Unexpected Embryonic Stem (ES) Cell Mutations Represent a Concern in Gene Targeting: Lessons From “Fierce” Mice

Ravinesh A. Kumar,1,2 Ka Ling Chan,2 Ambrose H.W. Wong,2 Ken Q. Little,2 Evica Rajcan-Separovic,3 Brett S. Abrahams,2,4 and Elizabeth M. Simpson2*

1Doctorate Program in Medical Genetics, University of British Columbia, Vancouver, British Columbia, V5Z 4H4 Canada
2Centre for Molecular Medicine and Therapeutics, British Columbia Research Institute for Children’s & Women’s Health, and Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, V5Z 4H4 Canada
3Department of Pathology, Cytogenetics, University of British Columbia, Centre for Children’s and Women’s Health, Vancouver, British Columbia, V6H 3V4 Canada
4Graduate Program in Neuroscience, University of British Columbia, Vancouver, British Columbia, V5Z 4H4 Canada

Received 17 July 2003; Accepted 16 October 2003

Summary: The exceptional value of gene targeting technology to generate mouse models of human disease exists under the shadow of potential genetic errors. We previously observed an unexpected brain-behavior phenotype that resulted from a gene-targeting experiment designed to delete the Zfa gene. Given that the transcription of Zfa is restricted to the germ cell lineage of adult testis, it was both a surprise and a concern when the resulting mice had a phenotype present in both sexes that included abnormal brains and violent behavior. We hypothesized that an unrelated mutation may have been responsible for the unexpected phenotype. Here we show that the single gene mutation, Nr2e1frc (fierce), which was responsible for the brain-behavior phenotype, existed in the embryonic stem (ES) cell even before the derivation of the Zfa knockout mice. Our work thus highlights a concern in gene targeting, namely, that ES cells can harbor unexpected mutations, which can lead to genotype-phenotype misattribution. Based on our findings, we caution the gene-targeting community to use low-passage ES cells, to characterize mice derived from more than one independently targeted ES cell clone, and to backcross mice to allow for segregation of distant but linked mutations. genesis 38:51–57, 2004. © 2004 Wiley-Liss, Inc.

Key words: homologous recombination; tailless; mouse models; E14TG2a; violence; aggression; brain development; nuclear receptor; Nr2e1

The need for caution when interpreting the results of gene-targeting experiments was underscored nearly a decade ago by one of the leading toolmakers (Smithies, 2001) of homologous recombinant technology (Shehee et al., 1993). Indeed, the occurrence of errors near sites of homologous recombination has been well documented (Doetschman et al., 1988; Hasty et al., 1991; Schwartzberg et al., 1990; Shehee et al., 1993; Thomas and Capecchi, 1990; Zheng et al., 1991). This is because standard molecular biology screening protocols (i.e., polymerase chain reaction (PCR) and Southern analysis) can typically detect anomalous homologous or illegitimate recombination events near the intended target site (Matise et al., 2000). More elusive are linked “distant” mutations in embryonic stem (ES) cells that may interfere with genotype-phenotype interpretation, and for which no means of standard molecular detection is available. Evidence for these mutations is indirect and anecdotal: for example, the presence of unexpected ES cell mutations has been used to explain the occurrence of early generation phenotypes in knockout mice that disappear after backcrossing (Moulson et al., 2003) (L.P. Kozak and B.H. Koller, pers. commun.). These events are too rapid to attribute to strain-specific modifier loci, but are attributed to unlinked or distant ES cell mutations. Here we demonstrate for the first time the validity of such speculations and make recommendations to avoid this serious problem.
The initial characterization of the Zfa knockout mice revealed an unexpected brain-behavior phenotype. Given that the transcription of Zfa is restricted to the germ cell lineage of adult testis (Ashworth et al., 1990), its targeted mutagenesis was hypothesized to result in failure of spermatogenesis. Thus, it was both a surprise and a concern when the resulting mice had a phenotype present in both sexes that included abnormal brains and violent behavior. The relatively low frequency of correct targeting at the Zfa locus raised the possibility that a repeat within the vector may have produced an anomalous integration event near Zfa (Banks et al., 2003). However, we had confirmed correct 5' and 3' targeting by Southern analyses in the ES cell clone that gave rise to the Zfa colony (Banks et al., 2003). Further Southern analyses using neomycin and thymidine kinase vector-specific probes did not identify unexpected bands in this clone (data not shown). Taken together, there was no evidence to support our initial hypothesis that misincorporation of the targeting vector into a locus “near” the Zfa gene was responsible for the phenotype observed in affected mice.

We then hypothesized that an unrelated “distant” mutation may have been responsible for the abnormal brain and violent behavior observed in these “fierce” mice, which we later demonstrated are deleted for, and fail to transcribe, the nuclear hormone receptor Nr2e1 (Young et al., 2002). Evidence that led us to suspect two segregating loci included a single mouse that was wildtype for Zfa but clearly displayed “hard-to-handle” and violent behaviors. However, the precise molecular nature of the spontaneous mutation remained unresolved. Specifically, the size, complexity, and location of breakpoints were unknown. Further, it was unclear whether more than one locus contributed to the phenotype. Importantly, its spontaneous occurrence during the gene targeting of the unrelated Zfa gene (Banks et al., 2003) left its origin uncertain. In particular, it was unclear whether the unexpected mutation occurred in an ES cell, or later in a mouse during the derivation of the Zfa colony. Indeed, the utility and reliability of “fierce” mice as a model for investigating the genetic basis of abnormal brain development and pathological violence was contingent on a thorough investigation of these questions.

We characterized the molecular nature of the Nr2e1frc allele using a BAC, bEMS4, which spans the wildtype Nr2e1 locus (Abrahams et al., 2002). We demonstrated that the Nr2e1frc deletion was contained within the region defined by bEMS4 using PCR assays specific to the BAC ends (data not shown). To further define the deletion, we designed a series of PCR assays from bEMS4 sequence and amplified each on wildtype and “fierce” (Nr2e1frc/frc) genomic DNA, which identified an ~40-kb deletion and the approximate breakpoints. Comparison of wildtype sequence with genomic DNA sequence from Nr2e1frc/frc mice further defined the deletion as 44.4 kb, which included all Nr2e1 coding and untranslated regions but not exons of adjacent genes (Fig. 1a). Specifically, the 5' breakpoint resided 1.6 kb upstream of exon 1 and the 3' breakpoint 23.5 kb downstream of the terminal exon. Interestingly, an unknown 188-bp insert interrupted the deletion breakpoints.

To determine the identity of the 188-bp sequence, we first performed a masked query against the NCBI nonredundant (nr) database (BLASTN 2.2.6; April 09 2003), which did not recover any significant hits. Importantly, this demonstrated that the 188-bp insert does not represent vector material. We next searched the C57BL/6J Mouse Ensembl database (build 23), which localized the 188-bp sequence to intron 2 of Lace1 (ENS-MUSG0000038302; E-value 10^-12). The distance separating the 188-bp insert at Nr2e1frc from its cognate (i.e., endogenous) site in Lace1 was calculated as 139 kb using UCSC Genome Browser (http://genome.ucsc.edu/, mouse assembly Feb. 2003). These results supported our earlier radiation hybrid mapping data that placed the cognate 188 bp on Chromosome 10, proximal to Nr2e1 (data not shown).

To test the hypothesis that Lace1 may have also been mutated by the event that created the Nr2e1frc deletion-insertion, we performed direct genomic DNA sequencing across the 188-bp region of Lace1 in Nr2e1frc/frc mice. Note that selection during backcrossing for ES cell-derived Nr2e1frc would have maintained the tightly linked original Lace1 (Fig. 1b). We established that the 188 bp is intact at its cognate site, which indicates that a transposition event (duplication-insertion) was involved in the insertion of the 188 bp in the Nr2e1frc allele. However, the new copy is not exact: we detected a 3-bp mismatch between the 188 bp at Nr2e1frc and the wildtype sequence (Fig. 1b). We initially suspected that the 3-bp difference might reflect interstrain polymorphisms between the C57BL/6J and 129 strains, which were the two strains involved in the derivation of the Nr2e1frc mice. However, we recovered and sequenced a portion of a BAC, bEMS40, containing wildtype 129 Lace1, and did not detect any differences between the C57BL/6J, 129, and Nr2e1frc/frc cognate 188-bp sequences. Thus, we conclude that the 3-bp difference in the Nr2e1frc mutation arose due to errors during the transposition event. We also aligned and examined all breakpoint junctions for evidence of repetitive elements, microdomains, or homology tracts that may suggest recombinatorial processes leading to the deletion-insertion (Roth and Wilson, 1986; Zhu et al., 2002); however, we detected none (data not shown).

We also demonstrated the absence of the 44.4-kb material in Nr2e1frc/frc mice by performing FISH using probes from within and throughout the deletion (Fig. 2). We conclude that the 44.4-kb sequence is absent in Nr2e1frc/frc mice and find no evidence that this material is present elsewhere in the genome (Fig. 2).

Finally, to further support the conclusion that the abnormal brain development and violent behavior observed in Nr2e1frc/frc mice is attributable solely to the loss of Nr2e1, we demonstrated the integrity of the neighboring genes. This is particularly important given that 1) the Lace1 sequence participated in the event that
created the Nr2e1frc deletion-insertion, and 2) a position effect caused by the 44.4-kb deletion at Nr2e1frc may affect the expression of nearby genes. We therefore performed Northern analyses and demonstrated that the neighboring genes, Lace1 and Snx3, are transcribed normally in Nr2e1frc/frc mice (Fig. 3). Thus, we conclude that the brain and behavior phenotype seen in the Nr2e1frc/frc mice results only from the single-gene disruption of Nr2e1.

We next sought to determine when the Nr2e1frc mutation arose, namely, whether the mutation was present at the ES cell stage or if it arose during the derivation of the knockout colony. We therefore genotyped the parental E14TG2a ES cell line and three targeted clones using Nr2e1frc-specific PCR and Southern assays (Fig. 4a–d). We did not detect the Nr2e1frc mutation in the parental ES cells. Importantly, however, we did demonstrate that the mutation existed in the one targeted clone (mEMS4) that gave rise to the mutant mice. We considered the possibility that our Nr2e1frc assays were not sensitive enough to detect the Nr2e1frc mutation if it was present in the E14TG2a ES cells at a low frequency. Thus, we repeated our assays on Nr2e1frc/frc DNA that was serially diluted with wildtype DNA to determine the sensitivity level for detecting Nr2e1frc in the E14TG2a cells. We conclude that, if Nr2e1frc existed in E14TG2a cells, it was present in less than 1/10^5 cells (Fig. 4e,f). Thus, Nr2e1frc likely arose during electroporation but could have been present at a low frequency in the parental ES cell strain.
The delineation of the origin of the Nr2e1frcc allele in ES cells highlights an important concern in gene targeting, namely, that ES cells harbor distant and unexpected mutations. Such distant mutations can result in the misattribution of genotype to phenotype, such as in our case, where we initially attributed the “fierce” phenotype to the targeted Zfa allele. Remarkably, had it not been for a discrepancy between our expected phenotype and genotype, and the large numbers of mice generated and examined at each generation, we might have lost the Nr2e1frcc allele and the “fierce” mice. Our present understanding of the origin and history of Nr2e1frcc (Fig. 5) begets three recommendations to circumvent this potential concern.

Although our data does not prove that Nr2e1frcc preexisted in the E14TG2a cells, our work raises awareness of this possibility (Fig. 5a). Given that chromosomal abnormalities occur frequently in ES cells (Liu et al., 1997; Longo et al., 1997), and that these aberrations increase with extended culture (Cervantes et al., 2002; Nagy, 2000), ES cells may harbor point mutations, microdeletions, and other subchromosomal mutations that increase at each passage. Indeed, our E14TG2a cells were first described in 1987 and now are a high-passage ES cell stock (Hooper et al., 1987). Interestingly, the lethal mutation accompanying Ucp1 targeting (Enerback et al., 1997) came from the same ES cell stock that gave rise to Nr2e1frcc. Thus, at least two independent unexpected mutations have occurred with these ES cells. We therefore suggest that there may be an advantage to using fresh or low-passage ES cells.

We show that the correctly targeted ZfaKO allele and Nr2e1frcc were both on Chromosome 10 in the mEMS4 ES cell clone (Fig. 4). We deduce from the segregation data (see below) that they were in cis on this chromosome (Fig. 5a) and were similarly positioned in the chimera and initial subsequent generations (Fig. 5b). We further deduce that had we generated mice from multiple correctly targeted ES cell clones they would not all have carried Nr2e1frcc and would have provided an early indication of a problem (Fig. 4). Thus, we strongly recommend the phenotypic characterization of mice derived from more than one independent ES cell clone, especially if the phenotype is inconsistent with the expression pattern or otherwise unanticipated.
Backcrossing, initially performed to establish a pure genetic background (Bucan and Abel, 2002; Gerlai, 1996; Simpson et al., 1997), ultimately led to the derivation of two distinct mutant mouse strains, \( \text{Nr2e1}^{frc/frc} \) and \( \text{Zfa}^{K0/K0} \) (Fig. 5c). However, the generation (N) at which the \( \text{Nr2e1}^{frc} \) mutation segregated from the \( \text{Zfa} \) knockout allele remained unknown. An extensive retrospective analysis of mouse breeding data including, where available, genotype and phenotype for all mice used, beginning with the initial chimera crosses, established that the two loci (10.1 Mb apart) segregated by at least N6 (C57BL/6J) and N5 (129) (Fig. 5b). We therefore recommend the use of backcrossing to "clean out" unexpected "distant" ES cell mutations. Inbreeding early, regardless of genetic background, will increase the probability of maintaining an unwanted mutation.

The molecular genetic characterization reported here demonstrates that \( \text{Nr2e1}^{frc} \) is a deletion-insertion that only affects a single gene. Further, our work establishes that the "fierce" phenotype can be attributable entirely to this unexpected mutation. Critically, we report that the deletion of an entirely unrelated gene in an ES cell may go undetected during gene targeting despite the use of rigorous molecular screening protocols. The serendipitous discovery of the "fierce" mice emphasizes that the integrity of ES cells cannot be taken for granted. In light of the recent first-draft publication of the mouse genome and subsequent identification of over 5,600 new predicted transcripts (Waterston et al., 2002), the genetics community can reasonably anticipate a systematic and global increase in the generation of knockout mouse models of human health and disease. Consequently, the need for caution in gene targeting has never been greater.

**MATERIALS AND METHODS**

**Mice, ES Cells, and Genomic DNA**

The derivation of the \( \text{Zfa}^{K0/K0} \) strain has been described previously (Banks et al., 2003). Approval for the use of all mice was obtained through the Office of Research Services and Administration, The University of British Columbia, and the Department of Medical Genetics (protocol numbers A99-0217 and A99-0275). DNA from 129 E14TG2a ES cells (Hooper et al., 1987) and targeted clones mEMS2, mEMS3, and mEMS4 was harvested as previously described (Laird et al., 1991).

**Nr2e1\(^{frc} \)** Microdeletion Characterization, Sequencing, and Analyses

We sequenced directly from genomic \( \text{Nr2e1}^{frc/frc} \) DNA using APA technology provided by BIO S&T (Montreal, PQ). The following primers were employed: oEMS2048 (5' - AAAAGCTTTGAGTGAGAAGGCTC - 3'), oEMS2049 (5' - GAGCTCTGCAGGAAAAGGAAAC - 3'), oEMS2050 (5' - CCTACGTCCAGGCAGCTCCAC - 3'), oEMS2051 (5' - ACCTCGGTTGCTCAGTAC - 3'), oEMS2052 (5' - AGAGTATTCTAGGAACACGGGAG - 3'), oEMS2053 (5' - CCGAGTTCTGAGCTGGGATCT - 3'), oEMS2054 (5' - GGGTTCTTCTTTTTTTTTCCCAGGTTCT - 3'), oEMS2055 (5' - CACAGTTCTTCTAAGGTTCT - 3'), oEMS2056 (5' - GAGCCCTCTGAGTCAGATGAGGTTGTC - 3'). The 188-bp \( \text{Lace1} \) sequence was localized using a masked (http://repeatmasker.genome.washington.edu/) query against the Mouse Ensemble database (http://www.ensembl.org, build 23) and bEMS4 (GenBank AF52042), bEMS40 (mouse CTIB BAC Clone; address 26C4) was recovered by screening Research Genetics Mouse BAC DNA Pools Release II under the following conditions: 30 cycles, 94°C for 30 s, 58°C for 30 s, and 72°C for 55 s.
using oEMS1651 (5’-TCCACCTTGCTTTCCTCGT-3’) and oEMS1652 (5’-CAGAGGGTTCTCTTTCTC-3’). Sequence alignment and comparisons were performed using Sequencher software (Gene Codes, Ann Arbor, MI). Physical map distances between Zfa, Nr2e1, Snx3, and Lace1 on Chromosome 10 were obtained using the UCSC Genome Browser (http://genome.ucsc.edu; mouse assembly Feb. 2003).

**Fluorescent In Situ Hybridization (FISH)**

Metaphase chromosome preparation, probe development, and FISH analyses were performed as previously described (Abrahams et al., 2003).

**Expression Analyses**

Total wildtype (WT) and Nr2e1frc/frc mouse RNA were prepared from brain, spleen, liver, and kidney using TRIzol Reagent (Invitrogen, Burlington, Ontario). Gel preparation, electrophoresis, and RNA transfer were performed using NorthernMax-Gly protocol according to the manufacturer’s instructions (Ambion, Austin, TX). We generated 32P-radiolabeled probes by random labeling using Ready-To-Go™ DNA Labeling Beads (Amersham, Piscataway, NJ) on the following templates: Lace1 (1030-bp insert from pEMS914-1, Mus musculus cDNA); Snx3 (RT-PCR amplification of total RNA from C57BL/6J brain, 403 bp, using oEMS1226 (5’-TCCTATCTTCAAGCTGAAGGAATC-3’/H11032) and oEMS1228 (5’-TGCGCAGCATGCTAGTTAGT-3’/H11032); Nr2e1 and GAPDH as previously described (Young et al., 2002). Hybridization using ULTRAHyb and washes using Low and High Stringency Wash Buffer were performed as per the manufacturer’s instructions (Ambion). Signals were detected using storage phosphor imaging screens (Bio-Rad, Hercules, CA).

**Genotyping, Southern Analyses, and Sensitivity Assays**

PCR products were generated from each of WT mice, Nr2e1frc/frc mice, and ES cell clones E14TG2a, mEMS2, mEMS3, and mEMS4 under the following conditions: 1) Nr2e1 assay (30 cycles): 94°C for 30 s, 58°C for 30 s, and 72°C for 55 s using oEMS296 (5’-CTCCATCTCACAAGCTGAGGAATC-3’) and oEMS298 (5’-CTCGACTGCGACAGCTGC-3’); and 2) Nr2e1frc assay (35 cycles): 94°C for 30 s, 60°C for 30 s, and 72°C for 55 s using oEMS650 (5’-GGGGGAGGGAGGGCTTAAATAG-3’) and oEMS1957 (5’-GGGATTTCATCTCCCTCATAAATAG-3’). PCR products
were electrophoresed in 2% agarose gel, transferred to a positively charged nylon membrane (Ambion), and hybridized with labeled PCR products from the Nr2e1 and Nr2e1frc assays. Probes were 32P-labeled by random labeling using Ready-To-Go™ DNA Labeling Beads (Amerham). Hybridization, washes, and detection were performed as described above. The sensitivity assay was developed by serially diluting Nr2e1frc DNA into WT DNA, followed by PCR, Southern, and detection as described above.

ACKNOWLEDGMENTS

The authors thank Melissa L. Berry for assistance in defining the deletion interval and Tracey D. Weir for help in manuscript preparation.

LITERATURE CITED


