bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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,#b], 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 and Center for Data Science and Big Data Analysis, 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

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[7]Departments of Pediatrics, University of Michigan, Ann Arbor, Michigan, United States of America

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

[#b]Current Address: Department of Molecular Genetics and Genomics, Washington University in St. Louis, Missouri, United States of America

* Corresponding author

Email: ginsburg@umich.edu (DG)

Abstract

Although the Factor V Leiden (FVL) gene variant is the most prevalent genetic risk factor for venous thrombosis, only 10% of FVL carriers will experience such an event in their lifetime. To identify potential FVL modifier genes contributing to this incomplete penetrance, we took advantage of a perinatal synthetic lethal thrombosis phenotype in mice homozygous for FVL ($F5^{L/L}$) and haploinsufficient for tissue factor pathway inhibitor ($Tfpi^{+/-}$) to perform a sensitized dominant ENU mutagenesis screen. Linkage analysis conducted in the 3 largest pedigrees generated from the surviving $F5^{L/L}$ $Tfpi^{+/-}$ mice ('rescues') using ENU-induced coding variants as genetic markers was unsuccessful in identifying major suppressor loci. Whole exome sequencing was applied to DNA from 107 rescue mice to identify candidate genes enriched for ENU mutations. A total of 3,481 potentially deleterious candidate ENU variants were identified in 2,984 genes. After correcting for gene size and multiple testing, *Arl6ip5* was

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

identified as the most enriched gene, though not reaching genome-wide significance. Evaluation of CRISPR/Cas9 induced loss of function in the top 6 genes failed to demonstrate a clear rescue phenotype. However, a maternally inherited (not ENU-induced) *de novo* mutation (*Plcb4*^{R335Q}) exhibited significant co-segregation with the rescue phenotype (p=0.003) in the corresponding pedigree. Thrombosis suppression by heterozygous *Plcb4* loss of function was confirmed through analysis of an independent, CRISPR/Cas9-induced *Plcb4* mutation (p=0.01).

Author summary

Abnormal blood clotting in veins (venous thrombosis) or arteries (arterial thrombosis) are major health problems, with venous thrombosis affecting approximately 1 in every thousand individuals annually in the United States. Susceptibility to venous thrombosis is governed by both genes and environment, with approximately 60% of the risk attributed to genetic influences. Though several genetic risk factors are known, >50% of genetic risk remains unexplained. Approximately 5% of people carry the most common known risk factor, Factor V Leiden. However, only 10% of these individuals will develop a blood clot in their lifetime. Mice carrying two copies of the Factor V Leiden mutation together with a mutation in a second gene called tissue factor pathway inhibitor develop fatal thrombosis shortly after birth. To identify genes that prevent this fatal thrombosis, we studied a large panel of mice carrying inactivating gene changes randomly distributed throughout the genome. We identified several genes as potential candidates to alter blood clotting balance in mice and humans with predisposition to

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

thrombosis, and confirmed this protective function for DNA changes in one of these genes (*Plcb4*).

Introduction

Venous thromboembolism (VTE) affects 1:1000 individuals in the US each year and is highly heritable [1, 2]. A single nucleotide variant (SNV) in the *F5* gene, referred to as Factor V Leiden (FVL, p.R506G) is present in 5-10% of Europeans, conferring a 2-4 fold increased risk for VTE [3]. Although ~25% of VTE patients carry the FVL variant [4], only ~10% of individuals heterozygous for FVL develop thrombosis in their lifetime.

To identify genetic variants that could potentially function as modifiers for FVL-associated VTE risk, we recently reported 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 progeny 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 [8, 9]. Additional challenges of this 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.

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

High throughput sequencing methods have enabled the direct identification of ENU-induced mutations. Thus, mutation identification in ENU screens is no longer dependent upon an outcross strategy for gene mapping [10, 11]. We now report whole exome sequencing (WES) of 107 rescue mice (including 50 mice from the previously reported ENU screen [5]). Assuming loss of gene function as the mechanism of rescue, these WES data were analyzed gene-by-gene to identify genes enriched with mutations (mutation burden analysis). The *Arl6ip5* gene emerged as the top candidate suppressor locus from this analysis. However, an independent CRISPR/Cas9-generated *Arl5ip5* mutant allele failed to demonstrate highly penetrant rescue of the $F5^{L/L}$ $Tfpi^{+/-}$ lethal phenotype. Surprisingly, a maternally inherited (not ENU-induced) *de novo* mutation ($Plcb4^{R335Q}$) exhibited significant co-segregation with the rescue phenotype (p=0.003) in an expanded pedigree.

Results and discussion

Smaller rescue pedigrees on pure C57BL/6J background

In the previously reported ENU screen [5], viable $F5^{L/L}$ $Tfpi^{+/-}$ rescue mice were outcrossed to the 129S1/SvImJ strain to introduce the 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 $F5^{L/L}$ $Tfpi^{+/-}$ mice lacking the original rescue mutation). To eliminate confounding effects of these thrombosis strain modifiers, we generated an additional 2,834 G1 offspring exclusively maintained on the C57BL/6J background. Fifteen new rescue pedigrees were established from this screen (S1 Table). The frequency, survival,

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

weight, and sex distributions of identified rescues were consistent with our previous report (S1 Fig). Though many of the pedigrees previously generated on the mixed 129S1/SvImJ-C57BL/6J background generated >45 rescue progeny per pedigree (8/16) [5], all pedigrees on the pure C57BL/6J background yielded <36 rescue mice (most generating ≤5 rescues) (S1 Table). Significantly smaller pedigrees in comparison to the previous screen (p=0.010, S2 Fig) are likely explained by a generally positive effect of the hybrid 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. The C57BL/6J and 129S1/SvImJ strains have been shown to exhibit significant differences in a number of hemostasis-related parameters, including platelet count and TFPI and tissue factor expression levels [12], with the genetic variations underlying such strain specific differences likely contributing to the genetic mapping complexity noted in the previous report [5].

Linkage analysis using coding ENU variants fails to map suppressor loci

As the rescue pedigrees were maintained on a pure C57BL/6J background, the only genetic markers that could be used for mapping were ENU-induced variants. WES of one G1 or G2 member of the three largest pedigrees (1, 6, and 13, S2 Table), identified a total of 86 candidate ENU variants that were also validated by Sanger sequencing analysis (S3 Table). Of these 86 candidate genes, 69 were present in the G1 rescue but not its parents (G0), indicating that they were likely ENU-induced variants. These 69 variants were then further genotyped in all other rescue progeny in the respective pedigrees. Given the low number of identified genetic markers (20-26 per

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pedigree), these three pedigrees were poorly powered (29.6%, 21.7% and 39.4%, respectively) to identify the rescue variants by linkage analysis (S3-S5 Figs A). None of the 19 ENU variants tested in pedigree 1 (S3 Fig B), showed linkage with a LOD-score >1.5 (S3 Fig C). Similarly, 26 and 24 variants analyzed in pedigrees 6 and 13, respectively (S4, S5 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 insufficient marker coverage. However, in these analyses, we could not exclude the contribution from a non-ENU-induced variant [13] or an unexpectedly high phenocopy rate. While WES has been successfully applied to identify causal ENU variants within inbred lines [14] and in mixed background lines [15, 16], whole genome sequencing (WGS) provides much denser and more even coverage of the entire genome (~3,000 ENU variants/genome expected) and outperforms WES for mapping [11]. However, a WGS approach requires sequencing multiple pedigree members [10], or pooled samples at high coverage [11], resulting in considerably higher expense with current methods.

WES identifies 6,771 ENU-induced variants in 107 rescues

In order to identify exonic ENU mutations, a total of 107 G1 rescues (57 from the current ENU screen and an additional 50 rescues with available material from the previous screen [5]), were subjected to WES (S2 Table). From ~1.5 million initially called variants, 6,735 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

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

synonymous SNVs (15%). The remaining variants (7%) were classified as splice site altering, stoploss, stopgain, or INDELs (Fig 1A). 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 1B), consistent with previously reported ENU studies [17, 18]. 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 single 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 1C). The average number of exonic ENU mutations for G1 rescues was ~65 SNV per mouse (Fig 1D), consistent with expected ENU mutation rates [10, 18]. 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 (3,481 out of 6,771 variants in 2,984 genes, S4 Table). Similar mutation burden analyses have been used to identify genes underlying rare diseases caused by *de novo* loss-of-function variants in humans [19-22]. In our study, 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 1E). After adjusting for coding region size and multiple testing (for

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2,984 genes), the ENU-induced mutation burden of potentially deleterious variants 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). Sanger sequencing validated 36 of the 37 variants in these 12 candidate genes (S4 Table). While in this study, stringent correction for multiple testing suggested no significant enrichment (*Arl6ip5* FDR=0.68, Fig 2), the potential power of this burden analysis is highly dependent on the number of possible genes that could result in a viable rescue. If there were 30 such genes in the genome and every one of the 107 rescue mice carried a mutation in one of these 30 genes, each gene would be, on average, represented by ~3.5 mutations (107/30), with >7 genes expected to carry 5 or more mutations, which should have been sufficient to distinguish from the background mutation rate. However, if 500 genes could rescue the phenotype, sequencing close to a thousand mice would be required to achieve sufficient mapping power. The power could be further compromised by modifier genes with incomplete penetrance, imperfect predictions for potentially harmful mutations, and by the previously reported background survival rate for the rescue mice [6]. Due to the uncertainty of the power of these analyses, we proceeded to experimentally test the thrombosupressive effects of loss of function mutations in the genes identified by mutation burden analysis.

Independent alleles for 6 candidate genes fail to replicate thrombosis suppression

Independent null alleles were generated with CRISPR/Cas9 for the top candidate genes (*Arl6ip5, C6, Itgb6, Cpn1, Sntg1 and Ces3b;* Fig 2) to test for

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

thrombosuppression. From 294 microinjected zygotes with pooled guide RNAs targeting these 6 genes, we obtained 39 progeny. CRISPR/Cas9 genome editing was assessed by Sanger sequencing of the sgRNA target sites. Approximately 190 independent targeting events were observed across the 6 genes in 36 of the 39 mice including small INDELs, single nucleotide changes, and several large (>30bp) deletions or inversions. Targeted alleles were either homozygous, heterozygous, or mosaic, with the number of editing events varying greatly for different sgRNAs (2.5-85%). Two or more different CRISPR/Cas9-induced alleles for each of the candidate genes (S5 Table) were bred to isolation but maintained on the $F5^{L}$ background for subsequent test crossing. The progeny of $F5^{L/L}$ $Tfpi^{+/+}$ mice crossed with $F5^{L/+}$ $Tfpi^{+/-}$ mice (one of these parental mice also carrying the CRISPR/Cas9-induced allele) were monitored for survival of $F5^{L/L}$ $Tfpi^{+/-}$ offspring (Table 1, S6 Table).

Table 1. Testing for rescue effect with CRISPR/Cas9-induced alleles

Gene	Total mice tested	No. of rescues w/o allele	No. of rescues with allele	P-value*
Arl6ip5	205	1	5	0.21
Itgb6	154	1	1	1
C6	106	0	0	1
Cpn1	139	0	1	1
Sntg1	223	3	4	1
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 it remained non-significant after surveying 205 offspring

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(p=0.21, Table 1). Nonetheless, rescue of the *F5*$^{L/L}$ *Tfpi*$^{+/-}$ phenotype by *Arl6ip5* haploinsufficiency cannot be excluded, particularly at reduced penetrance. Of note, rescue of *F5*$^{L/L}$ *Tfpi*$^{+/-}$ lethality by haploinsufficiency for *F3* (the target of TFPI) only exhibits penetrance of ~33% [5], a level of rescue which current observations cannot exclude for *Alr6ip5* and *Sntg1*. For most of the other candidate genes, the number of observed *F5*$^{L/L}$ *Tfpi*$^{+/-}$ mice did not differ from the expected background survival rate for this genotype (~2%) [6]. Though higher numbers of rescues were observed for offspring from the *Sntg1* cross, these were equally distributed between mice with and without the *Sntg1* loss-of-function allele.

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 3A. Though most of the 182 G0 males yielded few or no G1 rescue offspring, a single G0 produced 6 rescues out of a total of 39 offspring (Fig 3A), including the founder G1 rescue for the largest pedigree (number 13). This observation suggested a potential shared rescue variant rather than 6 independent rescue mutations from the same G0 founder. Similarly, another previously reported ENU screen identified 7 independent ENU pedigrees with an identical phenotype mapping to the same genetic locus, also hypothesized to result from a single shared mutation [7]. While rescue siblings could theoretically originate from the same mutagenized spermatogonial stem cell and share ~50% of their induced mutations [23], such a common stem cell origin

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was excluded by exome sequence analysis in the rescue G1 sibs identified here (see Materials and methods).

Analysis of WES for 3 of the G1 rescues originating from this common G0 founder male (Fig 3B, S2 Table) identified 3 protein-altering variants (*Plcb4*^{R335Q}, *Pyhin1*^{G157T}, and *Fignl2*^{G82S}) shared among 2 or more of the 6 G1 rescues (S7 Table). *Plcb4*^{R335Q} was detected as a *de novo* mutation in one of the non-mutagenized G0 females in phase with the *Tfpi* null allele (Fig 3B) 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 (predicted recombination between *Plcb4* and *Tfpi* ~14.1%) (Fig 3C) [24, 25]. While non-rescue littermates exhibited the expected rate of recombination between the *Plcb4*^{R335Q} and *Tfpi* loci (20.2%), 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*^{R335Q} variant and the rescue phenotype is statistically significant (p=0.003; Fig 3C). *Plcb4*^{R335Q} lies within a highly conserved region of *Plcb4* (Fig 3D) and is predicted to be deleterious by Polyphen-2 [26]. The other identified non-ENU variants (*Pyhin1*^{G157T} and *Fignl2*^{G82S}) did not segregate with the rescue phenotype (S6 Fig).

Although the estimated *de novo* mutation rate for inbred mice (~5.4 x 10^{-9} bp/generation) is 200X lower than our ENU mutation rate, other *de novo* variants have coincidentally been identified in ENU screens [27]. Mutations identified by DNA sequencing of offspring from ENU screens 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

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 was very efficient for removing false positive variants in our screen and others [16]. However, our findings illustrate the risk for potential false negative results that this approach confers.

Independent mutant allele for *Plcb4* recapitulates the rescue phenotype

An independent *Plcb4* null allele was generated by CRISPR/Cas9. Three distinct INDELs were identified by Sanger sequencing in the 25 progeny obtained from the CRISPR/Cas9-injected oocytes. One of these alleles introduced a single nucleotide insertion at amino acid 328, resulting in a frameshift in the protein coding sequence (*Plcb4*^{ins1}, Fig 4A-B). A total of 169 progeny from a *F5*$^{L/L}$ *Plcb4*$^{+/ins1}$ X *F5*$^{L/+}$ *Tfpi*$^{+/-}$ cross yielded 11 *F5*$^{L/L}$ *Tfpi*$^{+/-}$ rescue progeny surviving to weaning (Fig 4C, S8 Table). Ten of these 11 rescues carried the *Plcb4*^{ins1} allele, consistent with significant rescue (p=0.01, Fig 4C) with reduced penetrance (~40%). *Plcb4* encodes phospholipase C, beta 4 and has been recently associated with auriculocondylar syndrome in humans [28]. No role for PLCB4 in the regulation of hemostasis has been previously reported, and the underlying mechanism for suppression of the lethal *F5*$^{L/L}$ *Tfpi*$^{+/-}$ phenotype is unknown.

The above rescue of the *F5*$^{L/L}$ *Tfpi*$^{+/-}$ phenotype by an independent *Plcb4* mutant allele, strongly supports the identification of the *de novo Plcb4*^{R355Q} mutation as the causal suppressor variant for Pedigree 13. These findings are also most consistent with

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a loss-of-function mechanism of action for the *Plcb4*^{R355Q} mutation. The lack of a positive signal from this genomic region by the linkage analysis described above (S5 Fig) is likely explained by the absence of a nearby genetically informative ENU variant (the closest, *Abca2* is located >50 Mb downstream from both *Tfpi* and *Plcb4* (S3 Table, S5 Fig)). Of note, 4 of the 107 rescue mice in the WES mutation burden analysis also carried a *Plcb4* mutation consistent with its suppressor function, though below the level of statistical significance. Nonetheless, these findings highlight the feasibility of our approach, given sufficient power.

In conclusion, we performed a dominant, sensitized ENU mutagenesis screen for modifiers of thrombosis. Analysis of extended pedigrees identified *Plcb4* as a novel thrombosis modifier. Though mutation burden analysis suggested several other potential modifier loci, including *Arl6ip5*, incomplete penetrance and the background phenocopy rate significantly limited the power to detect additional thrombosis suppressor genes. Future applications of this approach will likely require significantly larger sample sizes and/or a more stringent sensitized genotype for screening. Nonetheless, our findings demonstrate the power of a sensitized ENU screen and mutation burden analysis to identify novel loci contributing to the regulation of hemostatic balance and candidate modifier genes for thrombosis and bleeding risk in humans.

Materials and methods

Mice

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mice carrying the murine homolog of the FVL mutation [29] (*F5*^{L}; also available from Jackson Laboratories stock #004080) or the TFPI Kunitz domain deletion (*Tfpi*$^{-}$) [30] were genotyped using PCR assays with primers and conditions as previously described [29, 30], 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. The Institutional Animal Care and Use Committee at the University of Michigan has approved protocols PRO00005191 and PRO00007879 used for the current study and conforms to the standards of "The Guide for the Care and Use of Laboratory Animals" (Revised 2011).

ENU screen

ENU mutagenesis was performed as previously described [5], with all mice on the C57BL/6J genetic background. Briefly, 189 *F5*$^{L/L}$ male mice (6-8 weeks old) were administered 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 (backcrossed >20 generations) 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 [31].

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

338 **Whole exome sequencing**

339 Gender, age, WES details, and other characteristics for 108 rescue mice are
340 provided in S2 Table. Genomic DNA (gDNA) extracted from tail biopsies of 56 G1
341 offspring from the current ENU screen and from an additional 50 *F5*$^{L/L}$ *Tfpi*$^{+/-}$ mice on the
342 C57BL/6J background from the previous screen [5] were subjected to WES at the
343 Northwest Genomics Center, University of Washington. Sequencing libraries were
344 prepared using the Roche NimbleGen exome capture system. DNA from an additional
345 two rescue offspring was subjected to WES at Beijing Genomics Institute or Centrillion
346 Genomics Technologies, respectively (S2 Table). These two libraries were prepared
347 using the Agilent SureSelect capture system. 100 bp paired-end sequencing was
348 performed for all 108 exome libraries using Illumina HiSeq 2000 or 4000 sequencing
349 instruments. Two WES mice represented rescue pedigree 1: the G1 founder and a G2
350 rescue offspring. The latter was used for linkage analysis, but excluded from the burden
351 analysis (S2 Table).
352
353 **WES data analysis**
354 Average sequencing coverage, estimated by QualiMap software [32], was 77X,
355 and >96% of the captured area was covered by at least 6 independent reads (S2
356 Table). All generated fastq files have been deposited to the NCBI Sequence Read
357 Archive (Project accession number #PRJNA397141). A detailed description of variant
358 calling as well as in-house developed scripts for variant filtration are online as a GitHub
359 repository (github.com/tombergk/FVL_mod). In short, Burrows-Wheeler Aligner [33] was
360 used to align reads to the *Mus Musculus* GRCm38 reference genome, Picard [34] to

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

remove duplicates, and GATK [35] to call and filter the variants. Annovar software [36] 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. Among 107 whole exome sequenced G1 mice, 38 were siblings (13 sib-pairs and 4 trios, S2 Table). 190 heterozygous variants present in 2 or 3 mice (representing sibpairs or trios) out of 107 rescues were examined, with 15 found to be shared by siblings (S7 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 spermatogonial stem cell.

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 summed 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 [37].

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

384

385 **Variant validation by Sanger sequencing**

386 All coding variants in pedigrees 1, 6, and 13 as well as all variants in candidate
387 modifier genes from the burden analysis were assessed using Sanger sequencing.
388 Variants were considered ENU-induced if identified in the G1 rescue but not its parents.
389 All primers were designed using Primer3 software [38] and purchased from Integrated
390 DNA Technologies. PCR was performed using GoTaq Green PCR Master Mix
391 (Promega), visualized on 2% agarose gel, and purified using QIAquick Gel Extraction
392 Kit (Qiagen). Sanger sequencing of purified PCR products was performed by the
393 University of Michigan Sequencing Core. Outer primers were used to generate the PCR
394 product which was then sequenced using the internal sequencing primers. All outer
395 PCR primers (named: gene name+'_OF/OR') and internal sequencing primers (named:
396 gene name+'_IF/IR') are listed in S9 Table.

397

398 **Guide RNA design and *in vitro* transcription**

399 Guide RNA target sequences were designed with computational tools [39, 40]
400 (http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design or http://genome-
401 engineering.org) and top predictions per each candidate gene were selected for
402 functional testing (S10 Table). Single guide RNAs (sgRNA) for *C6*, *Ces3b*, *Itgb6*, and
403 *Sntg1* were *in vitro* synthesized (MAXIscript T7 Kit, Thermo Fisher) from double
404 stranded DNA templates by GeneArt gene synthesis service (Thermo Fisher) while the
405 4 sgRNAs for *Arl6ip5* were *in vitro* synthesized using the Guide-it sgRNA In Vitro
406 Transcription Kit (Clontech). The sgRNAs were purified prior to activity testing

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(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 assays was PCR amplified from genomic DNA isolated from JM8.A3 C57BL/6N mouse embryonic stem (ES) cells [41] with candidate gene specific primers (S10 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 (S7 Fig). *Arl6ip5* was assayed separately, with one out-of-four tested sgRNAs successfully cleaving the PCR template.

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 (*Cpn1*-g1, *Cpn1*-g2) were cloned into plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene.org Plasmid #42230) [42] and co-electroporated into JM8.A3 ES cells as previously described [43]. Briefly, 15 µg of a Cas9 plasmid and 5 µg of a PGK1-puro expression plasmid [44] 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;

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

then selection free media was applied for four days. Genomic DNA was purified from surviving ES cells. 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 instructions (New England Biolabs).

Generation of CRISPR/Cas9 gene edited mice

CRISPR/Cas9 gene edited mice were generated in collaboration with the 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/µl of pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid containing guide *Cpn1*-g2 targeting *Cpn1* (S10 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 of offspring was genotyped for the presence of gene editing. The *Plcb4* allele was targeted in a separate experiment in collaboration with the University of Michigan Transgenic Animal Model Core using a pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid that contained guide *Plcb4* (5 ng/µl).

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

453 **CRISPR allele genotyping**

454 Initially, sgRNA targeted loci were tested using PCR and Sanger sequencing
455 (primer sequences provided in S10 Table). Small INDELs were deconvoluted from
456 Sanger sequencing reads using TIDE software [45]. A selection of null alleles from >190
457 editing events were maintained for validation (S5 Table). Large (>30 bp) deletions were
458 genotyped using PCR reactions that resulted in two visibly distinct PCR product sizes
459 for the deletion and wildtype alleles. Expected product sizes and genotyping primers for
460 each deletion are listed in S5 Table. All genotyping strategies were initially validated
461 using Sanger sequencing.

462

463 **Picogreen DNA quantification and qPCR**

464 A qPCR approach was applied to exclude large on-target CRISPR/Cas9-induced
465 deletions. All DNA samples were quantified using the Quant-iT™ PicoGreen® dsDNA
466 Assay Kit (Life Technologies) and analyzed on the Molecular Devices SpectraMax® M3
467 multi-mode microplate reader using SoftMax® Pro software and diluted to 5ng/µl.
468 Primer pairs were designed for each gene using Primer Express 3.0 software (S9
469 Table) and samples were measured in triplicate using Power SYBR® Green PCR
470 Master Mix (Thermo Fisher Scientific) on a 7900 HT Fast Real-Time PCR System
471 (Applied Biosystems) with DNA from wildtype C57BL/6J mice used as a reference.
472 While large CRISPR/Cas9 induced deletions extending the borders of the PCR primers
473 have been reported [46], qPCR did not detect evidence for a large deletion in any of the
474 CRISPR targeted genes.

475

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

476 **Statistical analysis**

477 Kaplan-Meier survival curves and a log-rank test to estimate significant
478 differences in mouse survival were performed using the 'survival' package in R [47]. A
479 paired two-tailed Student's t-test was applied to estimate differences in weights between
480 rescue mice and their littermates. Fisher's exact tests were applied to estimate
481 deviations from expected proportions in mouse crosses. Mendelian segregation for
482 CRISPR/Cas9-induced alleles among non-rescue littermates was assessed in a subset
483 of mice by Sanger sequencing and then assumed for the rest of the littermates in the
484 Fisher's exact tests. Benjamini and Hochberg FDR for ENU burden analysis, Student's
485 t-tests, and Fisher's exact tests were all performed using the 'stats' package in R
486 software [48]. Linkage analysis was performed on the Mendel platform version 14.0 [49]
487 and LOD scores ≥3.3 were considered genome-wide significant [50].

488

489 ## Acknowledgments

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

496

497 ## Figure captions

498 **Fig 1. Distribution of ENU-induced mutations in WES data from 107 G1 rescues**

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A) Overview of mutation types for the 6,771 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) screen [5]. E) Number of genes (x-axis) sorted by the number of protein-altering ENU-induced mutations observed per gene (y-axis). Most genes (2,567) carry only 1 mutation. In contrast, the ~0.1 megabase coding region of *Ttn* carries a total of 15 independent ENU variants.

Fig 2. 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 the 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. Gray dotted lines represent FDR values of 0.1 and 0.25 (normalized to 2,984 genes carrying mutations). Red dotted line represents FDR value 0.8 from a more stringent test (normalized to all 20,586 genes in the simulation).

Fig 3. *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 the *Plcb4* gene that was inherited in phase with the *Tfpi*$^{-}$ allele that was inherited by 3 of the G1 rescues. C) The *Plcb4* gene is loosely linked to the *Tfpi* locus on chromosome 2,

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

522 with a predicted recombination rate of 14.1%. No recombination was observed in 40
523 rescues from pedigree 12 and 13, while their littermates (n=149) exhibited close to the
524 expected recombination rate. D) The *Plcb4*^{R335Q} mutation lies in a highly conserved
525 region in exon 13 (data from Multiz alignment on UCSC Genome Browser).
526
527 **Fig 4. An independent CRISPR/Cas9 induced *Plcb4* allele validates the rescue**
528 **phenotype**
529 A) The CRISPR/Cas9-induced *Plcb4*^{ins1} allele (insertion of the nucleotide 'A' at amino
530 acid 328) results in a frameshift to the protein coding sequence leading to a premature
531 stop codon. B) Sanger sequencing analysis of a wildtype mouse and a heterozygous
532 mouse for the *Plcb4*^{ins1} allele C) 169 progeny genotyped from a validation cross of *F5*$^{L/L}$
533 *Plcb4*$^{+/ins1}$ mice with *F5*$^{L/+}$ *Tfpi*$^{+/-}$.
534
535 ## Supporting information
536 **S1 Fig. A sensitized ENU suppressor screen for thrombosis modifiers**
537 A) The ENU screen strategy is depicted here, along with the total numbers of G1
538 offspring observed by genotype. B) Survival curves for G1 rescue mice. Approximately
539 50% of the rescue mice died by 6 weeks of age, with no significant survival difference
540 observed between females and males (p=0.077), though females were
541 underrepresented compared to males during the initial genotyping (28 females
542 compared to 48 males, p=0.022). C-D) Weight at genotyping (at 14-21 days) was on
543 average 25-30% smaller for G1 rescues than their littermates ($p=7x10^{-13}$). E) Survival of
544 rescue mice beyond G1 (≥G2) is also reduced, with worse outcome in females

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(p=0.002). Across all pedigrees, mice beyond G1 (≥G2) continued to exhibit reduced survival with more pronounced underrepresentation of females (p=0.002), and F) an average ~22% lower body weight compared to littermates (mean defined as 100%) at the time of genotyping ($p=2x10^{-16}$).

S2 Fig. Size distribution of ENU pedigrees

The ENU rescue pedigrees from the previous screen (n=16, [5]) are significantly larger than the ENU rescue pedigrees observed in the current screen (p=0.010, n=15, S1 Table).

S3 Fig. Genetic mapping of ENU-induced variants in pedigree 1

A) Overview of pedigree 1 (only rescue mice displayed). B) All coding ENU-induced mutations identified by WES were genotyped in all rescues from the pedigree by Sanger sequencing. Blue boxes indicate presence and red boxes indicate absence of the mutation. P1-P3 refers to 3 parental genotypes (G0 male and 2 untreated females). C) Linkage analysis using the ENU-induced variants from (B) as genetic markers.

S4 Fig. Genetic mapping of ENU-induced variants in pedigree 6

A) Overview of pedigree 6 (only rescue mice displayed). B) All coding ENU-induced mutations identified by WES were genotyped in most rescues from the pedigree by Sanger sequencing. Blue boxes indicate presence and red boxes indicate absence of the mutation. P1-P3 refers to 3 parental genotypes (G0 male and 2 untreated females). C) Linkage analysis using the ENU-induced variants from (B) as genetic markers.

S5 Fig. Genetic mapping of ENU variants in pedigree 13

A) Overview of pedigree 13 (only rescue mice displayed). B) All coding ENU-induced mutations identified by WES were genotyped in all rescues from the pedigree if present

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

in key mice 3 and 5 by Sanger sequencing. Blue boxes indicate presence and red boxes indicate absence of the mutation. P1-P3 refers to 3 parental genotypes (G0 male and 2 untreated females). C) Linkage analysis using the ENU-induced variants from (B) as genetic markers.

S6 Fig. Segregation analysis for *Pyhin1* and *Fignl2* in pedigree 13

Segregation analysis in pedigree 13 for A) *Pyhin1* and B) *Fignl2* variants. Blue boxes indicate presence and red boxes indicate absence of the mutation. White boxes indicate untested mice, while light red boxes indicate untested mice with assumed absence of the mutation.

S7 Fig. *In vitro* cleavage assay for sgRNAs

A) sgRNA+Cas9 targeting created double strand breaks in DNA templates obtained from genomic DNA by PCR. Expected sizes after sgRNA+Cas9 endonuclease activity: 430bp/240bp (*Ces3b*), 334bp/273bp (*Sntg1*), 530bp/275bp (*Itgb6*), and 383bp/296bp (*C6*). B) sgRNA+Cas9 complexes targeting *Cpn1* using two different guides (g1, g2) failed to generate detectable double strand breaks. Positive control (P.C.) was added to ensure Cas9 protein activity, with expected sizes after cleavage (390bp/140bp) indicated by white stars.

S1 Table. Overview of successfully progeny tested rescues

S2 Table. Rescue mice subjected to WES

S3 Table. Variants identified for pedigrees 1, 6, and 13

S4 Table. ENU-induced coding variants in WES data

S5 Table. CRISPR/Cas9 induced alleles

S6 Table. Validation crosses with CRISPR/Cas9 induced alleles

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S7 Table. Shared variants between 2-3 mice in WES data

S8 Table. *Plcb4*^{ins1} validation cross

S9 Table. Primer sequences

S9 Table. Sequences and genotyping data for gRNAs

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.

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

614 5. Westrick RJ, Tomberg K, Siebert AE, Zhu G, Winn ME, Dobies SL, et al. Sensitized
615 mutagenesis screen in Factor V Leiden mice identifies thrombosis suppressor loci. Proc
616 Natl Acad Sci U S A. 2017;114(36):9659-64. doi: 10.1073/pnas.1705762114. PubMed
617 PMID: 28827327; PubMed Central PMCID: PMCPMC5594664.
618 6. Eitzman DT, Westrick RJ, Bi X, Manning SL, Wilkinson JE, Broze GJ, Jr., et al. Lethal
619 Perinatal thrombosis in mice resulting from the interaction of tissue factor pathway
620 inhibitor deficiency and factor V Leiden. Circulation. 2002;105:2139-42.
621 7. Wansleeben C, van Gurp L, Feitsma H, Kroon C, Rieter E, Verberne M, et al. An ENU-
622 mutagenesis screen in the mouse: identification of novel developmental gene functions.
623 PLoS One. 2011;6(4):e19357. doi: 10.1371/journal.pone.0019357. PubMed PMID:
624 21559415; PubMed Central PMCID: PMCPMC3084836.
625 8. Montagutelli X. Effect of the genetic background on the phenotype of mouse
626 mutations. J Am Soc Nephrol. 2000;11 Suppl 16:S101-5. Epub 2000/11/07. PubMed PMID:
627 11065339.
628 9. Papathanasiou P, Tunningley R, Pattabiraman DR, Ye P, Gonda TJ, Whittle B, et al. A
629 recessive screen for genes regulating hematopoietic stem cells. Blood. 2010;116(26):5849-
630 58. doi: 10.1182/blood-2010-04-269951. PubMed PMID: 20610815.
631 10. Bull KR, Rimmer AJ, Siggs OM, Miosge LA, Roots CM, Enders A, et al. Unlocking the
632 bottleneck in forward genetics using whole-genome sequencing and identity by descent to
633 isolate causative mutations. PLoS Genet. 2013;9(1):e1003219. Epub 2013/02/06. doi:
634 10.1371/journal.pgen.1003219
635 PGENETICS-D-12-02054 [pii]. PubMed PMID: 23382690; PubMed Central PMCID:
636 PMC3561070.

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

637 11. Gallego-Llamas J, Timms AE, Geister KA, Lindsay A, Beier DR. Variant mapping and
638 mutation discovery in inbred mice using next-generation sequencing. BMC Genomics.
639 2015;16:913. doi: 10.1186/s12864-015-2173-1. PubMed PMID: 26552429; PubMed
640 Central PMCID: PMCPMC4640199.
641 12. White TA, Pan S, Witt TA, Simari RD. Murine strain differences in hemostasis and
642 thrombosis and tissue factor pathway inhibitor. Thromb Res. 2010;125(1):84-9. doi:
643 10.1016/j.thromres.2009.03.006. PubMed PMID: 19398123; PubMed Central PMCID:
644 PMCPMC2826594.
645 13. Tomberg K, Khoriaty R, Westrick RJ, Fairfield HE, Reinholdt LG, Brodsky GL, et al.
646 Spontaneous 8bp Deletion in Nbeal2 Recapitulates the Gray Platelet Syndrome in Mice.
647 PLoS One. 2016;11(3):e0150852. doi: 10.1371/journal.pone.0150852. PubMed PMID:
648 26950939; PubMed Central PMCID: PMCPMC4780761.
649 14. Li Y, Klena NT, Gabriel GC, Liu X, Kim AJ, Lemke K, et al. Global genetic analysis in
650 mice unveils central role for cilia in congenital heart disease. Nature. 2015;521(7553):520-
651 4. doi: 10.1038/nature14269. PubMed PMID: 25807483; PubMed Central PMCID:
652 PMCPMC4617540.
653 15. Fairfield H, Gilbert GJ, Barter M, Corrigan RR, Curtain M, Ding Y, et al. Mutation
654 discovery in mice by whole exome sequencing. Genome Biol. 2011;12(9):R86. Epub
655 2011/09/16. doi: 10.1186/gb-2011-12-9-r86
656 gb-2011-12-9-r86 [pii]. PubMed PMID: 21917142; PubMed Central PMCID: PMC3308049.
657 16. Andrews TD, Whittle B, Field MA, Balakishnan B, Zhang Y, Shao Y, et al. Massively
658 parallel sequencing of the mouse exome to accurately identify rare, induced mutations: an

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

659 immediate source for thousands of new mouse models. Open Biol. 2012;2(5):120061. Epub
660 2012/06/23. doi: 10.1098/rsob.120061
661 rsob120061 [pii]. PubMed PMID: 22724066; PubMed Central PMCID: PMC3376740.
662 17. Justice MJ, Noveroske JK, Weber JS, Zheng B, Bradley A. Mouse ENU mutagenesis.
663 Human Molecular Genetics. 1999;8(10):1955-63.
664 18. Arnold CN, Barnes MJ, Berger M, Blasius AL, Brandl K, Croker B, et al. ENU-induced
665 phenovariance in mice: inferences from 587 mutations. BMC Res Notes. 2012;5:577. Epub
666 2012/10/26. doi: 10.1186/1756-0500-5-577
667 1756-0500-5-577 [pii]. PubMed PMID: 23095377; PubMed Central PMCID: PMC3532239.
668 19. Gibson WT, Hood RL, Zhan SH, Bulman DE, Fejes AP, Moore R, et al. Mutations in
669 EZH2 cause Weaver syndrome. Am J Hum Genet. 2012;90(1):110-8. Epub 2011/12/20. doi:
670 10.1016/j.ajhg.2011.11.018
671 S0002-9297(11)00496-4 [pii]. PubMed PMID: 22177091; PubMed Central PMCID:
672 PMC3257956.
673 20. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo
674 mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet. 2010;42(6):483-5. Epub
675 2010/05/04. doi: 10.1038/ng.581
676 ng.581 [pii]. PubMed PMID: 20436468.
677 21. Riviere JB, van Bon BW, Hoischen A, Kholmanskikh SS, O'Roak BJ, Gilissen C, et al. De
678 novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. Nat
679 Genet. 2012;44(4):440-4, S1-2. Epub 2012/03/01. doi: 10.1038/ng.1091
680 ng.1091 [pii]. PubMed PMID: 22366783; PubMed Central PMCID: PMC3677859.

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

681 22. Tsurusaki Y, Okamoto N, Ohashi H, Kosho T, Imai Y, Hibi-Ko Y, et al. Mutations
682 affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat Genet.
683 2012;44(4):376-8. Epub 2012/03/20. doi: 10.1038/ng.2219
684 ng.2219 [pii]. PubMed PMID: 22426308.
685 23. Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, Phipps EL. Specific-
686 locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc Natl
687 Acad Sci U S A. 1979;76(11):5818-9. Epub 1979/11/01. PubMed PMID: 293686; PubMed
688 Central PMCID: PMC411742.
689 24. Sang F, Jiang P, Wang W, Lu Z. ReDB: A meiotic homologous recombination rate
690 database. Chinese Science Bulletin. 2010;55(27-28):3169-73. doi: 10.1007/s11434-010-
691 3029-3.
692 25. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, et al.
693 Comparative recombination rates in the rat, mouse, and human genomes. Genome Res.
694 2004;14(4):528-38. doi: 10.1101/gr.1970304. PubMed PMID: 15059993; PubMed Central
695 PMCID: PMCPMC383296.
696 26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A
697 method and server for predicting damaging missense mutations. Nat Methods.
698 2010;7(4):248-9. doi: 10.1038/nmeth0410-248. PubMed PMID: 20354512; PubMed
699 Central PMCID: PMCPMC2855889.
700 27. Bannister LA, Pezza RJ, Donaldson JR, de Rooij DG, Schimenti KJ, Camerini-Otero RD,
701 et al. A dominant, recombination-defective allele of Dmc1 causing male-specific sterility.
702 PLoS Biol. 2007;5(5):e105. doi: 10.1371/journal.pbio.0050105. PubMed PMID: 17425408;
703 PubMed Central PMCID: PMC1847842.

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

704 28. Rieder MJ, Green GE, Park SS, Stamper BD, Gordon CT, Johnson JM, et al. A human
705 homeotic transformation resulting from mutations in PLCB4 and GNAI3 causes
706 auriculocondylar syndrome. Am J Hum Genet. 2012;90(5):907-14. doi:
707 10.1016/j.ajhg.2012.04.002. PubMed PMID: 22560091; PubMed Central PMCID:
708 PMCPMC3376493.
709 29. Cui J, Eitzman DT, Westrick RJ, Christie PD, Xu ZJ, Yang AY, et al. Spontaneous
710 thrombosis in mice carrying the factor V Leiden mutation. Blood. 2000;96(13):4222-6.
711 30. Huang ZF, Higuchi D, Lasky N, Broze GJ, Jr. Tissue factor pathway inhibitor gene
712 disruption produces intrauterine lethality in mice. Blood. 1997;90:944-51.
713 31. Boehnke M, Ploughman L. SIMLINK: A Program for Estimating the Power of a
714 Proposed Linkage Study by Computer Simulations. Version 4.12, April 2, 1997.
715 32. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, et al.
716 Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics.
717 2012;28(20):2678-9. doi: 10.1093/bioinformatics/bts503. PubMed PMID: 22914218.
718 33. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
719 transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi:
720 10.1093/bioinformatics/btp324
721 btp324 [pii]. PubMed PMID: 19451168; PubMed Central PMCID: PMC2705234.
722 34. Picard tools. Available from: http://picard.sourceforge.net.
723 35. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A
724 framework for variation discovery and genotyping using next-generation DNA sequencing
725 data. Nat Genet. 2011;43(5):491-8. Epub 2011/04/12. doi: 10.1038/ng.806
726 ng.806 [pii]. PubMed PMID: 21478889; PubMed Central PMCID: PMC3083463.

bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

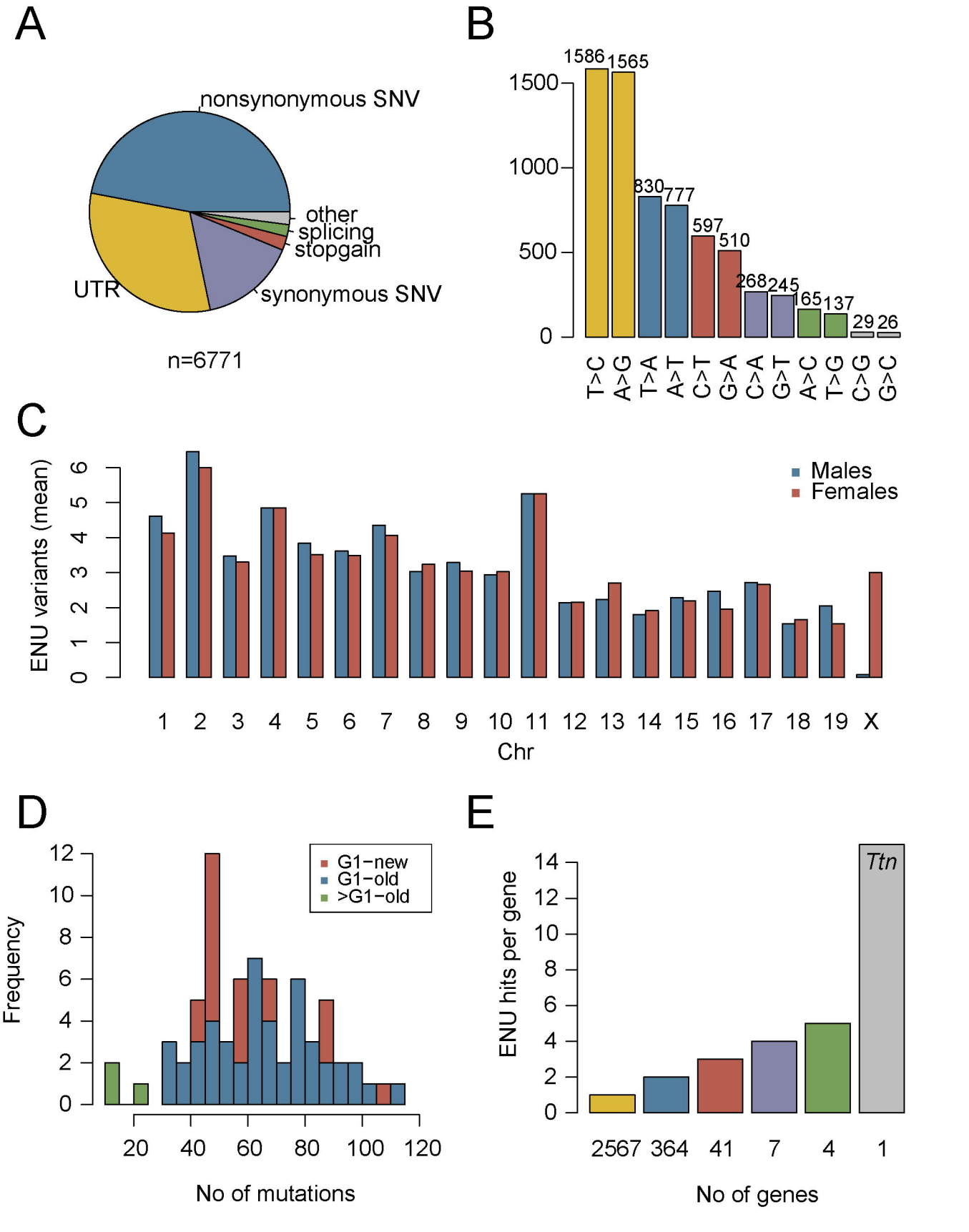
727 36. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants
728 from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164. Epub
729 2010/07/06. doi: 10.1093/nar/gkq603
730 gkq603 [pii]. PubMed PMID: 20601685; PubMed Central PMCID: PMC2938201.
731 37. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and
732 Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300. PubMed
733 PMID: WOS:A1995QE45300017.
734 38. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-
735 -new capabilities and interfaces. Nucleic Acids Res. 2012;40(15):e115. doi:
736 10.1093/nar/gks596. PubMed PMID: 22730293; PubMed Central PMCID:
737 PMCPMC3424584.
738 39. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational
739 design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat
740 Biotechnol. 2014;32(12):1262-7. doi: 10.1038/nbt.3026. PubMed PMID: 25184501;
741 PubMed Central PMCID: PMCPMC4262738.
742 40. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA
743 targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31(9):827-32.
744 doi: 10.1038/nbt.2647. PubMed PMID: 23873081; PubMed Central PMCID:
745 PMCPMC3969858.
746 41. Pettitt SJ, Liang Q, Rairdan XY, Moran JL, Prosser HM, Beier DR, et al. Agouti
747 C57BL/6N embryonic stem cells for mouse genetic resources. Nat Methods. 2009;6(7):493-
748 5. doi: 10.1038/nmeth.1342. PubMed PMID: 19525957; PubMed Central PMCID:
749 PMCPMC3555078.

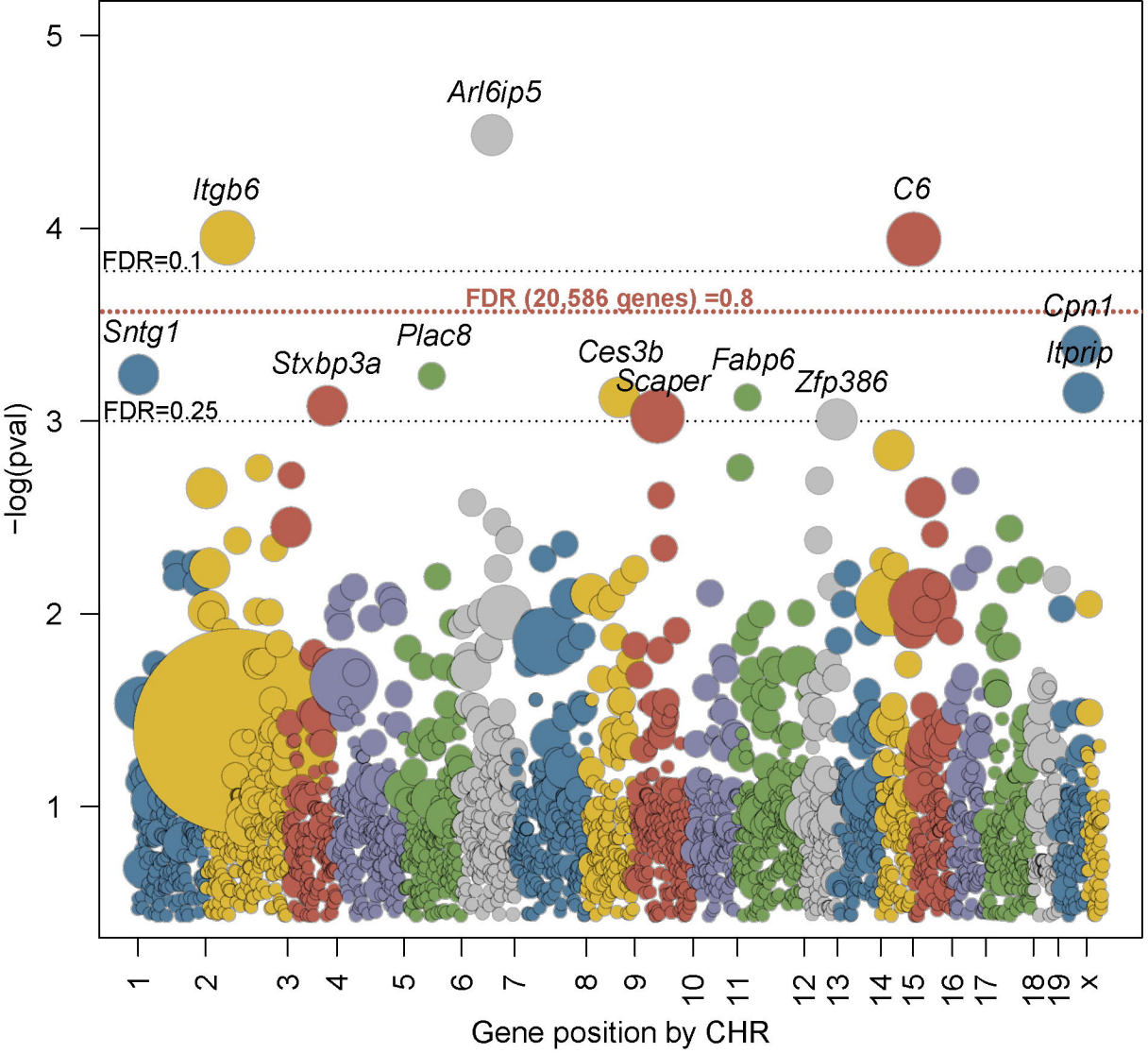
bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

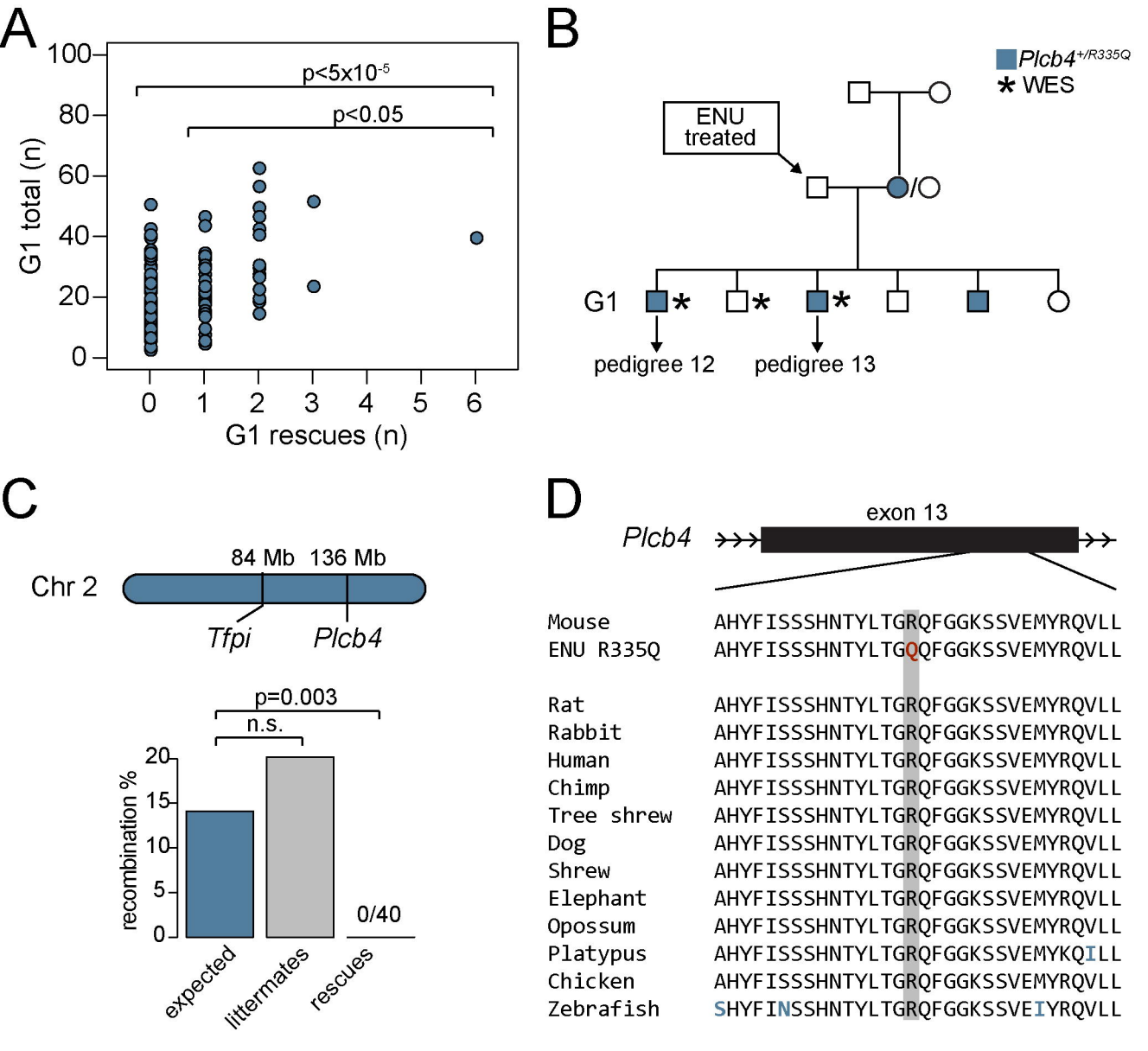
750 42. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome
751 engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23. doi:
752 10.1126/science.1231143. PubMed PMID: 23287718; PubMed Central PMCID:
753 PMCPMC3795411.
754 43. Pease S, Saunders TL, International Society for Transgenic Technologies. Advanced
755 protocols for animal transgenesis an ISTT manual. Heidelberg ; New York ;: Springer; 2011.
756 Available from:
757 http://proxy.lib.umich.edu/login?url=http://link.springer.com/openurl?genre=book&isbn
758 =978-3-642-20791-4.
759 44. McBurney MW, Fournier S, Jardine K, Sutherland L. Intragenic regions of the murine
760 Pgk-1 locus enhance integration of transfected DNAs into genomes of embryonal carcinoma
761 cells. Somat Cell Mol Genet. 1994;20(6):515-28. PubMed PMID: 7892649.
762 45. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of
763 genome editing by sequence trace decomposition. Nucleic Acids Res. 2014;42(22):e168.
764 doi: 10.1093/nar/gku936. PubMed PMID: 25300484; PubMed Central PMCID:
765 PMCPMC4267669.
766 46. Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, et al. CRISPR/Cas9 targeting
767 events cause complex deletions and insertions at 17 sites in the mouse genome. Nat
768 Commun. 2017;8:15464. doi: 10.1038/ncomms15464. PubMed PMID: 28561021; PubMed
769 Central PMCID: PMCPMC5460021.
770 47. Therneau TM, Grambsch PM. Modeling survival data : extending the Cox model. New
771 York: Springer; 2000. xiii, 350 p. p.

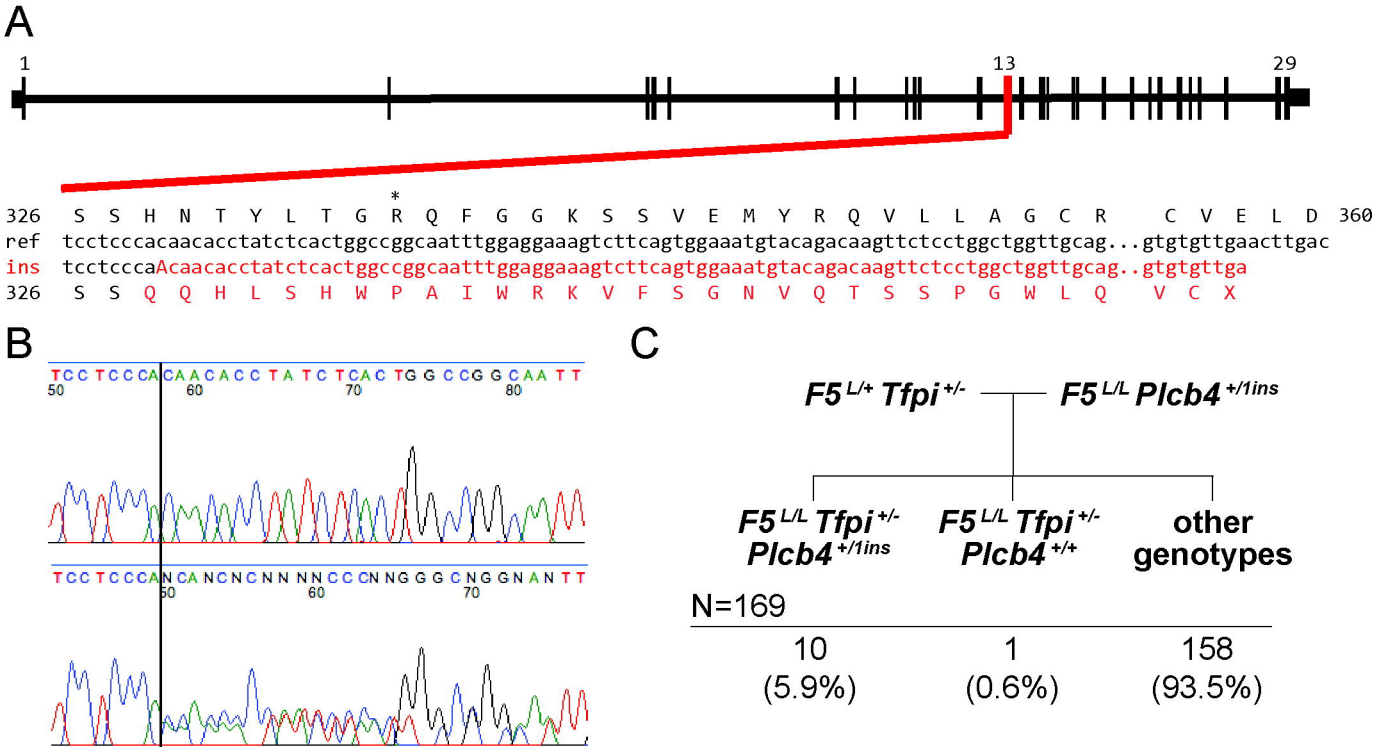
bioRxiv preprint doi: https://doi.org/10.1101/174086; this version posted March 24, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

772 48. R: A Language and Environment for Statistical Computing Vienna, AustriaR Core
773 Team. Available from: http://www.R-project.org/.
774 49. Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, Sobel EM. Mendel: the Swiss
775 army knife of genetic analysis programs. Bioinformatics. 2013;29(12):1568-70. doi:
776 10.1093/bioinformatics/btt187. PubMed PMID: 23610370; PubMed Central PMCID:
777 PMCPMC3673222.
778 50. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for
779 interpreting and reporting linkage results. Nat Genet. 1995;11(3):241-7. Epub
780 1995/11/01. doi: 10.1038/ng1195-241. PubMed PMID: 7581446.
781


A
nonsynonymous SNV
other
splicing
stopgain
UTR
synonymous SNV
n=6771
B
1586
1565
830
777
597
510
268
245
165
137
29
26
1500
1000
500
0
T>C
A>G
T>A
A>T
C>T
G>A
C>A
G>T
A>C
T>G
C>G
G>C
C
ENU variants (mean)
0 1 2 3 4 5 6
Males
Females
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
Chr
D
Frequency
12
10
8
6
4
2
0
G1−new
G1−old
>G1−old
20 40 60 80 100 120
No of mutations
E
ENU hits per gene
14
12
10
8
6
4
2
0
Ttn
2567 364 41 7 4 1
No of genes


Arl6ip5
Itgb6
C6
FDR=0.1
FDR (20,586 genes) =0.8
Cpn1
Sntg1
Plac8
Ces3b
Fabp6
Itprip
Stxbp3a
Scaper
Zfp386
FDR=0.25
−log(pval)
5
4
3
2
1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
x
Gene position by CHR


A
G1 total (n)
G1 rescues (n)
p<5x10-5
p<0.05
B
Plcb4+/R335Q
WES
ENU treated
G1
pedigree 12
pedigree 13
C
Chr 2
84 Mb
136 Mb
Tfpi
Plcb4
p=0.003
n.s.
recombination %
0/40
expected
littermates
rescues
D
Plcb4
exon 13
Mouse AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
ENU R335Q AHYFISSSHNTYLTGQQFGGKSSVEMYRQVLL
Rat AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Rabbit AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Human AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Chimp AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Tree shrew AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Dog AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Shrew AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Elephant AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Opossum AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Platypus AHYFISSSHNTYLTGRQFGGKSSVEMYKQILL
Chicken AHYFISSSHNTYLTGRQFGGKSSVEMYRQVLL
Zebrafish SHYFINSSHNTYLTGRQFGGKSSVEIYRQVLL


A
1
13
29
326 S S H N T Y L T G R Q F G G K S S V E M Y R Q V L L A G C R C V E L D 360
ref tcctcccacaacacctatctcactggccggcaatttggaggaaagtcttcagtggaaatgtacagacaagttctcctggctggttgcag...gtgtgttgaacttgac
ins tcctcccaAcaacacctatctcactggccggcaatttggaggaaagtcttcagtggaaatgtacagacaagttctcctggctggttgcag..gtgtgttga
326 S S Q Q H L S H W P A I W R K V F S G N V Q T S S P G W L Q V C X
B
50
60
70
80
50
60
70
C
F5 L/+ Tfpi +/-
F5 L/L Plcb4 +/1ins
F5 L/L Tfpi +/- Plcb4 +/1ins
F5 L/L Tfpi +/- Plcb4 +/+
other genotypes
N=169
10 (5.9%)
1 (0.6%)
158 (93.5%)