Biol. (Camb)* **2**, 77-93 (2010).

50. Chalfie, M. GFP: Lighting up life. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10073-10080 (2009).

51. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805 (1994).

52. SHIMOMURA, O., JOHNSON, F. H. & SAIGA, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **59**, 223-239 (1962).

53. Savage, C. *et al*. mec-7 is a beta-tubulin gene required for the production of 15-protofilament microtubules in Caenorhabditis elegans. *Genes Dev.* **3**, 870-881 (1989).

54. Hobert, O. The impact of whole genome sequencing on model system genetics: get ready for the ride. *Genetics* **184**, 317-319 (2010).

55. Giaever, G. *et al*. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387-391 (2002).

56. Moerman, D. G. & Barstead, R. J. Towards a mutation in every gene in Caenorhabditis elegans. *Brief Funct. Genomic Proteomic* **7**, 195-204 (2008).

57. Ballinger, D. G. & Benzer, S. Targeted gene mutations in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9402-9406 (1989).

58. Robert, V. J. & Bessereau, J. L. Manipulating the Caenorhabditis elegans genome using mariner transposons. *Genetica* **138**, 541-549 (2010).

59. van Luenen, H. G., Colloms, S. D. & Plasterk, R. H. The mechanism of transposition of Tc3 in C. elegans. *Cell* **79**, 293-301 (1994).

60. Vos, J. C., van Luenen, H. G. & Plasterk, R. H. Characterization of the Caenorhabditis elegans Tc1 transposase in vivo and in vitro. *Genes Dev.* **7**, 1244-1253 (1993).

61. Rosenzweig, B., Liao, L. W. & Hirsh, D. Sequence of the C. elegans transposable element Tc1. *Nucleic Acids Res.* **11**, 4201-4209 (1983).

62. Collins, J., Forbes, E. & Anderson, P. The Tc3 family of transposable genetic elements in Caenorhabditis elegans. *Genetics* **121**, 47-55 (1989).

63. Radice, A. D., Bugaj, B., Fitch, D. H. & Emmons, S. W. Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. *Mol. Gen. Genet.* **244**, 606-612 (1994).

64. Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J. & Skalka, A. M. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**, 851-860 (1991).

65. Fayet, O., Ramond, P., Polard, P., Prere, M. F. & Chandler, M. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol. Microbiol.* **4**, 1771-1777 (1990).

66. Emmons, S. W., Yesner, L., Ruan, K. S. & Katzenberg, D. Evidence for a transposon in Caenorhabditis elegans. *Cell* **32**, 55-65 (1983).

67. Emmons, S. W. & Yesner, L. High-frequency excision of transposable element Tc 1 in the nematode Caenorhabditis elegans is limited to somatic cells. *Cell* **36**, 599-605 (1984).

68. Sijen, T. & Plasterk, R. H. Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature* **426**, 310-314 (2003).

69. Ketting, R. F., Haverkamp, T. H., van Luenen, H. G. & Plasterk, R. H. Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133-141 (1999).

70. Vastenhouw, N. L. *et al*. A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. *Curr. Biol.* **13**, 1311-1316 (2003).

71. Barrett, P. L., Fleming, J. T. & Gobel, V. Targeted gene alteration in Caenorhabditis elegans by gene conversion. *Nat. Genet.* **36**, 1231-1237 (2004).

72. Tijsterman, M., Ketting, R. F., Okihara, K. L., Sijen, T. & Plasterk, R. H. RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs. *Science* **295**, 694-697 (2002).

73. Robert, V. & Bessereau, J. L. Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. *EMBO J.* **26**, 170-183 (2007).

74. Bessereau, J. L. Transposons in C. elegans. *WormBook*, 1-13 (2006).

75. Robert, V. J., Katic, I. & Bessereau, J. L. Mos1 transposition as a tool to engineer the Caenorhabditis elegans genome by homologous recombination. *Methods* **49**, 263-269 (2009).

76. Jacobson, J. W. & Hartl, D. L. Coupled instability of two X-linked genes in Drosophila mauritiana: germinal and somatic mutability. *Genetics* **111**, 57-65 (1985).

77. Jacobson, J. W., Medhora, M. M. & Hartl, D. L. Molecular structure of a somatically unstable transposable element in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 8684-8688 (1986).

78. Bessereau, J. L. *et al*. Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. *Nature* **413**, 70-74 (2001).

79. Clark, S. G., Lu, X. & Horvitz, H. R. The Caenorhabditis elegans locus lin-15, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* **137**, 987-997 (1994).

80. Williams, B. D., Schrank, B., Huynh, C., Shownkeen, R. & Waterston, R. H. A genetic mapping system in Caenorhabditis elegans based on polymorphic sequence-tagged sites. *Genetics* **131**, 609-624 (1992).

81. Granger, L., Martin, E. & Segalat, L. Mos as a tool for genome-wide insertional mutagenesis in Caenorhabditis elegans: results of a pilot study. *Nucleic Acids Res.* **32**, e117 (2004).

82. Preclin, V., Martin, E. & Segalat, L. Target sequences of Tc1, Tc3 and Tc5 transposons of Caenorhabditis elegans. *Genet. Res.* **82**, 85-88 (2003).

83. Duverger, Y. *et al*. A semi-automated high-throughput approach to the generation of transposon insertion mutants in the nematode Caenorhabditis elegans. *Nucleic Acids Res.* **35**, e11 (2007).

84. Flibotte, S. *et al*. Whole-genome profiling of mutagenesis in Caenorhabditis elegans. *Genetics* **185**, 431-441 (2010).

85. Yandell, M. D., Edgar, L. G. & Wood, W. B. Trimethylpsoralen induces small deletion mutations in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 1381-1385 (1994).

86. Frokjaer-Jensen, C. *et al*. Targeted gene deletions in C. elegans using transposon excision. *Nat. Methods* **7**, 451-453 (2010).

87. Culetto, E. *et al*. The Caenorhabditis elegans unc-63 gene encodes a levamisole-sensitive nicotinic acetylcholine receptor alpha subunit. *J. Biol. Chem.* **279**, 42476-42483 (2004).

88. Maduro, M. & Pilgrim, D. Identification and cloning of unc-119, a gene expressed in the Caenorhabditis elegans nervous system. *Genetics* **141**, 977-988 (1995).

89. Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. *Genetics* **157**, 1217-1226 (2001).

90. Colbere-Garapin, F., Horodniceanu, F., Kourilsky, P. & Garapin, A. C. A new dominant hybrid selective marker for higher eukaryotic cells. *J. Mol. Biol.* **150**, 1-14 (1981).

91. Giordano-Santini, R. *et al*. An antibiotic selection marker for nematode transgenesis. *Nat. Methods* **7**, 721-723 (2010).

92. Calixto, A., Ma, C. & Chalfie, M. Conditional gene expression and RNAi using MEC-8-dependent splicing in C. elegans. *Nat. Methods* **7**, 407-411 (2010).

93. Lundquist, E. A. *et al*. The mec-8 gene of C. elegans encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52 transcripts. *Development* **122**, 1601-1610 (1996).

94. Gengyo-Ando, K. & Mitani, S. Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode Caenorhabditis elegans. *Biochem. Biophys. Res. Commun.* **269**, 64-69 (2000).

95. Sarin, S. *et al*. Analysis of multiple ethyl methanesulfonate-mutagenized Caenorhabditis elegans strains by whole-genome sequencing. *Genetics* **185**, 417-430 (2010).

96. Irvine, D. V. *et al*. Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. *Genome Res.* **19**, 1077-1083 (2009).

97. Smith, D. R. *et al*. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res.* **18**, 1638-1642 (2008).

98. Blumenstiel, J. P. *et al*. Identification of EMS-induced mutations in Drosophila melanogaster by whole-genome sequencing. *Genetics* **182**, 25-32 (2009).

99. Srivatsan, A. *et al*. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* **4**, e1000139 (2008).

100. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851-1858 (2008).

101. Anderson, P. Mutagenesis. *Methods Cell Biol.* **48**, 31-58 (1995).

102. Sarin, S., Prabhu, S., O'Meara, M. M., Pe'er, I. & Hobert, O. Caenorhabditis elegans mutant allele identification by whole-genome sequencing. *Nat. Methods* **5**, 865-867 (2008).

103. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135-1145 (2008).

104. Ansorge, W. J. Next-generation DNA sequencing techniques. *N. Biotechnol.* **25**, 195-203 (2009).

105. Doetschman, T. *et al*. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**, 576-578 (1987).

106. Strange, K., Christensen, M. & Morrison, R. Primary culture of Caenorhabditis elegans developing embryo cells for electrophysiological, cell biological and molecular studies. *Nat. Protoc.* **2**, 1003-1012 (2007).

107. Strange, K. & Morrison, R. In vitro culture of C. elegans somatic cells. *Methods Mol. Biol.* **351**, 265-273 (2006).

108. Elliott, D. A. & Brand, A. H. The GAL4 system : a versatile system for the expression of genes. *Methods Mol. Biol.* **420**, 79-95 (2008).

109. Hertweck, M. & Baumeister, R. Automated assays to study longevity in C. elegans. *Mech. Ageing Dev.* **126**, 139-145 (2005).

110. Watson, D. A., Gaskill, D. F., Brown, L. O., Doorn, S. K. & Nolan, J. P. Spectral measurements of large particles by flow cytometry. *Cytometry A.* **75**, 460-464 (2009).

111. Aitman, T. J. *et al*. The future of model organisms in human disease research. *Nat. Rev. Genet.* **12**, 575-582 (2011).

112. Shao, H. *et al*. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19910-19914 (2008).