High resolution mapping and positional cloning of ENU-induced mutations in the Rw region of mouse chromosome 5

Yung-Hao Ching1, Robert J Munroe1, Jennifer L Moran2, Anna K Barker1, Evan Mauceli2, Tim Fennell2, Frederica diPalma2, Kerstin Lindblad-Toh2,24, Lindsay M Abcunas1, Joyanna F Gilmour3, Tanya P Harris1, Susan L Kloet1, Yunhai Luo1, John L McElwee1, Weipeng Mu1, Hyo K Park1, David L Rogal1, Kerry J Schimenti1, Lishuang Shen1, Mami Shindo1, James Y Shou1, Erin K Stenson1, Patrick J Stover3, John C Schimenti1*

Abstract

Background: Forward genetic screens in mice provide an unbiased means to identify genes and other functional genetic elements in the genome. Previously, a large scale ENU mutagenesis screen was conducted to query the functional content of a ~50 Mb region of the mouse genome on proximal Chr 5. The majority of phenotypic mutants recovered were embryonic lethals.

Results: We report the high resolution genetic mapping, complementation analyses, and positional cloning of mutations in the target region. The collection of identified alleles include several with known or presumed functions for which no mutant models have been reported (Tbc1d14, Nol14, Tyms, Cad, Fbd15, Haus3), and mutations in genes we or others previously reported (Tapt1, Rest, Ugdh, Paxip1, Hmx1, Otoe, Nsun7). We also confirmed the causative nature of a homeotic mutation with a targeted allele, mapped a lethal mutation to a large gene desert, and localized a spermiogenesis mutation to a region in which no annotated genes have coding mutations. The mutation in Tbc1d14 provides the first implication of a critical developmental role for RAB-GAP-mediated protein transport in early embryogenesis.

Conclusion: This collection of alleles contributes to the goal of assigning biological functions to all known genes, as well as identifying novel functional elements that would be missed by reverse genetic approaches.

Background

Since the human and mouse genomes were sequenced several years ago, it has been recognized that the next major challenge is to uncover the functional content of the genome. This was a motivation behind large scale projects such as ENCODE (Encyclopedia of DNA elements) and KOMP (Knockout Mouse Project; http://www.nih.gov/science/models/mouse/knockout/komp.html). The former used molecular strategies to identify features of human chromatin [1], and the next iteration (modENCODE) is geared to identifying functional genomic elements in vivo using non-vertebrate models http://www.genome.gov/26524507. KOMP uses a reverse genetic approach to identify the functions of all known and annotated mouse protein coding genes, by generating germline or conditional null alleles in embryonic stem (ES) cells.

Forward genetic mutagenesis in mice is most commonly conducted via whole animal mutagenesis with the point mutagen N-ethyl-N-nitrosourea (ENU) [2]. This approach has succeeded in identifying new genes that play important roles in biological processes and in modeling disease [3-8]. Two key advantages over reverse genetics are that: 1) ENU can cause non-null alleles, enabling the dissection of protein function and yielding variants that are more relevant to disease-causing mutations in humans; and 2) as a non-biased approach, genetic elements or genes can be identified that would
never have been implicated \textit{ab initio} to have specific activities or roles.

In multiple organisms, point mutagens have been employed in two basic strategies: random and region-directed. The former has the benefit of scanning the whole genome, whereas the latter allows one to focus resources on a microcosm of the genome. Several years ago, we initiated a larger scale regional screen of proximal mouse Chr 5 spanned by an inversion called rump-white (Rw). This recessive lethal balancer facilitated the isolation and maintenance of mutations recovered in the screen. Predominantly lethal mutations were recovered (a total of 37) \cite{9}. Here, we report on the mapping and cloning of much of this collection. Several mutations were identified in genes for which no models previously existed, providing new biological insights and models of interest. Intriguingly, we also mapped one mutation to a gene desert, in which the nearest protein-coding gene is over 1 megabase away.

\textbf{Results and Discussion}

\textbf{Mutation Mapping}

The rough locations of the \textit{Rw} region ENU-induced mutations were reported by Wilson et al \cite{9}, with most of them being localized to intervals that were too large to allow effective candidate gene selection. Additionally, since many of the mutations were located in common large intervals, allelism amongst them was possible. To map the mutations at higher resolution, we continued and expanded upon two strategies initiated by Wilson et al: deletion mapping and recombination mapping \cite{9}.

In the deletion mapping approach, complementation tests between nearly all the lethals and relevant \textit{Dpp6}, \textit{Hdhi} and \textit{Qdpr} deletions were completed (Figure 1). Various smaller deletions in each complex (e.g. \textit{Dpp6}^\textit{df40}, \textit{Hdhi}^\textit{df99}, \textit{Qdpr}^\textit{df60}, and others shown in Figure 1) were used to better refine locations of noncomplementing lethals. Importantly, complementation of mutations by deletions also provided useful information by excluding large regions. The map positions of \textit{L5cs6}, \textit{15}, \textit{16} and \textit{24} were determined in large part by the fact that they are complemented by both \textit{Dpp6}^\textit{df071} and \textit{Hdhi}^\textit{df71}. Recombination mapping of mutations was performed as described in Methods.

The mutation mapping is summarized in Figure 1. Several of the embryonic lethal mutations mapped near or beyond the distal end of the \textit{Rw} inversion, and these lines were not maintained (most are not shown here, but are presented in \cite{9}). Figure 1 and Table 1 also indicate those cases in which the mutated gene has been identified (or likely so). At present, 16 mutants have been identified from the original 37 that mapped within or near the \textit{Rw} inversion region. Not surprisingly, the mutations cluster to regions that are dense in RefSeq gene annotations (Figure 1). An exception is \textit{L5cs13}, which maps to gene desert (discussed below).

\textbf{Complementation analyses of the mutant collection}

Because ENU mutagenesis is random, some of the mutations may be alleles of the same gene. An allelic series can give insight into protein function. This proved to be the case with \textit{L5cs4} and \textit{L5cs36}, alleles of \textit{Paxip1} (\textit{Ptip}) with differing severity \cite{10}. Furthermore, multiple alleles can solidify evidence that the mutated gene indeed corresponds to the observed phenotype. Accordingly, we evaluated each mutation for potential allelism with the entire mutant collection. Rather than performing complementation tests between all possible pairwise combinations by breeding, we exploited the fact that mutation pairs with non-overlapping map positions cannot be alleles (unless a gene spans adjacent genetic intervals). This greatly reduced the number of potential complementation tests (37 \times \textit{36} = 1332) to a subset that co-localized at the time of analysis. Mutations that went extinct before positional cloning and completion of complementation analyses, and most of those that potentially map distal to the \textit{Rw} inversion, are not included. The results are summarized in Figure 2. In sum, we identified allelism between four pairs of mutations: \textit{Paxip1}^{\textit{L5cs4}} /\textit{Paxip1}^{\textit{L5cs36}} \cite{10}, \textit{Cad}^{\textit{L5cs24}} /\textit{Cad}^{\textit{L5cs27}} (Ching & Schimenti, in preparation), \textit{Nop14}^{\textit{L5cs7}} /\textit{Nop14}^{\textit{L5cs14}}, and \textit{L5cs6} /\textit{L5cs35} (gene not yet identified).

\textbf{Positional Cloning of Mutations}

Once mutations were mapped to intervals under 5 cM, or to regions containing a small number of annotated genes, we initiated searches for causative mutations. The strategies varied during the course of the project as technology advanced. In most cases, annotated genes were prioritized for mutational analysis by considering the following: 1) phenotypes of described knockouts; 2) whether the genes had orthologs in other species with mutant phenotype information; and 3) gene expression patterns from microarrays or EST library origins. Primer pairs were then designed to amplify exonic sequences from the prioritized candidate genes, using genomic DNA of mutant heterozygotes (or homozygous embryonic material in some cases) as templates. PCR products were analyzed by denaturing HPLC (to detect mismatches as heteroduplexes) or direct sequencing. This led to the identification of several mutations summarized below. The causative nature of the identified mutations is relatively confident in cases where two alleles were identified, or when the mutated gene has a similar knockout phenotype. Otherwise, follow-on confirmation of causality will be required by methods such as non-complementation with a targeted or gene-trap allele (as in the case of \textit{Tapt1}; see below).
This perinatal lethal mutation was identified previously as affecting a novel gene, Tapt1KO (Transmembrane anterior posterior transformation 1) [11]. Homozygous mutants exhibit homeotic transformation of the skeleton. To confirm that the point mutation is causative for the phenotype, we performed a complementation test between L5Jcs1 and a mice carrying a deletion allele of Tapt1 (Tapt1KO; see Methods). Intercrosses between Tapt1KO/Rw and L5Jcs1/Rw animals produced litters (N = 31 pups) in which only Rw offspring (21) survived to wean age. The remaining animals (10) died shortly after birth, before the Rw phenotype could be determined. Given that Rw/Rw causes embryonic lethality shortly after gastrulation, the expected number of non-Rw animals for those surviving to wean age is 7, and the lack thereof is significant (Chi square = 10.5; P = .0012). This confirms that the mutation in Tapt1 is causative for the L5Jcs1 lethal phenotype, and indicates that Tapt1L5Jcs1 is a null or severe hypomorph.

L5Jcs5 (Tbc1d14)

This mutation was mapped with a combination of F2 crosses and deletion non-complementation. It resides within the Hdh
df7J deletion region, but the proximal end was better defined by mapping of recombination breakpoints. It was originally identified as a pre-E9.5 embryonic lethal of undefined phenotype [9]. We performed timed matings to better characterize the developmental phenotype. Embryonic development was disrupted at a very early point post-implantation. At E7.5, when WT embryos have progressed to a late primitive streak stage, mutant embryos are smaller and appear arrested at an egg cylinder-like stage (Figure 3a).

L5Jcs5 was mapped to a 1.6 Mb region containing ~17 genes whose knockout phenotypes are not known. Sequencing of several of these genes (not all to completion) revealed only 1 mutation in Tbc1d14. This gene encodes a protein containing a TBC (Tre-2/Bub2/Cdc16) domain characteristic of RAB-GAPs, proteins that activate GTPase activity of RAB proteins [12,13].
RABs are guanine nucleotide binding proteins that mediate membrane-associated protein transport. The mutation causes a SER > GLY change in amino acid 433 (NP_001106833; or 413 for NP_598671), which resides in the TBC1 domain. This amino acid is conserved in mammals (dog; human), zebrafish and chickens. There have been no studies regarding the function of TBC1D14 in cells or animals, however the crystal structure of the human protein has been reported [14]. Interestingly, none of the 39 mouse genes annotated as a RAB-GAP (Gene Ontology term: Rab GTPase activator activity; GO:0005097) has been associated with a mutant phenotype of embryonic lethality. Thus, if indeed the *Tbc1d14* mutation proves to causative for the *L5Jcs5* phenotype, this would be the first implication of a critical developmental role for RAB-GAP-mediated protein transport in early embryogenesis.

L5Jcs7/L5Jcs14 (*Nop14*)

Pooled recombination mapping data of these two non-complementing mutations localized the culprit gene to a 600 kb interval containing 13 RefSeq genes. Mutations in *Nop14* were identified by next-gen sequencing of hybrid-selected exonic sequences (see Methods), and verified by Sanger sequencing (Figure 4).

L5Jcs7 alters the exon 3 splice acceptor sequence (AG|CG >GG|CG) of this 18 exon gene, whereas L5Jcs14 causes a premature stop codon in exon 3. Based on homology to yeast orthologs, this gene is annotated as "nucleolar protein homolog; ribosome biosynthesis." The pre- or peri-implantation lethality caused by these mutations [9] is consistent with such a fundamentally important cellular role.

---

**Table 1 Positional Cloning of Mutations Summary**

| Allele | Gene | Mutation | Genetic Interval | Gene Function | Refs |
|--------|------|----------|-----------------|---------------|------|
| L5Jcs1 | Tapt1 | ** | see ref | Mutations causes homeotic-like skeletal transformations and perinatal lethality. | [11] |
| L5Jcs4 | Paxop1 | ** | see ref | Vasculogenesis; DNA repair; epigenetic regulation | [10,36] |
| L5Jcs5 | Tbc1d14 | missense in TBC1 domain | M4-J57 (1.6 Mb) | TBC1 domain family; putative RAB GTPase activation protein. |
| L5Jcs7 | Nop14 | SA site, exon 3 nonsense, exon 3 | J36-M268 600 kb | Nucleolar protein homolog; ribosome biosynthesis. |
| L5Jcs37 | Rest | T > A (ASN > LYS) | J63-M176 (1.1 Mb) | Thymidylate synthase |
| Dumbo | Hmx1 | ** | see ref | Homeobox-containing gene | [38] |
| L5Jcs24 | Cad | See ref | M251-M334 M387-M353 | carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase. |
| L5Jcs31 | Haus3 | missense LEU > PRO in exon 4 | M388-J88 (1.8 Mb) | Subunit of augmin (HAUS) complex that regulates centrosome and spindle integrity. |
| L5Jcs32 | Fbx5 | missense MET > LYS | M105-J614 | Part of E3 ubiquitin ligase complex that regulates iron homeostasis. | [22,23] |
| L5Jcs8 | Ugdh | * nonsense | D5Ncnp1-J24 (1.3Mb) | UDP-glucose dehyrogenase; GAG biosynthesis; gastrulation. | [39] |
| Ste5Jcs1 | Nsun7 | ** | see ref | Required for normal sperm motility and male fertility. | [40] |
| Deaf5Jcs1 | Otof | ** | see ref | Otoferlin, inner hair cell neurotransmission | [41] |

*Separate alleles reported by others. ** Positional cloning of this allele previously reported. M = D5Mit; J = D5Jcs. Bolded alleles highlight that two alleles were identified. Ching & Schimenti et al will report on these alleles elsewhere. Alleles beginning with "L" are embryonic lethals; timing of lethality was reported previously [9].

---

Figure 2. Complementation and allelism analyses. The grid shows all pairwise comparisons of Rw-region lethal mutations for potential allelism. Original mutations located distal to Rw are not included, nor is L5Jcs17, which went extinct before mapping and allelism testing could be completed. Actual complementation tests were, in most cases, only performed when mutations co-mapped to overlapping genetic intervals (dark green or red). The majority of mutations were mapped to sufficient resolution to indirectly conclude they are not allelic (light green).

---
**L5Jcs11 (Tyms)**

*L5Jcs11* is an early postimplantation mutant overtly defective in gastrulation [9]. It was mapped genetically to a ~1.1 Mb interval by a combination of deletion mapping (*Dpp6 df4J* complemented *L5Jcs11*, defining the proximal end as being distal to *D5Jcs63*) and recombination mapping (defining the distal end of the critical region). Sequencing of candidate genes in the region identified a T > A transversion in the third nucleotide of the codon encoding amino acid 106 of TYMS. This causes a predicted asparagine to lysine change. The structure and function of thymidylate synthase has been characterized extensively in various eukaryotes. The protein is a symmetric dimer, with each subunit ranging from 30 to 35 kDa depending on the organism. However, in protozoa and plants, TYMS and DHFR are produced on the same polypeptide (reviewed in [16]).

The main role of TYMS is to convert deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP) using a 5,10-methylenetetrahydrofolate cofactor. dTMP is essential for the synthesis of DNA but not RNA, and therefore it is found at especially high levels in cells undergoing rapid cell division. Consequently, anti-TYMS drugs are used in cancer treatment, as are other chemicals that inhibit this pathway (such as methotrexate, a DHFR inhibitor) [17]. TYMS inhibitor drugs primarily target the dUMP substrate or the folate binding site on TYMS.

Since ASN106 is not specified as being a residue essential for TYMS function [18], we sought additional evidence that the mutation is responsible for the drastic phenotype of *L5Jcs11*. TYMS autoregulates its levels by binding its mRNA to decrease translation [19]. Consistent with a decrease in functional TYMS in cells bearing this allele, Western blots revealed that *L5Jcs11/+* heterozygotes have ~2.6 fold elevated TYMS protein in the liver relative to WT mice (Figure 3b). Mutant heterozygotes also exhibited 2.3 fold lower DHFR protein levels (Figure 3b). ASN106 is located close to the folate binding motif of TYMS, suggesting that the enzyme is defective in binding or effectively utilizing the major substrate N5, N10-Methylene H4 folate.

To further explore the effects of this mutation on mice, we performed inner cell mass (ICM) outgrowth assays. As shown in Figure 3c, although trophoblast cells from cultured blastocysts could grow, consistent with the ability to implant, mutant ICMs were unable to outgrow from hatched embryos. This indicates that the rapid cell division and DNA synthesis occurring in...
embryonic cells cannot be supported without normal TYMS function, consistent with failed postimplantation development of the embryo proper.

**L5Jcs31 (Haus3)**

This mutation was found by high-throughput sequencing of hybrid-selected exon DNA. It is a T > C transition resulting in a LEU > PRO change in the fourth of 5 exons. Currently, Haus3 is annotated as residing in the first intron of the *Poln* gene, in the same transcriptional orientation. However, this may be artifactual or not biologically relevant, since the putative exon 1 of *Poln* is: 1) non-coding; 2) shared with Haus3; 3) present in only 1 of >120 spliced ESTs aligned in the UCSC browser; 4) absent from human *POLN*; and 5) absent from other mammalian *Haus3* genes (Figure 5). However, the recovery of “full-length” cDNA clones indicates that *Poln* mRNAs can be produced that contain this mRNA, as do many Haus3 transcripts (ESTs not shown). It remains to be seen if these two genes share a promoter/enhancer, or even a common transcriptional start site.

Haus3 is a component of the multiprotein HAUS complex, homologous to the *Drosophila* Augmin complex. In mammalian cells, HAUS regulates centrosome and spindle integrity. Its disruption causes destabilization of kinetocore microtubules and centrosome disruption, and thus is critical for genome stability [20]. Interestingly, Haus3 was found to be mutated early in the development of a subset of lobular breast cancers [21]. Though we haven’t noted any obvious defects in heterozygous animals, this model might be useful for investigating the effects on breast cancer frequency or progression in tumor susceptible backgrounds. Homozygosity for L5Jcs31 causes pre- or peri-implantation lethality, indicating the essential nature of the augmin complex in mammalian cell growth or early development.

**L5Jcs32 (Fbxl5)**

This mutation was identified as a midgestation recessive lethal showing neural tube defects [9]. The allele was mapped by virtue of non-complementation with the *Hdhdf9J* deletion. Sequencing of candidate genes in the region led to the identification of a T to A transition in exon 3 of Fbxl5, resulting in the replacement of methionine by lysine at amino acid position 127.

Fbxl5 (F-box and leucine-rich repeat protein 5) is part of an SCF (SKP1-cullin-F-box) ubiquitin ligase complex that plays a key role in regulating iron homeostasis [22,23]. It serves as a sensor of iron and oxygen levels by tethering the iron regulatory proteins IRP1/2 to an E3 ligase complex for ubiquitination and proteasome degradation, or alternatively, increasing transferrin receptor transcription and inhibiting ferritin production under conditions of low intracellular iron. The mutation resides in the conserved iron-binding hemerythrin domain of Fbxl5, changing a highly conserved (invariant among vertebrates) MET immediately following a key histidine residue that is involved in coordinating Fe [22]. We postulate that this compromises or eliminates iron binding by the protein, thus disrupting its sensor function. Our findings that mutation of *Fbxl5* causes embryonic lethality underscores the known critical nature of iron homeostasis during development (reviewed in [24]).

The **L5Jcs13 embryonic lethal allele maps to a gene desert**

This mid-gestation embryonic lethal mutation was recombination-mapped to a 289 kb interval between SNP markers rs13478279 (60,593,878) and rs13478280 (60,882,847). No protein-coding genes or spliced ESTs are annotated in this region (Figure 6), nor are any known microRNAs. There appears to be an intronless, ORF-containing transcription unit (represented by two ESTs), with homology to transcripts in humans and weasel (Figure 6). However, no mutations were detected in the region spanned by the ORF in the L5Jcs13 allele (data not shown). We also sequenced an apparent processed pseudogene with homology to human POGK, and a conserved region corresponding to an unspliced EST EL606540, but again no mutations compared to the parental B6 sequence were found. Finally, approximately 59% of the critical region was scanned for mutations by shotgun sequencing of pooled PCR amplimers from L5Jcs13/Rw genomic DNA (see Methods). No de-novo mutations were found.
Remarkably, the L5Jcs13 critical region (289 kb) lies within a much larger gene desert that contains only 1 RefSeq gene (Pcdh7) in 7.7 Mb. The nearest gene is G6pd2, located 1.3 Mb distally, which marks the beginning of a gene rich region. The current data raises the possibility that the L5Jcs13 mutation affects a long range regulatory sequence, as appears to be the case with an ENU-induced allele of the quaking gene that is located 40-640 kb upstream [25], or the callipyge point mutation in sheep that affects transcription of genes up to several hundred kilobases away [26-28]. Alternatively, the mutation could disrupt the local chromatin structure, or interaction of chromatin with the nuclear matrix, in a manner that affects the expression of other genes. Finally, it is possible that there is an uncharacterized non-coding RNA in the critical region that is mutated. There is one reported case of a mouse ENU-induced mutation in a microRNA gene causing progressive hearing loss [29].

SteL5Jcs3 maps to a region lacking mutated genes with mutations or implicated spermatogenesis roles

This spermiogenesis mutation was localized to a 1.6 Mb region (Figure 1) containing 4 annotated genes: Tbc1d19, Rbpj, Cekar and Stim2. We sequenced all of these (cDNA or exons from genomic DNA) and found no mutations. Additionally, the latter 3 have been mutated by gene targeting to produce phenotypes not consistent with specific male infertility. CCKAR-deficient mice are viable with no fertility defects [30,31]. Mutation of Rbpj causes embryonic lethality. Stim2 null mice are slightly runted and die by 4-5 weeks of age [32]. That study reported T cell defects but did not address germ cell development. Tbc1d19 appears to be widely expressed based on EST library representation (Unigene), and, like Tbc1d14 (mutated in L5Jcs5), encodes a likely Rab-Gap containing the TBC domain. Further sequencing analyses will be required to identify the mutation and determine if it affects a novel element, the expression of any of these genes, or the expression of genes located outside the genetically defined interval.

Conclusions

This report summarizes work on a region-specific saturation mutagenesis project that was initiated before the mouse genome was sequenced. We reasoned that such regional screens (using deletions or balancer inversions such as Rw) provided powerful advantages for phenotype characterization and gene identification, including simplified stock maintenance using a single visibly marked balancer chromosome, and the knowledge of a mutation’s location. Coupled with nested deletions generated in the same region [33], it was possible to rapidly sublocalize mutations to regions upon which to concentrate recombination mapping and guide efficient complementation analyses. The sum of this work was the molecular identification of 16 mutant alleles representing 13 genes. For 10 of these genes, the mutants are/were the first reported in mice (Table 1), and 6 are presented here for the first time. The remaining mutations have been mapped to relatively small intervals.

Perhaps the most interesting mutation thus far is L5Jcs13, which is a recessive embryonic lethal allele mapping to a gene desert. Clearly, this region contains a sequence element of crucial importance to mouse development, and close scrutiny of this region for evidence of cryptic genes (using ab initio gene prediction programs), highly conserved elements or transcription units have not revealed any candidate elements that contain mutations (data not shown). In an era where all protein-coding genes will soon be knocked out in ES cells by the KOMP project, the identification of apparently novel, non-genic, essential genomic elements such as L5Jcs13 underscores the continuing value of unbiased forward genetic screens.

The drawback of forward genetic screens in mice has been the cost and time involved in identifying causative mutations. For genome-wide screens, genetic crosses must be conducted in order to localize mutations, followed by traditional positional cloning efforts guided by candidate gene prioritization. Region-specific screens ameliorate the mapping part, and in one study the approach was taken to directly resequence all exons (and their flanking sequences) in the target region to identify causative mutations [4]. Potentially causative mutations were identified in 31 of the 41 mutant lines examined, and most of the potentially causative
mutations were in non-coding regions within or near the transcription unit. The failure to identify the other 10 mutations amongst annotated genes, and the cases of L5Jcs13 and Ste5Jcs5, indicate that there is much we don’t know about the functional content of the genome. Furthermore, we must caution that for the potentially causative mutations presented here and in other studies such as the aforementioned, functional validation must be obtained. This caveat also underscores the value of genetic mapping, which contributes to validation of a mutation identified by sequencing.

The advent of high-throughput DNA sequencing technologies can be transformative for forward genetics in mice. Here, we identified three mutations by Solexa/Illumina resequencing of exon-enriched DNA from critical regions. However, the rapidly declining cost of whole genome sequencing will soon negate the benefit of such sequence enrichment, or even genetic mapping in advance of sequencing (bearing in mind the issue of validation mentioned above). Additionally, whole genome sequencing has the important advantage of enabling the identification of non-coding regulatory elements.

Methods

Sequence capture and next generation sequencing

A custom Agilent oligonucleotide array was designed to sequence exons, UTRs and promoters of all RefSeq genes in non-recombinant intervals of several mutations (from this and unrelated projects). The array included all RefSeq genes between D5Mit388 and D5Mit268 (mm8; NCBI Build 36). Solution-based hybrid capture of L5Jcs14/+B6, L5Jcs31/+B6 and C57BL/6J genomic DNAs to transcribed RNA oligonucleotide baits was performed as described [34]. Libraries made from captured DNA were sequenced using 51 or 76 bp paired-end reads with an Illumina Genome Analyzer. Candidate single nucleotide changes were identified by 1) aligning high quality sequence reads to the reference C57BL/6J mm8 genome sequence, 2) identification of single nucleotide polymorphisms (SNPs) between each sample against the reference, and 3) filtering for heterozygous SNPs unique to the mutant sample. In addition to Sanger sequencing, genotyping of L5Jcs14 SNPs with the Sequenom iPLEX technology was performed to confirm the L5Jcs14 mutation.

For L5Jcs13, 70 primer pairs were developed to amplify the 289 kb critical region. The products averaged a length of 4.5 kb and overlapped approximately 400 bp on each end. Amplification was initially attempted on all products using Roche Expand Long Range Polymerase or iProof Polymerase (Bio-Rad). Out of the 70 primer pairs, 51 showed some level of successful amplification. Amplimers were treated with FastAP Alkaline Phosphatase plus Exonuclease I, pooled, then sequenced on an Illumina GA genome analyzer using chemistry that generated 43 nucleotide, single-end reads. The sequence was aligned to C57BL/6J (“B6”) genomic sequence using Novoalign v2.07 and Stampy 1.0.9. Further comparison was done with the C3H/HeJ reference sequence from the Sanger Institute. Data was analyzed for single nucleotide polymorphisms (SNPs), insertions, and deletions using GATK v 1.4418 (The Genome Analysis Toolkit; http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit) and VarScan v 2.2.3 http://sourceforge.net/projects/varscan/. Candidate SNPs were compared to the NCBI SNP database dbSNP, and ultimately re-sequenced by standard methods to resolve whether a nucleotide difference was a SNP, a mutation, or a database or sequencing error.

Generation of Tapt1 mice

ES cells containing a deletion of the entire Tapt1 locus were obtained from the KOMP Repository (clone name Tapt1_AG12, produced by Regeneron, Inc.). The targeted cells were of strain C57BL/6N, and the official allele name is Tapt1tm1(KOMP)Vic. Cells were microinjected into albino CD1 blastocysts to produce chimeras.

Western blot analysis

Snap-frozen liver samples were lysed in 10 mM Tris, pH 7.4, 150 mM NaCl, 5 mM EDTA, 5 mM DTT, 1% Triton X-100, and Mammalian Protease Inhibitor Cocktail (Sigma). Tissue lysates were loaded onto 12% SDS-PAGE gels with 25 μg total protein/lane; protein concentrations were determined using the Lowry-Bensadoun method [35]. Proteins were then transferred to an Immobilon-P PVDF Membrane (Millipore). The membrane was blocked overnight at 4°C in phosphate-buffered saline with 10% nonfat dry milk and 1% NP40. The membrane was incubated for two hours at room temperature in one of two primary antibodies: 1:1000 -anti-TYMS (Zymed) or 1:2000 anti-DHFR (Sigma). After four washes (phosphate buffered saline with 0.1% Tween-20) of 10 minutes each, the membranes were incubated for 1 hour in 1:10,000 HRP-conjugated goat anti-mouse antibody for TYMS or 1:20,000 HRP-conjugated goat anti-rabbit antibody for DHFR (Pierce). After four washes of 10 minutes each, membranes were developed in SuperSignal West Pico Chemiluminescent Substrate (Pierce) and films exposed. As a loading control, membranes were also probed with 1:100,000 HRP-conjugated anti-beta actin (Abcam). Quantification of protein bands was done by digitizing the films, then analyzing the images with ImageJ software (NIH).
**Blastocyst Outgrowths**

Blastocysts (E3.5) were plated onto gelatinized tissue culture dishes, cultured for 5 days in ES cell medium, and harvested for genotyping as described previously [10].

**Mutation Mapping and screening**

For recombination mapping, \(Rw/Ls/cs#\) animals were crossed to \(C5\) and \(C5\) (a strain in which the proximal portion of Chr 5 from \(M\) was rendered partially congenic in strain C3HeB/FeJ), and non-

\(Rw\) progeny (\(Ls/cs#\)) were intercrossed to produce \(F2\)'s. Genomic DNA isolated from tails from \(F2\) animals were genotyped with polymorphic markers within the \(Rw\) region to identify recombinants. The locations of the mutations were determined indirectly, in that we excluded intervals in which a mutation resides by virtue of being able to obtain homozygosity for the parental \(B6\) allele in live offspring.

Markers not in the MGD database are:
- \(D5Jcs640\) GCCAGGTATAACAGCTCCA and TCTCCCTTCTTCTCTCTTCCTTC
- \(D5Jcs85\) GGGCTTTAGACGAGCAGAG and GGTTTACAAGGAGAAGGTTCTC
- \(D5Jcs24\) AACATGTCAGGGCGCAAGAG and TGTCGTTCCATTATATGGC

Mutation screening by dHPLC was performed on a Transgenomic Wave machine, designed to detect heteroduplexes. The screening was generally performed on heterozygous DNAs. Otherwise, amplified DNAs were sequenced by standard Sanger protocols in an automated sequencer.

All experiments using mice were conducted with approval of Cornell's Institutional Animal Care and Use Committee, protocol # 2004-0038.

**Acknowledgements**

This work was support by NIH grant HD35984 and a NHGRI medical sequencing grant to the Broad Institute. We thank Lauren Imbroglio, Stacy Gabriels, Ross Savford. Jeremy Johnson and the Broad Institute's Genetic Analysis Platform and Genome Sequencing Platform for SNP genotyping, hybrid capture and illumina sequencing.

**Authors' information**

 AKB, LMA, JYS, HKP, EKS, DLR and SLK were undergraduates at Cornell when they conducted the research.

Received: 1 October 2010 Accepted: 30 November 2010

Published: 30 November 2010

**References**

1. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al: Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007, 447(7146):799-816.

2. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A: Mouse ENU mutagenesis. Hum Mol Genet 1990, 8(18):1955-1963.

3. Herron BJ, Lu W, Rao C, Liu S, Peters H, Bronson RT, Justice MJ, McDonald JD, Beier DR: Efficient generation and mapping of recessive developmental mutations using ENU mutagenesis. Nature genetics 2002, 30(2):185-189.

4. Bolcs MK, Wilkinson BM, Wilming LG, Liu B, Probst FJ, Harrow J, Graafh D, Hentges KE, Woodward LP, Maxwell A, et al: Discovery of candidate disease genes in ENU-induced mouse mutants by large-scale sequencing, including a splice-site mutation in nucleoredoxin. PLoS genetics 2009, 5(12):e1000759.

5. Shima N, Hartford S, Duffy T, Wilson L, Schimenti K, Schimenti JC. Phenotype based identification of mouse chromosome instability mutants. Genetics 2003, 163:1031-1046.

6. Cordes SP: N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express. Microbial Mol Biol Rev 2005, 69(3):426-439.

7. Vitaterna M, King D, Chang A, Kornhause rJ, Lowrey PL, McDonald J, Dove W, Pinto L, Turek F, Takahashi J. Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. Science (New York, NY) 1994, 264:719-725.

8. Acevedo-Arozena A, Wells S, Potter P, Kelly M, Cox RD, Brown SD. ENU mutagenesis, a way forward to understand gene function. Annu Rev Genomics Hum Genet 2008, 9:49-69.

9. Wilson L, Ching YH, Fanias M, Hartford S, Howell G, Shao H, Bucan M, Schimenti J. Random mutagenesis of proximal mouse Chromosome 5 uncovers predominantly embryonic lethal mutations. Genome research 2005, 15(8):1095-1105.

10. Mu W, Wang W, Schimenti JC. An allelic series uncovers novel roles of the BIRC domain-containing protein PTP1B in mouse embryonic vascular development. Mol Cell Biol 2008, 28(20):6439-6451.

11. Howell GR, Shindo M, Murray S, Gridley T, Wilson LA, Schimenti JC. Mutation of a ubiquitously expressed mouse transmembrane protein (Tapt1) causes specific skeletal homeotic transformations. Genetics 2007, 175(2):699-707.

12. Ishibashi K, Kanno E, Itoh T, Fukuda M: Identification and characterization of a novel Tre-2/Bub2/Cdc16 (TBC) protein that possesses Rab3A-GAP activity. Genes Cells 2009, 14(1):41-52.

13. Richardson PM, Zon LI: Molecular cloning of a cDNA with a novel domain present in the tre-2 oncogene and the yeast cell cycle regulators BUB2 and cdc16. Oncogene 1995, 11(6):1139-1148.

14. Tempel W, Tong Y, Dimov S, Bockxvear A, Park H: First crystallographic models of human TBC domains in the context of a family-wide structural analysis. Proteins 2008, 71(1):497-502.

15. Liu PC, Drieje DJ: Novel stress-responsive genes EGM1 and NOP14 encode conserved, interacting proteins required for 40S ribosome biogenesis. Molecular biology of the cell 2001, 12(11):3644-3657.

16. Carreras CW, Santi DV: The catalytic mechanism and structure of thymidylate synthase. Annual review of biochemistry 1995, 64:721-762.

17. Costi MP, Ferrari S, Venturelli A, Calo S, Tonidi D, Barlaco D: Thymidylate synthase structure, function and implication in drug discovery. Curr Med Chem 2005, 12(19):2241-2258.

18. Perry KM, Fauman EB, Finer-Moore JS, Montfort WR, Maley GF, Maley F, Stoud RM: Plastic adaptation toward mutations in proteins: structural comparison of thymidylate synthases. Proteins 1990, 8(4):315-333.

19. Chu E, Koeller DM, Casey JL, Drake JC, Chabner BA, Elwood PC, Zinn S, Allegra CJ: Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase. Proceedings of the National Academy of Sciences of the United States of America 1991, 88(20):8977-8981.
20. Lawo S, Bashkurov M, Mullin M, Ferriera MG, Kittler R, Habermann B, Tagliaferro A, Poser I, Hutchins JR, Hegemann B, et al. HAUS, the 8-subunit human Avgumin complex, regulates centrosome and spindle integrity. Curr Biol 2009, 19(10):816-826.
21. Shah SP, Morin RD, Khatri J, Princete L, Pugh T, Burleigh A, Delaney A, Gelmon K, Gullany R, Seriz J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009, 461(7265):809-813.
22. Salahudeen AA, Thompson JW, Ruiz JC, Ma HW, Kinch LN, Li Q, Grishin NV, Bruck RK. An E3 ligase possessing an iron-responsive hemerythrin domain is a regulator of iron homeostasis. Science (New York, NY) 2009, 326(5953):722-726.
23. Vashti AA, Zumbrennen KB, Huang X, Powers DN, Durazo A, Sun D, Bhaskaran N, Persson A, Uhlen M, Sangfelt O, et al. Control of iron homeostasis by an iron-regulated ubiquitin ligase. Science (New York, NY) 2009, 326(5953):718-721.
24. Kambe T, Weaver BP, Andrews GK. The genetics of essential metal homeostasis during development. Genesis 2008, 46(4):214-228.
25. Noveroske JK, Hardy R, Dapper JD, Vogel H, Justice MJ. A new ENU-induced allele of mouse quaking causes severe CNS dysmyelination. Mamm Genome 2005, 16(9):672-682.
26. Freking BA, Murphy SK, Wyley AA, Rhodes SJ, Keele JW, Leymaster KA, Jintle RL, Smith TP. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome research 2002, 12(10):1496-1506.
27. Smit M, Segers K, Carrascosa LG, Shay T, Baraldi F, Gyapay G, Snowder G, et al. Essential role of glycosaminoglycans in maintenance of body weight. J Clin Invest 1999, 103(4):8119-8124.
28. Lewis MA, Quint E, Glazer AM, Fuchs H, De Angelis MH, Langford C, van Dongen S, Abreu-Goodger C, Ripani M, Redshaw N, et al. An ENU-induced mutation of mir-96 associated with progressive hearing loss in mice. Nature genetics 2009, 41(5):614-618.
29. Takiguchi S, Suzuki S, Sato Y, Kanai S, Miyasaka K, Jimi A, Shinozaki H, Takata Y, Funakoshi A, Kono A, et al. Role of CCK-A receptor for pancreatic function in mice: a study in CCK-A receptor knockout mice. Pancreas 2002, 24(3):276-283.
30. Kopin AS, Mathes WF, McBride EW, Nguyen M, Al-Haidar W, Schmiz F, Bonner-Weir S, Kanarek R, Beinborn M. The cholecystokinin-A receptor mediates inhibition of food intake yet is not essential for the maintenance of body weight. J Clin Invest 1999, 103(3):363-391.
31. Oh-Hora M, Yamashita M, Hogan PG, Sharma S, Lamperti E, Chung W, Prakriya M, Feske S, Rao A. Dual functions for the endoplasmic reticulum calcium sensors STIM1 and STIM2 in T cell activation and tolerance. Nat Immunol 2008, 9(4):432-443.
32. Schimenti JC, Libby B, Bergstrom R, Wilson L, Naf D, Tarantino L, Avizadel A, Lengeling A, Bucan M. Interdigitated deletion complexes on mouse chromosome 5 induced by irradiation of embryonic stem cells. Genome Res 2000, 10(1):85-90.
33. Gnjirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature biotechnology 2009, 27(2):182-189.
34. Bensadoun A, Weinstein D. Assay of proteins in the presence of interfering materials. Anal Biochem 1976, 70(1):241-250.
35. Munroe RJ, Prabhu V, Acland GM, Johnson KR, Harris BS, OBrien TP, Welsh IC, Noden DM, Schimenti JC. Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. BMC Dev Biol 2009, 9:27.
36. Garcia-Garcia MJ, Anderson KV. Essential role of glycosaminoglycans in Fgf signaling during mouse gastrulation. Cell 2003, 114(6):727-737.
37. Harris T, Marquez B, Suarez S, Schimenti J. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009, 461(7265):809-813.
38. Munroe RJ, Prabhu V, Acland GM, Johnson KR, Harris BS, OBrien TP, Welsh IC, Noden DM, Schimenti JC. Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. BMC Dev Biol 2009, 9:27.