Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Megabase deletions of gene deserts result in viable mice

Permalink

https://escholarship.org/uc/item/4fd0v6b0

Authors

Nobrega, Marcelo A.
Zhu, Yiwen
Plajzer-Frick, Ingrid
et al.

Publication Date

2004-05-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Megabase Deletions of Gene Deserts Result in Viable Mice

Marcelo A. Nobrega*, Yiwen Zhu*, Ingrid Plajzer-Frick, Veena Afzal and Edward M. Rubin

DOE Joint Genome Institute

Walnut Creek, CA 94598

and

Genomics Division

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

Correspondence and requests for materials should be addressed to E.M.R. (emrubin@lbl.gov).

* These authors contributed equally to this work

The functional importance of the approximately 98% of mammalian genomes not corresponding to protein coding sequences remain largely un-scrutinized [1]. To test experimentally whether some extensive regions of non-coding DNA, referred to as gene deserts [2-4], contain critical functions essential for the viability of the organism, we deleted two large non-coding intervals, 1,511 kb and 845 kb in length, from the mouse genome. Viable mice homozygous for the deletions were generated and were indistinguishable from wild-type littermates with regards to morphology, reproductive fitness, growth, longevity and a variety of parameters assaying general homeostasis. Further in-depth analysis of the expression of genes bracketing the deletions revealed similar expression characteristics in homozygous deletion and wild-type mice. Together, the two deleted segments harbour 1,243 non-coding sequences conserved between humans and rodents (>100bp, 70% identity). These studies demonstrate that some large-scale deletions of non-coding DNA can be well tolerated by an organism, bringing into question the role of many human-mouse conserved sequences [5,6], and further supports the existence of potentially "disposable DNA" in the genomes of mammals.

The genome of an organism is frequently referred to as its "book of life" [7]. It remains unclear, however, whether this is an information dense book, where every page is required for the proper telling of the story, or whether some of it is disposable, without an impact on the story line. While, in general, the requirements for the majority of coding regions of the mammalian genome have been established in a large number of studies, the activity of nearly the entire genome corresponding to non-coding sequences have only been minimally examined [1]. In order to explore the activity of non-coding DNA on a large scale, we removed extended regions of non-coding DNA from the mouse genome, to determine its impact on the organism.

We selected two regions for deletion, a 1,880 kb gene desert mapping to mouse chromosome 3, and a second region, 960 kb in length, mapping to mouse chromosome 19 (Fig. 1a). Orthologous gene deserts of approximately the same size are present on human chromosomes 1p31 and 10q23, respectively. No striking sequence signatures such as repeat content, nucleotide composition or substitution rate distinguish these two selected gene deserts from other regions of the genome, except for their lack of annotated genes and evidence of transcription. Together, the two selected regions contain 1,243 human-mouse conserved non-coding elements (>100bp, 70% identity), also similar to genome averages, while no ultra-conserved elements [8] or sequences conserved to fish are present.

We generated the deletions by coupling genomic targeting and Cre-lox technologies [9,10] (Fig. 1b). Vectors carrying loxP sites were consecutively integrated into the mouse genome at the boundaries of each gene desert, in embryonic stem cells [11]. The boundaries for the deletions ranged from 50 to 250 kb from the nearest gene bracketing the desert, allowing for proximate regulatory elements nearby the flanking genes to remain intact. Cre-mediated recombination in embryonic stem cells resulted in the deletion of a 1,511 kb segment on the mouse chromosome 3 desert (termed del^{Mm3}), and a 845 kb deletion on the mouse chromosome 19 desert (termed del^{Mm19}) (Fig. 1c and d). We subsequently generated mice carrying the deletions, first as chimeras and, after successfully transmitting the deletions through the germline, as heterozygous del^{Mm3} /+ and del^{Mm19} /+ mice. The heterozygous deletion mice, appearing phenotypically normal, were intercrossed to generate homozygous deletion mice (del^{Mm3} / del^{Mm3} and del^{Mm19} / del^{Mm19}). Since the homozygous deletion mice for both deletions were viable, we next examined a large cohort of live-born offspring from heterozygous deletion intercrosses, to determine whether the deletions had an impact on

embryonic development. Genotyping 292 offspring of the del^{Mm3} /+ intercross and 318 offspring of the del^{Mm19}/+ intercross revealed an approximate 1.0:2.0:1.0 ratio of wild-type to heterozygous to mutant homozygous mice obtained for both the Mm3 and Mm19 deletions, indicating that the deletions do not result in embryonic lethality.

We next explored the overall fitness and a variety of phenotypic parameters measured in the homozygous deletion mice, compared to controls. Monitoring post-natal survival rates for 25 weeks for each of the deletions revealed no differences between del^{Mm3} / del^{Mm3} or del^{Mm19} / del^{Mm19} and their respective wild-type littermates (Fig. 2). Neither genomic deletion elicited measurable growth retardation either, as shown by similar growth curves obtained for each of the 4 groups, del^{Mm3} / del^{Mm3}, del^{Mm19} / del^{Mm19} and their wild-type littermates (Fig. 2). In addition, we measured 18 different clinical chemistry tests representing general as well as specific plasma parameters in the same groups of mice, also with no significant differences detected between del^{Mm3} / del^{Mm3} or del^{Mm19} / del^{Mm19} and their wild-type littermate controls. We then carried out visual and pathological examinations on multiple organs from the various groups of mice at 6 months of age, including brain, thymus, heart, lungs, liver, spleen, stomach, intestine, kidney, urinary bladder, uterus, ovaries and testicles. No morphological abnormalities, evidence of abnormal growth or tissue degeneration were observed in del^{Mm3} / del^{Mm3} or del^{Mm19} / del^{Mm19} mice. Organ mass was similar in both groups of deletion mice and their wild-type littermates.

Finding no differences in any of the whole organism phenotypes tested, we subsequently explored the impact of the deletions at a molecular level. The expression

levels of multiple genes flanking the boundaries of the two deletions were determined in 12-week old mice. We assayed four genes bracketing the Mm19 deletion and five genes bracketing the Mm3 deletion by real-time quantitative PCR, in a panel of 12 tissues representing the overall expression patterns of each assayed gene. The tissue specificity of expression for all the genes tested was similar in homozygous deletion mice compared to their wild-type littermates (Fig.3). Out of the 108 quantitative expression assays (12 tissues for 9 genes), only 2 revealed detectable alterations in levels of expression. The expression of Prkacb was reduced in the heart of del^{Mm3} / del^{Mm3} mice and Rpp30 was reduced in intestine of del^{Mm19} / del^{Mm19} mice, compared to wild-type littermates (Fig. 3).

.

Given the small number of expression changes that we did observe, coupled with the limited set of tissues used for the expression assay, we carried out a series of reporter assays in transgenic mice to further explore the possible existence of regulatory sequences in the deleted segments [12-14]. The elements selected for testing in this gain-of-function assay correspond to those with the highest degree of sequence conservation extending over the largest intervals between humans and mice. We also sought those elements with conservation extending over the longest evolutionary time scale. While non-coding conservation in the Mm19 desert was shared only amongst mammals, the Mm3 desert contained human-mouse conserved elements also shared with chicken (75 elements) and frog (5 elements). From the Mm19 desert, we picked five human-mouse conserved elements representing the most conserved sequences between humans and mice (>180bp, 90% identity) for the in vivo assay. The ten elements chosen from desert Mm3 (>400bp, 90% identity) included all five sequences that are conserved across humans, rodents, chicken and frog, and five that are conserved among humans, rodents and chicken only (Fig. 4).

We cloned each element in a reporter vector [15], injected it into fertilized mouse oocytes and assayed for the presence of beta-galactosidase activity in the resulting embryos at 14.5 days post-coitum . Eight to sixteen independent transgenic micethat were generated for each of the 15 elements were examined . Of the elements tested, only one, located within the desert Mm3, reproducibly drove beta-galactosidase expression in a set of tissues that include mammary glands and abdominal muscles (Fig. 4). It is noteworthy that this element is conserved deep in the vertebrate lineage, including human, mouse, chicken and frog. The small fraction of elements with enhancer activity in this study (1 out of 15) contrasts with the results obtained when human-fish conserved non-coding sequences were previously tested using the same *in vivo* assay [16,17]. In those studies a significant fraction of human-fish conserved non-coding sequences present in gene deserts were shown to be functional, with 5 out of 7 elements in one study [13] and 22 out of 29 in a second study (E.M.R., personal communication) demonstrating enhancer activity.

The deletions carried out in this study represent the largest reported viable homozygous deletions in mice and supports the existence of potentially "disposable DNA" in mammalian genomes. In assessing the impact of these deletions on the engineered mice, it is important to acknowledge that our ability to phenotype an organism will always miss some features no matter how detailed. It is possible, even likely, that the animals carrying the megabase-long genomic deletions do harbour abnormalities undetected in our assays, which might impact their fitness, in some other time scale or setting than the ones assayed in this study. Nonetheless, the lack of detectable phenotypes in these mice raises the possibility that the mammalian genome is

not densely encoded and that significant reductions in genome size may be tolerated. Linked to this, the extensive degree of non-coding conservation in the deleted intervals brings into question the functionality, if any, of many of the large number of non-coding sequences shared among mammals.

Methods

Generation of the targeting vectors. Two targeting vectors were generated, in house, for each deletion. One targeting vector consists of a loxP site coupled to a neomycin-resistance gene and hsv-tk (ploxPneoTK-2-BRI) and the other vector consists of a loxP site coupled to a hygromycin-resistance gene and hsv-tk gene (plocPhygTK-BH). Homologous arms, 5,700 to 7,000 bp in length, were generated by amplification from 129Sv mouse genomic DNA, with the PCR products cloned initially into a pCR2.1 vector (Invitrogen TOPO XL PCR cloning kit), and subsequently cloned into each corresponding targeting vector.

Creation of the genomic deletions *in vitro*. Constructs were linearized and electroporated (20 μg) into mouse ESVJ embryonic stem cells (129Sv genetic background). Cells were electroporated with either ploxPneoTK-2-BRI or ploxPhygTK-BH constructs and selected under G418 (180 μg/ml) or hygromycin B (150 μg/ml), respectively, for 8 days. Resistant colonies were picked and expanded on 96-well plates and screened by both by PCR and southern hybridisation. Clones that were correctly targeted (double-targeted) were electroported with 20 μg of the Cre recombinase-expressing plasmid pBS185 [18]. After selection with FIAU, which selects for colonies in which the deletion occurred removing the hsv-TK genes, resistant colonies were

screened both by PCR and Southern blot to identify those carrying the designed genomic deletion (Fig. 1).

Generation of mice carrying the genomic deletions. The selected mouse embryonic stem cells were injected into C57BL/6 blastocysts, and mice generated as previously described [19]. Male chimeric mice were bred to 129/SvEv females to generate heterozygous deletion mice. No subsequent backcrosses were performed, and all mice reported in this study carry a mixed C57BL/6 and 129/SvEv backgrounds, minimizing potentially confounding factors due to genetic background.

Real-time quantitative PCR. Total RNA was extracted from each of the 12 organs tested, and pooled (4 males and 4 females). Following reverse transcription, real-time PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems). An 18S RNA standard internal control (Ambion) was used. A ratio of 4:6 of 18S primer pair to 18S competimers was adopted. All procedures and calculations were carried out according to manufacturer's recommendations.

Mouse transgenic reporter assay. Conserved non-coding sequences were PCR-amplified from human DNA (Invitrogen), with PstI, HndIII or XhoI linkers, and cloned into phsp68/LacZ. The constructs were purified and injected into pronuclei as previously described[20]. Transgenic embryos were identified by PCR of beta-galactosidase, using DNA from yolk sac, with the following primers: F-TTTCCATGTTGCCACTCGC and R- AACGGCTTGCCGTTCAGCA. Expression of beta-galactosidase was assayed in all embryos, using X-gal, (Sigma), as previously described[21].

Acknowledgements We thank Ivan Ovcharenko, Gabriela Loots and Jody Schwartz for help with the identification and annotation of the gene deserts; Dario Boffelli, Len Pennacchio, Nadav Ahituv, James Bristow and other Rubin lab member for suggestions and criticisms to the manuscript; Howard Jacob for kindly providing the clinical chemistry assays. Research was conducted at the E.O. Lawrence Berkeley National Laboratory and at the Joint Genome Institute, with support by grants from the Programs for Genomic Application, NHLBI, and D.O.E.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.M.R (e-mail: emrubin@lbl.gov)

Figure Legends

Figure 1. Design and generation of genomic deletions. **A,** the two gene deserts selected for deletions are illustrated, with the distances (in kb) form the boundary of each deletion and the closest gene denoted in parenthesis. **B,** General strategy of deletions. Consecutive targeting of the deserts' ends is followed by Cre-mediated deletion of one copy of the desert. **C,** PCR screening of cell colonies using primer pairs from regions bracketing the deletions identifies colonies carrying the deletion. The distance separating the primer pairs decrease from 1,514 and 848 kb in wild-type chromosomes from deserts Mm3 and Mm19, respectively, to 3.0 kb in chromosomes carrying the deletions (3 positive heterozygous colonies shown for each desert). **D,**

Southern blot of the cell colonies shown in (C) confirm the presence of heterozygous deletions in 3 colonies of each desert deletion.

Figure 2. Survival and growth curves for each deletion.

Figure 3. Expression characteristics of the genes bracketing the deletions in a panel of 12 tissues. Asterisks denote statistically significant differences between homozygous deletion mice compared to wild-type littermates. Distances of each gene to the closest end of the deletion are given in parenthesis. Vertical axes correspond to the ratio of 18S RNA expression to the expression of the corresponding gene at each given tissue.

Figure 4. Conservation profile across vertebrates of each gene desert. Alignments were obtained using Vista (http://www-gsd.lbl.gov/VISTA), using human sequence as reference. Exons of the genes bracketing the deserts are denoted in blue, while non-coding sequence conservation is shown in red. A black bar over each plot illustrates the deleted segments form each desert. A thin bar at the bottom corresponds to 200 kb. The locations of the sequence elements selected for the in vivo mouse transgenic assay are given by arrows at the bottom of the conservation plots. The element which corresponded to an enhancer at E14.5 is highlighted by a red arrow, with a representative embryo showing the expression pattern of the element displayed.

Figure 1.

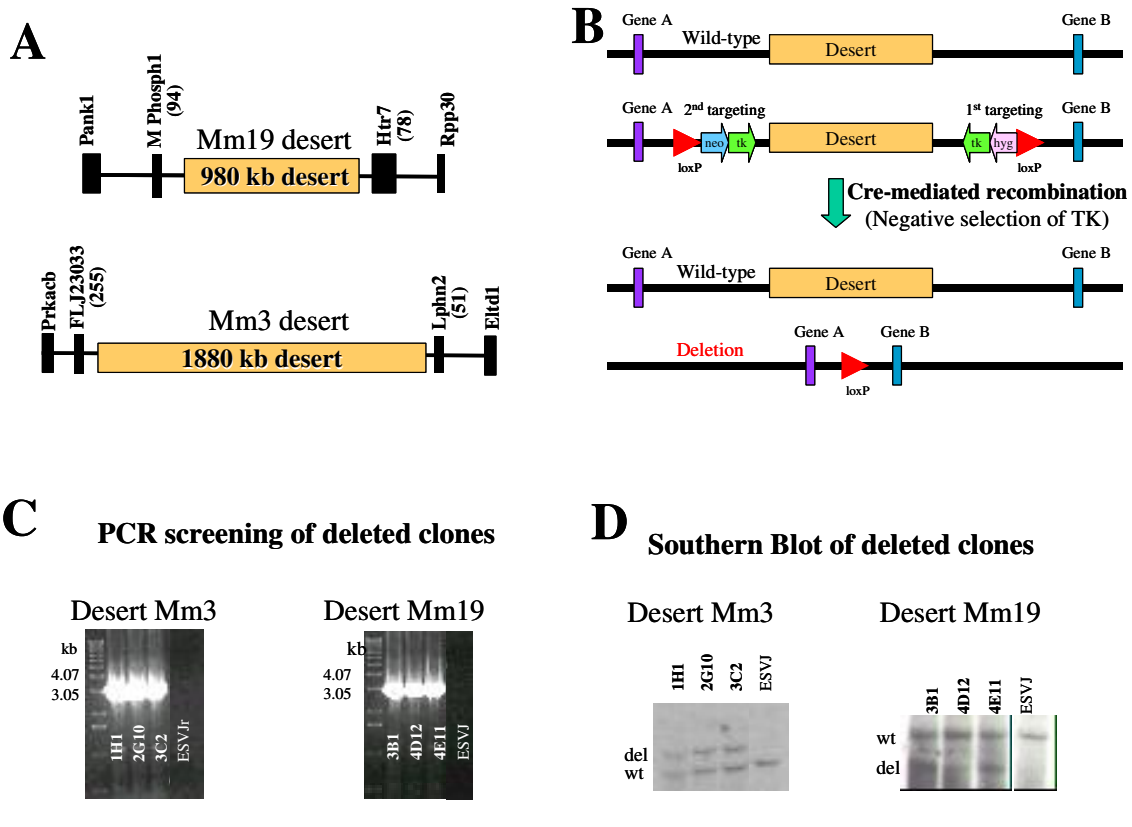

A
Pank1
M Phosph1 (94)
Mm19 desert
980 kb desert
Htr7 (78)
Rpp30
Prkacb
FLJ23033 (255)
Mm3 desert
1880 kb desert
Lphn2 (51)
Eltd1
B
Gene A
Wild-type
Desert
Gene B
2nd targeting
neo
tk
loxP
1st targeting
tk
hyg
loxP
Cre-mediated recombination
(Negative selection of TK)
Deletion
C
PCR screening of deleted clones
Desert Mm3
Desert Mm19
kb
4.07
3.05
1H1
2G10
3C2
ESVJr
3B1
4D12
4E11
ESVJ
D
Southern Blot of deleted clones
Desert Mm3
Desert Mm19
del
wt

Figure 2.

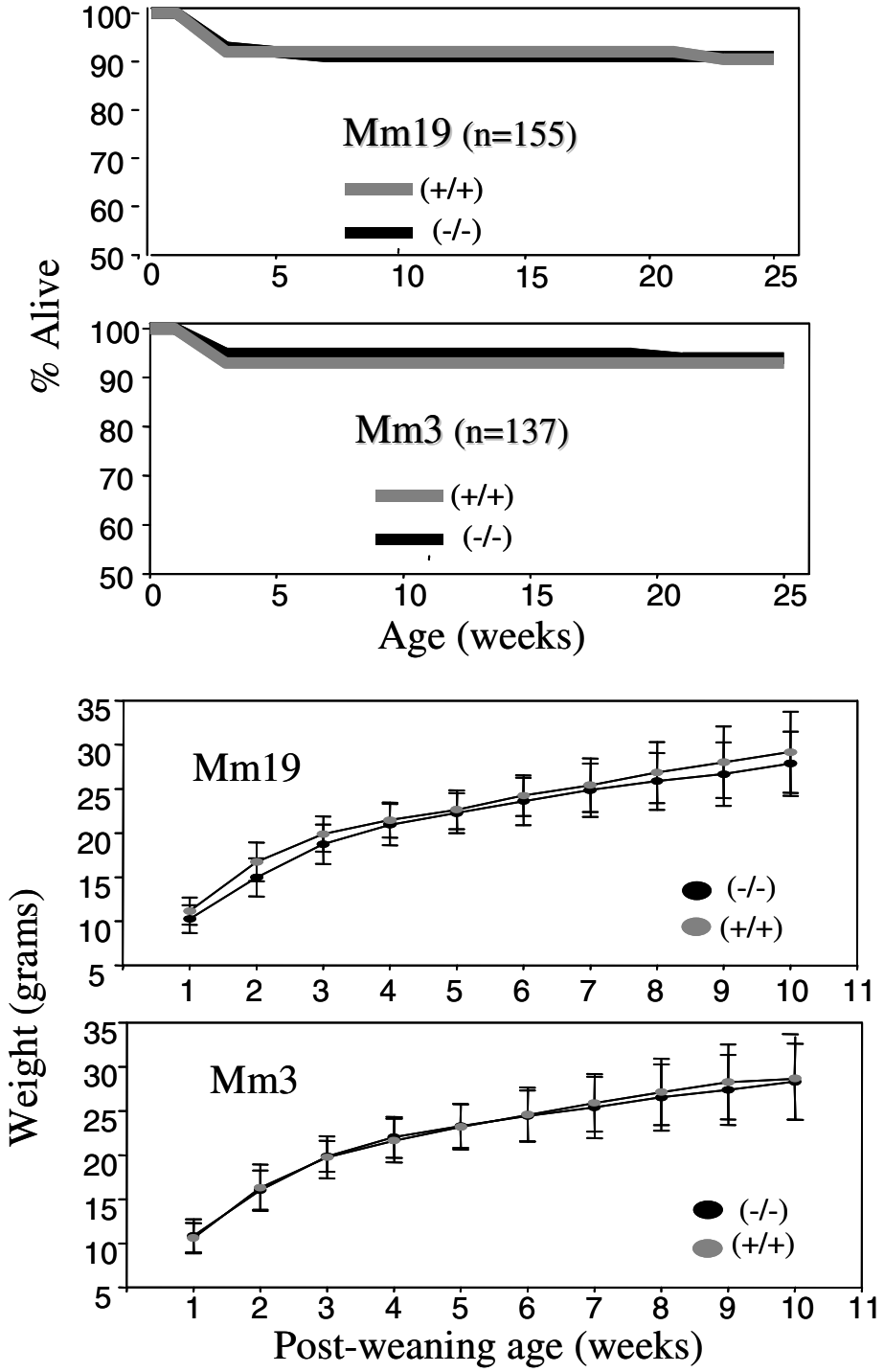

% Alive
100
90
80
70
60
50
Mm19 (n=155)
(+/+)
(-/-)
0
5
10
15
20
25
Mm3 (n=137)
(+/+)
(-/-)
Age (weeks)
Weight (grams)
35
30
25
20
15
10
5
Mm19
(-/-)
(+/+)
1
2
3
4
5
6
7
8
9
10
11
Mm3
(-/-)
(+/+)
Post-weaning age (weeks)

Figure 3.

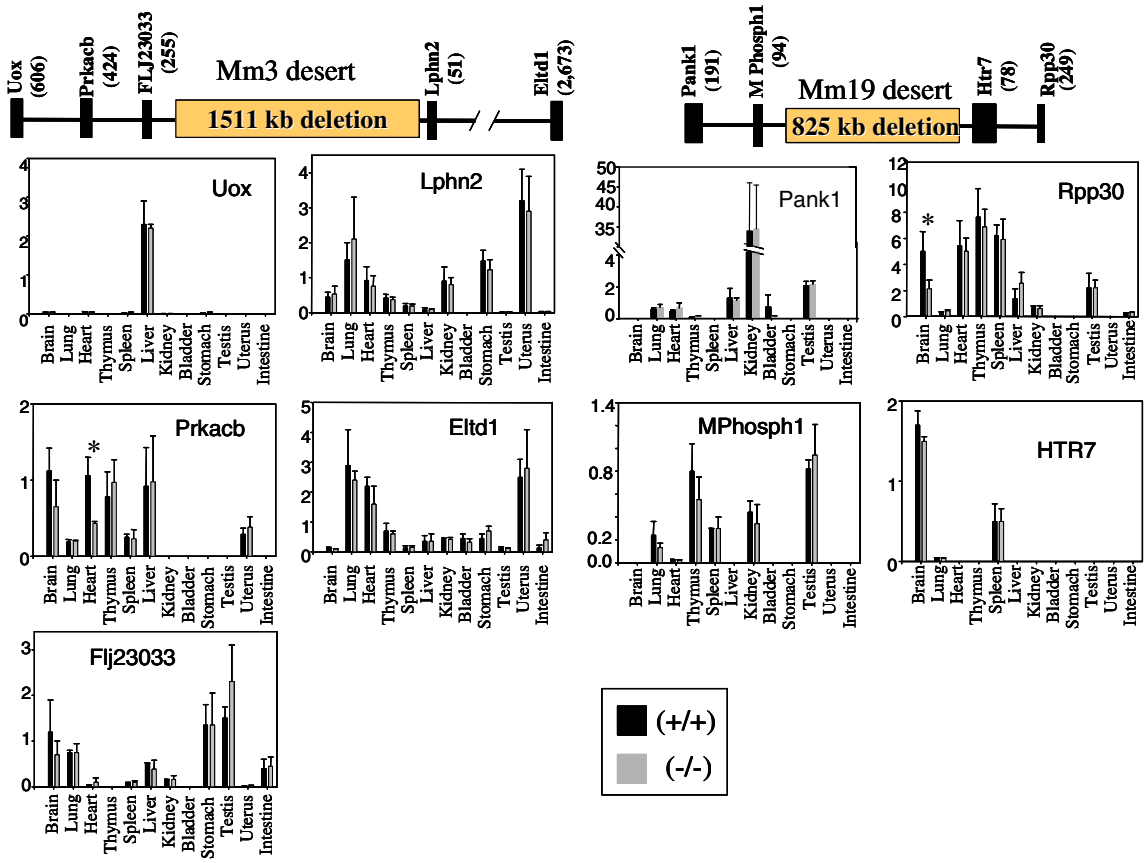
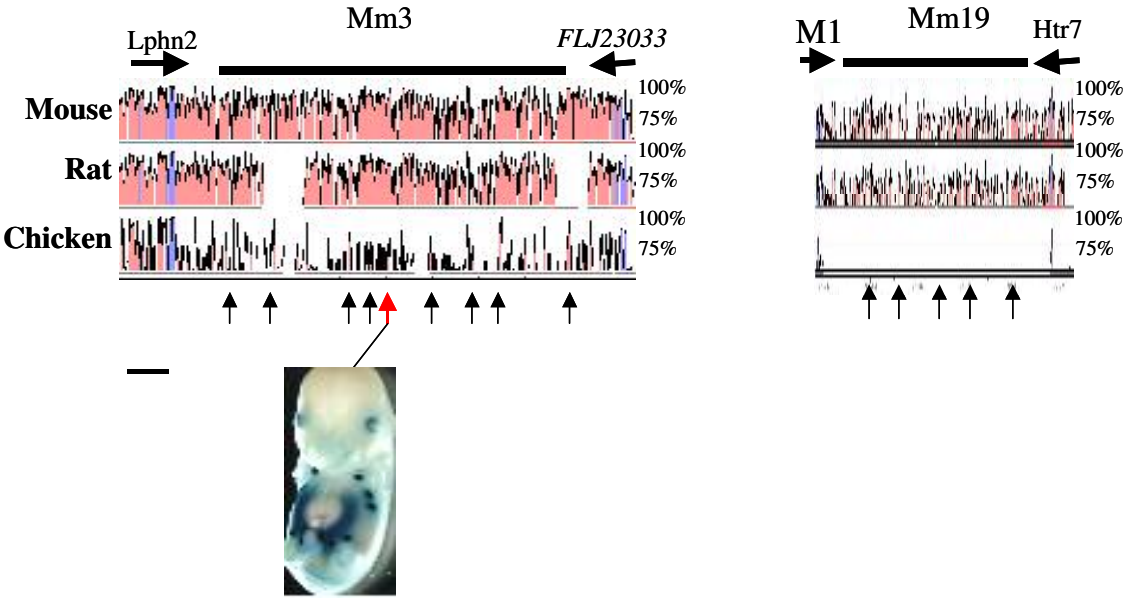

Uox (606)
Prkacb (424)
FLJ23033 (255)
Mm3 desert
1511 kb deletion
Lphn2 (51)
Eltd1 (2,673)
Pank1 (191)
M Phosph1 (94)
Mm19 desert
825 kb deletion
Htr7 (78)
Rpp30 (249)
Uox
Lphn2
Pank1
Rpp30
Prkacb
Eltd1
MPhosph1
HTR7
Flj23033
Brain
Lung
Heart
Thymus
Spleen
Liver
Kidney
Bladder
Stomach
Testis
Uterus
Intestine
(+/+)
(-/-)

Figure 4.


Mm3
Lphn2
FLJ23033
Mouse
Rat
Chicken
100%
75%
100%
75%
100%
75%
M1
Mm19
Htr7
100%
75%
100%
75%
100%
75%

Supplemental material

Genomic coordinates and annotation of gene deserts

All coordinates are given using the Oct. 2003 Mouse genome assembly at UCSC (http://genme.cse.ucsc.edu) as reference.

Coordinates for desert Mm19 deleted segment: chr19:35,033,162-35,878,431

Coordinates for Mm3 deleted segment: chr3:151,157,238-152,668,920

Annotation of the deleted segments:

Desert mm19:

Two mRNAs map to this desert:

U6662 – this single-exon mRNA has no counterpart identified in any other species tested. It also fully overlaps with a LTR repeat, with several hundred identical hits (100% identity) throughout the mouse genome. It was deemed, thus, a probable contamination sequenced in a single cDNA library, and discarded as a putative gene.

AK080606- This mRNA corresponds to a RefSeq gene (A830019P07Rik). It also has no counterpart in any other species, including rats. From the 6 exons of this gene, five overlap fully or partially with repetitive elements from the LINE and LTR classes, also suggesting that this may be a relic of an ancient transposon-associated gene. The only mRNA identified only in mouse was cloned from an E10.5 embryonic library. We

attempted to identify this transcript by RT-PCR and failed. Finally, the putative protein encoded by this mRNA does not have any domain match with any known protein. Therefore, we also decided to exclude this as a mouse gene.

Desert Mm3:

Five mRNAs map to this desert:

AK041984 - this single-exon mRNA has no counterpart identified in any other species tested and no sequence conservation. The putative protein has no similarity to any known protein. It also partially overlaps with a LTR repeat, with several hundred identical hits (100% identity) throughout the mouse genome. It was deemed, thus, a probable contamination sequenced in a single cDNA library, and discarded as a putative gene.

AK029582- this double-exon mRNA has no counterpart identified in any other species tested. The putative protein has no similarity to any known protein. RT-PCR from an adult mouse testis cDNA library (from where the mRNA was identified) failed to identify a similar trascript. Therefore, while we cannot rule this sequence out as a putative gene, we find the evidence pointing to that unconvincing.

AK085270 – this three-exon mRNA has no counterpart identified in any other species tested. One of the three exons fully overlaps with repeats form the SINE and LTR classes, with several hundred identical hits (100% identity) throughout the mouse genome. No similarities between the putative protein and known protein databases was identified. It was deemed, thus, as an unlikely candidate for being characterized as a gene.

AK015113- this single-exon mRNA has no counterpart identified in any other species tested and no sequence conservation. The putative protein has no similarity to any known protein. It also partially overlaps with a LTR repeat, with several hundred identical hits (100% identity) throughout the mouse genome. It was deemed, thus, a probable contamination sequenced in a single cDNA library, and discarded as a putative gene.

S74315 – This mRNA corresponds to a copy of a mouse retrotransposon gene, encoding the retroviral nucleocapsid protein Gag; there are hundreds of similar copies of this retrotransposon in the mouse genome, and despite the evidence that this is indeed a coding gene, was disregarded from our catalogue of mammalian genes.

KO vector construction:

Two targeting vectors were generated for each deletion. One targeting vector, containing homologous region at one end of a desert to be deleted, was in ploxPneoTK-2-BRI vector backbone. The second targeting vector, containing homologous region at another end of a desert to be deleted, was in ploxPhygTK-BH vector backbone. Both ploxPneoTK-2-BRI and ploxPhygTK-BH were generated in this lab.

Homologous arms were generated by PCR from 129Sv mouse genomic DNA. The PCR products of homologous arms were first cloned into TA vector (Invitrogen TOPO XL PCR cloning Kit), and then sub-cloned into

ploxPneoTK-2-BRI right next to loxP site or ploxPhygTK-BH at the 3' side of PGK terminator of PGKtk cassette.

Table 1: Summary of targeting vector construction:

Desert		Neo^r Constructs	Hrg^r Constructs
1p21	primers for homologous arm	1p21F5: 5' AGGTGTTAGCAATTATGCTCCTCTG 3' 1p21R5: 5' GCATGCTCTTACAGAGGCTACCTTA 3'	1p21F3b: 5' GACCATAGCTTGGGTTCTAATCTCA 3' 1p21R3b: 5' CAGGAGAAGAAATAGTCCCTCCAAAC
	product size	7000 bp	7000 bp
	vector	ploxPneoTK-2-BRI	ploxPhygTK-BH
	Targeting location	Position of homologous arm at 5' end of the desert	Position of homologous arm at 3' end of the de
	Linearized at	BamHI	?
	Selection	G418, 180 μg/ml active G418 for 8 days	Hygromycin, 150 μg/ml hygromycin B for 8 d
10q23	primers for homologous arm	10q23F5g: 5'TCCTTCTCTACAATGCTTGTTCCTT3' 10q23R5g: 5'GCTCATCTCCATTTCCTCTGTGTAT3'	10q23F3b:CTTTGTAGCTCCACCTGACTTCTTCCTGC 10q23R3: 5' TCTCAGGAAGTTCACCTGCTAGTTT 3'
	product size	6500 bp	5700 bp
	vector	ploxPneoTK-2-BRI	ploxPhygTK-BH
	Targeting location	Position of homologous arm at 5' end of the desert	Position of homologous arm at 3' end of the de
	Linearized at	PmlI	?
	Selection	G418, 180 μg/ml active G418 for 8 days	Hygromycin, 150 μg/ml hygromycin B for 8 d

Gene Targeting and Screening:

A) Gene targeting to introduce loxP sites to both ends of the regions to be deleted

Basic technology used for gene targeting and screening has been described previous[1)]. About 20 μg of linearized Hyg^r construct of each desert was electroporated into ESVJ ES cells (Genomesystems). Cells were under hygromycin B (150 μg/ml) selection for 8 days. Hyg resistant colonies were picked and expanded on 96-well plates. Genomic DNA was extracted, digested with restriction enzyme and screened by Southern hybridization. This targeting introduced first loxP site into 3' end of the desert to be deleted (Figure 1).

Correct targeted Hyg^r clones from each desert were polled together, expanded and electroporated with about 20 μg of linearized Neo^r construct for the same desert. Cells were under G418 (180 μg/ml) selection for 8 days. Neo resistant colonies were picked and screened by Southern hybridization. This targeting introduced second loxP site into 5' end of the desert to be deleted.

Southern probes were generated by PCR, and were located near but outside of the homologous region.

Table 2: Southern screening of targeting events of Hyg^r constructs and Neo^r constructs

Desert		Neo^r Constructs	Hyg^r Constructs
1p21	primers for Southern probe	5' CCAGCACATAGTAGTCTCATACAGCT 3' 5' CTATTCCCCTCCTAGCTTGCAT 3'	5' TTTTTTCTGTAAGTCGATACA 3' 5' ACAGTCATTCACAGCAGCTTCC 3'
	probe size	236 bp	190 bp
	probe location	46 bp 3' downstream of the homologous arm	217 bp 5' upstream of the homologous arn
	restriction enzyme	EvoRV	BglII
	expected fragment size	wt --- 13.5 kb, targeted --- 9 kb	wt --- 16.5 kb, targeted --- 12.9 kb
	No. correct targeted clone	28	12
	No. clones screened	265	204
	efficiency	10.60%	6%
10q23	primers for Southern probe	5' TGGGAAGAAAAGCAGCTGTG 3' 5' GCCTAAGTGGACTTACAGTTATTCCA 3'	5' TGGACTATACAGTTATTGGTGGATTTG 3' 5' CTGTCACTAGTTTCACAGGTTCATTCT 3'
	probe size	199 bp	407 bp
	probe location	1563 bp 5' upstream of the homologous arm	11 bp 3' downstream of the homologous a
	restriction enzyme	XbaI	BglII
	expected fragment size	wt --- 12.4 kb, targeted --- 9.3 kb	wt --- 15.2 kb, targeted --- 11.7 kb
	No. correct targeted clone	13	10
	No. clones screened	254	201
	efficiency	5.10%	5%

B) Cre-recombinase-mediated *loxP* recombination

Double targeted clones (Hyg^r/ Neo^r, note the two loxP sites could be in *cis* or in *trans*) from each desert were polled together, expanded and electroporated with about 20 μg of Cre recombinase-expressing plasmid *pBS185*[2]. When loxP sites presence in *cis*, the

loxP bracketed sequence, which includes desert to be deleted, the PGKhyg, PGKneo and *HSV-tk* cassettes, can be deleted by Cre recombinase-mediated *loxP* recombination. While loxP sites presence in *trans* can only result in translocation. ES cells that had undergone recombination and deletion were identified by selecting for hygromycin sensitivity and 1-(2-deoxy-2-fluoro-b-D-arabinofuransyl)-5-iodouracil (FIAU) resistance. Cells surviving this selection were screened by PCR for the predicted *cis* deletion using a primer outside the deleted region and T7 primer within the vector backbone left on chromosome after deletion. The predicted *cis* deletion was further confirmed by positive PCR using primers outside the deleted region, negative PCR of Neo or Hyg primers, and by Southern blot analysis using a probe outside the deletion region.

Table 3: Primers and probes used for screening *cis* deletion event

Desert		1p21	10q23
Primers for screening deletion	desert.fwd	5' GAGTAATTAAGGTAGCCTCTGTAAGAGCA3'	5' TTCAATAGCGATCACAACCTCTGT 3'
	T7 primer	5' GCGTAATACGACTCACTATAGGGC 3'	same as 1p21
	product size	~ 140 bp	~ 140 bp
Primers outside deleted region	desert.fwd	same as above desert.fwd primer	same as above desert.fwd primer
	desert.rev	5' AGGGTGAGATTAGAACCCAAGCTAT 3'	5' CAGGAAGAAGTCAGGTGGAGCTA 3'
	product size	~ 2.9 kb	~ 2.9 kb
Neo primers	Neo.fwd	5' GATGGATTGCACGCAGGT 3'	same as 1p21
	Neo.rev	5' GGCAGGAGCAAGGTGAGA 3'	
	product size	318 bp	
Hyg primers	Hyg.fwd	5' CGGAAGTGCTTGACATTGG 3'	same as 1p21
	Hyg.rev	5' GTATTGACCGATTCCTTGCG 3'	
	product size	209 bp	
Primers for Southern probes	forward	5' CATTCAGAGAAGGAGTCTGGCTT 3'	5' TGGACTATACAGTTATTGGTGGATTTG 3'
	reverse	5' CTGTTCCTATCTGTCCCCATTG 3'	5' CTGTCACTAGTTTCACAGGTTCATTCT 3'
	probe location	546 bp 5' upstream of (Neo) homologous arm	11 bp 3' downstream of (Hyg) homologous arm
	probe/enzyme	195 bp probe, Bbs I digestion	407 bp probe, Bgl II digestion
	expected bands	wt - 9.8 kb, del - 13 kb, Neo targeted no deletion - 24.5 kb	wt - 15.2 kb, del - 9.9 kb, Hyg targeted no deletion - 11.7 k
No. deletion clones/No. screened		3 deletion clones, 44 screened	3 deletion clones, 75 screened
Efficiency		6.80%	4%

Blast Injection and Mouse Line Establishment:

The clones heterozygous for the deleted chromosome were injected into blastocysts to generate chimeric mice, and germline transmission of the deletion for both 1p21 and 10q23 deserts were obtained. Mice heterozygous for the deletion were intercrossed to produce the wild-type, heterozygote, and homozygote littermates used in this study.

Mouse tail genotyping (screening of wild-type, heterozygous deletion and homozygous deletion mice)

Genomic DNA was extracted from mouse tail. Genotyping of mice was performed by using the following PCR screening scheme.

Table 4: Primers used for genotyping

1p21		10q23	
desert.fwd	5' GAGTAATTAAGGTAGCCTCTGTAAGAGCA 3'	desert.fwd	5' TTCAATAGCGATCACAACCTCTGT 3'
1p21(neo arm) 3'	5' CTATTCCCCTCCTAGCTTGCAT 3'	T7 primer	5' GCGTAATACGACTCACTATAGGGC 3'
T7 primer	5' GCGTAATACGACTCACTATAGGGC 3'	10q23.fwd	5' CCTAGTTTTACTGAGGAATATTGACGAG 3'
		desert.rev	5' CAGGAAGAAGTCAGGTGGAGCTA 3
Primer ratio	desert.fwd : 1p21(neo arm) 3' : T7 = 2 : 1 : 1	Primer ratio	desert.fwd : T7 : 10q23.fwd : desert.rev = 0.6 : 0.6 :1 :1
PCR bands	wt - 314 bp; (+/-) del - 140 bp, 314 bp; (-/-) del - 140 bp	PCR bands	wt - 375 bp; (+/-) del - 140 bp, 375 bp; (-/-) del - 140 b

Quantitative Real-time RT-PCR using ABI Prism 7700 Sequence Detection System

A) Total RNA extraction:

Total RNA was extracted from tissues using Invitrogen's TRIzol reagent. Total RNA for brain, lung, heart, thymus, spleen, liver, kidney, intestine, stomach and skeletal muscle

were extracted from a pool of 4M, 4F mice. Total RNA for bladder, aorta, testis, uterus and ovaries were extracted from a pool of more mice (10-16). Total RNA from ovaries includes RNA from oviducts. See DesertRNAprep file for detail.

B) DNase treatment of total RNA:

Total RNA was treated with Premega RQ1 RNase-Free DNase (cat# M6101) before used for reverse transcription.

C) Reverse transcription:

DNase-treated RNA was reverse-transcribed using Invitrogen life technologies SuperScript™ First-Strand Synthesis System (cat# 11904018).

D) Quantitative real-time PCR:

PCR was performed using SYBR Green PCR Master Mix (Applied Biosystems) on an ABI Prism 7700 Sequence Detection System.

E) Primers for PCR

The 18S internal standard control was from Ambion (315 bp product). A ratio of 4:6 of 18S primer pair : 18S competimers was used. All of the procedures and calculation of the results were carried out according to manufacturer's recommendations.

Gene-specific primers are listed below.

Table 5: Gene-specific primers for real-time RT-PCR

Gene	Primer sequence	PCR product size (bp)	Distance on genomic DNA (bp)
1p21 desert:			
FLJ23033	GAAGATTACAGACCTGCAGAACTACC TCCATAGGCCGAGACTTGATC	120	5,156
Prkacb	AGCCTAGCCAAAGTTGCTGC AAATCCTCAAGCCCAGCATTAC	224	57,261
Uox	AAGAACACTGTGCACGTCCTG TGATTCCAGAGTGAATGACGG	259	13,257
Lphn2	CACTCTGAAACCAGATTCTAGCAGG AACGCGCCAAGTACCCAA	159	2,003
Eltd1	AATGGTGTGGAAGCCTGCTT CAGAAGACTCGCTGCACTCATC	91	53,207
10q23 desert:			
MPHOSPH1	AGTTCAGCAAGCTTCAGGACG TTCTCTGCCAGCTGTGCTGT	215	8,766
Pank1	CTGGTTACCTCTTTGTGGGTGTC CCAGTTTTGCCGTAGGCAGT	212	37,094
Htr7	TGAGGACCACCTATCGTAGCCT GCTCTGGATCATGTATCATGACCT	175	9,007
Rpp30	CAAGGGTCCGGCTGTATGATAT CCATAGACAAGTTCAAAGCCCAG	187	11,058

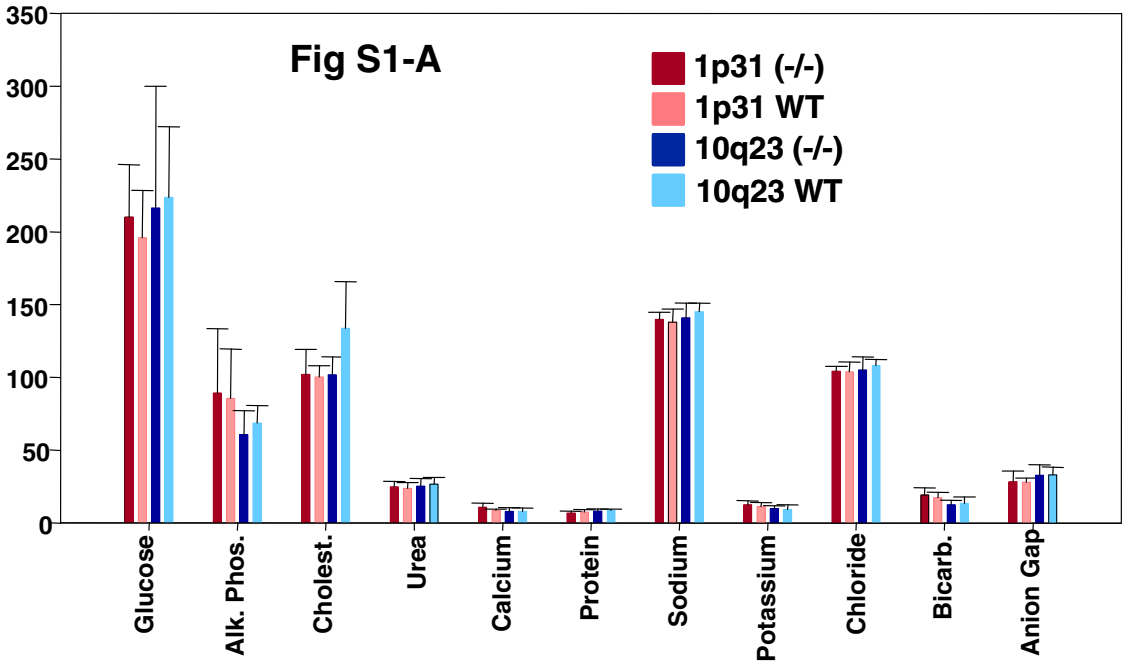
Reproductive fitness and genotypes of offspring from heterozygous deletion intercrosses:

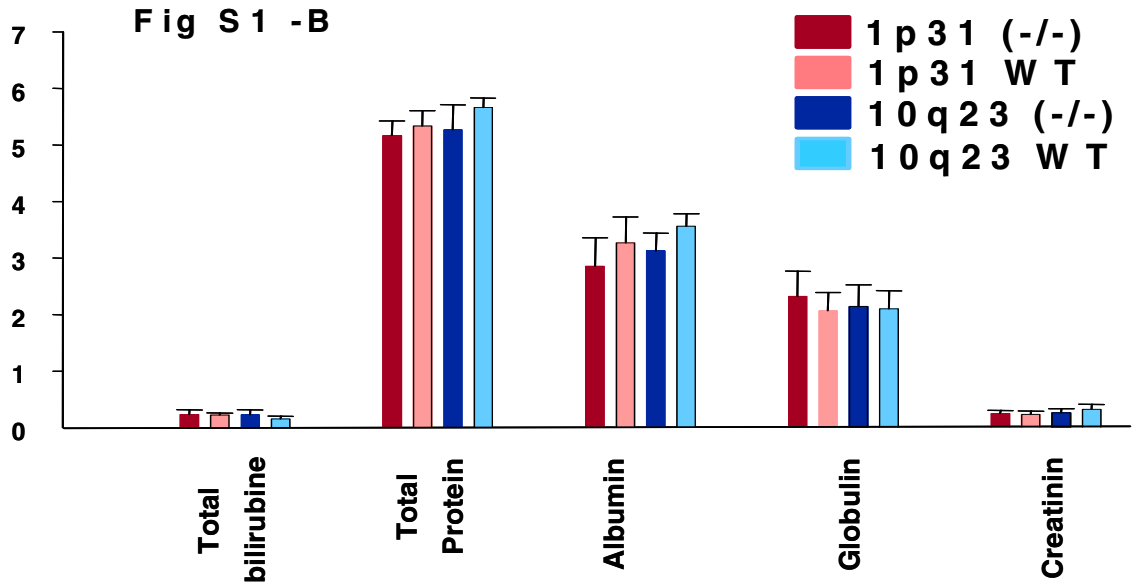
Table 6. Reproductive fitness, litter size, neonatal death and genotype at weaning

Summary --- deletion KO heterozygous breeding

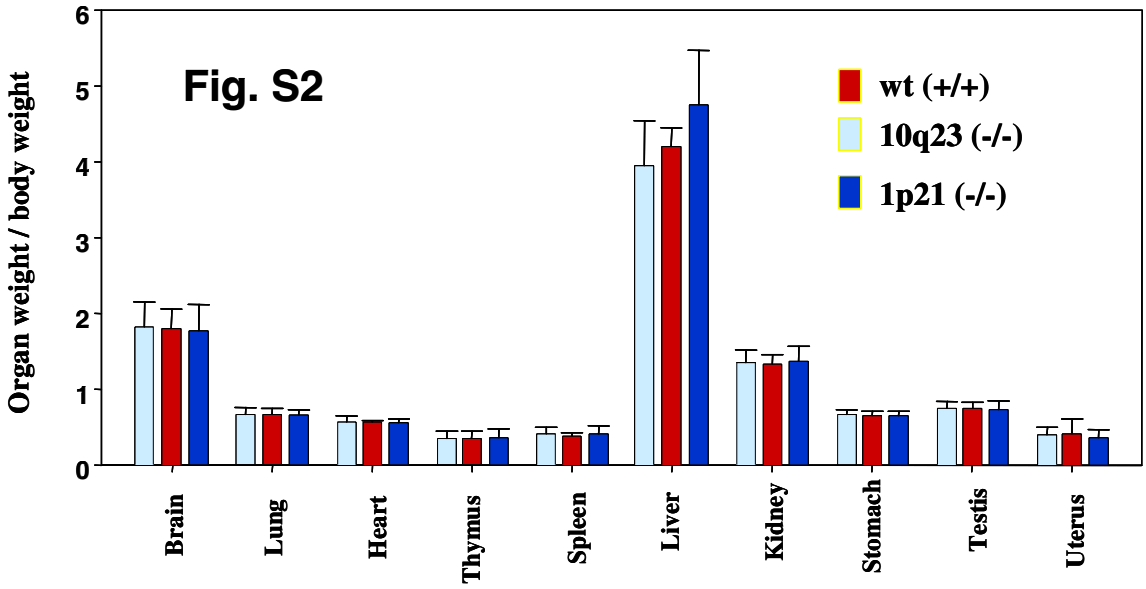
	1p21 deletion	10q23 deletion
No. of pups born	302	343
Litter No.	37	45
Average litter size	8.2	7.6
No. of pups at wean	292	318
No. of pups died	10	25
Average litter size at wean	7.8	7.1
No. male pups	148	160
No. female pups	144	158
Ratio Male : Female	1 : 1	1 : 1
No. (-/-) pups	35M + 33F = 68	40M + 33F = 73
No. (+/-) pups	74M + 79F = 153	77M + 86F = 163
No. wt pups	39M + 32F = 71	43M + 39F = 82
Ratio (-/-) : (+/-) : wt	0.94 : 2.09 : 0.97	0.92 : 2.05 : 1.03

Clinical Chemistry





Pathology – Organ Mass



References

1. Bentley, D. R. Genomes for medicine. *Nature* **429**, 440-445 (2004).
2. Dunham, A. et al. The DNA sequence and analysis of human chromosome 13. *Nature* **428**, 522-8 (2004).
3. Scherer, S. W. et al. Human chromosome 7: DNA sequence and biology. *Science* **300**, 767-72 (2003).
4. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
5. Dermitzakis, E. T. et al. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* **302**, 1033-5 (2003).
6. Thomas, J. W. et al. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788-93 (2003).
7. Jasny, B. R. & Kennedy, D. The human genome. *Science* **291**, 1153 (2001).
8. Bejerano, G. et al. Ultraconserved Elements in the Human Genome. *Science* (2004).
9. Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720-4 (1995).
10. Yu, Y. & Bradley, A. Engineering chromosomal rearrangements in mice. *Nat Rev Genet* **2**, 780-90 (2001).
11. Zhu, Y. et al. Genomic interval engineering of mice identifies a novel modulator of triglyceride production. *Proc Natl Acad Sci U S A* **97**, 1137-42 (2000).
12. Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-35 (2003).
13. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
14. Kimura-Yoshida, C. et al. Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. *Development* **131**, 57-71 (2004).
15. Kothary, R. et al. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707-14 (1989).
16. Boffelli, D., Nobrega, M. A. & Rubin, E. M. Comparative genomics at the vertebrate extremes. *Nat Rev Genet* **5**, 456-65 (2004).
17. Zerucha, T. et al. A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J Neurosci* **20**, 709-21 (2000).
18. Sauer, B. & Henderson, N. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol* **2**, 441-9 (1990).
19. Hogan B., B. R., Costantini F., Lacy E. *Manipulating the mouse embryo: a laboratory manual* (Cold Spring Harbor Lab. Press, Plainview, New York, 1994).
20. Anderson, J. P. et al. HRC is a direct transcriptional target of MEF2 during cardiac, skeletal, and arterial smooth muscle development in vivo. *Mol Cell Biol* **24**, 3757-68 (2004).

21. Dodou, E., Xu, S. M. & Black, B. L. mef2c is activated directly by myogenic basic helix-loop-helix proteins during skeletal muscle development in vivo. *Mech Dev* **120**, 1021-32 (2003).

1. Bentley, D. R. Genomes for medicine. *Nature* **429**, 440-445 (2004).
2. Dunham, A. et al. The DNA sequence and analysis of human chromosome 13. *Nature* **428**, 522-8 (2004).
3. Scherer, S. W. et al. Human chromosome 7: DNA sequence and biology. *Science* **300**, 767-72 (2003).
4. Venter, J. C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
5. Dermitzakis, E. T. et al. Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* **302**, 1033-5 (2003).
6. Thomas, J. W. et al. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788-93 (2003).
7. Jasny, B. R. & Kennedy, D. The human genome. *Science* **291**, 1153 (2001).
8. Bejerano, G. et al. Ultraconserved Elements in the Human Genome. *Science* (2004).
9. Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720-4 (1995).
10. Yu, Y. & Bradley, A. Engineering chromosomal rearrangements in mice. *Nat Rev Genet* **2**, 780-90 (2001).
11. Zhu, Y. et al. Genomic interval engineering of mice identifies a novel modulator of triglyceride production. *Proc Natl Acad Sci U S A* **97**, 1137-42 (2000).
12. Lettice, L. A. et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-35 (2003).
13. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
14. Kimura-Yoshida, C. et al. Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. *Development* **131**, 57-71 (2004).
15. Kothary, R. et al. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707-14 (1989).
16. Boffelli, D., Nobrega, M. A. & Rubin, E. M. Comparative genomics at the vertebrate extremes. *Nat Rev Genet* **5**, 456-65 (2004).
17. Zerucha, T. et al. A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *J Neurosci* **20**, 709-21 (2000).
18. Sauer, B. & Henderson, N. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol* **2**, 441-9 (1990).
19. Hogan B., B. R., Costantini F., Lacy E. *Manipulating the mouse embryo: a laboratory manual* (Cold Spring Harbor Lab. Press, Plainview, New York, 1994).

20. Anderson, J. P. et al. HRC is a direct transcriptional target of MEF2 during cardiac, skeletal, and arterial smooth muscle development in vivo. *Mol Cell Biol* **24**, 3757-68 (2004).
21. Dodou, E., Xu, S. M. & Black, B. L. mef2c is activated directly by myogenic basic helix-loop-helix proteins during skeletal muscle development in vivo. *Mech Dev* **120**, 1021-32 (2003).