Whole exome sequencing of ENU-induced thrombosis modifier mutations in the mouse

Short title: Identifying thrombosis modifier genes

Kärt Tomberg[1,#a], Randal J. Westrick[2], Emilee N. Kotnik[3], David R. Siemieniak[3,4], Guojing Zhu[3], Thomas L. Saunders[5,6], Audrey C. Cleuren[3], and David Ginsburg[1,3,4,5,7,*]

[1]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, United States of America
[2]Department of Biological Sciences, Oakland University, Rochester, Michigan, United States of America
[3]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, United States of America
[4]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan, United States of America
[5]Departments of Internal Medicine, Division of Molecular Medicine and Genetics, University of Michigan, Ann Arbor, Michigan, United States of America
[6]Transgenic Animal Model Core Laboratory, University of Michigan, Ann Arbor, Michigan, United States of America
[7]Departments of Pediatrics, University of Michigan, Ann Arbor, Michigan, United States of America

[#a]Current Address: Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

* Corresponding author

Email: ginsburg@umich.edu

Abstract

Although the Factor V Leiden (FVL) gene variant is the most common genetic risk factor for venous thrombosis, only 10% of these individuals will experience such an event in their lifetime. In order to identify potential FVL modifier genes contributing to this incomplete penetrance, we performed a sensitized dominant ENU mutagenesis screen, based on the perinatal synthetic lethal thrombosis previously observed in mice homozygous for FVL ($F5^{L/L}$) and haploinsufficient for tissue factor pathway inhibitor ($Tfpi^{+/-}$). Whole exome sequencing was applied to DNA from 107 viable $F5^{L/L}$ $Tfpi^{+/-}$ progeny ('rescues') to identify candidate genes that are enriched for ENU mutations. A total of 3481 potentially deleterious candidate ENU variants were identified in 2984 genes. After correcting for gene size and multiple testing, *Arl6ip5* was identified as the most enriched gene but did not meet genome-wide significance and validation of the top 6 genes using independent CRISPR/Cas9 induced mutations failed to demonstrate clear rescue phenotype. Linkage analysis conducted in 3 largest pedigrees generated from the rescue mice using ENU-induced coding variants as genetic markers did not map the corresponding suppressor loci. However, in one of the pedigrees, a maternally inherited (not ENU-induced) *de novo* mutation ($Plcb4^{R335Q}$) exhibited significant co-segregation with the rescue phenotype (p=0.003).

Introduction

Venous thromboembolism (VTE) affects 1:1000 individuals in the US each year and is highly heritable [1, 2]. The most common known genetic risk factor for VTE is a single nucleotide variant (SNV) in the *F5* gene, referred to as Factor V Leiden (FVL, Arg506Gln) [3]. Although the FVL variant is present in ~25% of VTE patients [4], only ~10% of individuals heterozygous for FVL develop thrombosis in their lifetime.

To identify genetic variants potentially modifying the FVL, we recently employed a dominant ENU screen [5] in mice sensitized for thrombosis. Mice homozygous for the FVL mutation ($F5^{L/L}$) and haploinsufficient for tissue factor pathway inhibitor ($Tfpi^{+/-}$) die of perinatal thrombosis [6]. After ENU mutagenesis, 98 G1 $F5^{L/L}$ $Tfpi^{+/-}$ progeny survived to weaning ('rescues') and 16 of them exhibited successful transmission of the ENU-induced suppressor mutation. However, subsequent efforts to genetically map the corresponding suppressor loci were confounded by complex strain-specific differences introduced by the required genetic outcross [5]. Similar genetic background effects have complicated previous mapping efforts [7] and have been noted to significantly alter other phenotypes of interest [8, 9]. Additional challenges of such mapping approach include the requirement for large pedigrees and limited mapping resolution, with candidate intervals typically harboring tens to hundreds of genes and multiple closely linked mutations.

The emergence of high throughput sequencing methods has greatly enabled the direct identification of ENU-induced mutations and removed the necessity for outcrossing to introduce genetic markers for mapping [10, 11]. We now report the application of this approach to a total cohort of 107 rescue mice (including 50 mice from the previously reported ENU screen). Mutation burden analysis of whole exome sequencing (WES) in this cohort identified mutations in *Arl6ip5* as the top candidate suppressor locus, but an independent CRISPR/Cas9-generated *Arl5ip5* mutant allele did not demonstrate a clear rescue of $F5^{L/L}$ $Tfpi^{+/-}$ lethal phenotype. In addition, a maternally inherited (not ENU-induced) *de novo* mutation ($Plcb4^{R335Q}$) exhibited significant co-segregation with the rescue phenotype (p=0.02) in an expanded pedigree.

Results and discussion

$F5^{L/L}$ $Tfpi^{+/-}$ mice exhibit reduced survival and smaller size

A previously described [5] sensitized ENU-mutagenesis for thrombosis modifiers was extended to screen for dominant suppressors of the perinatal lethal $F5^{L/L}$ $Tfpi^{+/-}$ genotype (Fig 1A). 2834 G1 offspring were generated from mutagenized C57BL/6J $F5^{L/L}$ males crossed to unmutagenized C57BL/6J $F5^{L/+}$ $Tfpi^{+/-}$ females, with a total of 76 viable $F5^{L/L}$ $Tfpi^{+/-}$ rescue progeny identified at weaning. Approximately 50% of the rescue mice died by 6 weeks of age, with no significant survival difference observed between females and males (Fig 1B, p=0.077). Females were underrepresented compared to males during the initial genotyping (28 females compared to 48 males, p=0.022). $F5^{L/L}$ $Tfpi^{+/-}$ rescues were on average 25-30% smaller than their littermates at 2-3 weeks of age ($p=7x10^{-13}$, Fig 1C-D). The proportion of identified rescues among G1

offspring, their smaller weight compared to littermates, and fewer overall females among rescues, were all consistent with our previous report [5].

WES identifies 6771 ENU-induced variants in 107 rescues

In order to identify all/most exonic ENU mutations, a total of 107 G1 rescues (57 from the current ENU screen and an additional 50 rescues from the previous screen [5]) were subjected to WES (S1 Table). From ~1.5 million initially called variants, 6735 SNVs and 36 insertions-deletions (INDELs) within exonic regions were identified as potential ENU-induced mutations, using an in-house filtering pipeline (see Materials and methods). The most common exonic variants were nonsynonymous SNVs (47%), followed by mutations in 3' and 5' untranslated regions (31%) and synonymous SNVs (15%). The remaining variants (7%) were classified as splice site altering, stoploss, stopgain, or INDELs (Fig 2A). T/A -> C/G (47%), and T/A -> A/T (24%) SNVs were overrepresented, while C/G -> G/C (0.8%) changes were greatly underrepresented (Fig 2B), consistent with previously reported ENU studies [12, 13]. Since ENU is administered to the G0 father of G1 rescues, only female progeny are expected to carry induced mutations on the X chromosome, while males inherit their only X chromosome from the unmutagenized mother. Among the called variants, all chromosomes harbored a similar number of mutations in both sexes, with the exception of the X chromosome where a >35 fold increase in SNVs per mouse was observed in females (Fig 2C). The average number of exonic ENU mutations for G1 rescues from the current and previous screens was ~65 SNV per mouse (Fig 2D), consistent with expected ENU mutation rates [10,

13]. Sanger sequencing validated 36 of the 37 variants in 12 chosen genes (S2 Table). These data suggest that most called variants are likely to be of ENU origin.

Mutation burden analysis identifies potential candidate thrombosis suppressor genes

WES data for 107 independent rescue mice were jointly analyzed to identify candidate genes that are enriched for potentially deleterious ENU-induced variants including missense, nonsense, frameshift, and splice site altering mutations (3481 variants in 2984 genes, S2 Table). The majority of genes harbored only a single ENU-induced variant with 15 SNVs identified in *Ttn*, the largest gene in the mouse genome (Fig 2E). After adjusting for coding region size and multiple testing (for 2984 genes), the ENU-induced mutation burden was significantly greater than expected by chance for 3 genes (FDR<0.1, *Arl6ip5*, *Itgb6*, *C6*) and suggestive for 9 additional genes (FDR<0.25). Yet a more stringent multiple testing approach accounting also for genes that did not get hit by ENU in those 107 mice (20,586 genes used in the simulation) suggested no significant enrichment (*Arl6ip5* FDR=0.68, Fig 3). 107 rescue mice (1.9-2.7% of expected mice, Fig 1A) represent approximately 900-1400 (25%) tested $F5^{L/L}$ $Tfpi^{+/-}$ conceptuses. Assuming ~65 exonic mutations per mouse (Fig. 2D), this correlates with approximately 3-4X genome coverage which is further reduced when considering only potentially damaging mutations. The power to map the underlying genes is further compromised by the previously reported background survival rate for the rescue mice [6].

Simultaneous targeting of 6 genes using CRISPR-Cas9 resulted in >100 new alleles.

To test if the above-identified candidates would validate as thrombosis suppressor genes, independent null alleles were generated with CRISPR-Cas9. From 294 microinjected zygotes with pooled guide RNAs against top six candidate genes (*Arl6ip5, C6, Itgb6, Cpn1, Sntg1 and Ces3b;* Fig 3) we obtained 39 progeny. CRISPR-Cas9 genome editing was assed by Sanger sequencing around the sgRNA target sites. Approximately 190 independent targeting events were observed across the 6 genes in 36 mice including small INDELs, single nucleotide changes, and several large (>30bp) deletions or inversions. Targeted alleles were either homozygous, heterozygous, or mosaic. While the number of editing events varied greatly for different sgRNAs (2.5-85%) the strategy to simultaneously target multiple genes [14] proved successful and cost-effective.

Independent alleles for the candidate genes did not validate as thrombosis suppressors

Two or more different CRISPR-Cas9-induced alleles for each of the candidate genes (S3 Table) were bred to isolation but maintained on $F5^{L}$ background for validation. The progeny of $F5^{L/L}$ $Tfpi^{+/+}$ mice crossed with $F5^{L/+}$ $Tfpi^{+/-}$ mice (one additionally carrying the CRISPR-Cas9-induced allele) were monitored for survival of $F5^{L/L}$ $Tfpi^{+/-}$ offsprings (Table 1, S4 Table).

Table 1. Validation of survival with CRISPR-Cas9-induced alleles

Gene	Total mice tested	No. of rescues w/o allele	No. of rescues with allele	P-value*
Arl6ip5	179	1	5	0.36
Itgb6	144	1	1	1
C6	105	0	0	1

Cpn1	122	0	1	0.46
Sntg1	210	3	4	0.68
Ces3b	219	2	1	1

*Fisher's exact test

Over 100 progeny were generated for each of the candidate genes with no obvious rescue effect. A slight increase in rescues carrying $F5^{L/L}$ $Tfpi^{+/-}$ $Arl6ip5^{+/-}$ genotype was noted, although after surveying 179 mice it remained non-significant (Table 1). Still, we cannot exclude the possibility of our top candidate gene rescuing the lethal phenotype at lower penetrance (<55%). Our previously mapped thrombosis modifier *F3* also rescued the phenotype with incomplete penetrance [5]. For most of the other candidate genes, the number of observed $F5^{L/L}$ $Tfpi^{+/-}$ mice reflected the expected background survival rate for this genotype (2 $F5^{L/L}$ $Tfpi^{+/-}$ mice from 94 total) [6]. Slightly higher numbers of rescues was observed for the *Sntg1* gene, both with and without the *Sntg1* loss of function allele. As 6 of the seven $F5^{L/L}$ $Tfpi^{+/-}$ rescues were descendants of the same CRISPR-Cas9 edited mouse, we entertained the possibility that we had missed identifying a large deletion (extending the borders of the PCR primers) in the original mouse in one of the other candidate genes (e.g. *Arl6ip5*) that could explain the higher number of rescues. All gRNA editing sites were assessed using qPCR to capture discrepancies in expected copy-numbers, yet none were found. Nevertheless, a CRIPSR/Cas9-induced variation may still exist beyond the qPCR primer sequences. While the median deletion size using single sgRNA is 9bp, larger deletions as well as more complex rearrangements have been introduced by CRISPR/Cas9 in mice and would be missed with conventional genotyping approaches [15]. Alternatively, the slightly higher background in rescues for that founder mouse could be explained by an off-target effect or just normal variation.

Rescue pedigrees exhibit reduced fertility and incomplete penetrance

In addition to focusing on the variation in G1 progeny, the 42 G1 *F5*$^{L/L}$ *Tfpi*$^{+/-}$ mice alive at 6 weeks of age were mated to *F5*$^{L/L}$ mice to test the heritability of the survival phenotype. Twenty-one of these 42 mice generated at least one litter and 15 (1 female, 14 males) produced ≥1 offspring with the *F5*$^{L/L}$ *Tfpi*$^{+/-}$ genotype (Table 2). Across all pedigrees, mice beyond G1 (≥G2) continued to exhibit reduced survival with more pronounced underrepresentation of females (p=0.002; Fig 1E), and an average ~22% lower body weight compared to littermates at the time of genotyping ($p=2x10^{-16}$; Fig 1F). In the previous screen [5], rescues were outcrossed to the 129S1/SvImJ strain to introduce genetic diversity required for subsequent mapping experiments. However, complex strain modifier gene interactions confounded this analysis and resulted in a large number of "phenocopies" (defined as viable rescues despite lacking the original rescue mutation). To minimize this problem in the current screen, rescue pedigrees were maintained exclusively on the C57BL/6J background. While half of the pedigrees (8/16) previously generated on the mixed 129S1/SvImJ-C57BL/6J background generated >45 rescue progeny per pedigree [5] all pedigrees on the pure C57BL/6J background in the current study yielded <36 rescue mice, with the majority of pedigrees generating ≤5 rescues (Table 2). Significantly smaller pedigrees in comparison to the previous screen (p=0.010, S1 Fig) is likely explained by a general positive effect of mixing 129S1/SvImJ-C57BL/6J strain background either directly on rescue fertility (hybrid vigor) or indirectly by reducing the severity of the *F5*$^{L/L}$ phenotype. C57BL/6J and 129S1/SvImJ strains have been shown to exhibit significant differences in a number

of hemostasis-related parameters including platelet count, TFPI and TF expression levels [16]. Variation in genes underlying such strain specific differences may have contributed as modifiers to the rescue pedigrees in the previous screen [5].

Table 2. Overview of successfully progeny tested rescues

Pedigree	Mouse ID	Sex	Total # progeny	# Rescues	Penetrance*
1	60654	M	208	23	24.9%
2	82147	M	34	2	12.5%
3	82723	M	91	5	11.6%
4	83071	F	50	2	8.3%
5	83217	M	4	1	66.7%
6	83457	M	191	22	26.0%
7	83737	M	18	3	40.0%
8	83796	M	19	3	37.5%
9	83875	M	39	8	51.6%
10	83882	M	4	1	66.7%
11	10382	M	25	3	27.3%
12	11241	M	32	5	37.0%
13	11954	M	157	35	57.4%
14	19193	M	29	2	14.8%
15	20000	M	16	3	46.1%

*Penetrance is calculated as follows: #Rescues / (Total # of progeny - #Rescues) / 2

Linkage analysis with coding ENU variants fails to map suppressor loci

The three largest pedigrees (1, 6, and 13) were still poorly powered (29.6%, 21.7% and 39.4%, respectively) to identify the rescue variants by linkage analysis (S2-S4 Figs A). A total of 86 candidate ENU variants across the three pedigrees were validated by Sanger sequencing (S5 Table). Sixty-nine variants present in G1 rescue but not in their parents (G0) were further genotyped in all other rescue progeny in respective pedigrees. As expected from the power estimations, none of the 19 ENU variants tested in pedigree 1 (S2 Fig B), showed linkage with a LOD-score >1.25 (S3 Fig C). Similarly,

26 and 24 variants analyzed in pedigrees 6 and 13, respectively (S3,S4 Figs B) also failed to demonstrate a LOD-score >1.5 (S4,S5 Figs C). Failure to map the causal loci in any of these pedigrees was likely due to the lack of genetic power for mapping. However, we cannot exclude contribution from a non-ENU-induced variant [17] or insufficient marker coverage. While WES has been successfully applied to identify causal ENU variants within inbred lines [18] and in mixed background lines [19, 20], ~3000 ENU variants identified by whole genome sequencing (WGS) provide a much denser and more even coverage of the entire genome and expectedly outperforms WES in mapping [11]. On the other hand, WGS approach requires sequencing multiple pedigree members [10], or pooled samples at high coverage [11] and may present the challenge of interpreting causality from many closely linked non-coding variants.

A *Plcb4* mutation co-segregates with the rescue phenotype in 3 G1 siblings and their rescue offspring

The number of G1 rescues produced from each ENU-treated G0 male is shown in Fig 4A. Though most of the 182 G0-s yielded no or very few G1 rescue offspring, a single G0 produced 6 rescues out of a total of 39 offspring ($p<5x10^{-5}$, Fig 4A). This observation suggests a potential shared 'rescue' variant rather than 6 independent rescue mutations in the same G0 founder. A similar observation of 7 independent ENU pedigrees with an identical cardiac edema phenotype mapped to the same genetic locus was previously reported and hypothesized to result from a single shared mutation [7].

Our in-house pipeline for ENU-induced variant analysis (see Methods) filters out all variants shared between 2 or more G1 rescue mice based on the assumption that ENU-induced variants should be unique in each individual G1. However, rescue siblings could theoretically originate from the same mutagenized spermatogonial stem cell and share ~50% of their induced mutations [21]. Among 107 whole exome sequenced G1 mice, 38 were siblings (13 sib-pairs and 4 trios, S1 Table). 190 heterozygous variants present in 2-3 mice (representing sibpairs or trios) out of 107 rescues were examined, with 15 found to be shared by siblings (S6 Table). Of the 7 sibs/trios sharing an otherwise novel variant, none shared >10% of their identified variants – inconsistent with the expected 50% for progeny originating from the same ENU-treated stem cell.

Three protein-altering variants (*Plcb4*^{R335Q}, *Pyhin1*^{G157T}, and *Fignl2*^{G82S}) were shared among 2 or more of the 6 G1 rescues originating from the common G0 founder male (S6 Table). *Plcb4*^{R335Q} was detected as a *de novo* mutation in one of the G0 females (Fig 4B) and was present in 3 out of 6 G1 rescue siblings. *Plcb4* is located approximately 50 megabases upstream of the *Tfpi* locus on chromosome 2, with predicted recombination rate of ~14.1% (Fig 4C) [22, 23]. While non-rescue littermates exhibited close to the expected rate of recombination (20.2%) between the *Plcb4*^{R335Q} and *Tfpi* loci, all 43 rescue mice (3 G1s and their 40 ≥G2 progeny) were non-recombinant and carried the *Plcb4*^{R335Q} variant. This co-segregation between the *Plcb4*^{R335} variant and the rescue phenotype is statistically significant (p=0.003; Fig 4C). *Plcb4*^{R335Q} lies within a highly conserved region of *Plcb4* (Fig 4D) and is predicted to be deleterious by Polyphen-2 [24]. The other identified non-ENU variants (*Pyhin1*^{G157T} and

Fignl2^{G82S}) did not segregate with the rescue phenotype. *F5*$^{L/L}$ *Tfpi*$^{+/+}$ *Plcb4*$^{+/R335Q}$ mice we additionally crossed to *F5*$^{L/+}$ *Tfpi*$^{+/-}$ to ensure independent segregation of the *Tfpi* and *Plcb4* alleles. Out of 228 progeny, we genotyped 9 rescues carrying the *Plcb4* allele and 4 that did not (p=0.15, S7 Table). While an excess number of *F5*$^{L/L}$ *Tfpi*$^{+/-}$ *Plcb4*$^{+/R335Q}$ rescues was observed throughout the experiment (at times the difference being statistically significant), the final numbers were not as optimistic and an additional >100 mice would have been required to establish clear statistical difference from the background survival rate.

Although the estimated *de novo* mutation rate for inbred mice (~5.4 x 10^{-9} bp/generation) is >50X lower than the ENU mutation rate [25], other *de novo* variants have coincidentally been identified in ENU screens [26]. Mutations identified by DNA sequencing of offspring from ENU screen will not distinguish between an ENU-induced and *de novo* origin, though the former is generally assumed, given its much higher prevalence in the setting of a mutagenesis screen. *De novo* mutations originating in the G0 paternal or maternal lineages will be identified by analysis of parental genotypes, as was the case for the *Plcb4*^{R355Q} variant. However, this variant was originally removed from the candidate list by a filtering step based on the assumption that each ENU-induced mutation should be unique to a single G1 offspring. This filtering algorithm has been very efficient for removing false positive variants in our and previous screens [20]. However, our findings illustrate the risk for potential false negative results that this approach confers.

In conclusion, we performed a dominant, sensitized ENU mutagenesis screen for modifiers of thrombosis. Mutation burden analysis of 107 WES-d rescue mice highlighted *Arl6ip5* as a potential thrombosis suppressor gene. While the burden analysis lacked sufficient power to reach genome-wide significance for any of the candidates, it demonstrated a promising approach for mapping in the future using larger sample sizes.

Materials and methods

Mice

Mice carrying the murine homolog of the FVL mutation ($F5^{L}$; B6.129S2-$F5^{tm2Dgi}$/J, Jackson Laboratories stock #004080) [27] or the TFPI Kunitz domain deletion (*Tfpi*$^{-}$) [28] were generated as previously described. Mice were genotyped using PCR assays with primers and conditions as previously described [27, 28], and maintained on the C57BL/6J background (Jackson Laboratories stock #000664). All animal care and procedures were performed in accordance with the Principles of Laboratory and Animal Care established by the National Society for Medical Research. Institutional Animal Care and Use Committee (IACUC) at University of Michigan has approved the protocol number 05191 used for current study. The University of Michigan is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International and the animal care and use program conforms to the standards of “The Guide for the Care and Use of Laboratory Animals” (Revised 2011). All animal procedures were approved by the University of Michigan IACUC.

ENU screen

ENU mutagenesis was performed as previously described [5], with all mice on C57BL/6J genetic background. Briefly, 189 *F5*$^{L/L}$ male mice (6-8 weeks old) were administrated three weekly intraperitoneal injections of 90 mg/kg of ENU (N-ethyl-N-nitrosourea, Sigma-Aldrich). Eight weeks later, 182 surviving males were mated to *F5*$^{L/+}$ *Tfpi*$^{+/-}$ females and their G1 progeny were genotyped at age 2-3 weeks to identify viable *F5*$^{L/L}$ *Tfpi*$^{+/-}$ offspring ('rescues'). *F5*$^{L/L}$ *Tfpi*$^{+/-}$ G1 rescues were crossed to *F5*$^{L/L}$ mice on the C57BL/6J genetic background and transmission was considered positive with the presence of one or more rescue progeny. Theoretical mapping power in rescue pedigrees was estimated by 10,000 simulations using SIMLINK software [29].

Whole exome sequencing

Gender, age, whole exome sequencing (WES) details, and other characteristics for 108 rescue mice are provided in S1 Table. Genomic DNA (gDNA) extracted from tail biopsies of 56 G1 offspring from the current ENU screen and from an additional 50 *F5*$^{L/L}$ *Tfpi*$^{+/-}$ mice on the C57BL/6J background from the previous screen [5] were subjected to WES at the Northwest Genomics Center, University of Washington. Sequencing libraries were prepared using the Roche NimbleGen exome capture system. DNA from an additional two rescue offspring was subjected to WES at Beijing Genomics Institute or Centrillion Genomics Technologies, respectively (S1 Table). These two libraries were prepared using the Agilent SureSelect capture system. 100 bp paired-end sequencing was performed for all 108 exome libraries using Illumina HiSeq 2000 or 4000 sequencing instruments. Two WES mice represented rescue pedigree 1: the G1

founder and a G2 rescue offspring. The latter was used for linkage analysis, but excluded from the burden analysis (S1 Table).

WES data analysis

Average sequencing coverage, estimated by QualiMap software [30], was 77X, and >96% of the captured area was covered by at least 6 independent reads (S1 Table). All generated fastq files have been deposited to the NCBI Sequence Read Archive (Project accession number #SUB2922422). A detailed description of variant calling as well as in-house developed scripts for variant filtration are online as a GitHub repository (github.com/tombergk/FVL_mod). In short, Burrows-Wheeler Aligner [31] was used to align reads to the *Mus Musculus* GRCm38 reference genome, Picard [32] to remove duplicates, and GATK [33] to call and filter the variants. Annovar software [34] was applied to annotate the variants using the Refseq database. All variants within our mouse cohort present in more than one rescue were declared non-ENU induced and therefore removed. Unique heterozygous variants with a minimum of 6X coverage were considered as potential ENU mutations.

Mutation frequency estimations

All ENU-induced variants predicted to be potentially harmful within protein coding sequences including missense, nonsense, splice site altering SNVs, and out-of-frame insertions-deletions (INDELs), were totaled for every gene. The number of potentially damaging variants per gene was compared to a probability distribution of each gene being targeted by chance. Probability distributions were obtained by running 10 million

random permutations using probabilities adjusted to the length of the protein coding region. A detailed pipeline for the permutation analysis is available online (github.com/tombergk/FVL_mod). Genes that harbored more potentially damaging ENU-induced variants than expected by chance were considered as candidate modifier genes. FDR statistical correction for multiple testing was applied as previously described [35].

Variant validation by Sanger sequencing

All coding variants in pedigrees 1, 6, and 13 as well as all variants in candidate modifier genes from the burden analysis were assed using Sanger sequencing. Variants were considered ENU-induced if identified in the G1 rescue but not its parents. All primers were designed using Primer3 software [36] and purchased from Integrated DNA Technologies. PCR was performed using GoTaq Green PCR Master Mix (Promega), visualized on 2% agarose gel, and purified using QIAquick Gel Extraction Kit (Qiagen). Sanger sequencing of purified PCR products was performed by the University of Michigan Sequencing Core. All PCR primers (named: gene name+'_OF/OR') and internal sequencing primers (named: gene name+'_IF/IR') are listed in S8 Table.

Guide RNA design and *in vitro* transcription

Guide RNA target sequences were designed with computational tools [37, 38] (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design or http://genome-engineering.org) and top predictions per each candidate gene were selected for functional testing (S9 Table). Single guide RNA (sgRNA) for *C6*, *Ces3b*, *Itgb6*, and

Sntg1 were *in vitro* synthesized (MAXIscript T7 Kit, Thermo Fisher) from double stranded DNA templates by GeneArt gene synthesis service (Thermo Fisher) while sgRNA for *Arl6ip5* was *in vitro* synthesized using Guide-it sgRNA In Vitro Transcription Kit (Clontech). The *Cpn1* sgRNA target was cloned into plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene.org Plasmid #42230) [14]. The sgRNAs were purified prior to activity testing (MEGAclear Transcription Clean-Up Kit, Thermo Fisher). Both the Wash and Elution Solutions of the MEGAclear Kit were pre-filtered with 0.02 µm size exclusion membrane filters (Anotop syringe filters, Whatman) to remove particulates from zygote microinjection solutions, thus preventing microinjection needle blockages.

in vitro Cas9 DNA cleavage assay

Target DNA for the *in vitro* cleavage assay was PCR amplified from genomic DNA isolated from JM8.A3 C57BL/6N mouse embryonic stem (ES) cells [39] with candidate gene specific primers (S9 Table). *In vitro* digestion of target DNA was carried out by complexes of synthetic sgRNA and *S. pyogenes* Cas9 Nuclease (New England BioLabs) according to manufacturer's recommendations. Agarose gel electrophoresis of the reaction products was used to identify sgRNA molecules that mediated template cleavage by Cas9 protein (S5 Fig). *Arl6ip5* was assayed separately, with one out of four tested sgRNAs successfully cleaving the PCR template (data not shown).

Cell culture DNA cleavage assay

Synthetic sgRNAs that targeted *Cpn1* were not identified by the *in vitro* Cas9 DNA cleavage assay. As an alternative assay, sgRNA target sequences were subcloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 and co-electroporated into JM8.A3 ES cells as previously described [40]. Briefly, 15 µg of a Cas9 plasmid and 5 µg of a PGK1-puro expression plasmid [41] were co-electroporated into $0.8x10^7$ ES cells. On days two and three after electroporation media containing 2 µg/ml puromycin was applied to the cells; then selection free media was applied for four days. Surviving ES cells were collected and genomic DNA was purified. The *Cpn1* region targeted by the sgRNA was PCR amplified and tested for the presence of indel formation with a T7 endonuclease I assay according to the manufacturer's directions (New England Biolabs).

Generation of CRISPR-Cas9 gene edited mice

CRISPR-Cas9 gene edited mice were generated in collaboration with University of Michigan Transgenic Animal Model Core. A premixed solution containing 2.5 ng/µl of each sgRNA for *Arl6ip5*, *C6*, *Ces3b*, *Itgb6*, *Sntg1*, and 5 ng/µl of Cas9 mRNA (GeneArt CRISPR Nuclease mRNA, Thermo Fisher) was prepared in RNAse free microinjection buffer (10 mM Tris-Hcl, pH 7.4, 0.25 mM EDTA). The mixture also included 2.5 ng/µl of pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid containing guide *Cpn1*-g1 and a 2.5 ng/ul of pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid containing guide *Cpn1*-g2 targeting *Cpn1* (S9 Table). The mixture of sgRNAs, Cas9 mRNA, and plasmids was microinjected into the male pronucleus of fertilized mouse eggs obtained from the mating of stud males carrying the $F5^{L/+}$ $Tfpi^{+/-}$ genotype on the C57BL/6J background with superovulated C57BL/6J female mice. Microinjected eggs were transferred to

pseudopregnant B6DF1 female mice (Jackson Laboratories stock #100006). DNA extracted from tail biopsies was genotyped for the presence of gene editing.

Genotyping CRISPR alleles

Initially, sgRNA targeted loci were tested using PCR and Sanger sequencing (primer sequences provided in S9 Table). Small INDELs were deconvoluted from Sanger sequencing reads using TIDE software [42]. A selection of null alleles from >190 editing events were maintained for validation (S3 Table). Large (>30 bp) deletions were genotyped using PCR reactions that resulted in two visibly distinct PCR product sizes for the deletion and wildtype alleles. Expected product sizes and genotyping primers for each deletion are listed in S3 Table. All genotyping strategies were initially validated using Sanger sequencing. Should we describe the qpcr here as well? Need Emilee to write that up.

Statistical analysis

Kaplan-Meier survival curves and log-rank test to estimate significant differences in mouse survival were performed using the 'survival' package in R [43]. A paired two-tailed Student's t-test was applied to estimate differences in weights between rescue mice and their littermates. Fisher's exact tests were applied to estimate deviations from expected proportions in mouse crosses. Benjamini and Hochberg FDR for ENU burden analysis, Student's t-tests, and Fisher's exact tests were all performed using the 'stats' package in R software [44]. Linkage analysis was performed on the Mendel platform version 14.0 [45] and LOD scores ≥3.3 were considered genome-wide significant [46].

Acknowledgments

We acknowledge Wanda Filipiak, Galina Gavrilina, Elizabeth Hughes, and Michael Zeidler for assistance in the preparation of CRISPR-Cas9 reagents and gene edited transgenic mice in the Transgenic Animal Model Core of the University of Michigan's Biomedical Research Core Facilities. We thank the University of Michigan DNA Sequencing Core and the Northwest Genomics Center at the University of Washington, Department of Genome Sciences for sequencing services.

Figure Captions

Figure 1. A sensitized ENU suppressor screen for thrombosis modifiers

A) The ENU screen strategy is depicted here, along with the total numbers of G1 offspring by genotype. B) Survival curves for G1 rescue mice. Survival for females is slightly reduced compared to males (p=0.077). C-D) Weight at genotyping (at 14-21 days) for G1 rescues compared to control littermates. E) Survival of rescue mice beyond G1 (≥G2) is also reduced, with worse outcome in females (p=0.002). F) ≥G2 rescue offspring also exhibit reduced weights compared to their littermates at genotyping.

Figure 2. Distribution of ENU-induced mutations in WES data from 107 G1 rescues

A) Overview of mutation types for the 6771 observed ENU-induced exonic variants. B) Distribution of missense mutations by nucleotide substitution type. C) Distribution of ENU-variants by chromosome. D) The average number of exonic SNVs is ~65 for both

the current (G1-new) and previous (G1-old) screens. E) Number of genes (x-axis) sorted by the number of protein-altering ENU-induced mutations observed per gene (y-axis). Most genes (2567) carry only 1 mutation. In contrast, the ~0.1 megabase coding region of *Ttn* carries a total of 15 independent ENU variants.

Figure 3. Mutation enrichment per gene in WES data from 107 G1 rescues

All genes with potentially deleterious ENU mutations are sorted by their chromosomal position on x-axis, with the y-axis indicating the statistical significance (negative log of the p-value) of each gene's enrichment based on 10^6 permutations normalized to coding region size. Each dot represents a gene and the diameter is proportional to the number of mutations observed. Grey dotted lines represent FDR values of 0.1 and 0.25 (normalized to 2984 genes carrying mutations). Red dotted line represents FDR value 0.8 (normalized to all 20,586 genes in the simulation).

Figure 4. *Plcb4*^{R335Q} co-segregates with the rescue phenotype in pedigrees 12 and 13

A) One ENU mating exhibited a significantly higher number of rescue progeny (n=6) compared to all ENU matings ($p<5x10^{-5}$) and compared to ENU matings with ≥1 rescue progeny ($p<0.05$). B) One female in this ENU mating carried a *de novo* SNV (R335Q) in *Plcb4* gene that was inherited by 3 of the G1 rescues. C) *Plcb4* gene is loosely linked to the *Tfpi* locus on chromosome 2 and the expected recombination rate is 14.1%. No recombination was observed in 40 rescues from pedigree 12 and 13, while their littermates (n=149) showed close to expected recombination levels. D) The *Plcb4*^{R335Q} mutation lies in a highly conserved region in exon 13.

REFERENCES

1. Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. AmJHumGen. 2000;67(6):1452-9.
2. Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. J Thromb Haemost. 2004;2(5):731-6. Epub 2004/04/22. doi: 10.1111/j.1538-7933.2004.00660.x
JTH660 [pii]. PubMed PMID: 15099278.
3. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. Am J Hum Genet. 2015;96(4):532-42. doi: 10.1016/j.ajhg.2015.01.019. PubMed PMID: 25772935; PubMed Central PMCID: PMCPMC4385184.
4. Roldan V, Lecumberri R, Munoz-Torrero JF, Vicente V, Rocha E, Brenner B, et al. Thrombophilia testing in patients with venous thromboembolism. Findings from the RIETE registry. Thromb Res. 2009;124(2):174-7. Epub 2008/12/23. doi: 10.1016/j.thromres.2008.11.003
S0049-3848(08)00550-1 [pii]. PubMed PMID: 19101711.
5. Westrick RJ, Tomberg K, Siebert AE, Zhu G, Winn ME, Dobies SL, et al. A sensitized mutagenesis screen in Factor V Leiden mice identifies novel thrombosis suppressor loci. Proc Natl Acad Sci U S A. (In Press)
6. Eitzman DT, Westrick RJ, Bi X, Manning SL, Wilkinson JE, Broze GJ, Jr., et al. Lethal Perinatal thrombosis in mice resulting from the interaction of tissue factor pathway inhibitor deficiency and factor V Leiden. Circulation. 2002;105:2139-42.
7. Wansleeben C, van Gurp L, Feitsma H, Kroon C, Rieter E, Verberne M, et al. An ENU-mutagenesis screen in the mouse: identification of novel developmental gene functions. PLoS One. 2011;6(4):e19357. doi: 10.1371/journal.pone.0019357. PubMed PMID: 21559415; PubMed Central PMCID: PMCPMC3084836.
8. Montagutelli X. Effect of the genetic background on the phenotype of mouse mutations. J Am Soc Nephrol. 2000;11 Suppl 16:S101-5. Epub 2000/11/07. PubMed PMID: 11065339.
9. Papathanasiou P, Tunningley R, Pattabiraman DR, Ye P, Gonda TJ, Whittle B, et al. A recessive screen for genes regulating hematopoietic stem cells. Blood. 2010;116(26):5849-58. doi: 10.1182/blood-2010-04-269951. PubMed PMID: 20610815.
10. Bull KR, Rimmer AJ, Siggs OM, Miosge LA, Roots CM, Enders A, et al. Unlocking the bottleneck in forward genetics using whole-genome sequencing and identity by descent to isolate causative mutations. PLoS Genet. 2013;9(1):e1003219. Epub 2013/02/06. doi: 10.1371/journal.pgen.1003219
PGENETICS-D-12-02054 [pii]. PubMed PMID: 23382690; PubMed Central PMCID: PMC3561070.
11. Gallego-Llamas J, Timms AE, Geister KA, Lindsay A, Beier DR. Variant mapping and mutation discovery in inbred mice using next-generation sequencing. BMC Genomics. 2015;16:913. doi: 10.1186/s12864-015-2173-1. PubMed PMID: 26552429; PubMed Central PMCID: PMCPMC4640199.

12. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A. Mouse ENU mutagenesis. Human Molecular Genetics. 1999;8(10):1955-63.
13. Arnold CN, Barnes MJ, Berger M, Blasius AL, Brandl K, Croker B, et al. ENU-induced phenovariance in mice: inferences from 587 mutations. BMC Res Notes. 2012;5:577. Epub 2012/10/26. doi: 10.1186/1756-0500-5-577
1756-0500-5-577 [pii]. PubMed PMID: 23095377; PubMed Central PMCID: PMC3532239.
14. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23. doi: 10.1126/science.1231143. PubMed PMID: 23287718; PubMed Central PMCID: PMCPMC3795411.
15. Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, et al. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat Commun. 2017;8:15464. doi: 10.1038/ncomms15464. PubMed PMID: 28561021; PubMed Central PMCID: PMCPMC5460021.
16. White TA, Pan S, Witt TA, Simari RD. Murine strain differences in hemostasis and thrombosis and tissue factor pathway inhibitor. Thromb Res. 2010;125(1):84-9. doi: 10.1016/j.thromres.2009.03.006. PubMed PMID: 19398123; PubMed Central PMCID: PMCPMC2826594.
17. Tomberg K, Khoriaty R, Westrick RJ, Fairfield HE, Reinholdt LG, Brodsky GL, et al. Spontaneous 8bp Deletion in Nbeal2 Recapitulates the Gray Platelet Syndrome in Mice. PLoS One. 2016;11(3):e0150852. doi: 10.1371/journal.pone.0150852. PubMed PMID: 26950939; PubMed Central PMCID: PMCPMC4780761.
18. Li Y, Klena NT, Gabriel GC, Liu X, Kim AJ, Lemke K, et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature. 2015;521(7553):520-4. doi: 10.1038/nature14269. PubMed PMID: 25807483; PubMed Central PMCID: PMCPMC4617540.
19. Fairfield H, Gilbert GJ, Barter M, Corrigan RR, Curtain M, Ding Y, et al. Mutation discovery in mice by whole exome sequencing. Genome Biol. 2011;12(9):R86. Epub 2011/09/16. doi: 10.1186/gb-2011-12-9-r86
gb-2011-12-9-r86 [pii]. PubMed PMID: 21917142; PubMed Central PMCID: PMC3308049.
20. Andrews TD, Whittle B, Field MA, Balakishnan B, Zhang Y, Shao Y, et al. Massively parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an immediate source for thousands of new mouse models. Open Biol. 2012;2(5):120061. Epub 2012/06/23. doi: 10.1098/rsob.120061
rsob120061 [pii]. PubMed PMID: 22724066; PubMed Central PMCID: PMC3376740.
21. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, Phipps EL. Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl Acad Sci U S A. 1979;76(11):5818-9. Epub 1979/11/01. PubMed PMID: 293686; PubMed Central PMCID: PMC411742.
22. Sang F, Jiang P, Wang W, Lu Z. ReDB: A meiotic homologous recombination rate database. Chinese Science Bulletin. 2010;55(27-28):3169-73. doi: 10.1007/s11434-010-3029-3.
23. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, et al. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 2004;14(4):528-38. doi: 10.1101/gr.1970304. PubMed PMID: 15059993; PubMed Central PMCID: PMCPMC383296.

24. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9. doi: 10.1038/nmeth0410-248. PubMed PMID: 20354512; PubMed Central PMCID: PMCPMC2855889.
25. Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, et al. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. Genome Res. 2015;25(8):1125-34. doi: 10.1101/gr.186148.114. PubMed PMID: 26129709; PubMed Central PMCID: PMC4509997.
26. Bannister LA, Pezza RJ, Donaldson JR, de Rooij DG, Schimenti KJ, Camerini-Otero RD, et al. A dominant, recombination-defective allele of Dmc1 causing male-specific sterility. PLoS Biol. 2007;5(5):e105. doi: 10.1371/journal.pbio.0050105. PubMed PMID: 17425408; PubMed Central PMCID: PMC1847842.
27. Cui J, Eitzman DT, Westrick RJ, Christie PD, Xu ZJ, Yang AY, et al. Spontaneous thrombosis in mice carrying the factor V Leiden mutation. Blood. 2000;96(13):4222-6.
28. Huang ZF, Higuchi D, Lasky N, Broze GJ, Jr. Tissue factor pathway inhibitor gene disruption produces intrauterine lethality in mice. Blood. 1997;90:944-51.
29. Boehnke M, Ploughman L. SIMLINK: A Program for Estimating the Power of a Proposed Linkage Study by Computer Simulations. Version 4.12, April 2, 1997.
30. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 2012;28(20):2678-9. doi: 10.1093/bioinformatics/bts503. PubMed PMID: 22914218.
31. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324
btp324 [pii]. PubMed PMID: 19451168; PubMed Central PMCID: PMC2705234.
32. Picard tools. Available from: http://picard.sourceforge.net.
33. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8. Epub 2011/04/12. doi: 10.1038/ng.806
ng.806 [pii]. PubMed PMID: 21478889; PubMed Central PMCID: PMC3083463.
34. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. Epub 2010/07/06. doi: 10.1093/nar/gkq603
gkq603 [pii]. PubMed PMID: 20601685; PubMed Central PMCID: PMC2938201.
35. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300. PubMed PMID: WOS:A1995QE45300017.
36. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115. doi: 10.1093/nar/gks596. PubMed PMID: 22730293; PubMed Central PMCID: PMCPMC3424584.
37. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol. 2014;32(12):1262-7. doi: 10.1038/nbt.3026. PubMed PMID: 25184501; PubMed Central PMCID: PMCPMC4262738.

38. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31(9):827-32. doi: 10.1038/nbt.2647. PubMed PMID: 23873081; PubMed Central PMCID: PMCPMC3969858.
39. Pettitt SJ, Liang Q, Rairdan XY, Moran JL, Prosser HM, Beier DR, et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nat Methods. 2009;6(7):493-5. doi: 10.1038/nmeth.1342. PubMed PMID: 19525957; PubMed Central PMCID: PMCPMC3555078.
40. Pease S, Saunders TL, International Society for Transgenic Technologies. Advanced protocols for animal transgenesis an ISTT manual. Heidelberg ; New York ;: Springer; 2011. Available from: http://proxy.lib.umich.edu/login?url=http://link.springer.com/openurl?genre=book&isbn=978-3-642-20791-4.
41. McBurney MW, Fournier S, Jardine K, Sutherland L. Intragenic regions of the murine Pgk-1 locus enhance integration of transfected DNAs into genomes of embryonal carcinoma cells. Somat Cell Mol Genet. 1994;20(6):515-28. PubMed PMID: 7892649.
42. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 2014;42(22):e168. doi: 10.1093/nar/gku936. PubMed PMID: 25300484; PubMed Central PMCID: PMCPMC4267669.
43. Therneau TM, Grambsch PM. Modeling survival data : extending the Cox model. New York: Springer; 2000. xiii, 350 p. p.
44. R: A Language and Environment for Statistical Computing Vienna, AustriaR Core Team. Available from: http://www.R-project.org/.
45. Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, Sobel EM. Mendel: the Swiss army knife of genetic analysis programs. Bioinformatics. 2013;29(12):1568-70. doi: 10.1093/bioinformatics/btt187. PubMed PMID: 23610370; PubMed Central PMCID: PMCPMC3673222.
46. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11(3):241-7. Epub 1995/11/01. doi: 10.1038/ng1195-241. PubMed PMID: 7581446.

Figure 1.

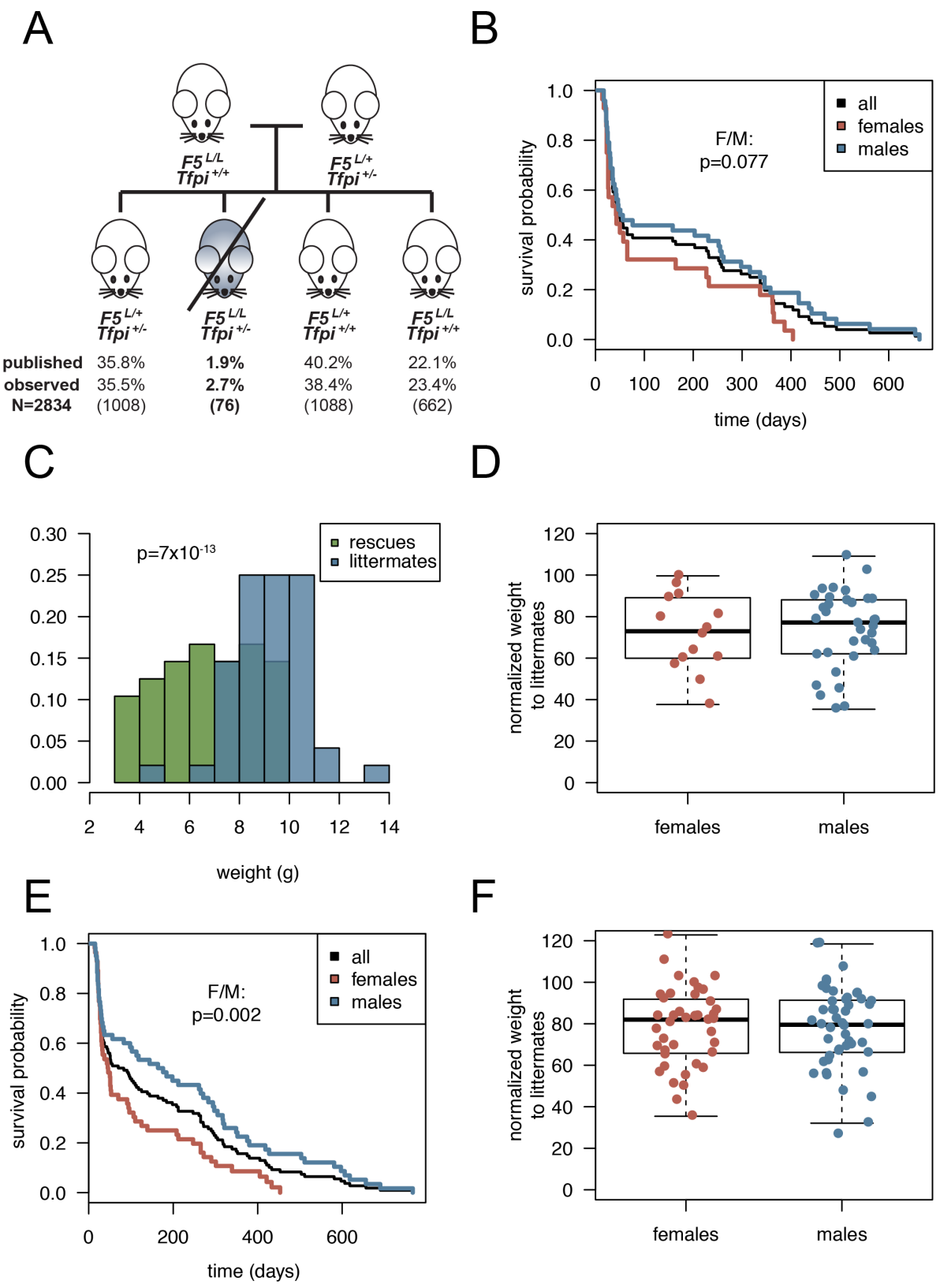

Figure 2.

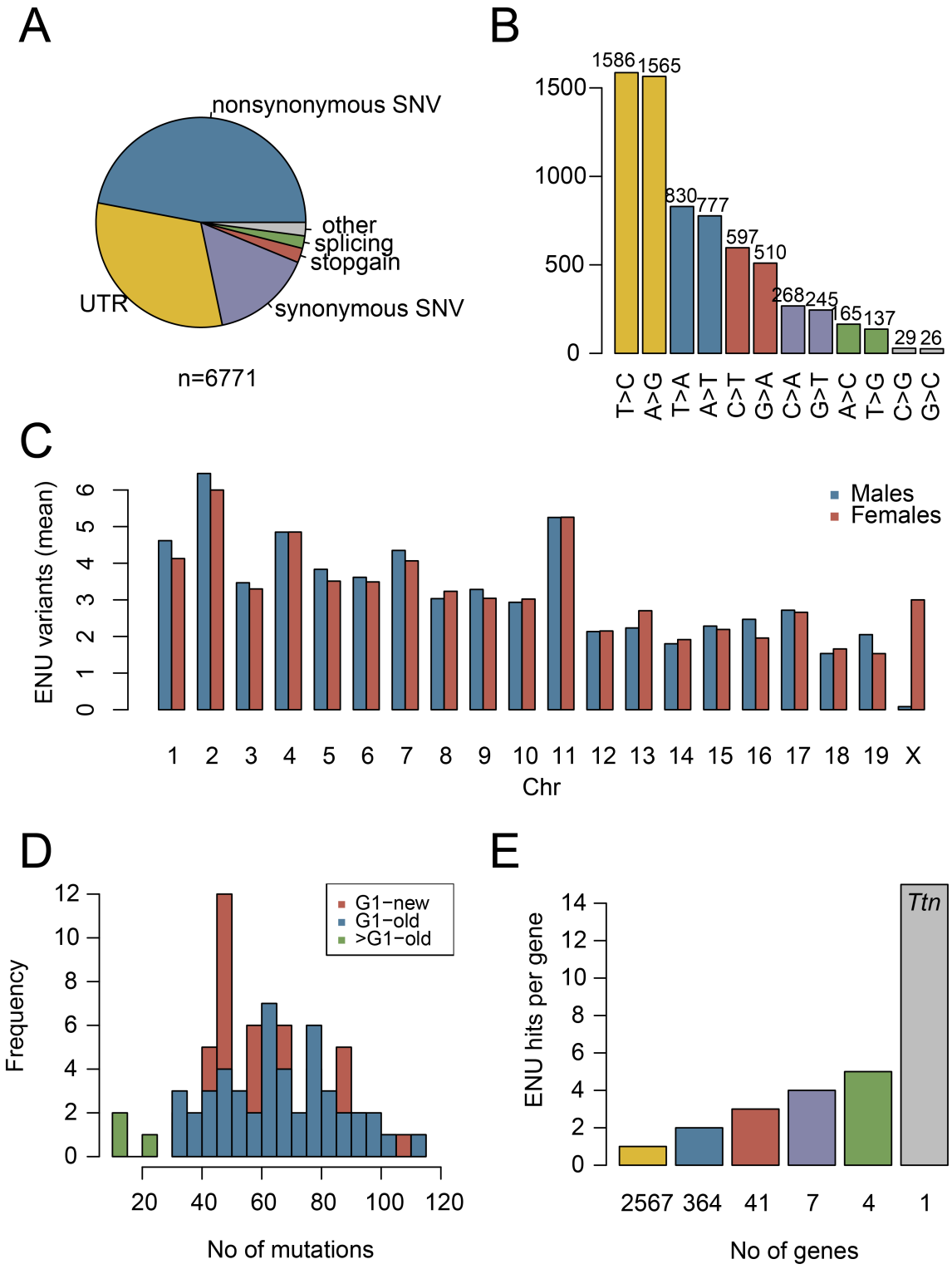

Figure 3.

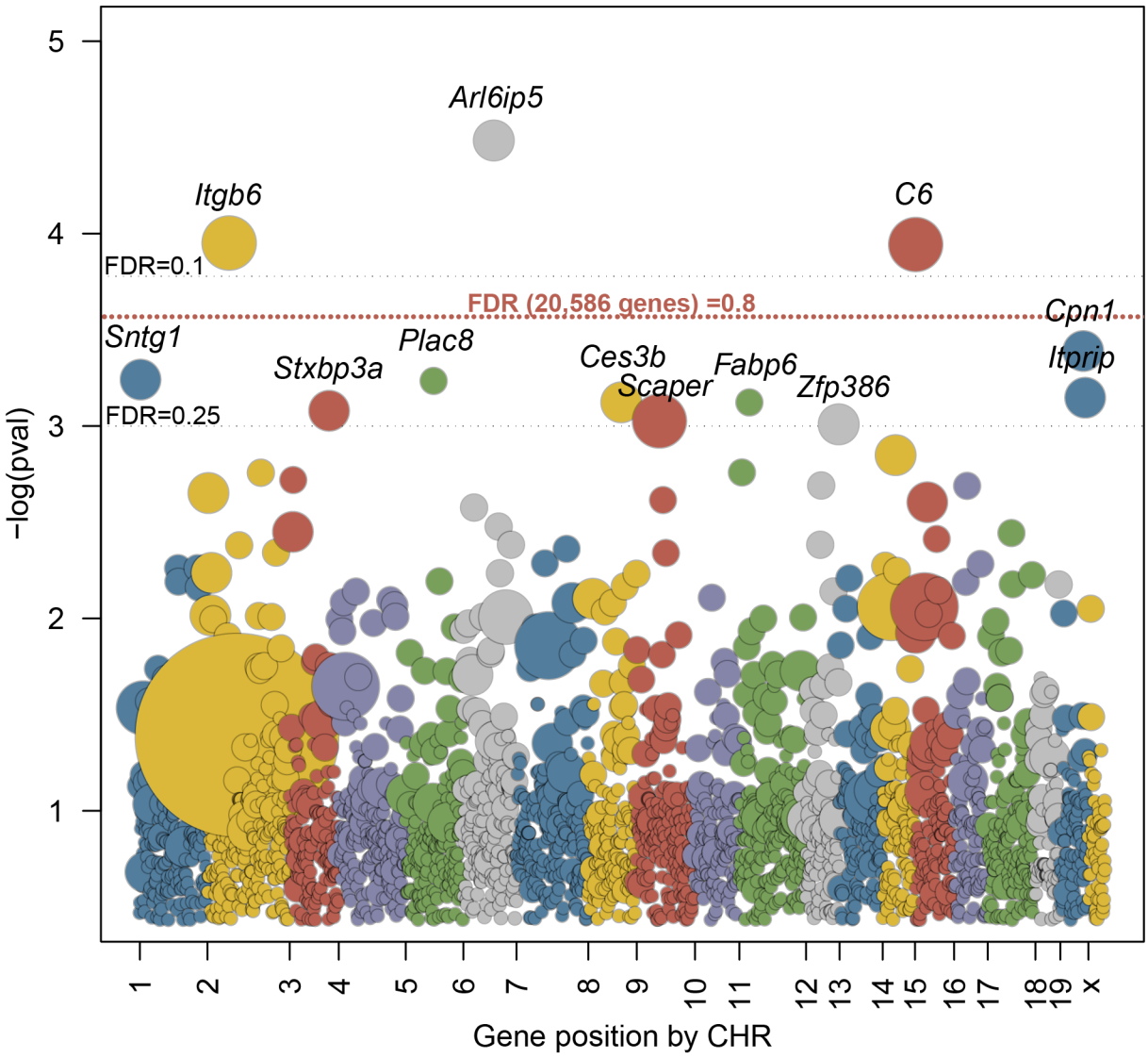

Figure 4.