REVIEW ARTICLE

Homologous Recombination in Human Embryonic Stem Cells: A Tool for Advancing Cell Therapy and Understanding and Treating Human Disease

Andrew D. Leavitt, M.D.[1,2] and Isla Hamlett, Ph.D.[1]

Abstract

Human embryonic stem cells (hESCs) hold great promise for ushering in an era of novel cell therapies to treat a wide range of rare and common diseases, yet they also provide an unprecedented opportunity for basic research to yield clinical benefit. HESCs can be used to better understand human development, to model human diseases, to understand the contribution of specific mutations to the pathogenesis of disease, and to develop human cell-based screening systems to identify novel therapeutic agents and evaluate potential toxicity of therapeutic agents under development. Such basic research will benefit greatly from efficient methods to perform targeted gene modification, an area of hESC investigation that is currently in its infancy. Moreover, the reality of hESC-based cellular therapies will require improved methods for generating the specific cells of interest, and reporter cell lines generated through targeted gene modifications are expected to play an important role in developing optimal cell-specific differentiation protocols. Herein, we review the current status of homologous recombination in hESCs, a gene targeting technique that is sure to continue to improve, and to play an important role in realizing the maximal human benefit from hESCs. Clin Trans Sci 2011; Volume 4: 298–305

Keywords: human embryonic stem cells, gene targeting, homologous recombination, RMCE, recombinase, reporter cell lines

Introduction

The first reported isolation of a human embryonic stem cell (hESC) in 1998[1] triggered an explosion of research and general interest in stem cell biology within the scientific community and the public at large. Much of the excitement has focused on the potential for developing novel cellular therapies to treat a wide range of diseases of the nervous system, cardiovascular system, endocrine system, and beyond. However, hESCs also offer the unprecedented opportunity to better understand human development and to establish disease models that will inform on the biology of disease while also providing useful systems to test the efficacy and toxicity of new therapeutics under development. Gene modification technology will play an important role in establishing such *in vitro* models, as well as new systems to screen and identify novel therapeutic agents and to improve the *in vitro*, directed hESC differentiation required to realize hESC-derived cellular therapies.

It was only 30 years ago, in 1981,[2] that Gail Martin first coined the term embryonic stem cell for pluripotent cells derived directly from blastocysts. The name helped distinguish them from embryonal carcinoma cells, which are derived from teratocarcinomas and used at the time to study early embryo development.[2,3] Mouse genetics over the past 25 years has remarkably advanced our understanding of mammalian development and helped establish animal models of human diseases. Techniques that modify specific genetic loci within murine embryonic stem cells (mESCs), in particular targeted knockout and knockin mutations generated through homologous recombination (HR), have provided the engine driving these advances.[4] The broad scientific importance of these technical advances was rightfully acknowledged by awarding the 2007 Nobel Prize in Medicine to three pioneers in the field. The application of similar targeted gene modification techniques to hESCs will be essential if we are to fully realize their research and therapeutic potential.[5] This review summarizes the current state-of-the-field for hESC gene targeting through HR, providing a centralized resource for those wishing to apply this technology to their own basic or translational research.
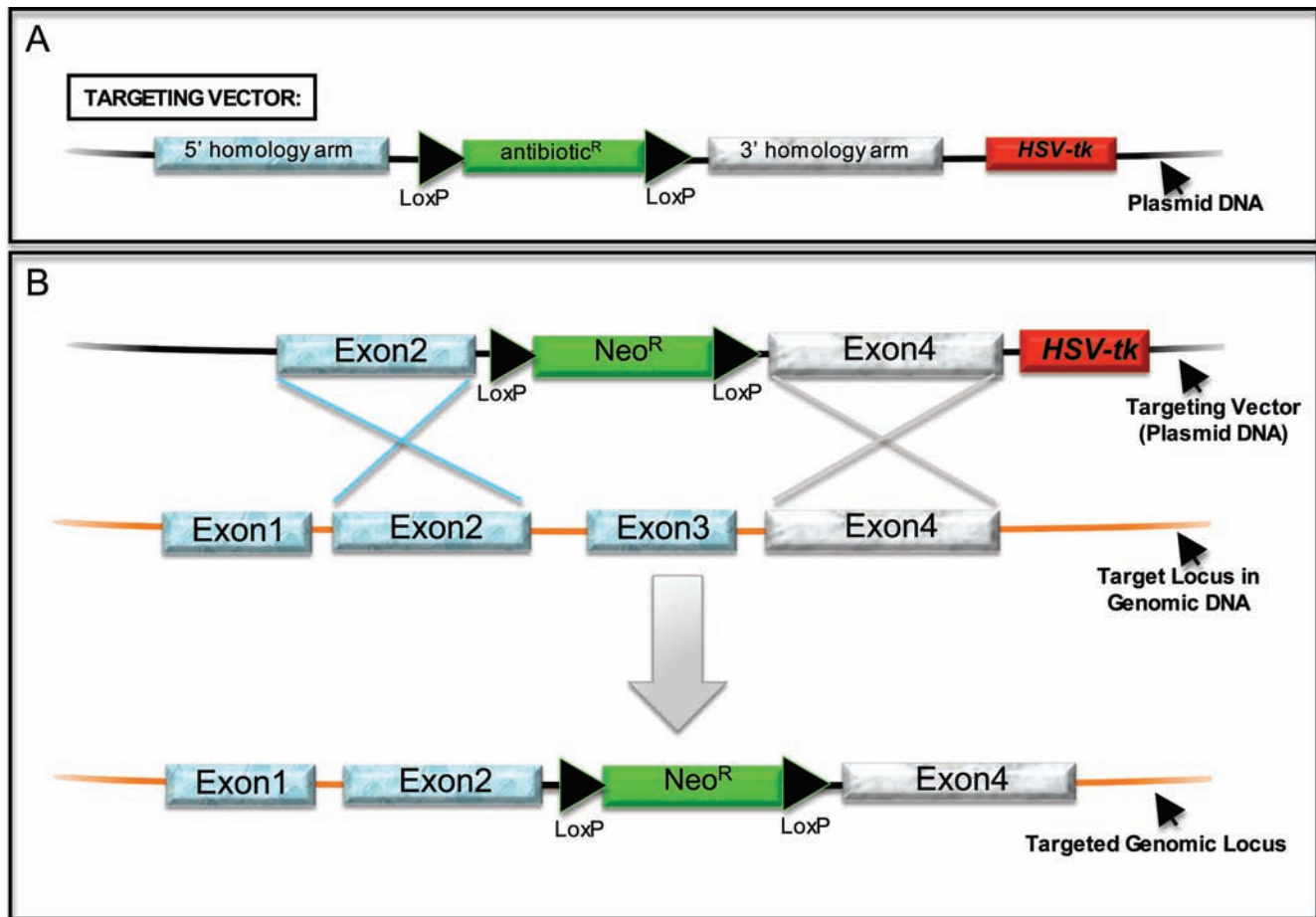
Gene Targeting and HR

Gene targeting through HR, which can be performed to achieve various goals,[5–7] relies on the cell's innate DNA repair machinery and requires homologous DNA sequences shared by the targeting vector and the genomic locus of interest (*Figure 1*). HR can be used to inactivate a gene by disrupting its coding sequence, that is, a "knockout" (*Figure 1*). Such targeted ESCs are heterozygous for the inactivated allele and can be used in blastocyst injections to generate heterozygous mice.[4] These mice can then be interbred to yield mice lacking both alleles, thereby providing powerful murine models to study a specific gene's function during development and within selected cells of interest. Obviously, hESCs will not be used in blastocyst injections, and other approaches will be needed to generate genetically null hESCs.

HR can also be used to introduce a specific mutation into a desired gene allowing one to express a mutant gene product under the control of the gene's endogenous regulatory elements (i.e., a "knockin"). This is undertaken in a manner similar to generating a knockout (*Figure 1*), but one must design the targeting vector so that the introduced cDNA is expressed from the endogenous promoter in place of the native coding sequence.[4] hESCs harboring such knockin mutations, combined with *in vitro* differentiation protocols to develop cells of interest, will provide unprecedented opportunities to define the role of specific mutations in human disease pathogenesis. They will also provide a more ideal system to test the potential efficacy and toxicity of therapeutics developed to target disease-related, pathogenic mutations.

A second type of knockin gene modification involves introducing reporter gene cDNA, typically for a fluorescent protein, into a desired genetic locus. Cells fluoresce only if they have developed into the cell type that naturally expresses the gene of interest, and such fluorescently labeled cells can be directly

Figure 1. Model for gene targeting using homologous recombination. (A) This diagram of a targeting vector highlights its basic essential components: (1) an antibiotic (positive) selectable marker flanked by *LoxP* sites and located between 5′ and 3′ homology arms, and (2) a negative selection marker located outside the two homology arms, in this example the herpes simplex virus-thymidine kinase (*HSV-tk*) gene expressed by a constitutive promoter. The example provided is for disrupting a gene of interest (a "knockout"). The antibiotic expression cassette contains its own promoter in most targeting vectors but can be promoterless if targeting a gene expressed in the undifferentiated hESC. The latter approach requires a design that ensures that the endogenous promoter will express the antibiotic resistance gene. Targeting vectors can also be designed to introduce a reporter gene that will be under the control of the endogenous regulatory elements, in which case the reporter cDNA is located between the homology arms but outside of the *LoxP*-flanked antibiotic selection cassette. The positive and negative selection cassettes can be oriented in either direction except when the positive selection cDNA is promoterless. Flp recombinase and *Frt* sites can be used similarly to Cre recombinase and *LoxP* sites. (B) This diagram outlines the process of homologous recombination (HR) using a targeting vector (top) designed to remove Exon3 of a target locus. The crossed lines indicate the crossing over event between the targeting vector and the genomic locus, which is mediated by host DNA repair machinery and requires homologous sequences. HR introduces the positive selectable (antibiotic) marker in place of Exon3. The negative selection marker, *HSV-tk* in this example, will not be incorporated into the target DNA in cells that have properly undergone HR, allowing gancyclovir resistance to be used to select against recombination events that occurred through mechanisms other than HR. The *LoxP*-flanked *Neo*[R] expression cassette can then be removed by expressing Cre recombinase, which can be introduced either by transient transfection of a Cre expression plasmid or by using Tat-Cre protein.

identified by microscopy, isolated by fluorescence-activated cell sorting (FACS), or tracked *in vivo* without requiring an antibody. The fluorescent reporter cDNA expression is controlled by the endogenous promoter and associated regulatory elements of the desired target gene, thus providing an accurate readout of endogenous gene expression. Such hESC reporter lines, with their ability to specifically mark cells of desired lineages, and by permitting rapid screening of live cells over a wide range of conditions, will provide invaluable tools for researchers wanting to optimize differentiation protocols to yield specific cell types for therapeutic purpose.[5]

The targeting vector is the critical reagent required to perform HR (*Figure 1*A). Its basic design includes (1) a positive selectable marker, typically an antibiotic resistance gene expressed by a strong constitutive promoter, to select for embryonic stem cells that have incorporated the DNA sequence of interest, (2) a negative selectable marker, herpes simplex virus thymidine kinase (*HSV-tk*) in *Figure 1*A, to select against cells that incorporated the targeting vector through non-HR at an undesired chromosomal location, and (3) 5′ and 3′ flanking sequences homologous to the desired genetic locus, which help direct recombination at the desired chromosomal location.

HR in hESCs

The first successful report of HR in hESCs appeared in 2003,[8] 5 years after hESCs were first described.[1] As with mESCs, hypoxanthine-guanine phosphoribosyl transferase (*HPRT*) was the first gene targeted using HR. To date, there have been only 12 reports of successful HR in hESCs, yet the basic and translational potential of the application of this technology is well defined.[5] Herein, we review those studies to help bring this technology to more investigators. The application of HR to generate faithful hESC reporter lines and human disease models is expected to meaningfully increase the return on our basic and applied research investment in hESCs.

A few key problems stalled the initial progress for HR in hESCs. First, HR requires single-cell cloning to identify the rare

targeted cell selected through antibiotic resistance. However, unlike mESCs, most hESC lines are intolerant of being dissociated to single cells. Consequently, the first report of HR in hESCs was performed using hESCs in small clumps.[8] Three different approaches have been used to overcome this problem. One group pioneered the use of repeated single-cell passaging in an effort to condition the hESCs. This generates extensive cell death in the initial passages, but ultimately yields cells that are more tolerant of the single-cell replating.[9,10] A second independent advance was the observation that the p160-Rho-associated coiled kinase inhibitor, Y-27632, markedly improves single-cell replating of dissociated hESCs.[11] Its use has rapidly become incorporated into many protocols that require recovery of single-cell dissociated hESCs, including after single-cell FACS sorting.[12] Third, some hESC lines, including members of the HUES series,[13,14] are more tolerant of single-cell passaging.[14–18] However, some HUES lines have known karyotypic abnormalities.[13]

Beyond the problem of single-cell replating, there were problems with hESC transfection efficiency, including the ineffectiveness of mESC electroporation protocols for introducing targeting vectors into hESCs.[8] The reports of successful HR in hESCs reviewed herein demonstrate that this problem has clearly been overcome, but not necessarily optimized. Except for a single inefficient example,[19] chemical transfection reagents have failed to effectively introduce targeting vectors into hESCs, and none of the reported HR-based gene targeting papers has used nucleofection. For the moment, electroporation is the preferred method for introducing targeting constructs into hESCs.

A third problem involved the method used to dissociate hESCs during passage. Bulk enzymatic dissociation of hESCs, including trypsin, was reported to predispose to gross chromosomal abnormalities,[20,21] but the laborious nature of physically harvesting multiple individual colonies made it difficult to expand adequate numbers of cells needed for gene targeting. Fortunately, more gentle forms of bulk colony harvesting and disruption using accutase®, collagenase, and dispase®, or short exposures to low concentrations of trypsin have greatly improved this problem. However, it remains mandatory that one routinely karyotype all cells that have been taken through single-cell replating to confirm stable and normal chromosome content.

Gene Targeting in hESC

Successfully targeted loci

Twelve reports present 15 successful examples of hESC gene targeting using HR when this manuscript was submitted.[8,9,14–19,22–25] They include a total of 11 different genetic loci, seven that are expressed in undifferentiated hESCs and four that are not (*Table 1*). Five of the 15 (33%) examples target *HPRT*, a particularly appealing target because it is expressed in hESCs, is x-linked so that targeting a single allele in male cells generates a cell functionally null for *HPRT* activity, and there is established media to rapidly select for cells that lack *HPRT* activity. Consequently, *HPRT* targeting statistics do not necessarily apply to other loci. Regardless, the initial targeting of the *HPRT* locus in 2003 was a welcome demonstration that HR can be performed in hESCs, and it provided key observations for the field. First, the authors identified electroporation settings that improved hESC transfection by >100-fold relative to conventional mESC settings: approximately 1 Neomycin (Neo)^R colony/1 × 10^7 starting hESCs versus 350 Neo^R colonies/1.5 × 10^7 starting hESCs. Second, they

demonstrated that *HSV-tk* negative selection (i.e., gancyclovir^R) reduced by approximately 85% the number of Neo^R colonies they needed to screen, more than validating the importance of incorporating negative selection into hESC targeting vector design. Third, they observed that electroporation in room temperature protein-rich media worked better than phosphate buffered saline.

HR conditions, reagents, and vector design

Table 1 summarizes many of the key features of reported successful HR in hESCs. Nine hESC lines have been successfully targeted. Beyond the suggestion that HUES-3 worked better than HUES-1 for HR at the *NANOG* locus, there is no evidence to suggest that any given hESC line is particularly inferior or superior to another. This is encouraging given that one might want to target more than one line when testing for the biology of various specific mutations within a gene, or when developing reporter lines for high-throughput screening experiments to help ensure that a result is not cell line specific. However, with more experience, certain lines may prove to be more robust. Both exons and introns have been targeted, including part of the 5′-untranslated region (*Table 1*).

The EF1α, PGK, CAG, and SV40 promoters have been used successfully to express antibiotic resistance genes within the targeting vectors (*Table 1*). Experience with mESCs shows that such strong constitutive promoters used in antibiotic selectable marker cassettes can affect local gene expression, thereby altering the biology observed. Consequently, it has become routine to have the antibiotic resistance expression cassette in the targeting vector flanked by recombination sites, most commonly *LoxP* sites, to allow one to remove the expression cassette after the targeted embryonic stem cell clone is generated (*Figure 1*). This is of particular importance when generating a reporter cell line so as to prevent the promoter within the antibiotic resistance expression cassette from interfering with expression of the reporter cDNA.[25] Both Cre and Flp recombinase have been used successfully to remove or rearrange parts of the targeting vectors following HR in hESCs (*Table 1*).

HSV-tk and diphtheria toxin A (*DTA*) negative selection have each been shown to improve targeting efficiency, more than justifying inclusion of a negative selection cassette when designing hESC targeting vectors. While various electroporation settings have been used successfully, and one report claims no significant difference when comparing 230 versus 320 V and 200 versus 500 μF,[18] the settings of 320 V and 200 μF first reported in 2003[8] have been the most common settings used to date (*Tables 2 and 3*).

Of the 11 successfully targeted loci, seven are expressed in hESCs (*Table 2*). Antibiotic resistance used to identify the rare hESC that has incorporated a targeting vector is typically achieved by including in the targeting vector an expression cassette that contains a strong promoter and the resistance gene cDNA (*Figure 1*). However, targeting vectors for loci expressed in undifferentiated hESCs can use promoterless selectable markers to enhance targeting efficiency (no. of correctly targeted clones/no. of antibiotic resistant clones). Such promoterless antibiotic resistance genes will not be expressed in most off-target recombination sites; expression only works if the vector lands in its intended, in-frame, location, or if by chance, the off-target site properly links the resistance gene cDNA to a functional promoter.

For the *HPRT* locus, a promoterless selectable marker yielded 2%–14% targeting efficiency, while a promoter-driven

Targeted locus*	Year[REF]	Expressed in hESC	Cell line (sex)	Homology arms 5′/3′ (Kb)	Targeting location	Positive selection[†]	Negative selection[‡]	Reporter cell line[§]	RMCE	Recombinase
HPRT	2003[8]	Yes	H1(XY)	1.9/10	Exons 7–9	Neo	TK and 6-TG			
	2004[19]	Yes	NR[¶](XY)	8.5 total	Exon 2	Hygro	6-TG			
	2008[22]	Yes	H1(XY)	4.8/2.1	Exon 3	EF1α-Neo	6-TG		YES	CAG-Cre
	2010[24]	Yes	KhES1(XX)	7/2	Exon 1	PGK-Neo	–		YES	EF1α-Cre
	2010[14]	Yes	HUES-3(XY) and HUES-8(XY)	118/77	Intron 6	CAG-Neo-IRES-Puro	–			
OCT4	2003[8]	Yes	H1.1(XY)	6.3/1.6 and 6.3/6.5	Exons 3–5	Neo	–	YES (EGFP)		
ROSA26	2007[23]	Yes	HES-2(XX)	1.0/4.1	Intron 1	Neo	DTA		YES	Cre
MIXL1	2008[9]	No	HES-3(XX)	9.4/1.9	Exon 1	PGK-Neo	–	YES (GFP)		Cre
FEZF2	2009[18]	No	HUES-9(XX)	4.2/4.1	Exon 2	PGK-Neo	TK	YES (EYFP)		Cre
OLIG2	2009[25]	No	BG01(XY)	2/5.3	Exon 2	PGK-Neo	TK	YES (EGFP)		Cre
ISL1	2009[16]	No	H9(XX) HUES-3(XY)	9.2/5.5	Exon 1	Puro	–	YES (dsRED)		Cre
ATM	2010[14]	Yes	HUES-9(XX) and H9(XX)	9.0/2.0 and 103/6.5	Intron 40	CAG-Neo CAG-Neo-IRES-Puro	–			
p53	2010[14]	Yes	Not stated	77/5.5	Exons 2–6	Same as ATM	–			
BAF250a	2010[15]	Yes	H9(XX) HUES-8(XY)	3.9/3.7	Intron 8	β-geo	–			CAG-Flp Cre-[ERT2]
NANOG	2010[17]	Yes	HUES-1(XX) HUES-3(XY)	12.5/3.5	Exon 1; includes 5′UTR	SV40-Neo	–	YES (EGFP)		

*HPRT = hypoxanthine polyribosyl transferase. [†]Promoters indicated if used: Neo = neomycin (G418); Hygro = hygromycin; CAG = cytomegalovirus early enhancer element and chicken β-actin promoter; β-geo = β-galactosidase-neomycin resistance fusion gene. [‡]TK = thymidine kinase; 6-TG = 6-thioguanine; DTA = diphtheria toxin A. [§]EGFP = enhanced green fluorescent protein; EYFP = enhanced yellow fluorescent protein; dsRED = red fluorescent protein. [¶]NR = not reported. RMCE = recombinase-mediated cassette exchange.

Table 1. Genetic loci successfully targeted in hESCs.

selectable marker yielded only 0.16%–1.4% (*Table 2*). For *OCT4*, a promoterless selectable marker yielded 27%–40% targeting efficiency, while the *ROSA26* targeting vector with a promoterless selectable marker averaged a 2.3% targeting efficiency (*Table 2*). Of note, the *ROSA26* efficiency was achieved with homology arms of only 1.0 and 4.1 Kb, but the vector did include *DTA* for negative selection (*Table 2*). In aggregate, these data support using a promoterless selectable marker when targeting loci expressed in undifferentiated hESCs. However, alternate approaches may be more productive depending on one's final objective.

Most reporter cell lines will require targeting loci not expressed in hESCs, but to date, successful targeting has been reported at only four such loci (*Table 3*). Inadequate data are provided to determine targeting efficiency for two of them, *MIXL1* and *ISL1*, and neither used a negative selection marker in their targeting vector. The other two both used *HSV-tk* negative selection, which increased targeting efficiency fivefold for human *FEZF2* (0.3% vs. 1.5% of Neo[R] colonies) and yielded 5.7% targeting efficiency for *OLIG2*, strongly supporting the inclusion of a negative selection cassette in a targeting vector design.

Expressing multiple gene products
It is sometimes desirable to express more than one cDNA from a targeted locus, and an internal ribosome entry sequence (IRES) has been used successfully in hESCs to express multiple

gene products from a single polycistronic message encoded by a targeting vector.[8,14] However, in each case the downstream cDNA encoded an antibiotic selection marker, allowing the authors to select for cells with higher levels of downstream gene expression. Multiple independent gene products have also been expressed from a single cistronic message in hESCs using foot and mouth virus 2A sequences that allow a single ribosome to yield multiple independent protein products.[26] There are inadequate published data to know whether an IRES or a 2A sequence is preferred when wanting to express multiple gene products from a HR targeting vector, and there may be no single correct answer, but 2A-sequences are being employed more and more by investigators, and one study suggests that they may be superior to IRES sequences in hESCs, at least in certain settings.[26]

Introns and targeting vectors
The murine ESC literature is rife with articles demonstrating that introns help expression of gene products expressed from the targeting vectors used to generate transgenic mice.[27–29] There are no published reports directly addressing the importance of introns in targeting vectors when performing HR in hESCs. However, it is important to note that the rabbit β-globin intron 2 was used successfully in the *NANOG* reporter construct that targeted exon 1, including the 5′-UTR.[17]

Targeted locus*	Year[REF]	Electroporation settings	No. of antibiotic[R] colonies/starting cell no.	No. of properly targeted colonies/total	Positive selection	Negative selection	Homology arms 5'/3' (Kb)
HPRT*	2003[8]	220V 960 μF 320V 200 μF	~1/10^7 350 Neo^R/1.5 × 10^7; 50/350 Gancyclovir^R; 7/50 6-TG^R	7/350 (2%) 7/50 (14%) 7/7 (100%)	Neo Neo Neo Neo	TK & 6-TG TK TK and 6-TG	1.9/10
	2004[19]	ExGen 500 reagent	2/1 × 10^8	2/2 (100%)	Hygro	6-TG	8.5 total
	2008[22]	300V 100μSec 320V 200μF	488 Neo^R/1 × 10^6 cells 233 Neo^R/1 × 10^6 cells 600 Neo^R/1 × 10^6 cells	1/488 (0.2%) 1/233 (0.43%) 1/600 (0.16%)	EF1α-Neo EF1α-Neo EF1α-Neo	–	4.8/2.1
	2010[24]	250V 4 ms × 2	424 Neo^R/2–4 × 10^6 cells	6/424 (1.4%)	PGK-Neo	–	7/2
	2010[14]	320V 200μF	Data not provided	Data not provided	CAG-Neo-IRES-Puro	–	118/77
OCT4	2003[8]	320V 200μF	103 Neo^R/1.5 × 10^7 cells	28/103 (27%) 22/56 (40%)	Neo	–	6.3/1.6 6.3/6.5
ROSA26	2007[23]	250V 500μF	24/5 × 10^6 cells 64/5 × 10^6 cells	1/24 (4.1%) 1/64 (1.5) Avg: 2/88 (2.3%)	Neo	DTA	1.0/4.1
ATM	2010[14]	320V 200μF	10–50 Neo^R/2 × 10^7 cells	<0.5% 3/14(21%); 10/37 (27%) 2/10 (20%) with second allele	CAG-Neo CAG-Neo-IRES-Puro	–	9.0/2.0 103/6.5
P53	2010[14]	320V 200μF	~30 Neo^R/2 × 10^7 cells (done twice; yielded 60 Neo^R colonies)	2/60 (3.3%) 7/32 (22%)	CAG-Neo & CAG-Neo-IRES-Puro	–	77/5.5 80/81
BAF250a	2010[15]	320V 200μF	22 Neo^R/1 × 10^7 cells	16/22 (73%)	β-geo	–	3.9/3.7
NANOG	2010[17]	250V 500μF or 800V 10μF	637 Neo^R/1.9 × 10^7 cells 167 Neo^R/0.9 × 10^7 cells	4/637 (0.6%) HUES-1 19/40 (11.4%) HUES-3	SV40-Neo	–	12.5/3.5

*See Table 1.

Table 2. Transfection conditions and targeting efficiency for loci expressed in undifferentiated hESCs.

Targeted locus*	Year[REF]	Electroporation settings	Antibiotic[R] colonies/starting cell no.	No. of properly targeted colonies/total	Positive selection	Negative selection	Homology arms 5'/3' (Kb)
MIXL1	2008[9]	Data not provided	Data not provided	Data not provided	PGK-Neo		9.4/1.9
FEZF2	2009[18]	320V 200 μF	128 Neo^R/1.5 × 10^7 145 Neo^R/1.05 × 10^7	2/675 (0.3%) 2/130 (1.5%)	PGK-Neo	– TK	4.2/4.1 4.2/4.1
OLIG2	2009[25]	250V 250 μF	106 Neo^R/0.5–1 × 10^7	6/106 (5.7%)	PGK-Neo	TK	2/5.3
ISL1	2009[16]	320V 200 μF	Data not provided	Data not provided	Puro		9.2/5.5

*See Table 1.

Table 3. Transfection conditions and targeting efficiency for loci not expressed in undifferentiated hESCs.

Homology arms

As the name implies, HR requires 5' and 3' DNA sequences in the targeting vector that are homologous to sequences in the target locus (*Figure 1*). Successful hESC targeting vector homology arm length has ranged from 1 to 103 Kb (*Table 1*). Murine ESC targeting data suggest that homology arms should be at least approximately 2 Kb in length, and that increasing homology arm length up to approximately 14 Kb yields progressively more effective gene targeting.[30,31] All but one of the hESC targeting vectors met the approximately 2 Kb minimum rule, but many had arms less than 14 Kb (*Table 1*).

Gene targeting at three loci specifically addressed the effect of homology arm length on HR efficiency in hESCs.[8,14] For *OCT4*, increasing the 3' homology arm from 1.6 to 6.5 Kb, while keeping the 5' homology arm fixed at 6.3 Kb, increased targeting efficiency by almost 50% (27% vs. 40%). Two targeting vectors with different homology arm lengths were also used to target the *ATM* locus. One with 5' and 3' homology arms of 9 and 2 Kb yielded correct targeting in <0.5% of the Neo^R colonies, while a vector with 5' and 3' homology arms of 103 and 6.5 Kb yielded 21%–27% properly targeted colonies.[14] Two vectors were also compared when targeting the *p53* locus, with one having

5′ and 3′ homology arms of 77 and 5.5 Kb; the other contained 80 and 81 Kb homology arms. Targeting efficiency was roughly sevenfold greater (3.3% vs. 22%) using the vector with longer arms of homology.

In all three examples, longer homology arms improved targeting efficiency, which allows one to screen fewer colonies, and can thereby translate into significantly less work. The findings also show that electroporation can readily introduce targeting vectors of well over 100 Kb into hESCs, but it remains to be seen if homology arms greater than approximately 14 Kb provide added value when targeting hESCs. However, with the ready availability of human bacterial artificial chromosomes (BACs) and the power of recombineering technology to generate targeting vectors,[32–34] it has become much simpler to incorporate long 5′ and 3′ homology arms. In fact, BAC recombineering was used to make three of the targeting vectors in *Table 1*.[14,17] While optimal homology arm length has not been defined, the summary of arm lengths and targeting success presented in *Tables 2 and 3* should provide a useful guide for designing a targeting vector.

Generating reporter hESC lines

Reporter cell lines generated through HR express a readily detectable marker, typically a fluorescent protein, under the control of a selected endogenous promoter and associated regulatory sequences. Reporter lines can facilitate identification of specific subsets of cells in a complex cell mixture. Six hESC reporter lines have so far been generated through HR (*Table 1*). In two, reporter expression reflects pluripotency by having targeted the *NANOG* locus[17] or the *OCT4* locus,[8] providing cell lines that can be used to screen for molecules that release cells from pluripotency. In four lines, reporter expression turns on when specific lineage-related genes are expressed, so these lines can be used to monitor the development of a specific cell type in culture based on the targeted locus, *MIXL1*[9] (mesoderm; primitive hematopoietic precursors), human *FEZF2*[18] (corticospinal neurons), *OLIG2*[25] (neuroglial cells), and *ISL1*[16] (cardiac cells) (*Table 1*). Like their murine counterparts,[35] such reporter lines can be used to identify specific cell types, making them powerful tools for developing novel *in vitro* conditions to optimize production of specific human cell types, a key step toward establishing hESC-derived cellular therapies.[5]

Disease and development models through HR

HR in hESCs allows one to introduce a specific mutation into a designated endogenous locus to develop models of human disease. The same cell lines can also be used to test the efficacy and toxicity of potential new therapeutic agents. However, unlike the mouse system, independent mutations introduced into hESCs cannot be combined through the generation of mice and interbreeding. Recombinase-mediated cassette exchange (RMCE) in hESCs[36] provides a powerful approach for using a single targeted cell line to evaluate the biology of multiple mutations at a given locus by simply exchanging one mutation for another (*Figure 2*).

Cre- (or Flp-) recombinase-based RMCE relies on heterotypic *Lox* (or *Frt*) sites that inefficiently react with each other due to differences in their sequences. When a chromosomal cassette is flanked by two such mutually incompatible *Lox* (or *Frt*) sites, it can readily be exchanged for another cassette (located on a plasmid) by a double reciprocal recombination[37, 38] (*Figure 2*).

RMCE has been used successfully in hESCs in combination with HR for different purposes in three studies to date (*Table 1*), including delivery of the necessary genetic machinery to create doxycycline-regulated conditional gene expression in hESCs.[24] Others have used RMCE to replace an initial targeting construct and selectable cassette with a reporter transgene,[22] while others have demonstrated the versatility of RMCE to readily introduce different reporter transgenes into a targeted locus in hESCs.[23] Additional versatility can be achieved by combining both Cre- and Flp-mediated recombination sites into a single targeting event.[15] The initial use of RMCE in hESCs has been more related to proof-of-principle than to addressing a specific biological question, but we expect this to change in the very near future. We also envision in the near future the use of RMCE to create expression systems that allow for cell tracking in *in vitro* developmental culture systems, an advance that will open hESCs to the study of human development much like lineage tracking has done for advancing our understanding of mouse development.

Targeted mESCs are often used to generate mouse lines that can be crossed to generate mice with cells deleted for both alleles of a given locus. Such genetically null mice have proven quite informative for understanding the role of specific gene products during development and within specific cellular functions. Generating genetically null hESC lines, from which to derive genetically null cells of interest, will require one to independently target both loci within the same hESC. Two groups have achieved this goal using innovative but different approaches to inactivate both alleles of a single gene in a single hESC line, including *p53*,[14] *ATM*,[14] and *BAF250a*.[15] In each case, essentially proof-of-principle experiments, the authors were able to demonstrate an expected functional defect associated with the loss of gene activity. To date, this has only been reported for a locus that is expressed in hESCs, not for one that is silent in hESCs and that initiates expression later during hESC *in vitro* differentiation. However, regardless of when the gene is expressed, any attempt to sequentially target both alleles in the same cell must use an approach that retains the initial targeting event when performing selection during the second round of HR. The *BAF250a*[−/−] hESC[15] line incorporates a 4-OH tamoxifen-regulatable Cre that allows for the added utility of a conditional null system. These and future null cell lines will provide new and potentially powerful systems to understand the role of critical gene products in human biology and disease.

Summary

HR is the gold standard for site-specific genetic modification in mESCs,[7] but it is still in its infancy when being applied to hESCs. However, the successes to date (*Table 1*) suggest that its application to hESCs will soon expand considerably as investigators apply this technology to develop improved *in vitro* systems for deriving cells for therapeutic use, to model human disease in human cells, to screen for new therapeutic agents, and to evaluate the toxicity of therapeutic agents under development.[5] It is hoped that this review will help bring HR technology in hESCs to more investigators. At this time, there is not a clear "best" targeting vector design, and different designs may be optimal for vectors targeting genes expressed in undifferentiated hESCs compared with genes not expressed in undifferentiated hESCs. However, the details provided herein should help orient investigators as to where to start and what hurdles they face, and the tables provide direction regarding electroporation conditions, useful promoters for antibiotic positive selection cassettes, and the range of successful
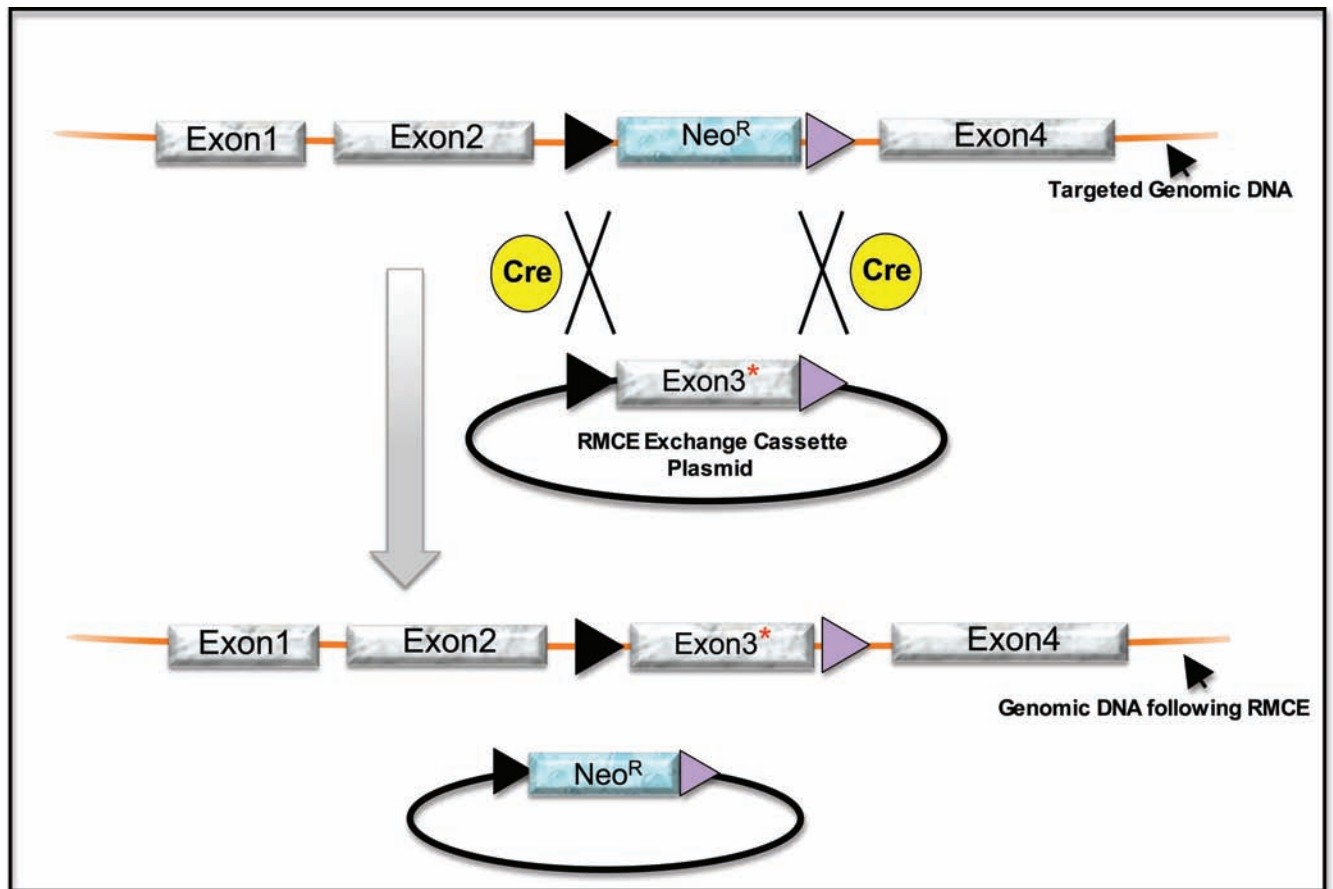
Figure 2. Model for recombinase-mediated cassette exchange (RMCE). The upper linear figure depicts a site in a genomic locus that has previously undergone HR to replace Exon3 with a *Neo*[R] expression cassette flanked by two heterotypic *LoxP* sites (one a black and one a purple triangle) that cannot recombine with each other. One can then introduce a donor plasmid carrying the cDNA of interest flanked by the same heterotypic *LoxP* sites (the RMCE exchange cassette plasmid) into the targeted hESCs. Cre recombinase (yellow circles) then catalyzes an exchange between the RMCE exchange cassette plasmid and the targeted locus (black "X"s), removing the antibiotic selection marker and introducing the sequence flanked by the heterotypic *LoxP* sites into the targeted genomic DNA. Cre can be introduced by using an expression vector or by using Tat-Cre. In the example shown, a specific point mutation in Exon3 (indicated by the red asterisk) has been introduced into the cell to produce a model for a disease-initiating mutation. RMCE allows one to introduce various sequences between the heterologous *LoxP* sites. For example, one can introduce mutations, correct mutations, or, as has been demonstrated, one can introduce all of the components necessary to establish a conditional gene expression system,[15] giving tremendous research potential to a single targeted cell line.

homology arm lengths that have been reported to date. The more recent examples of HR in hESCs have incorporated increasingly sophisticated designs, including RMCE and conditionally activated expression cassettes. Further advances in the design and implementation, including tissue-specific conditional gene expression and lineage tracing, will further exploit the benefit of HR technology, and we anticipate that a sharp increase in its use in hESCs is just around the corner.

Acknowledgment

References

1. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergie JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science.* 1998; 282(5391): 1145–1147.

2. Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A.* 1981; 78(12): 7634–7638.

3. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. *Nature.* 1981; 292(5819): 154–156.

4. Hall B, Limaye A, Kulkarni AB. Overview: generation of gene knockout mice. *Curr Protoc Cell Biol.* 2009; Chapter 19: Unit 19:12, 11–17.

5. Giudice A, Trounson A. Genetic modification of human embryonic stem cells for derivation of target cells. *Cell Stem Cell.* 2008; 2(5): 422–433.

6. Vasquez KM, Marburger K, Intody Z, Wilson JH. Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci U S A.* 2001; 98(15): 8403–8410.

7. Mortensen R. Overview of gene targeting by homologous recombination. *Curr Protoc Mol Biol.* 2006; Chapter 23: Unit 23.1, pp. 1–12.

8. Zwaka TP, Thomson JA. Homologous recombination in human embryonic stem cells. *Nat Biotechnol.* 2003; 21(3): 319–321.

9. Davis RP, Ng ES, Costa M, Mossman AK, Sourris K, Elefanty AG, Stanley EG. Targeting a GFP reporter gene to the MIXL1 locus of human embryonic stem cells identifies human primitive streak-like cells and enables isolation of primitive hematopoietic precursors. *Blood.* 2008; 111(4): 1876–1884.

10. Costa M, Dottori M, Sourris K, et al. A method for genetic modification of human embryonic stem cells using electroporation. *Nat Protoc.* 2007; 2(4): 792–796.

11. Watanabe K, Ueno M, Kamiya D, et al. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat Biotechnol.* 2007; 25(6): 681–686.

12. Emre N, Vidal JG, Elia J, et al. The ROCK inhibitor Y-27632 improves recovery of human embryonic stem cells after fluorescence-activated cell sorting with multiple cell surface markers. *PloS One.* 2010; 5(8): e12148.

13. Cowan CA, Klimanskaya I, McMahon J, Atienza J, Witmyer J, Zucker JP, Wang S, Morton CC, McMahon AP, Powers D, et al. Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med.* 2004; 350(13): 1353–1356.

14. Song H, Chung SK, Xu Y. Modeling disease in human ESCs using an efficient BAC-based homologous recombination system. *Cell Stem Cell.* 2010; 6(1): 80–89.

15. Bu L, Gao X, Jiang X, Chien KR, Wang Z. Targeted conditional gene knockout in human embryonic stem cells. *Cell Res.* 2010; 20(3): 379–382.

16. Bu L, Jiang X, Martin-Puig S, Caron L, Zhu S, Shao Y, Roberts DJ, Huang PL, Domian IJ, Chien KR. Human ISL1 heart progenitors generate diverse multipotent cardiovascular cell lineages. *Nature.* 2009; 460(7251): 113–117.

17. Fischer Y, Ganic E, Ameri J, Xian X, Johannesson M, Semb H. NANOG reporter cell lines generated by gene targeting in human embryonic stem cells. *PloS one.* 2010; 5(9): e12533.

18. Ruby KM, Zheng B. Gene targeting in a HUES line of human embryonic stem cells via electroporation. *Stem Cells.* 2009; 27(7): 1496–1506.

19. Urbach A, Schuldiner M, Benvenisty N. Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. *Stem Cells.* 2004; 22(4): 635–641.

20. Draper JS, Smith K, Gokhale P, Moore HD, Maltby E, Johnson J, Meisner L, Zwaka TP, Thomson JA, Andrews PW. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol.* 2004; 22(1): 53–54.

21. Mitalipova MM, Rao RR, Hoyer DM, Johnson JA, Meisner LF, Jones KL, Dalton S, Stice SL. Preserving the genetic integrity of human embryonic stem cells. *Nat Biotechnol.* 2005; 23(1): 19–20.

22. Di Domenico AI, Christodoulou I, Pells SC, McWhir J, Thomson AJ. Sequential genetic modification of the hprt locus in human ESCs combining gene targeting and recombinase-mediated cassette exchange. *Cloning Stem Cells.* 2008; 10(2): 217–230.

23. Irion S, Luche H, Gadue P, Fehling HJ, Kennedy M, Keller G. Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol.* 2007; 25(12): 1477–1482.

24. Sakurai K, Shimoji M, Tahimic CG, Aiba K, Kawase E, Hasegawa K, Amagai Y, Suemori H, Nakatsuji N. Efficient integration of transgenes into a defined locus in human embryonic stem cells. *Nucleic Acids Res.* 2010; 38(7): e96.

25. Xue H, Wu S, Papadeas ST, Sputsa S, Swistowska AM, MacArthur CC, Mattson MP, Maragakis NJ, Capecchi MR, Rao MS, et al. A targeted neuroglial reporter line generated by homologous recombination in human embryonic stem cells. *Stem Cells.* 2009; 27(8): 1836–1846.

26. Hasegawa K, Cowan AB, Nakatsuji N, Suemori H. Efficient multicistronic expression of a transgene in human embryonic stem cells. *Stem Cells.* 2007; 25(7): 1707–1712.

27. Brinster RL, Allen JM, Behringer RR, Gelinas RE, Palmiter RD. Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A.* 1988; 85(3): 836–840.

28. Choi T, Huang M, Gorman C, Jaenisch R. A generic intron increases gene expression in transgenic mice. *Mol Biol Biol.* 1991; 11(6): 3070–3074.

29. Nott A, Meislin SH, Moore MJ. A quantitative analysis of intron effects on mammalian gene expression. *RNA.* 2003; 9(5): 607–617.

30. Deng C, Capecchi MR. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol.* 1992; 12(8): 3365–3371.

31. Hasty P, Rivera-Perez J, Bradley A. The length of homology required for gene targeting in embryonic stem cells. *Mol Cell Biol.* 1991; 11(11): 5586–5591.

32. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev.* 2001; 2(10): 769–779.

33. Malureanu LA. Targeting vector construction through recombineering. *Methods Mol Biol.* 2010; 693: 181–203.

34. Ohtsuka M, Kimura M, Tanaka M, Inoko H. Recombinant DNA technologies for construction of precisely designed transgene constructs. *Curr Pharm Biotechnol.* 2009; 10(2): 244–251.

35. Borowiak M, Maehr R, Chen S, Chen AE, Tang W, Fox JL, Schreiber SL, Melton DA. Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell.* 2009; 4(4): 348–358.

36. Du ZW, Hu BY, Ayala M, Sauer B, Zhang SC. Cre recombination-mediated cassette exchange for building versatile transgenic human embryonic stem cells lines. *Stem Cells.* 2009; 27(5): 1032–1041.

37. Feng YQ, Seibler J, Alami R, Eisen A, Westerman KA, Leboulch P, Fiering S, Bouhassira EE. Site-specific chromosomal integration in mammalian cells: highly efficient CRE recombinase-mediated cassette exchange. *J Mol Biol.* 1999; 292(4): 779–785.

38. Seibler J, Schubeler D, Fiering S, Groudine M, Bode J. DNA cassette exchange in ES cells mediated by Flp recombinase: an efficient strategy for repeated modification of tagged loci by marker-free constructs. *Biochemistry.* 1998; 37(18): 6229–6234.