TECHNIQUES IN GENETIC ENGINEERING

Işıl KURNAZ

CHAPTER 8

Genetic Manipulation of Stem Cells and Animals

"In the beginning there is the stem cell; it is the origin of an organism's life. It is a single cell that can give rise to progeny that differentiate into any of the specialized cells of embryonic or adult tissue."

Stewart Sell, Stem Cells Handbook

Stem cells have been known for decades, and they have been used for knockout animal production or transgenics for over a score.

Mario Capecchi, Sir Martin Evans and Oliver Smithies were awarded the Nobel Prize for Physiology or Medicine in 2007 for their discoveries of "**principles for introducing specific gene modifications in mice by the use of embryonic stem cells"**, which they had been working on since late 1980s.

And in 2012, Nobel Prize for Physiology or Medicine was awarded to Sir John Gurdon and Shinya Yamanaka "for the discovery that mature cells can be reprogrammed to become pluripotent" (see http://www.nobelprize.org/nobel_prizes/medicine/laureates/2012/press.html).

So the excitement around stem cells in late 1990s till now was not for their identification, but rather for the potential of genetically manipulating and thereafter using them for other purposes, medical or commercial.

The term “stem cell” was actually coined back in 1908 by the Russion histologist Alexander Maximov to describe what we today refer to as “hematopoietic stem cells”.

In 1960s, Joseph Altman and Gopal Das presented the scientific evidence of adult neurogenesis (and thus the existence of adult neural stem cells), and in 1992 neural stem cells were cultured *in vitro*.

In 1963, presence of stem cells was shown in the bone marrow, and presence of hematopoietic stem cells in the human cord blood was discovered in 1978.

What was exciting was that the first human stem cell line was established in 1998 by Thomson and his team, and from 2000 onwards the applications of stem cell technology was on the rise:

In 2001, Advanced Cell Technology cloned the first early human embryos for the purpose of generating human embryonic stem cells, and in 2003 adult stem cells were discovered in children's primary teeth, which was a relief in the midst of heated discussions on the ethics of using embryonic stem cells.

In 2006, pluripotent cells were induced from adult cells (induced pluripotent stem, iPS, cells), later repeated for reprogramming of mouse skin cells and of human fibroblasts in 2007.

Stem cell technology and Knock-out cells

The widely accepted standard definition of stem cells relies on two major properties:

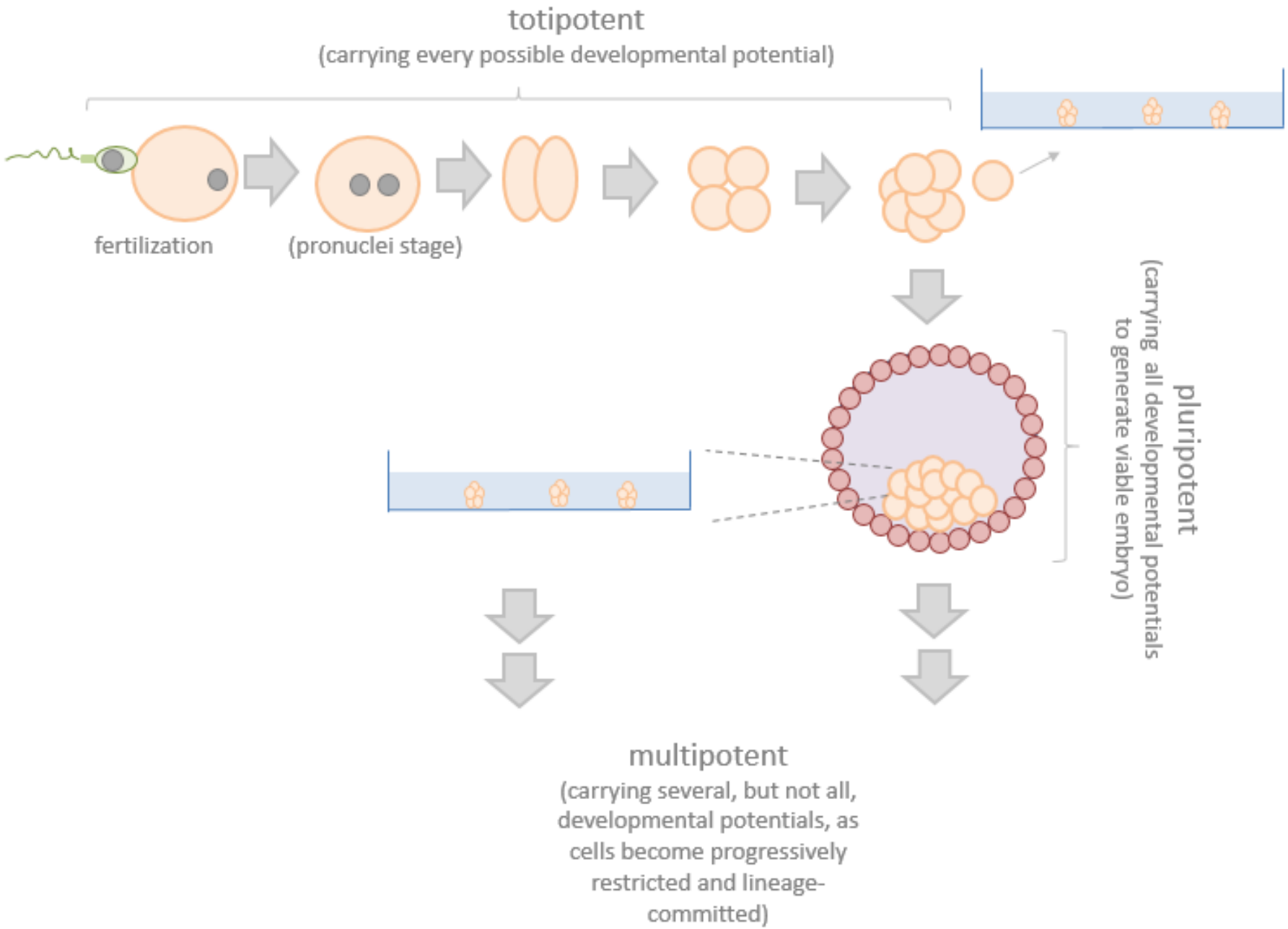
(1) self-renewal, ie the capacity to generate more stem cells like itself, and

(2) potency, ie the capacity to give rise to various differentiated cell types

The ultimate “stem cell” in that respect is the fertilized egg, the zygote, which gives rise to an entire embryo with all the necessary extraembryonic tissues necessary for its survival, which therefore is called **totipotent** (total potential).

The *embryonic stem cells,* or ES cells, however, are typically isolated from the inner cell mass of the blastocyst and are merely **pluripotent**, meaning they can give rise to many different cell types of the embryo, but cannot produce any extraembryonic tissues such as the placenta.

From gastrulation onwards the embryo already begins to differentiate, forming initially the three embryonic layers, ectoderm, mesoderm and endoderm, and thus the capacity of these cells are already restricted – these cells progressively become more and more restricted in their respective lineages, going from **multipotent** to **oligopotent, bipotential** or **unipotential** precursors, sometimes also referred to as stem cells.


totipotent
(carrying every possible developmental potential)
fertilization
(pronuclei stage)
pluripotent
(carrying all developmental potentials to generate viable embryo)
multipotent
(carrying several, but not all, developmental potentials, as cells become progressively restricted and lineage-committed)

For genetic manipulations, researchers either use embryonic stem cells or ES-like cells such as embryonic germ (EG) cells, embryonic carcinoma (EC) cells, fetal stem cells, umbilical stem cells, or induced pluripotent stem (iPS) cells.

Human ES cells can be obtained from the inner cell mass (ICM) of blastocysts from surplus IVF (in vitro fertilization) embryos that are donated for research. Alternatively, pluripotent ES-like cells can be obtained from terminated pregnancies. **Both of these procedures have been banned in many countries due to ethical considerations.**

For clinical use, ie transplantation of stem cells to patients, the general transplantation terminologies apply: stem cells to be transplanted could be

(a) autologous, meaning the patient's own stem cells are removed, stored, manipulated if necessary, and given back to the same person,

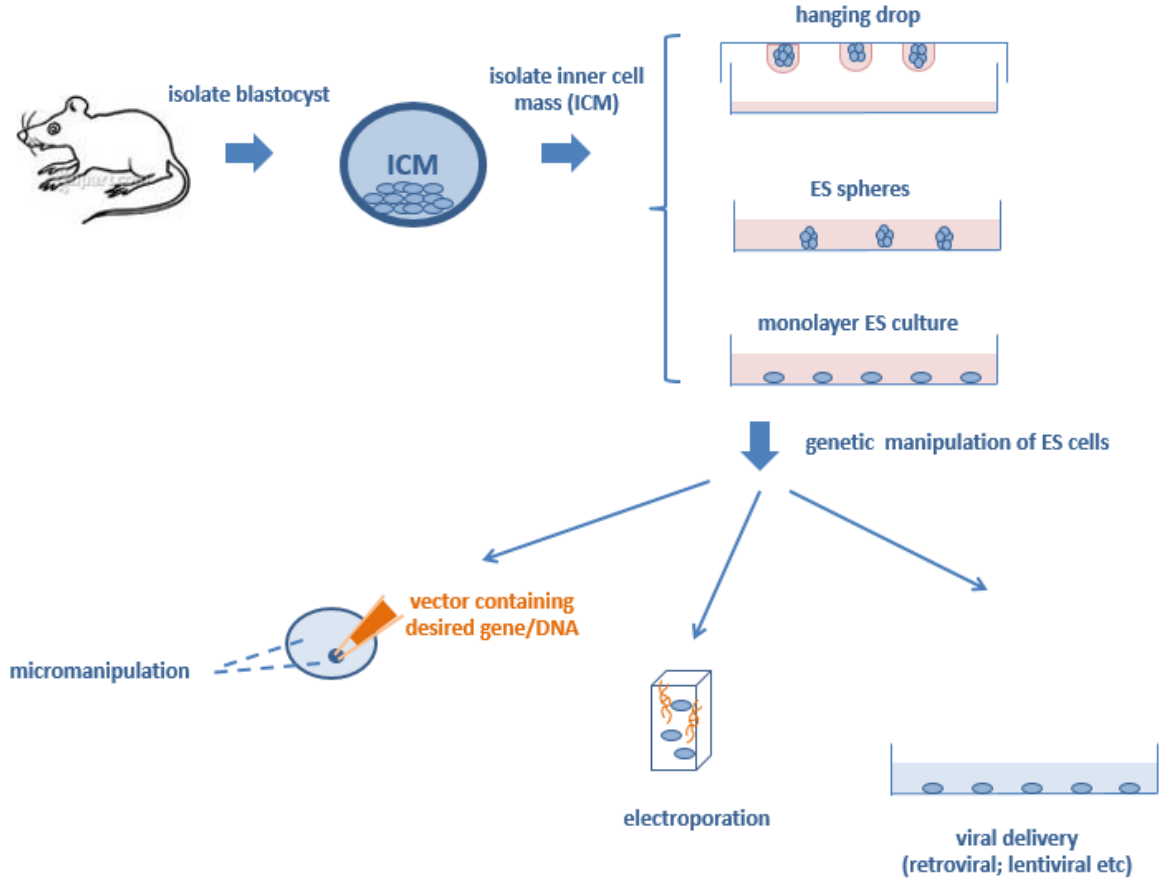
(b) allogeneic, meaning cells are isolated from a genetically non-identical person, stored, manipulated if necessary, and transplanted to the patient,

(c) syngeneic, meaning stem cells to be stored, manipulated and used are from a genetically identical or else immunologically compatible person, such as a relative of the patient, or

(d) xenogeneic, meanining stem cells to be used are from an immunologically as compatible as possible but non-human species (such as primates or pigs).

Genetic manipulation of embryonic stem cells

Since embryonic stem (ES) cells can be cultured as monolayers or expanded to form embyroid spheres, or clones, *in vitro,* they can be manipulated to stably express exogenous genes, introduced either by viral vectors, by electroporation, micromanipulation, or transfection (see Chapter 7). Random integration (or insertion) is commonly used to overexpress or mutate genes for large-scale screening purposes or to integrate reporter genes for monitoring differentiation or other developmental events; gene targeting in ES cells is employed to exchange endogenous genes with engineered ones (either to knock-out a functional gene by homologous recombination with a non-functional version, or vice versa).



Homologous recombination can be exploited when the transgene is to be targeted to the location of its endogenous counterpart, especially for knock-out or gene therapy applications. It was shown back in 1987 by Thomas and Capecchi that transgenes could integrate to the mouse ES genome through homologous recombination. Essentially, the homologous recombination targeting vector contains regions homologous to the targeted gene, between which a marker is inserted that would disrupt the endogenous gene (Fig.8.1)

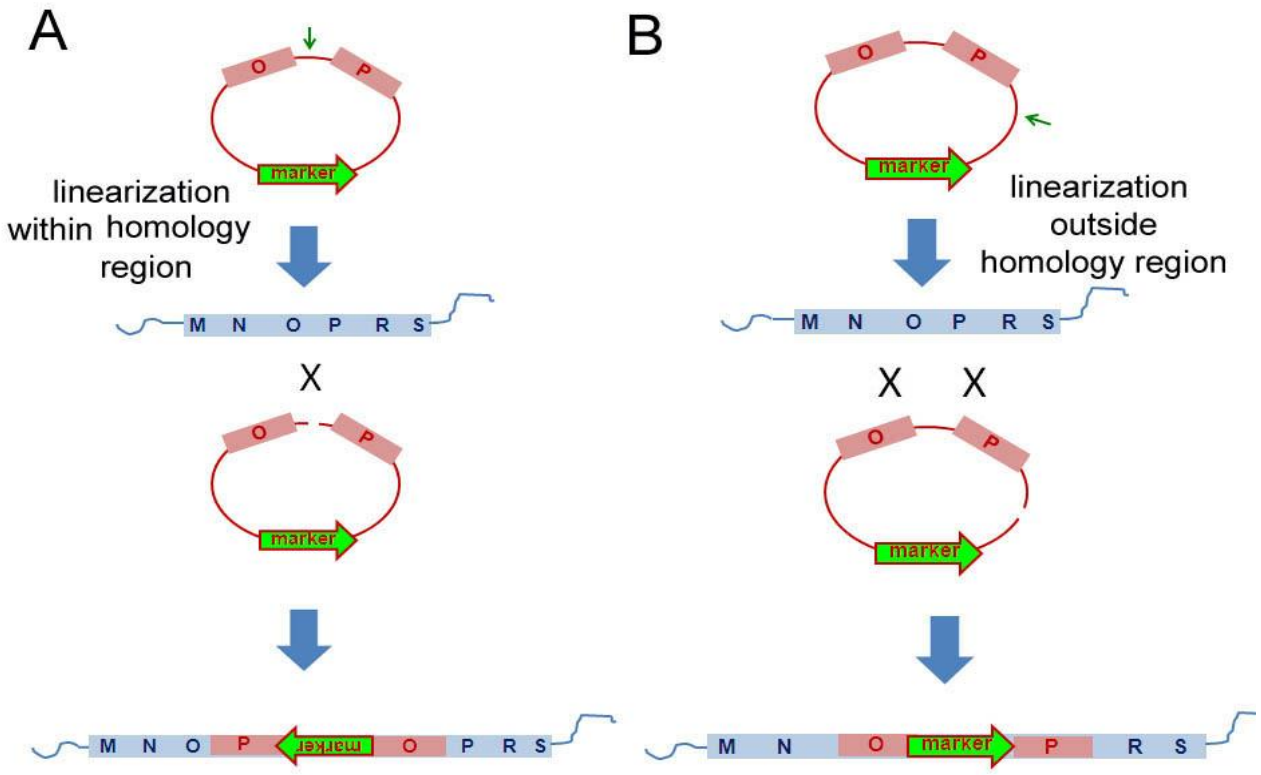


Figure 8.1. Schematic diagram of the basis of homologous recombination used for transgene insertion. (a) if the vector is linearized within the region of homology, then the marker is inserted into the target region in opposite orientation; (b) if the vector is linearized outside the region of homology, then the marker is inserted into the target region in the same orientation.

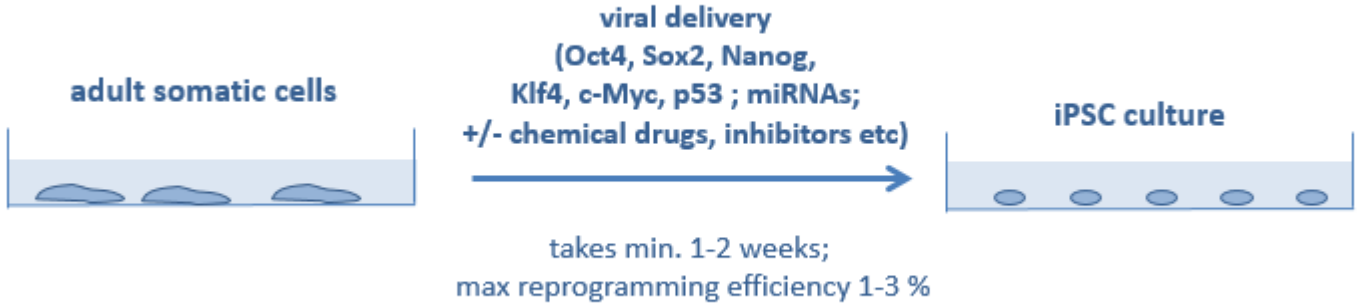
Ethical and Technical considerations behind ES cells & Induced pluripotent stem cells (iPSCs)

The use of ES cells for clinical purposes has been a subject of debate and criticism from many aspects… Therefore an alternative to ES cells was long needed.

This alternative came to existence thanks to work from Yamanaka's group in 2006, where they have converted adult somatic cells to ES-like cells through genetic manipulation and «reprogramming», which was termed **induced pluripotent stem cells (iPSCs)**.

Yamanaka was awarded the 2012 Nobel Prize for Physiology or Medicine, along with Sir John Gurdon, «for the discovery that mature cells can be reprogrammed to become pluripotent».

This reprogramming required viral delivery of «stemness factors» to the adult somatic cells, including Oct4, Sox2, Klf4 and c-Myc. Later on, a number of other factors, such as Nanog, was also added to the reprogramming coctail, whereas c-Myc was taken out, since it icreased the possibility of cancer in the iPSC recipient animals. Low efficiency of reprogramming or incomplete reprogramming are other disadvantages of iPSC production.



Transgenic animals

A transgenic animal is defined as an animal carrying a foreign gene that is inserted into its genome. This genetic modification is usually done at the ES cell stage, or at the pronucleus stage, although other methods are also available.

The result of this genetic manipulation may be either

- a **knock-in** or **transgenic** (where a new gene is introduced to the animal, bringing a new function to the animal),

- a **knock-out** (where a marker is introduced within a protein-coding genomic segment, rendering the gene in question non-functional), or

- a **conditional knock-out** (where the knock-out is switched on in certain tissues or in response to certain drugs / chemicals etc).

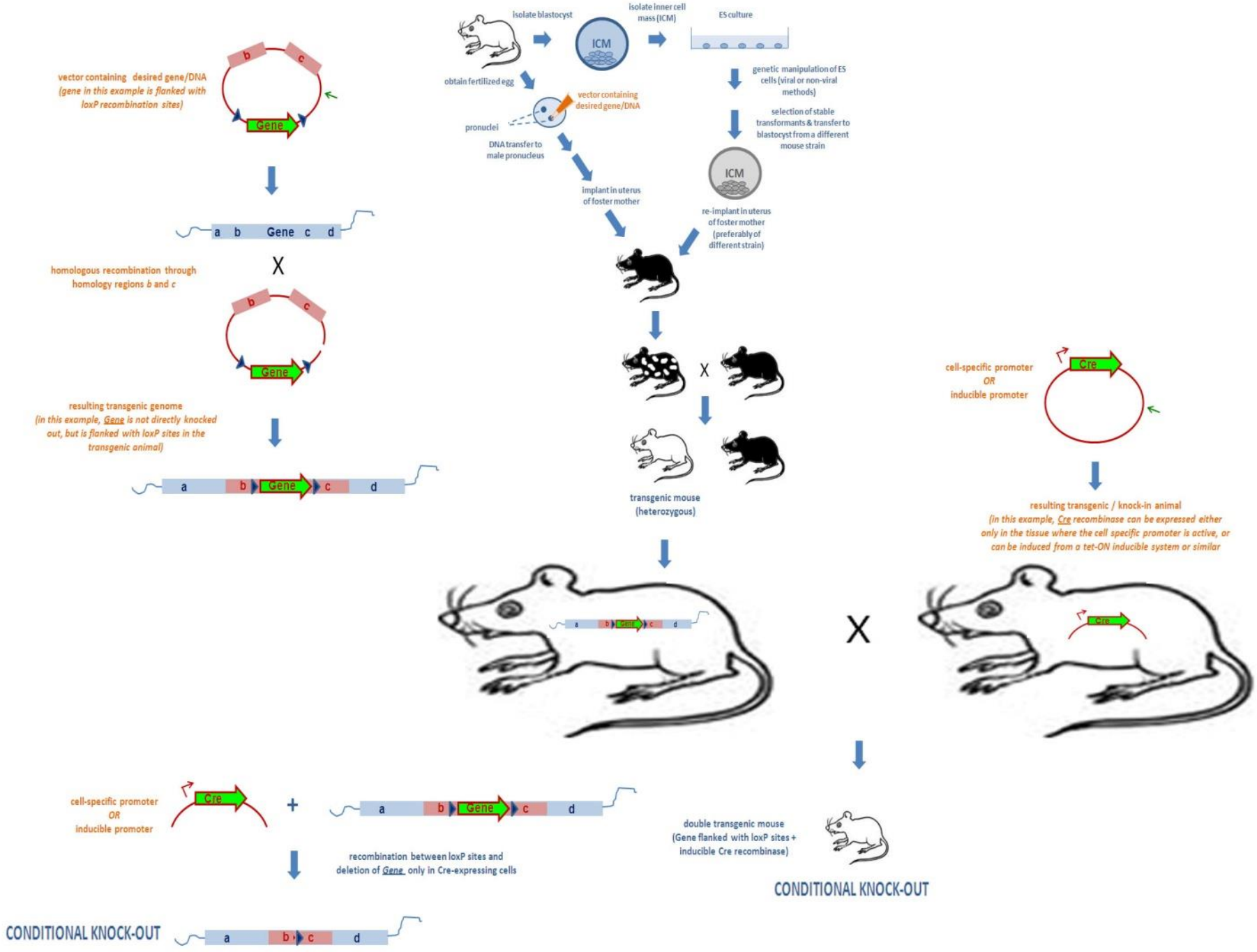


Figure 8.2. Schematic diagram of conditional knock-out generation.

The first step is to insert the loxP sites to either side of the target gene (which is embryonic lethal); in order to do that, homologous recombination can be used to replace the endogenous Gene with Gene flanked with loxP sites (see panel on the left). This vector can either be introduced either to the male pronucleus of the fertilized egg, or the Embryonic Stem (ES) cells, and the embryo that is thus generated is transferred to a foster mother from a different mouse strain (middle panel).

The chimeric mice born to the foster mother are further crossed to the mouse of the same strain as the foster mother, until pure heterozygous transgenic mouse of the original donor strain is obtained. This transgenic mouse still has the gene and expresses it, therefore can survive the embryonic period.

If this transgenic animal is then crossed to a second transgenic that expresses Cre recombinase either from a cell-specific promoter or from an inducible promoter (right panel), the double transgenic obtained from this cross can be considered a conditional knock-out: Cre recombinase will cause site-specific recombination between loxP sites and cause removal of the Gene from the genome either in a specific cell type (if Cre is expressed from a cell-specific promoter), or only upon induction by an agent at a desired embryonic or fetal or postnatal period (if Cre is expressed from an inducible promoter).

BAC genomic libraries have also been used in association with transgenic animal models, since they were found to reduce positional effects of transgenes, and since transgenes were expressed at physiological levels and as such exhibited developmental timing and expression patterns as in source organism.

Although BAC transgenesis has largely replaced conventional transgenesis methods due to its above-mentioned advantages, it also comes with its own set of disadvantages. BAC transgenesis occurs through random integration, and while positional effects are reduced, the number of copies can vary. The relatively low efficiency of generating transgenic founders means that more pronuclei have to be injected with BAC constructs to achive the same number of transgenic animals. And constructing a BAC library might take weeks to months, and when combined with the time required to inject more pronuclei (which may take much longer than in traditional transgenesis methods) the setup can be more time-consuming (Biel et al, 2012).

Gene Trapping is a random knock-in that is essentially used to randomly interfere with different genes' expression so as to identify genes associated with a particular phenotype, in large-scale and high throughput genetic screens. The gene trap vector can be of two versions, as discussed below, but ultimately is designed to prevent RNA splicing of genes into which the transgene is inserted.

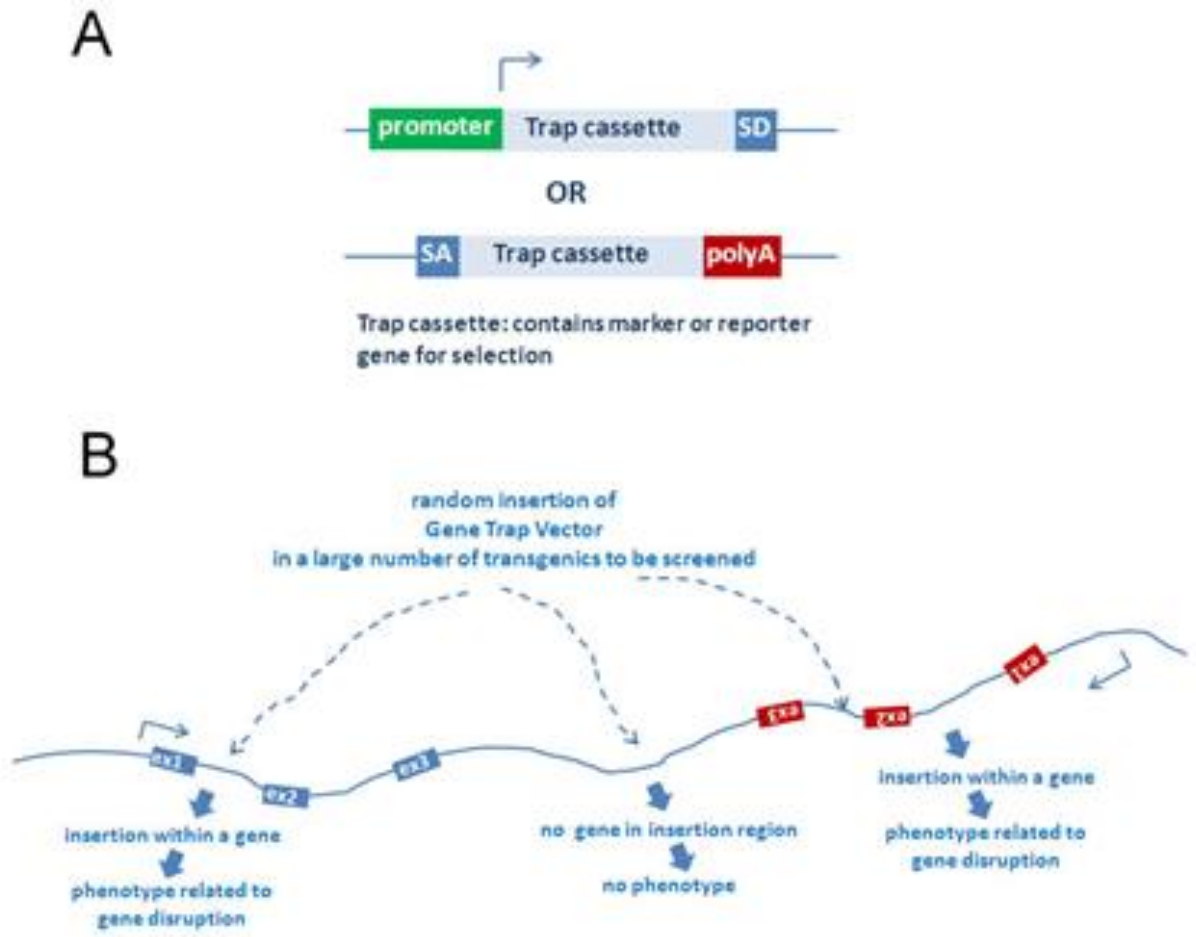


Figure 8.3. Schematic diagram of Gene Trapping principle.

(a) The two types of Gene Trap vectors.

(b) Random integration of the Gene Trap vector to the genome of mice; large number of such random transgenics will be analyzed for the desired phenotype. If the GeneTrap vectors integrate within a gene, expression from the gene will be disrupted to various degrees of severity (see Text for details); if, however, the vector is integrated in a chromosomal region that harbours no gene, then no gene expression will be detected, hence desired phenotype will not be observed.

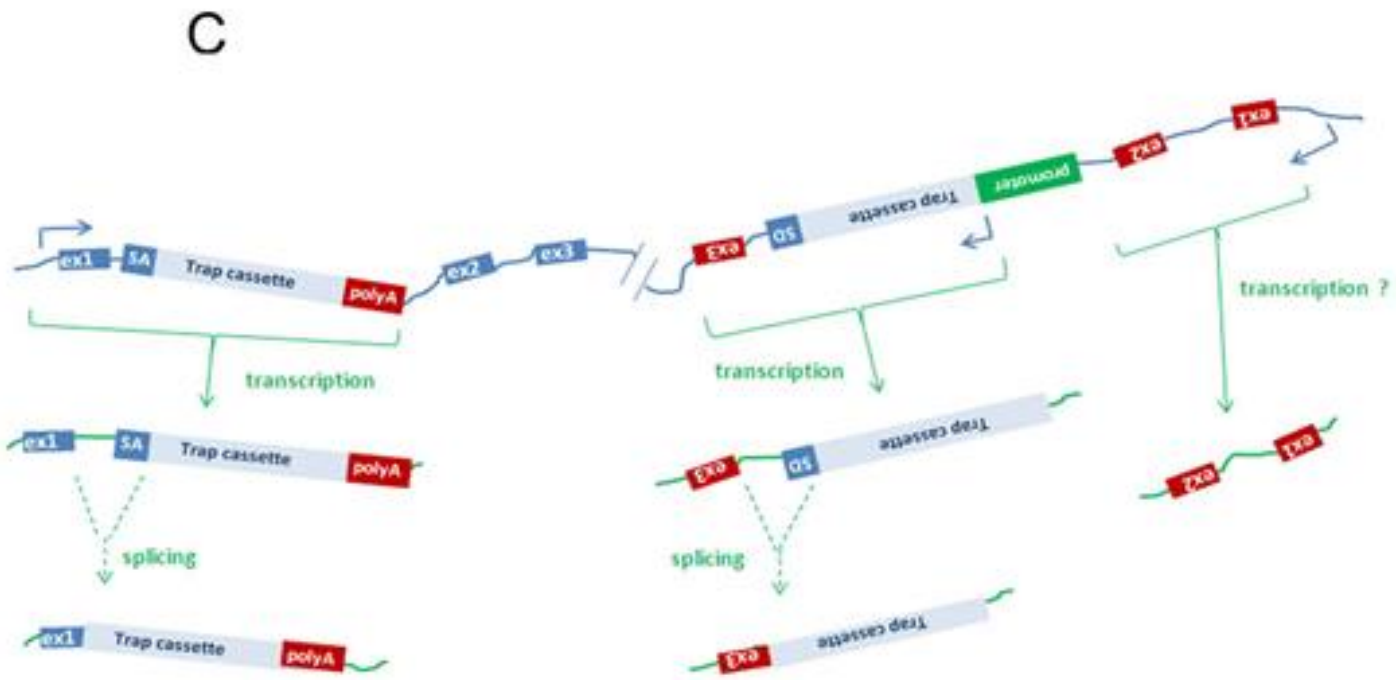


Figure 8.3. Schematic diagram of Gene Trapping principle.

(c) Two different types of Gene Trap vectors integrated to different gene regions are shown here to represent the functioning principle; if the first type of Gene Trap vector that contains an SA motif, a Trap cassette and a polyA signal is integrated within an intronic region, it will get transcribed into a primary transcript, and will be spliced into the mature mRNA through the SA site (left side). If the second type of Gene Trap vector that contains a promoter, a Trap cassette and an SD motif is integrated within an intronic region, it will interfere with the transcription of the trapped gene (ex1 and ex2 may or may not be transcribed properly; at best, the primary transcript will include the entire Gene Trap sequence), but also the Trap cassette will be transcribed from its own promoter and spliced to ex3 of the trapped gene (right side). (SA: splice acceptor; SD: splice donor; polyA: polyadenylation signal; ex: exon)

RNA interference and microRNAs

Antisense RNA has long been used to inhibit protein synthesis, hence study function of genes.

In an interesting study performed by Andrew Fire and Craig Mello in 1998, however, an interesting phenomenon was consistently observed: the double-stranded RNA (sense and antisense together) that was supposed to be "negative control" of the antisense RNA, the *C. elegans* muscle started some twitching motion, more prominent than the antisense-injected worm (Fire et al, 1998).

Fire and Mello have conducted a series of very elegant experiments and showed that the **RNA interference** mechanism was very effective in silencing gene expression, which led to their Nobel prize (http://www.nobelprize.org/nobel_prizes/medicine/laureates/2006/press.html).

RNA interference was later shown to also exist in *Drosophila*, *C. elegans*, and mammalian systems. It has then become a handy tool in studying gene expression, since it was much easier (took relatively less time) than generating knock-out mice.

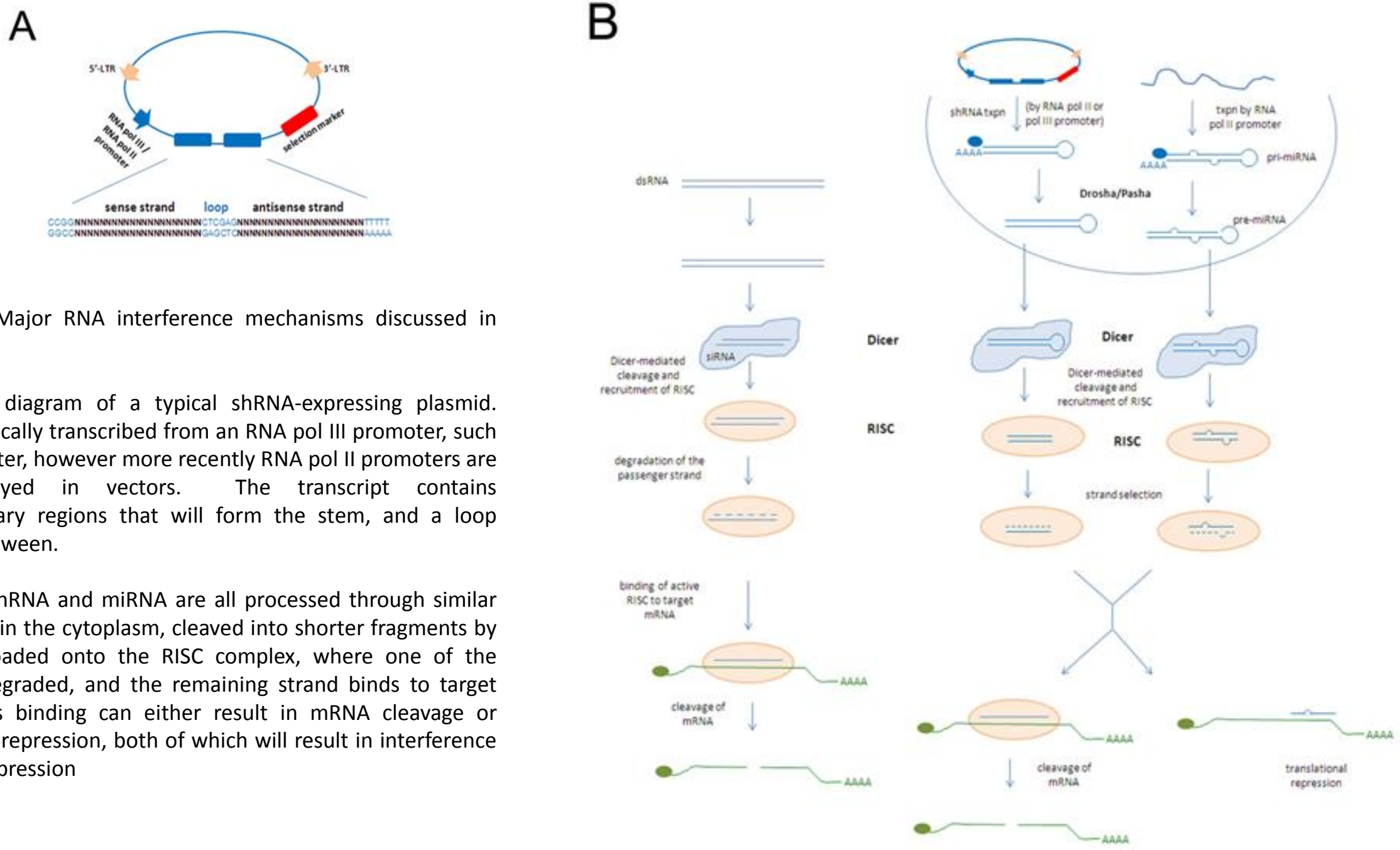


Figure 8.4. Major RNA interference mechanisms discussed in text.

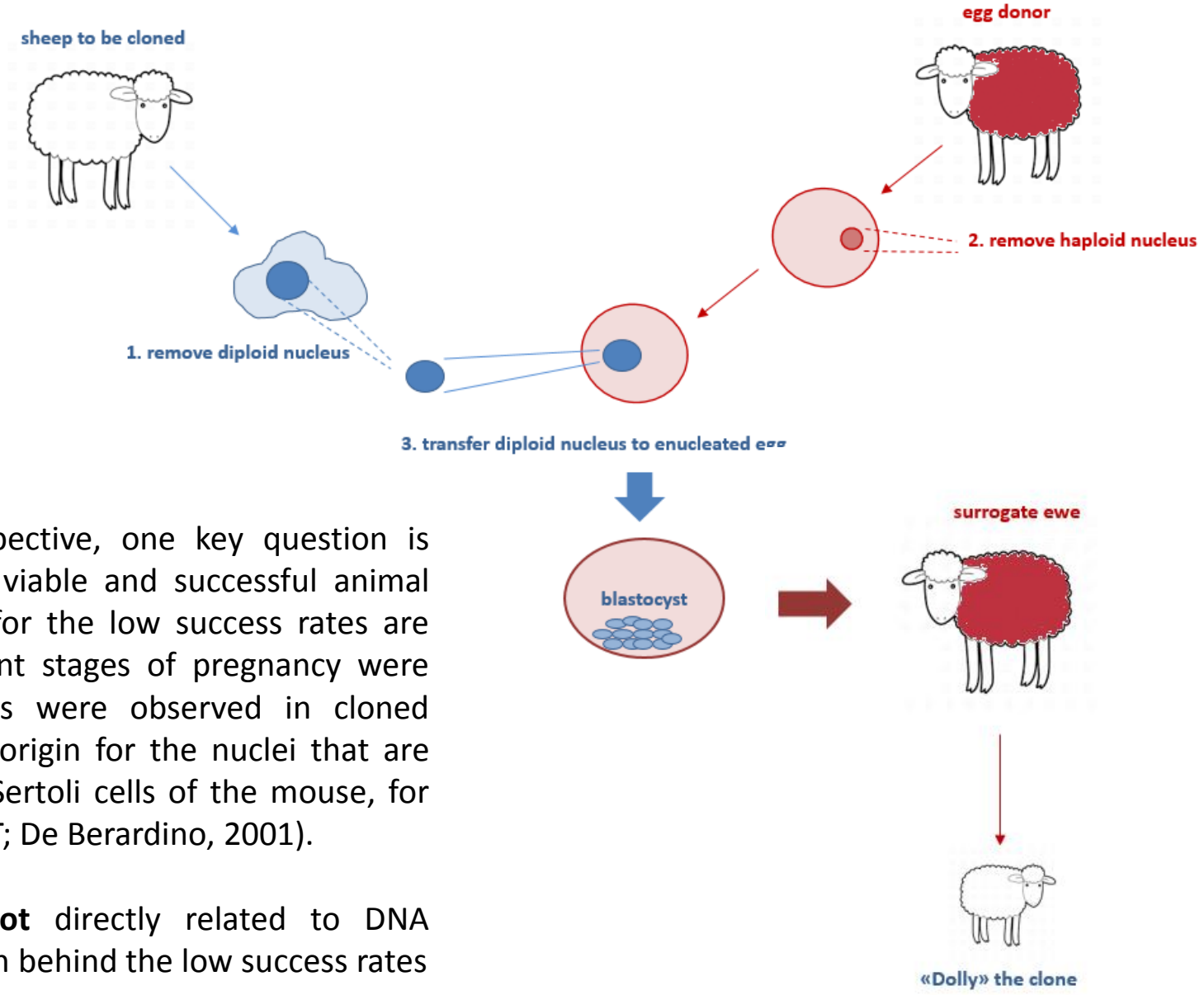
(a) Cartoon diagram of a typical shRNA-expressing plasmid. shRNA is typically transcribed from an RNA pol III promoter, such as U6 promoter, however more recently RNA pol II promoters are also employed in vectors. The transcript contains complementary regions that will form the stem, and a loop stretch in between.

(b) dsRNA, shRNA and miRNA are all processed through similar mechanisms in the cytoplasm, cleaved into shorter fragments by Dicer and loaded onto the RISC complex, where one of the strands is degraded, and the remaining strand binds to target mRNA. This binding can either result in mRNA cleavage or translational repression, both of which will result in interference with gene expression

Animal cloning

The concept of cloning was in a way discovered by Spemann back in 1938, when he used a nuclear transfer technique to split a salamander embryo into two. But cloning as we know it today was first generated for mammals in 1997, when researchers in Roslin Institute cloned Dolly, and thereafter cloning of Rhesus monkeys (1997), calves (1998), pigs (2001) and cat (2002), among others, were reported.

The use of **Somatic cell nuclear transfer (SCNT)**, is another method to generate pluripotent stem cells, and is what the Roslin scientists used to clone Dolly. Essentially, a tissue biopsy is taken from the donor, cells are grown in tissue culture and the nuclei are isolated, carrying the genetic information of the donor. These nuclei are then transferred to enucleated eggs (ie, the haploid nuclei of the eggs are removed) of the recipient surrogate mother; thus the egg is now a hybrid of the recipient egg's cytoplasm, and donor cell's nucleus with 2n chromosomes. These genetically altered eggs are then grown *in vitro*, some producing embryoid bodies, and these cloned embryos are then transferred to the surrogate mother, until babies are born.



From an ethical as well as technical perspective, one key question is whether all clones are “normal”. Rates of viable and successful animal clones are extremely low, and the reasons for the low success rates are numerous: high rates of abortion at different stages of pregnancy were observed; in addition various abnormalities were observed in cloned animals, depending mostly on the tissue of origin for the nuclei that are transferred (cumulus cells of the cattle and Sertoli cells of the mouse, for example, proved to have a better yield in SCNT; De Berardino, 2001).

Epigenetics, heritable changes that are **not** directly related to DNA sequence, is assumed to be yet another reason behind the low success rates

Pharm animals

Whether through cloning or through transgenics, commercialization and application of genetically modified animals, particularly for use in the pharmaceutical industry, has given way to the term **pharm animals**.

There may, in fact, be several different reasons for transgenic or cloned animal generation:

(a) production of superior livestock (in terms of breeding, nutritional value, taste or other),
(b) production of human therapeutic proteins or other biopharmaceuticals,
(c) use as organ source in transplantations,
(d) resurrection of beloved yet diseased or dead pet animals, or
(e) resurrection of extinct species and/or recovery of endangered species (serious ethical debates around these last two issues).

Use of transgenic or cloned animals as bioreactors for the production of biopharmaceuticals has attracted a lot of attention due to the market value of these products. Dairy industry here provides a valuable tool for large scale production and purification, hence much attention focuses on transgenic cattle for the production of such biopharmaceuticals in milk. Human lactoferrin is one such clinically valuable macromolecule, particularly in the treatment of infectious diseases. van Berkel and colleagues have created a transgenic cattle that produced human lactoferrin in milk from a bovine b-casein promoter (van Berkel et al, 2002).

Gene Therapy

Gene therapy studies in fact date back to 1990s, with the first gene therapy applied to Ashanti DeSilva in the laboratory of Dr. W. French Anderson of the National Heart, Lung, and Blood Institute.

Ashanti, then a 4-year old, had a defect in the gene coding for the enzyme adenosine deaminase (ADA), which results in immune system deficiency (Severe Combined Immunodeficiency syndrome, SCID for short; for a brief timeline of gene therapy trials, visit NIH history page: http://history.nih.gov/exhibits/genetics/sect4.htm#2).

Gene therapy strategies generally fall into 2 major categories:

- **Direct gene therapy**, where the functional genes are directly transferred to the patient (most commonly as an extra functional copy, although direct correction of the mutated gene is also under investigation; will be discussed in genome editing approaches),

- **Cell-based gene therapy**, where live cells (either patient's somatic cells from target tissue, or stem cells) are themselves used as vehicles to deliver functional genes into the patient's body (cell-and-gene therapy).

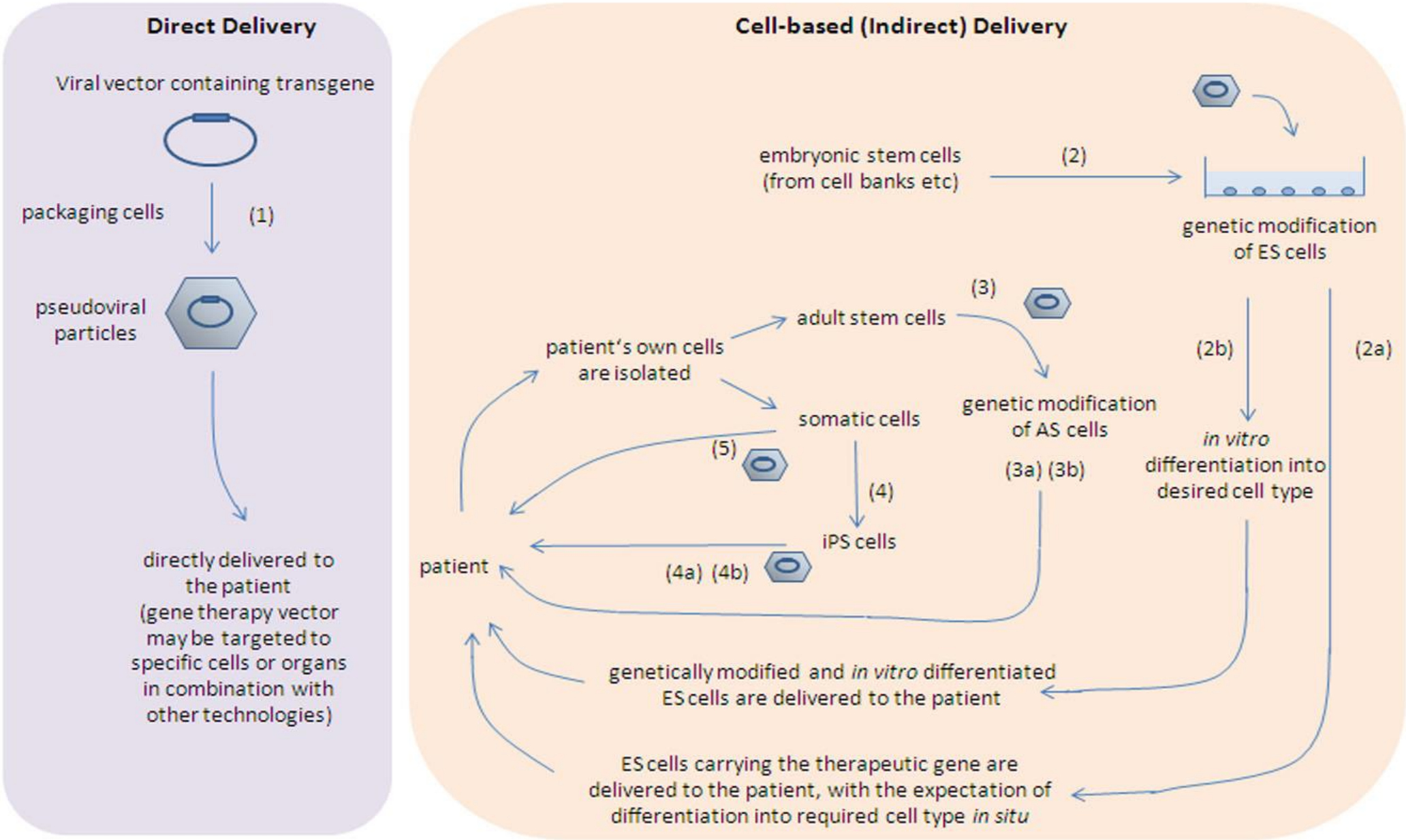


Figure 8.5. Brief outline of gene therapy strategies.

Genome Editing

Genome editing, elected the Method of the Year 2011 by *Nature Methods*, is a method whereby a piece of DNA is inserted, replaced or removed from the organism's genome, with the help of genetically engineered nucleases and the host cell's own double-stranded break repair machinery.

Homologous recombination takes place two DNA molecules that harbour significant sequence homology (see Chapter 7.c.3 for a brief overview). Double-stranded breaks stimulate homologous recombination machinery, which is the basis for genome editing approaches. The double-stranded breaks in this case are generally achieved by nucleases that were engineered to recognize specific target sequences.

There are currently four families of engineered nucleases used for genome editing purposes:

(1) Zinc finger nucleases (ZFNs),
(2) Transcription Activator-Like Effector Nucleases (TALENs),
(3) the CRISPR/Cas system, and
(4) engineered meganucleases

Zinc-finger nucleases (**ZFN**s) are based most typically on Fok I restriction enzyme that is fused to a zinc finger DNA-binding domain engineered to target specific DNA sequence.

The **TALENs** are similar to the ZFNs; each DNA-binding domain of TALENS can recognize a different single DNA base, hence a combination of different TALENs can in practice be used to target any specific sequence on the genome; the endonuclease activity again is through FokI restriction enzyme. TALENs have major advantages over ZFNs, firstly off-target mutation rates are generally lower, and they can be designed to target virtually any genomic sequence.

The **CRISPR/Cas system**, also discussed above with respect to primate genome editing, has generated a lot of excitement in that this system can achieve a relatively higher mutation rate (here is a catch: the off-target mutation rates are also higher) and is relatively easy and cheap. It consists of a target-specific "guide" RNA, and a non-target-specific nuclease. CRISPR stands for *clustered regularly interspaced short palindromic repeats*, and cas genes are *CRISPR-associated genes*.

Meganucleases of microorganisms have naturally long recognition sequences (>14bp), and with protein engineering various meganuclease variants have been generated to cover a large plethora of unique sequence combinations; in addition meganucleases have are known to cause less toxicity in cells compared to ZFNs or TALENs.

While traditional gene targeting methods in embryonic stem cells were shown to introduce genetic variations in roughly 1 in 106 cells, nuclease-mediated knock-out or knock-in trials were shown to significantly increase targeting rate to around 1 in 100 to 1 in 2 cells (http://ko.cwru.edu/services/directtargeting.shtml).

And while traditional targeting methods takes 6 to 12 months, nuclease-mediated genome editing approaches have been reported to last only several months, in some cases even 4 weeks (Yang et al, 2014), significantly reducing the time it takes to generate the transgenic cell or organism.

“Science never solves a problem without creating ten more.”

George Bernard Shaw