Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled with machine learning

Darui Xu[a], Stephen Lyon[a], Chun Hui Bu[a], Sara Hildebrand[a], Jin Huk Choi[a,b], Xue Zhong[a], Aijie Liu[a], Emre E. Turer[a,c], Zhao Zhang[a], Jamie Russell[a], Sara Ludwig[a], Elena Mahrt[a], Evan Nair-Gill[a], Hexin Shi[a], Ying Wang[a], Duanwu Zhang[a,1], Tao Yue[a], Kuan-wen Wang[a], Jeffrey A. SoRelle[a], Lijing Su[a], Takuma Misawa[a,2], William McAlpine[a], Lei Sun[a,3], Jianhui Wang[a], Xiaoming Zhan[a], Mihwa Choi[a], Roxana Farokhnia[a], Andrew Sakla[a], Sara Schneider[a], Hannah Coco[a], Gabrielle Coolbaugh[a], Braden Hayse[a], Sara Mazal[a], Dawson Medler[a], Brandon Nguyen[a], Edward Rodriguez[a], Andrew Wadley[a], Miao Tang[a], Xiaohong Li[a], Priscilla Anderton[a], Katie Keller[a], Amanda Press[a], Lindsay Scott[a], Jiexia Quan[a], Sydney Cooper[a], Tiffany Collie[a], Baifang Qin[a], Jennifer Cardin[a], Rochelle Simpson[a], Meron Tadesse[a], Qihua Sun[a], Carol A. Wise[d,e,f,g], Jonathan J. Rios[d,e,f,g], Eva Marie Y. Moresco[a], and Bruce Beutler[a,4]

[a]Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX 75390; [b]Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX 75390; [c]Department of Internal Medicine, Division of Gastroenterology, University of Texas Southwestern Medical Center, Dallas, TX 75390; [d]Center for Pediatric Bone Biology and Translational Research, Scottish Rite for Children, Dallas, TX 75219; [e]McDermott Center for Human Growth & Development, University of Texas Southwestern Medical Center, Dallas, TX 75390; [f]Department of Orthopaedic Surgery, University of Texas Southwestern Medical Center, Dallas, TX 75390; and [g]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75390

Forward genetic studies use meiotic mapping to adduce evidence that a particular mutation, normally induced by a germline mutagen, is causative of a particular phenotype. Particularly in small pedigrees, cosegregation of multiple mutations, occasional unawareness of mutations, and paucity of homozygotes may lead to erroneous declarations of cause and effect. We sought to improve the identification of mutations causing immune phenotypes in mice by creating Candidate Explorer (CE), a machine-learning software program that integrates 67 features of genetic mapping data into a single numeric score, mathematically convertible to the probability of verification of any putative mutation–phenotype association. At this time, CE has evaluated putative mutation–phenotype associations arising from screening damaging mutations in ~55% of mouse genes for effects on flow cytometry measurements of immune cells in the blood. CE has therefore identified more than half of genes within which mutations can be causative of flow cytometric phenovariation in *Mus musculus*. The majority of these genes were not previously known to support immune function or homeostasis. Mouse geneticists will find CE data informative in identifying causative mutations within quantitative trait loci, while clinical geneticists may use CE to help connect causative variants with rare heritable diseases of immunity, even in the absence of linkage information. CE displays integrated mutation, phenotype, and linkage data, and is freely available for query online.

ENU mutagenesis | automated meiotic mapping | machine learning | flow cytometry | immune cells

Forward genetics begins with a phenotype, often induced by a random germline mutagen, and ends with the discovery of a causative mutation. We developed a process for rapid identification of causative mutations in mice carrying *N*-ethyl-*N*-nitrosourea (ENU)-induced germline mutations (1, 2). Our pipeline involves mutagenizing male C57BL/6J (G0) mice and breeding them on the C57BL/6J background to create G1 male pedigree founders, G2 daughters, and G3 mice of both sexes (*SI Appendix*, Fig. S1). We sequenced the exomes of all G1 founders of pedigrees, achieving >99% 10X coverage over the targeted exome. Identified variants (with respect to the C57BL/6J reference genome) are genotyped in G2 and G3 mice in advance of phenotypic screening. Using a variety of phenotypic screens, G3 mice are then tested for phenovariance with respect to C57BL/6J mice or a control population of G3 mice. Demonstrating linkage between a mutant

phenotype detected in screening and a particular mutation is accomplished by automated meiotic mapping (AMM) performed by the Linkage Analyzer software, which tests the null hypothesis for every mutation in the pedigree (i.e., "mutation A is unrelated to

Significance

We developed software called Candidate Explorer (CE) that uses a machine-learning algorithm to identify chemically induced mutations that are causative of screened phenotypes. CE determines the probability that a mutation will be verified as causative for a phenotype if the gene is independently targeted for knockout or recreation of the mutation. CE uses 67 parameters from the mapping data—including gene, mutation, genotype, allelism, and phenotype information—to determine the CE Score and verification probability. We used CE to evaluate ~87,000 mutation/phenotype associations identified by flow cytometry screening of circulating immune cells from mutagenized mice: 1,279 genes representing 2,336 mutations were rated good or excellent candidates for causation of phenotypes. Many of these genes were not previously implicated in immunity.

GENETICS

www.manaraa.com

phenotypic performance in screen α") (1). In contrast, a mutation associated with the mutant phenotype at a frequency greater than predicted by chance alone is likely to confer the phenotype. Rejection of the null hypothesis with a $P \leq 0.05$, with Bonferroni correction for multiple comparisons, has generally been considered suggestive of causation. Verification by an independently generated allele is necessary to confirm the association.

Experience with many thousands of mutation–phenotype associations identified by AMM and either verified or excluded by testing CRISPR/Cas9-targeted alleles, has shown that the P value determined by AMM is not the sole indicator of causation. That is, a mutation linked to a phenotype with a $P < 0.05$ is sometimes not the causative mutation. Many other factors, such as the nature of the mutation (benign, damaging, null), the essentiality of the gene for survival prior to weaning, pedigree size, the number of homozygotes tested, the magnitude of phenotypic effect, data variance characteristics of the screen in question, the number of distinct phenotypes caused by the mutation, the presence or absence of cosegregating mutations, and the observation of other alleles with similar effects, influence correct selection of an authentic causative mutation. These numerous considerations, not readily integrated into a decision by human observers, impelled us to develop Candidate Explorer (CE), a software tool employing a supervised machine-learning algorithm to estimate the likelihood of verification of any putative mutation–phenotype association implicated by AMM.

In this study, we focused on changes in immune cell populations, specifically B cells, T cells, conventional and plasmacytoid dendritic cells (DC), macrophages, neutrophils, natural killer (NK) cells, and NK1.1$^+$ T cells. Cell populations and subpopulations were detected and measured by flow cytometric analysis of peripheral blood leukocytes from G3 mutant mice carrying ENU-induced mutations. We present CE assessments of 87,795 mutation–phenotype associations ($P < 0.05$). CE has identified more than 1,270 genes with a high and defined probability of verifiable importance in leukocyte development or maintenance. Many of these genes were not previously known to be important in immune function.

Results

CE Overview. The purpose of CE is to aid the researcher in predicting whether a mutation associated with a phenotype by AMM is a truly causative mutation. CE evaluates mutation–phenotype associations that pass specific basal filters for conventionally good candidates. In this paper, we use as the default filters $P < 0.05$ (Bonferroni corrected), ≥ 10 mice in the tested pedigree, and ≥ 2 homozygous reference mice screened; however, more stringent criteria can be set by the user. The core of CE is a supervised machine-learning algorithm that outputs a numerical score (CE score), a categorical assessment (candidate status), and verification probability for each mutation–phenotype association based on input phenotype data (from screening), mutation data, gene data, and meiotic mapping data (Fig. 1A). CE is trained based on phenotypic assessment of mice carrying targeted null or replacement alleles of candidate genes (see below). In predicting, performed four times per day because of the dynamic status of the database, CE uses all defined features of the original pedigree screening data to estimate the probability of candidate verification. CE is publicly available for querying mutation–phenotype associations identified in flow cytometry screens, as well as radiographic screens of bone (DEXA scanning). An example of the use of CE is presented in Movie S1. In this paper, we present the results of flow cytometry screening.

CE Training. At present, the CE training set contains 1,903 verified and 3,013 excluded mutation–phenotype associations (4,916 assessments in all), based on germline retargeting of 514 genes. Germline retargeting was performed using CRISPR/Cas9 to

generate knockout alleles of the candidate genes in mice on a pure reference background (C57BL/6J or C57BL/6N). Alternatively, when evidence for homozygous lethality of null alleles existed (see *Essentiality score*, below) or the ENU mutation was suspected to cause hypermorphic, neomorphic, or antimorphic effects, the original ENU allele was recreated by CRISPR/Cas9 targeting (designated "replacement" allele). Mice carrying targeted germline knockout or replacement alleles were expanded to form pedigrees containing mice homozygous for the reference allele (REF), heterozygous (HET), and homozygous for the variant allele (VAR). Compound heterozygous mice with two or more variant alleles of a gene were sometimes also generated. Fresh pedigrees of mice carrying the CRISPR-targeted alleles were subjected to the phenotypic screens in which the original ENU mutations scored as hits. CRISPR-targeted mutations were considered verified according to the criteria:

1. Observation of the same phenotype with the same directionality of change as observed for the original ENU allele with a P value better than 0.01, or
2. Observation of the same phenotype with the opposite directionality of change as observed for the original ENU allele with a P value better than 0.001, or
3. De novo observation of a phenotype (not seen in the original screen) with a P value better than 0.001.

CE Output and Performance.

CE score and candidate status. The CE Score (range 0 to 1) is a class probability related by a polynomial function to the actual probability of verification by CRISPR-targeted alleles, as determined by regression analysis (Fig. 1B). In conjunction with the algorithmic score, it is used by CE to designate one of four possible candidate statuses for each mutation–phenotype association (excellent, good, potential, or not good) as follows:

1. Excellent candidate: CE score ≥ 0.39 and algorithmic score ≥ -0.5,
2. Good candidate: CE score ≥ 0.39 and $-4.5 \leq$ algorithmic score < -0.5,
3. Potential candidate: CE score ≥ 0.39 and algorithmic score < -4.5 OR CE score < 0.39 and algorithmic score ≥ -0.5
4. Not good candidate: CE score < 0.39 and algorithmic score < -0.5.

We generally choose good or excellent candidates for CRISPR/Cas9 targeting and further study. However, CE scores are not strictly proportional to the probability of verification (Fig. 1B), and some "good" or "excellent" candidates fail to verify. Conversely, "potential" and "not good" candidates will sometimes verify as true positive associations. We take it as a truism that authentic candidates will achieve strong CE scores as more alleles are obtained and tested (approaching saturation) and will therefore eventually be verified.

Performance. The performance of the CE prediction model established using the training set was assessed using the repeated 10-fold cross-validation method. The receiver operating characteristic (ROC) curve has an area under the curve (AUC) of 0.943 (Fig. 1C and *SI Appendix*, Fig. S2); the current cutoff is 0.39, corresponding to the point with the minimum distance to the upper left corner of the ROC curve. CE ranking of good or better corresponds to ~80% precision (correctly calling a verified candidate "true;" i.e., a 20% false-discovery rate) and 87% recall (true positive rate) (Table 1).

CE is also often capable of correctly identifying which mutation is causative when two or more mutations cosegregate (see also below, *Driven By status*). Among 961 such cases, CE correctly identified on average 76.5% of causative mutations as the top CE scorer, with generally better performance when fewer
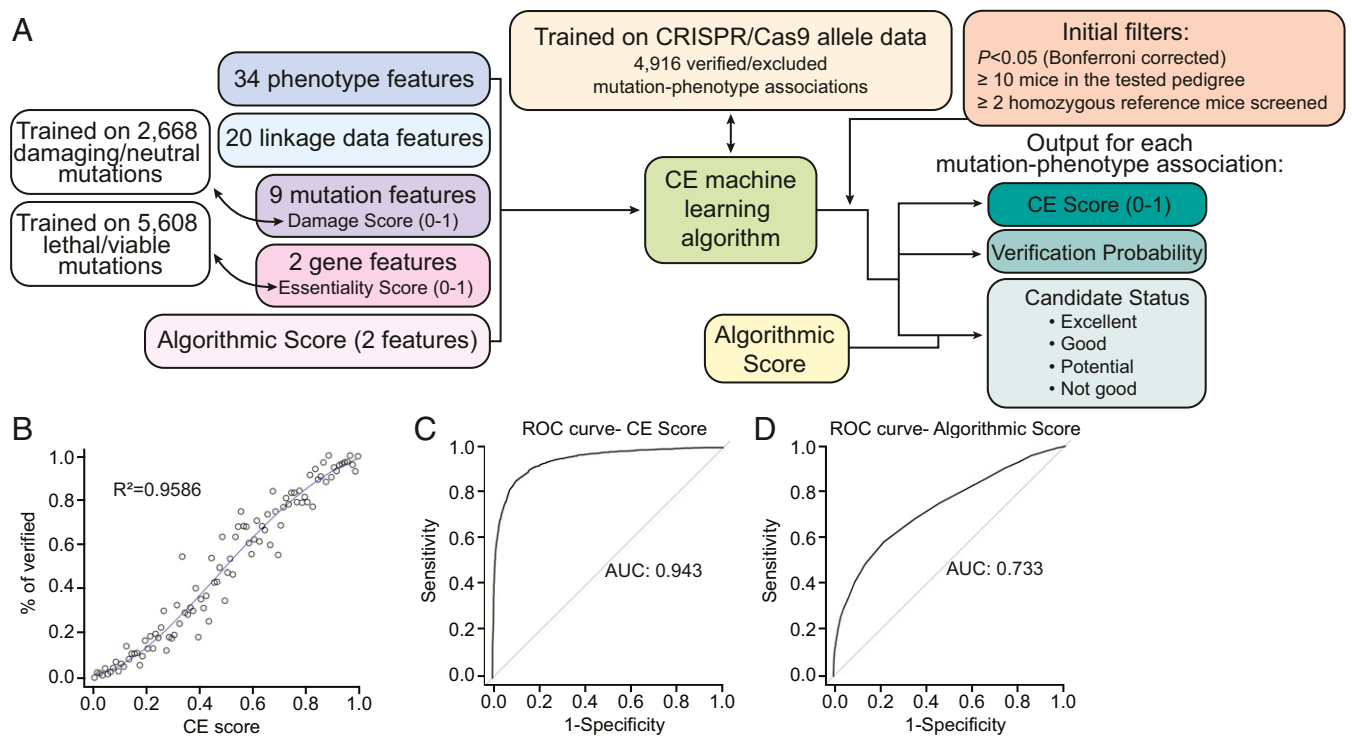
Fig. 1. CE overview and performance. (*A*) Schematic of input features and outputs from CE. (*B*) Polynomial regression analysis of CE score and average percentage of verified mutation–phenotype associations. Each data point represents a group of mutation–phenotype associations. The percentage of verified associations (*y* axis) is plotted versus CE Score range (*x* axis) in bins of 0.01 (e.g., 0.35 to 0.36, 0.37 to 0.38, and so forth). *n* = 4,916 mutation–phenotype associations and 514 CRISPR/Cas9-targeted genes. (*C* and *D*) ROC curves for CE Score (*C*) and algorithmic score (*D*).

mutations cosegregated (Table 2). As further training is performed, and as the total volume of screening data increases (with an attendant increase in the number of genes with allelism and the overall density of allelic series), CE performance will continue to improve.

Verification probability. Multiple alleles of a given gene may be subjected to a given phenotypic screen, resulting in several mutation–phenotype associations for the same gene and phenotype. Each mutation–phenotype association is independently accorded an allele verification probability (AVP) estimate for the mutation in question, extrapolated from the polynomial regression analysis of CE score and the average percentage of verified mutation–phenotype associations (Fig. 1*B*). In addition,

the composite estimate that one or more mutations within a certain gene will be verified as the source of a certain phenotype (gene verification probability, GVP) is given by:

$$GVP = 1 - (1 - AVP_1)(1 - AVP_2)(1 - AVP_3)\ldots(1 - AVP_N).$$

AVPs of alleles causing the same direction of phenotypic change in a given screen are included in the calculation.

Input Data Features. The CE prediction model currently incorporates 67 features of the input data (34 phenotype features, 20 linkage analysis features, 9 mutation features, 2 gene features, and 2 other features) (Table 3). The 20 most important features

Table 1. CE performance for flow cytometry phenotypes

Cutoff	True positives	False positives	True negatives	False negatives	Recall, %	Accuracy, %	Precision, %
Mutation–phenotype associations (*n* = 96,022; 1,310 verified, 1,757 excluded, 92,955 untested)							
Excellent	838	176	1,581	472	63.97	78.87	82.64
Good and above	1,140	286	1,471	170	87.02	85.13	79.94
Potential and above	1,209	669	1,088	101	92.29	74.89	64.38
Not good and above	1,310	1,757	0	0	100	42.71	42.71
Alleles (*n* = 3,020 named alleles; 275 verified, 112 excluded, 2,633 untested)							
Excellent	215	14	98	60	78.18	80.88	93.89
Good and above	246	21	91	29	89.45	87.08	92.13
Potential and above	257	59	53	18	93.45	80.10	81.33
Not good and above	275	112	0	0	100	71.06	71.06
Genes (*n* = 15,313; 146 causative, 108 noncausative, 15,059 untested)							
Excellent	114	19	89	32	78.08	79.92	85.71
Good and above	132	24	84	14	90.41	85.04	84.62
Potential and above	139	69	39	7	95.21	70.08	66.83
Not good and above	146	108	0	0	100	57.48	57.48

Xu et al.
Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled with machine learning

PNAS | 3 of 11
https://doi.org/10.1073/pnas.2106786118
www.manaraa.com

GENETICS

Table 2. CE performance in scoring colocalizing mutations

No. of colocalized genes	Total cases	CE ranked correctly	% Correct
2	591	514	87.0
3	233	208	89.3
4	91	76	83.5
5	32	22	68.8
6	6	5	83.3
7	6	6	100
8	1	0	0
10	1	1	100

are ranked in Table 4. The damage score and essentiality score (E-score) result from independent machine-learning programs. The rule-based algorithmic score results from computational execution of a fixed algorithm that was human designed.

Damage score. The damage score (range 0 to 1), a mutation feature, is the fifth most important feature overall in the CE algorithm, and has important biological relevance. The damage score denotes the likelihood that a protein is functionally impaired and is determined by a machine-learning algorithm that integrates 37 independent prediction scores from the human database for Nonsynonymous Functional Prediction (dbNSFP) and the probability of protein damage to phenovariance caused by mouse mutations (3). A higher score suggests a mutation is more likely to be deleterious, and therefore more likely to be causative (although not always the case). The current damage score prediction model was trained on 871 known deleterious mutations and 1,797 known neutral mutations; 666 mutations with known effects were used to test the performance of the established model, which yielded an ROC curve with AUC of 0.852 (*SI Appendix*, Fig. S3).

Essentiality score. The E-score (range 0 to 1) is a gene feature and denotes the likelihood of lethality prior to weaning age (4 wk postpartum) in mice homozygous for a robust knockout allele of a gene. The E-score is calculated using a machine-learning algorithm incorporating various independent features of genes, including gene conservation, protein–protein interaction network, expression stage, and viability/proliferative ability of human cell lines in which the gene is mutated. The E-score prediction model is trained at monthly intervals. The current training dataset consists of 3,538 known nonessential genes (E-score = 0) and 2,070 known essential genes (E-score = 1), determined based on annotations in the Mouse Genome Informatics (MGI) database and observed effects of CRISPR-targeted null mutations we generated in C57BL/6J mice. The current cutoff values are >0.5 for essential genes and <0.5 for nonessential genes, and are used to inform gene-targeting efforts, in which either a knockout allele or a replacement identical to the original ENU allele is created for verification of phenotype; 1,401 genes with known effects on viability were used to test the performance of the established model, yielding an ROC curve with AUC of 0.894 (*SI Appendix*, Fig. S4).

Algorithmic score. Assessments of mutation–phenotype associations are made using a human-developed algorithm that outputs a points-based score called the algorithmic score (current range −13.5 to 3.5). The algorithmic score appears twice among the most important features contributing to the CE algorithm (first and third in importance) (Table 4), and provides an overall assessment based on our (human researcher) experience of how likely the mutation is to be causative. The algorithm consists of a set of rules based on empirical observations (Table 5). For each feature supporting or opposing the authenticity of a mutation–phenotype association, respectively, the algorithmic score is increased or decreased. The features used in the algorithmic score calculation are similar to those used in the CE machine-learning algorithm, but static (not influenced by exposure to new training data), and the performance of the rule-based algorithm by itself falls short of the performance of the CE prediction model (Fig. 1D).

Driven By status. Another input feature to the CE algorithm is generated by a software program called Driven By, which evaluates both linked and unlinked candidate mutations to determine the best candidate. At times a cluster of linked mutations fails to undergo meiotic separation; hence, more than one mutation may stand as a candidate for causation of a phenotype. On other occasions, as a matter of happenstance, homozygotes for a noncausative, unlinked mutation may also be homozygous for a causative mutation. Usually this occurs when the number of homozygotes for the noncausative mutation is small. The Driven By program omits all instances of shared zygosity for both mutations and recomputes *P* values testing departure from the null hypothesis in recessive, additive, and dominant models of transmission, and determines which mutation is the more robust causation candidate. This mutation is assigned "driver" status. Based on driver status together with other factors (e.g., which mutation is the most damaging, which mutation is the most essential for survival to weaning age, and which mutation has evidence of other alleles with a similar phenotype), CE may be able to identify the causative mutation out of a set of colocalizing mutations, giving it a markedly superior CE score.

Finally, an allelic series probed with a phenotypic screen provides an extremely important clue to causation and is considered in CE assessments (Tables 3 and 4, multiple rows). If multiple alleles of the same gene are associated with the same phenotype, it is a strong indication that a mutation in this gene caused the observed phenotype. Superpedigrees—composites of multiple pedigrees assayed in the same screen—are of three types. Gene superpedigrees pool different and identical alleles of a given gene, subjected to the same screen. Position superpedigrees pool identical alleles only. Identical alleles may result from: 1) chance mutation of the same nucleotide, 2) transmission of a single mutation to multiple G1 descendants of a single G0 mouse, and 3) a background mutation present in mutagenized stock and shared by multiple G0 mice. Selective gene superpedigrees incorporate only alleles associated with *P* values < 0.05 with a common direction of effect in a given phenotypic screen, and thus give an intentionally biased view of mutation effects. Because many (but not all) ENU-induced mutations are functionally hypomorphic, a selective gene superpedigree for a set of mutations in a particular gene can strongly implicate that gene in the phenotype probed by the screen in question. The number of pedigrees (and alleles) tested is also important; for very large genes, hundreds of alleles may have been tested, and the finding that two or three alleles score in a particular screen may be due to chance alone. CE takes account of this in computing probability of causation (e.g., Table 4, #8 and #9).

CE Assessments of 87,795 Mutation–Phenotype Associations Identified by AMM in Flow Cytometry Screens. The flow cytometry screens survey 42 parameters of peripheral blood cells, measuring the frequencies of various immune cell populations and expression levels of several cell surface markers (Table 6). Of 7,109,669 mutation–phenotype associations tested by AMM in the flow cytometry screens, 87,795 passed the default initial filters, permitting analysis by CE. These putative mutation–phenotype associations emanated from 39,685 mutations in 14,809 genes, resident in 142,653 G3 mice from 3,987 pedigrees. Restriction to good or excellent candidates reduced the number of mutation–phenotype associations to 7,676, emanating from 2,336 mutations in 1,279 genes, resident in 1,634 pedigrees (Dataset S1; see also CE online for the most updated dataset). Gene–phenotype associations for the 1,279 genes (those with at least one good/excellent mutation–phenotype association) are displayed in a heatmap in Dataset S2.

We could make several observations concerning gene–phenotype associations (Dataset S2). First, mutations in the majority (872

Table 3. Features of input data in CE prediction algorithm

Input data	Features
Phenotype data	**The percentage of VAR mice whose screen results overlap with those of B6 mice**
	The percentage of VAR mice whose screen results overlap with those of REF mice
	Difference between HET and VAR results
	Direction of the results (whether the average of VAR screening results is greater or less than the average of REF screening results)
	Difference between REF and VAR results
	Number of female HET mice
	Number of female REF mice
	Number of male REF mice
	Number of male HET mice
	Number of male VAR mice
	Number of female VAR mice
	The identity of the phenotype (e.g., FACS T cell)
	The group identity of the phenotype (e.g., FACS screen or bone screens)
	The number of outliers in REF mice
	The number of outliers in HET mice
	The number of outliers in VAR mice
	Difference between REF and B6 results
	Difference between REF and HET results
	Whether the variance of REF is big
	Whether the variance of HET is big
	Whether the variance of VAR is big
	Whether the average age of the mice for this mutation/phenotype is older than the average age of all mice tested for this phenotype
	Whether the average age of the VAR mice is younger than the average age of the REF mice
	Number of pedigrees this gene/phenotype has
	The direction of the position superpedigree results for this mutation/phenotype
	Number of significant* single pedigrees in the significant position superpedigree for this mutation/phenotype
	Number of pedigrees included in the significant position superpedigree results for this mutation/phenotype
	The direction of the gene superpedigree results (null alleles) for this phenotype
	The direction of the gene superpedigree results (null+missense alleles) for this phenotype
	Whether there are corresponding trimmed results for the untrimmed data (only when VAR results are greater than REF results)†
	How closely VAR results resemble B6 results
	How closely HET results resemble B6 results
	How closely REF results resemble B6 results
	Whether REF and B6 results are different
Linkage data	**Average number of Linkage Analyzer runs with $P < 0.00005$ for each allele of this gene**
	Number of phenotypes with significant selective gene superpedigree results for this gene
	Number of Linkage Analyzer runs with $P < 0.00005$ for this gene
	Number of pedigrees in the selective gene superpedigree and whether the result is significant for this gene/phenotype
	Number of pedigrees contributing to a significant gene superpedigree result (null alleles)
	Number of pedigrees in a significant gene superpedigree result (null alleles)
	The minimum P value of single Linkage Analyzer result for this mutation/phenotype
	The percentage of body weight screens with $P < 0.0001$ for this mutation
	The percentage of FACS screens with $P < 0.0001$ for this mutation
	Whether the gene superpedigree results are significant (null+missense) for this phenotype
	Whether P value is significant in both raw and normalized assays for this mutation/phenotype
	Whether the minimum P value is for a recessive model of inheritance (rather than dominant or additive)
	Whether this phenotype is driven by another mutation
	The percentage of DSS screens with $P < 0.0001$ for this mutation
	Number of FACS phenotypes with $P < 0.0001$ for this mutation
	Number of DSS phenotypes with $P < 0.0001$ for this mutation
	Number of body weight phenotypes with $P < 0.0001$ for this mutation
	Whether the position superpedigree results are significant for this mutation/phenotype
	Whether the gene superpedigree results are significant (null alleles) for this phenotype
	Whether the gene superpedigree results are significant (missense alleles) for this phenotype
Mutation data	**Damage score for this mutation**
	Number of alleles this gene has
	Whether the mutation is autosomal
	Whether the mutation is colocalized with another mutation for this phenotype
	Whether the mutation is colocalized with a verified mutation for this phenotype

GENETICS

Xu et al.
Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled with machine learning

PNAS | 5 of 11
https://doi.org/10.1073/pnas.2106786118

www.manaraa.com

Table 3. Cont.

Input data	Features
	Whether the mutation is colocalized with an excluded mutation for this phenotype
	Whether the mutation is colocalized with a mutation of higher damage score
	The number of splice variants for the gene containing this mutation
	The ratio of number of named mutations vs. number of incidental mutations for this amino acid change
Gene data	The P value for a lethal phenotype
	The probability that the gene is an essential gene (E-score)
Other	**Number of phenotypes with algorithmic score greater or equal to −0.5 for this mutation**
	Algorithmic score for this mutation/phenotype

The top 20 most important features are in boldface. B6, C57BL/6J.
*Significant pedigree refers to linkage analysis of a pedigree or superpedigree by AMM in which $P < 0.05$ for a mutation-phenotype association.
†Trimmed results = raw data normalized for cell viability.

genes, 68.2%) of the 1,279 genes resulted in three or fewer good/excellent phenotype associations, with 533 genes (41.7%) having a single good/excellent phenotype association (Fig. 2A). In contrast, only 30 genes (2.3%) had at least 20 good/excellent phenotype associations, and among them 26 are well-known immune regulatory genes. Second, we found that the number of good/excellent gene associations varied widely depending on the cell-type affected, with B cell and T cell phenotypes associated with the most genes and conventional and plasmacytoid DC phenotypes associated with very few genes (Fig. 2B). Finally, 449 genes (35.1%) known or predicted to be essential for viability (E-score > 0.55 in this case) were associated with at least one flow cytometry phenotype, indicating that numerous developmentally important genes likely also have postnatal functions in leukocytes (Fig. 2C).

A total of 1,354 mutations in 667 genes rated good/excellent by CE and suspected or proven causative of flow cytometry phenotypes were given allele names and annotated as phenotypic mutations in the Mutagenetix database, irrespective of present candidate status (Dataset S3). While we consider that named alleles are very likely causative, we cannot be certain that unnamed alleles are not also causative; indeed, 27% of named alleles had AVP ≤ 0.5. Some of the unnamed alleles are designated as "linked to" or "driven by" another mutation in the same pedigree. This may indicate that they are not causative, but does not always guarantee it, and in some cases two named alleles are linked, suggesting that we have declared both mutations to be causative (even though they may cosegregate). Definitive evidence for such dual causation can only be adduced by CRISPR/Cas9 targeting.

We searched for highly represented gene ontology (GO) annotations associated with the 667 genes with named alleles (Datasets S4 and S5). As expected, the biological process annotations were most highly enriched for terms related to immune system processes (211 genes, $P = 9.82e-42$), lymphocyte activation (113 genes, $P = 5.21e-39$), immune system development (117 genes, $P = 9.73e-36$), and other immune development/regulatory processes, which was consistent with our manual evaluation identifying 281 (42.1%) of the 667 genes as previously known immune regulators (Datasets S4 and S5). By manual evaluation, 386 genes represented "new" immunologically important genes, each necessary for a normal flow cytometry profile. For many of these genes, mutant alleles were not previously available in mice and no primary immunological or other phenotypic data were available. This may be due in part to known or predicted lethality caused by null alleles of 146 of these 386 genes (E-score > 0.5). Enriched GO terms associated with the 386 new immunologically important genes were dominated by metabolic process terms, including cellular metabolic process (232 genes, $P = 3.73e-12$), organic substance metabolic process (240 genes, $P = 8.50e-12$), cellular macromolecule metabolic process (178 genes, $P = 3.59e-7$), and protein metabolic process (127 genes, $P = 0.000264$) (Dataset S6). We also assigned the 386

genes to a defined set of broad GO annotations for biological processes without regard for enrichment (Dataset S7). Based on its granular GO annotations, each gene was assigned to any of 70 parent GO terms to which it was related. Notably, 31 of the 386 genes were associated with the term "immune system process," based upon genetic interactions, an immune system association of an ancestral gene, sequence orthology to another gene associated with immune system process, or association of the orthologous human gene with an immune system process. In addition, 300 of the 386 genes were detected by RNA-sequencing with medium (11 to 1,000 transcripts per million) or high (>1,000 transcripts per million) expression in the spleen and/or thymus (4).

At present, a total of 603 genes implicated in flow cytometry phenotypes show a GVP > 0.5, 332 genes show a GVP > 0.8, and 222 genes show a GVP > 0.95. The genes with GVP > 0.95, from which flow cytometry phenotypes are nearly certain to emanate, are listed in Dataset S8; 121 (55%) of these genes are known to affect flow cytometry measurements and 101 (45%) are novel.

Discussion

CE allows rapid examination of mutations and genes strongly predicted to affect (or not to affect) phenotypes of interest measured in forward genetic screening. In general, CE is superior to the human researcher in evaluating mutation–phenotype associations because of its ability to integrate parameters not intuitively favorable or detrimental with respect to linkage analysis, and because it can perform this evaluation more rapidly on a large scale. Using the numerical CE score and categorical assessment given by CE, it is simple to rank mutations into priority lists for further in-depth study. In addition, causative mutations can frequently be discerned among several colocalizing mutations. As millions of coding/splicing mutations are introduced into the mouse genome pedigree by pedigree, more extensive allelic series will result, and nearly all genes in which causative loss-of-function mutations can exist will be identified with high confidence. CE is a tool necessary to deconvolute causation and permit this to occur.

Beyond its use as a tool for rapid identification of the mutations responsible for ENU-induced phenotypes, CE should be exceptionally useful to mouse geneticists studying complex traits (e.g., the Collaborative Cross). Meiotic mapping may confine phenotypes to a relatively large genomic interval, within which many candidate genes with mutational differences exist. If the phenotype is immunologic, knowledge of all genes from which flow cytometric phenotypes emanate is an important starting point for studies of causation, wherein these genes can be targeted.

CE will also have value to clinical geneticists seeking to identify causes of human disease. For patients with immunopathology and flow cytometric anomalies—but no mutation in a "classic" causative gene—other gene variants may be evaluated using CE. Mouse gene symbols corresponding to all loci mutated in the

Table 4. Top 20 most important features of input data in CE prediction algorithm

Rank	Feature
1	Number of phenotypes with algorithmic score greater or equal to −0.5 for this mutation
2	Average number of Linkage Analyzer runs with $P < 0.00005$ for each allele of this gene
3	Algorithmic score for this mutation/phenotype
4	Number of Linkage Analyzer runs with $P < 0.00005$ for this gene
5	Damage score for this mutation
6	Number of pedigrees in the selective gene superpedigree and whether the result is significant* for this gene/phenotype
7	Number of phenotypes with significant selective gene superpedigree results for this gene
8	Number of pedigrees contributing to a significant gene superpedigree result (null alleles)
9	Number of pedigrees in a significant gene superpedigree run (null alleles)
10	The percentage of FACS screens with $P < 0.0001$ for this mutation
11	The minimum P value of single Linkage Analyzer result for this mutation/phenotype
12	The percentage of VAR mice whose screen results overlap with those of B6 mice
13	Whether the gene superpedigree results are significant (null+missense alleles)
14	Whether the gene superpedigree results are significant (null alleles)
15	The percentage of VAR mice whose screen results overlap with those of REF mice
16	Difference between HET and VAR results
17	Number of female REF mice
18	The percentage of body weight screens with $P < 0.0001$ for this mutation
19	Number of female HET mice
20	Difference between REF and VAR results

Ranked from most (1) to least (20) important.
*Significant pedigree refers to linkage analysis of a pedigree or superpedigree by AMM in which $P < 0.05$ for a mutation-phenotype association.

patient (identified by whole-genome or whole-exome sequencing) can be entered into CE and searched as a batch. Those found to cause a flow cytometric abnormality in the mouse evocative of that in the patient may be considered prime candidates. If genetic mapping has been performed in a human family and a particular chromosomal region has been identified, the identification of a candidate gene can be made with even higher confidence using CE, which also accepts human chromosome coordinates as search input. By using CE in conjunction with analyses of large human genome/phenotype datasets (e.g., UK Biobank), CE may also facilitate and accelerate identification of causal variants within disease-associated loci found by genome-wide association studies (GWAS). CE could be queried for relevant mutation–phenotype associations for each candidate gene within a locus identified by GWAS; a mouse gene variant associated with a phenotype similar to the human phenotype under study would suggest causality. Moreover, a mutant mouse can, in most cases, be ordered immediately from the Mutant Mouse Resource and Research Centers, providing a model of the human disease for laboratory study. Because the majority of mutations cause loss-of-function (rather than gain-of-function or new functions), and the majority of mouse genes have human orthologs or homologs, many such cases might quickly be solved. Thus, CE is a powerful resource that addresses the question of "missing heritability" associated with immune abnormalities, and as noted for the 386 new genes (Dataset S6),

genes that regulate or mediate cellular metabolic processes may be prime candidates for consideration.

In this paper, we evaluated mutation–phenotype associations representative of 14,809 genes with one or more variant alleles and 42 flow cytometric parameters of peripheral blood leukocytes. Flow cytometric analyses allow detection and measurement of immune cell populations with specific functional correlates, and provide insight into the developmental stages cells traverse. Abnormal flow cytometry patterns are often associated with immune dysfunction, and many immunodeficiency and autoimmune phenotypes were initially detected not by functional screens per se, but by analyzing the peripheral blood with flow cytometry. Human disease states, exhibiting similar or identical flow cytometry phenotypes, attest to the clinical relevance of many mouse flow cytometry abnormalities (5–12). We have to date achieved ~55% genome saturation in screening 42 flow cytometry parameters, from which we identified 1,004 genes with good/excellent phenotype associations not previously associated with immune function (from GO analysis of the 1,279 genes, which found that 275 were associated with "immune system process"). Thus, even with a false-discovery rate up to 20%, we expect that about 456 more new immunologically important genes remain to be found.

In broadly surveying all 1,279 genes with at least one good/excellent phenotype association, we observed that a far greater percentage of genes had one, two, or three good/excellent phenotype associations (68.2%) compared to the percentage with many (≥20) good/excellent phenotype associations (2.3%). These findings suggest that the majority of genes affecting immune cell populations in the blood carry out cell type- or phenotype-specific functions. We are investigating the hypothesis that identical or similar combinations of phenotypes affected by two or more genes can indicate the functioning of those genes in a common molecular pathway. We also observed that good/excellent gene associations did not affect cell populations with equal frequency despite uniform phenotypic testing across all screens. For example, T cells had 4.8-fold more gene associations than conventional DC, 12.5-fold more than plasmacytoid DC, and 4.4-fold more than neutrophils. While a trivial explanation is that significant phenotypic differences are detected less often for rarer blood cell populations, another possibility reflecting the biology of the cells is that T cells are intrinsically less tolerant of genetic variation than conventional DC, plasmacytoid DC, or neutrophils, at least with respect to the numbers of these cells represented in the peripheral blood. An understanding of individual protein function and the pathways they regulate is critical to gain insight into these issues.

The vast majority of mice phenotyped by flow cytometry were also phenotyped in other screens, among them screens measuring responses to immunization, innate immune responses, body weight, blood pressure, heart rate, dextran sodium sulfate (DSS) sensitivity, circadian rhythms, and motor coordination. Data from screens for skeletal phenotypes detected by DEXA scanning are currently publicly accessible. In the future, the data from other screens will be released for public users of CE to interpret a wide range of phenotypic consequences that emanate from each mutation. All biomedically relevant phenotypic screens may ultimately enlighten the study of human phenotype and help to distinguish mechanisms of phenotypes caused by certain alleles, as many mutations score in disparate screens (for example, immune function and body weight, or immune function and neurobehavioral function).

Materials and Methods

Mice. Eight- to 10-wk-old C57BL/6J males purchased from The Jackson Laboratory were mutagenized with ENU, as described previously (13). Mutagenized G0 males were bred to C57BL/6J females, and the resulting G1 males were crossed to C57BL/6J females to produce G2 mice. G2 females were backcrossed to their G1 sires to yield G3 mice, which were screened for phenotypes.

Xu et al.
Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled with machine learning

PNAS | 7 of 11
https://doi.org/10.1073/pnas.2106786118

Table 5. Rules for algorithmic score determination

Feature	Points
REF outliers*	−1
HET outliers*	−1
VAR outliers*	−1
HET results have big variance[†]	−1
VAR results have big variance[†]	−1
REF and B6 results are different[‡]	−1
REF and VAR results overlap[§]	−1 to −3
B6 and VAR results overlap[§]	−0.5
HET results more similar than REF results to B6 results[¶]	−1 to −2
VAR results more similar than REF results to B6 results[#]	−1 to −2
Magnitude of change less than 2-fold for FACS B-1 B cell phenotype	−0.5
Magnitude of change less than 1.5-fold for FACS B-1 B cell phenotype	−3
Magnitude of change less than 2-fold for FACS B-2 B cell phenotype	−0.5
Magnitude of change less than 1.5-fold for FACS B-2 B cell phenotype	−3
Magnitude of change less than 2-fold for FACS DC phenotype	−0.5
Insignificant position superpedigree result	−1
Significant[‖] position superpedigree result (only minority of pedigrees contributed)	−1
In opposite direction of significant position superpedigree result	−3
Significant position superpedigree result	1.5
In opposite direction of significant gene superpedigree result (null alleles)	−3
Insignificant gene superpedigree result (null allele)	−1
Significant gene superpedigree result (null alleles)	1
In opposite direction of significant gene superpedigree result (null+missense alleles)	−0.5
Insignificant gene superpedigree result (null+missense alleles)	−0.5
Significant gene superpedigree result (null+missense alleles)	0.5
In opposite direction of significant gene superpedigree result (missense alleles)	−0.5
Insignificant gene superpedigree result (missense alleles)	−0.5
Significant gene superpedigree result (missense alleles)	0.5
Significant selective gene superpedigree result with more than two pedigrees	3
Significant selective gene superpedigree result with two pedigrees	2
In opposite direction of insignificant selective gene superpedigree result	1
In opposite direction of significant selective gene superpedigree result with two pedigrees	−2
In opposite direction of significant selective gene superpedigree result with more than two pedigrees	−3
Insignificant selective gene superpedigree result	−1
Significant selective gene superpedigree result exists for other phenotypes	0.5
Mutation is linked to a more damaging mutation	−1
Mutation is linked to an excluded mutation	1
Mutation is linked to a verified mutation	−3
No corresponding trimmed result	−3
Phenotype is driven by another mutation	−1

B6, C57BL/6J.

*A screen result is considered as an outlier if its value is outside the range of mean $\pm$ 3 × standard deviation (SD).

[†]The variance of REF, HET, or VAR results is considered big if the SD is more than 30% of the absolute difference between the maximum screening result and the minimum screening result.

[‡]REF and B6 results are considered different if the absolute difference between the REF mean and B6 mean is more than 2 × REF SD and 2 × B6 SD.

[§]REF or B6 results overlap with VAR results if they are within the range of VAR mean $\pm$ 1 × VAR SD.

[¶]HET results are considered more similar than REF results to B6 results if the absolute difference between the B6 mean and REF mean is more than half of REF SD and more than the absolute difference between HET mean and B6 mean.

[#]VAR results are considered more similar than REF results to B6 results if the absolute difference between the B6 mean and REF mean is more than half of REF SD and more than the absolute difference between VAR mean and B6 mean.

[‖]Significant pedigree refers to linkage analysis of a pedigree or superpedigree by AMM in which $P < 0.05$ for a mutation-phenotype association.

Whole-exome sequencing and mapping were performed as described previously (1).

To generate mice carrying CRISPR/Cas9-targeted mutations, female C57BL/6J mice were superovulated by injection with 6.5 U pregnant mare serum gonadotropin (PMSG; Millipore), then 6.5 U human chorionic gonadotropin (hCG; Sigma-Aldrich) 48 h later. The superovulated mice were subsequently mated with C57BL/6J male mice overnight. The following day, fertilized eggs were collected from the oviducts and in vitro transcribed Cas9 mRNA (50 ng/μL) and small base-pairing guide RNA (50 ng/μL) were injected into the cytoplasm or pronucleus of the embryos. The injected embryos were cultured in M16 medium (Sigma-Aldrich) at 37 °C and 5% CO_2. For the production of mutant mice, two-cell stage embryos were transferred into the ampulla of the oviduct (10 to 20 embryos per oviduct) of pseudopregnant Hsd:ICR (CD-1) (Harlan Laboratories) females.

Mice were housed in specific pathogen-free conditions at the University of Texas Southwestern Medical Center and all experimental procedures were performed in accordance with the guidelines established by the Institutional Animal Care and Use Committee of the University of Texas Southwestern Medical Center and with the NIH *Guide for the Care and Use of Laboratory Animals* (14). Male and female mice were used in all experiments and data for males and females were combined for analysis.

Flow Cytometry. Peripheral blood was collected from G3 mice >6 wk old by cheek bleeding. Red blood cells (RBCs) were lysed with hypotonic buffer

Table 6. Flow cytometry screening parameters

	Parameter
1	B cells
2	B:T cell ratio
3	B-1 B cells
4	B-1a B cells
5	B-1a B cells in B-1 B cells
6	B-1b B cells
7	B-1b B cells in B-1 B cells
8	B-2 B cells
9	B220 MFI
10	CD11b$^+$ DC (gated in CD11c$^+$ cells)
11	CD11c$^+$ DC
12	CD4:CD8 T cell ratio
13	CD4$^+$ T cells
14	CD4$^+$ T cells in CD3$^+$ T cells
15	CD44 MFI on CD4$^+$ T cells
16	CD44$^+$ CD4$^+$ T cells
17	CD44 MFI on CD8$^+$ T cells
18	CD44$^+$ CD8$^+$ T cells
19	CD44$^+$ T cells
20	CD44 MFI on T cells
21	CD8$^+$ T cells
22	CD8$^+$ T cells in CD3$^+$ T cells
23	CD8α^+ DC (gated in CD11c$^+$ cells)
24	Central memory CD4$^+$ T cells in CD4$^+$ T cells
25	Central memory CD8$^+$ T cells in CD8$^+$ T cells
26	Effector memory CD4$^+$ T cells in CD4$^+$ T cells
27	Effector memory CD8$^+$ T cells in CD8$^+$ T cells
28	Effector T cells
29	IgD MFI
30	IgD$^+$ B cells
31	IgM MFI
32	IgM$^+$ B cells
33	Macrophages
34	Memory T cells
35	Naïve CD4$^+$ T cells in CD4$^+$ T cells
36	Naïve CD8$^+$ T cells in CD8$^+$ T cells
37	Naïve T cells
38	Neutrophils
39	NK cells
40	NK1.1$^+$ T cells
41	Plasmacytoid DC
42	T cells

Parameters represent frequencies unless otherwise indicated. MFI, mean fluorescence intensity.

(eBioscience). Samples were washed with FACS staining buffer (PBS with 1% [wt/vol] BSA) one time and then centrifuged at 500 × g for 5 min. The RBC-depleted samples were stained for 1 h at 4 °C, in 100 μL of a 1:200 mixture of fluorescence-conjugated antibodies to 15 cell surface markers encompassing the major immune lineages B220 (BD, clone RA3-6B2), CD19 (BD, clone 1D3), IgM (BD, clone R6-60.2), IgD (BioLegend, clone 11-26c.2a), CD3ε (BD, clone 145-2C11), CD4 (BD, clone RM4-5), CD8α (BioLegend, clone 53-6.7), CD11b (BioLegend, clone M1/70), CD11c (BD, clone HL3), F4/80 (Tonbo, clone BM8.1), CD44 (BD, clone 1M7), CD62L (Tonbo, clone MEL-14), CD5 (BD, clone 53-7.3), CD43 (BD, clone S7), NK 1.1 (BioLegend, clone OK136), and 1:200 Fc block (Tonbo, clone 2.4G2). Flow cytometry data were collected on a BD LSR Fortessa and the proportions of immune cell populations in each G3 mouse were analyzed with FlowJo software. The resulting phenotypic data were uploaded to Mutagenetix for automated mapping of causative alleles.

Automated Meiotic Mapping. AMM was performed as previously described (1). Briefly, genotypes at all mutation sites present in the exomes of G3 mice were determined prior to phenotypic screening: tail DNA from G1 males was subjected to whole-exome sequencing using an Illumina HiSEq. 2500 instrument; G2 and G3 mice were then genotyped at the identified mutation

sites using an Ion PGM (Life Technologies). Following phenotypic screening, linkage analysis using recessive, additive, and dominant models of inheritance was performed for every mutation in the pedigree using the program Linkage Analyzer; phenotypic data scatter plots and Manhattan plots were displayed using the program Linkage Explorer. The P values of association between genotype and phenotype were calculated using a likelihood ratio test from a generalized linear model or generalized linear mixed-effect model and Bonferroni correction applied.

Candidate Explorer. The CE prediction model was built using a random forest algorithm implemented in the R caret package. CE is publicly accessible at https://mutagenetix.utsouthwestern.edu/linksplorer/candidate.cfm. Linkage data obtained through screening will be released in phases according to phenotype. Blood cell flow cytometry screening data are currently available for search using CE and new data will be released as they are acquired after a 6-mo delay from the date of screening.

Damage Score. The damage score is an ensemble score that uses a logistic regression model to integrate 38 independent prediction scores. Thirty-seven prediction scores are retrieved from the human dbNSFP, and consist of scores from the following algorithms: SIFT, SIFT4G, Polyphen2-HDIV, Polyphen2-HVAR, LRT, MutationTaster2, MutationAssessor, FATHMM, MetaSVM, MetaLR, CADD, CADD_hg19, VEST4, PROVEAN, FATHMM-MKL coding, FATHMM-XF coding, fitCons (four scores), LINSIGHT, DANN, GenoCanyon, Eigen, Eigen-PC, M-CAP, REVEL, MutPred, MVP, MPC, PrimateAI, GEOGEN2, BayesDel_addAF, BayesDel_noAF, ClinPred, LIST-S2, and ALoFT. Our dataset uses the ranked scores of each algorithm transformed by dbNSFP. The 38th prediction score is the probability of protein damage to phenovariance caused by mouse mutations, calculated as previously described (3). Among the 38 prediction scores, the most important is the score from MutPred (mutpred.mutdb.org/), followed by the probability of protein damage to phenovariance caused by mouse mutations, and phastCons100way_vertebrate (a conservation score). Damage Score can be used as a quantitative prediction score to measure the likelihood of a mouse mutation being deleterious.

Our assumption is that if the mouse missense mutation is the same as the human mutation (both nucleotide and amino acid changes), then the mutation effect in human and mouse should be similar. We therefore use human scores to predict likelihood of damage in mice.

A set of mouse ENU mutations with class tags (known damaging or neutral) was retrieved from the Mutagenetix database. The known mutation class tags come from four sources: 1) Physically isolated mutations (of linkage with all other coding/splicing mutations in the pedigree) that fall within essential genes yet can be transmitted from heterozygous G2 females and their heterozygous G1 sire to homozygous G3 mice at a ratio that does not significantly depart from Mendelian expectation, are considered neutral. 2) Conversely, isolated mutations in essential genes that are not transmitted to homozygosity, to the extent that homozygotes are observed at frequencies significantly beneath the expected Mendelian ratio, are considered damaging. 3) Mutations that cause qualitative (usually visible) phenotypes are considered damaging. 4) Mutations that have been verified to be significant in phenotypic screening of CRISPR replacement alleles are also considered to be damaging.

The mutations tagged as damaging or neutral were lifted-over from mouse genome to human genome (translated to the equivalent amino acid) and kept for mutations that lead to the same nucleotide and amino acid changes in both genomes. About 4% of mouse mutations could not be mapped to the corresponding human mutations using the lift-over tool from the University of California, Santa Cruz (UCSC) (https://genome.ucsc.edu/cgi-bin/hgLiftOver). They were not included in the final dataset for model training and testing. A point-biserial correlation was used to estimate the relationship between the mouse mutations tagged damaging or neutral with the most important human mutation prediction score (MutPred score). The correlation coefficient was 0.525, with 95% CI: 0.50 to 0.55. Then we searched for corresponding human mutations in the dbNSFP database to obtain scores for all available prediction methods. The retrieved scores, combined with the probability of phenotypically detectable damage by the mutations in mice, were integrated with the input dataset and used to train and optimize a logistic regression model using the train function of the R caret package with 10-fold cross-validation (https://topepo.github.io/caret/model-training-and-tuning.html); this was repeated three times. The scaling of the data were performed by the preProcess function. The constructed model (classifier) was then used to compute the score of a set of mutations with unknown class membership. The dataset used for prediction was created in the same way as dataset used for modeling. The score predicted by
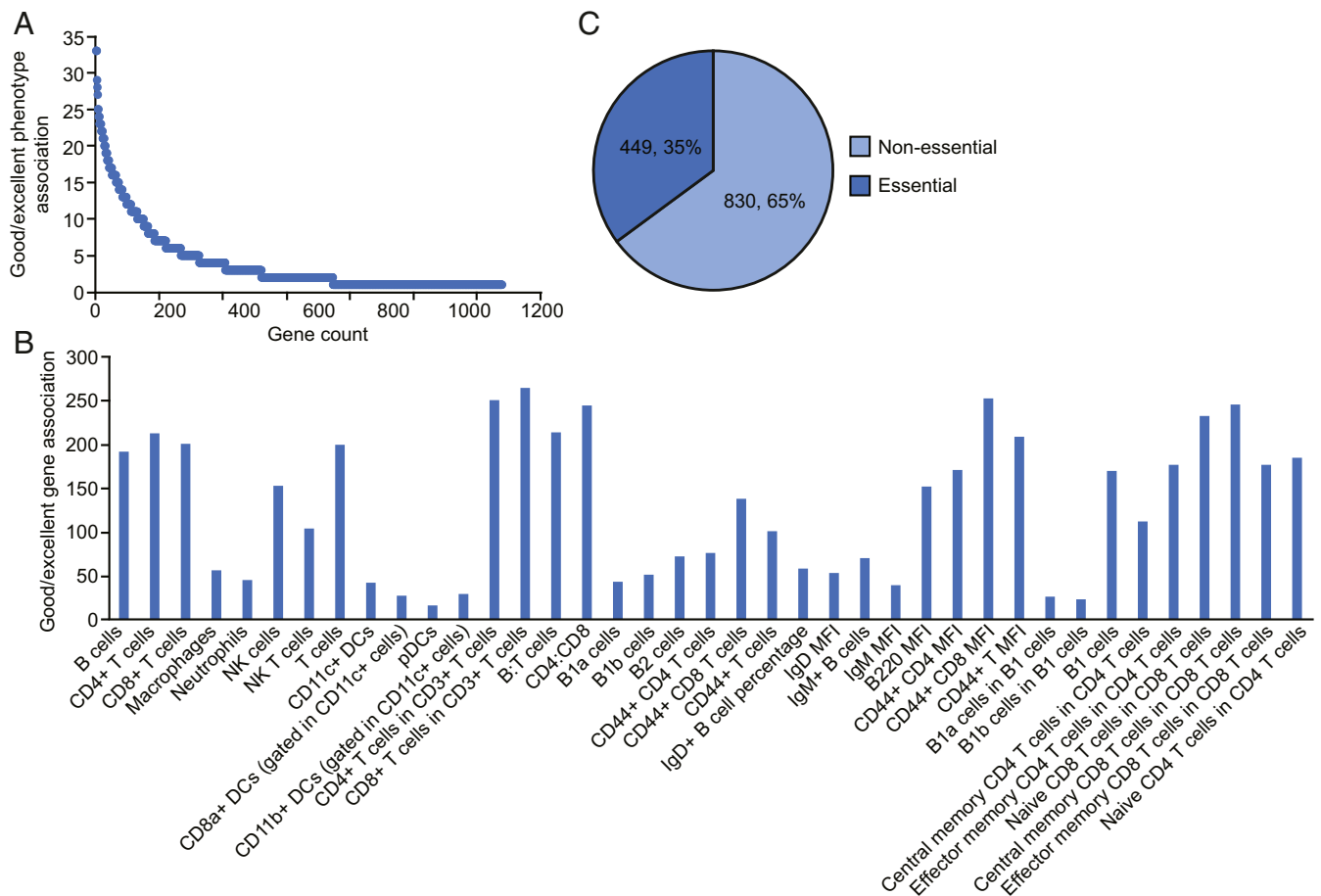
Xu et al.
Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled with machine learning

PNAS | 9 of 11
https://doi.org/10.1073/pnas.2106786118

GENETICS

www.manaraa.com

Fig. 2. Characteristics of gene–phenotype associations for 1,279 genes with at least one good/excellent mutation–phenotype association. (*A*) Number of good/excellent phenotype associations plotted versus gene count. (*B*) Number of good/excellent gene associations plotted versus flow cytometry parameter. Parameters are cell frequencies unless indicated. MFI, mean fluorescence intensity. (*C*) Number and percentage of essential and nonessential genes.

the model represents the probability of a mutation being in the damaging class. The higher the score, the more likely to be deleterious the mutation.

The input dataset contains 3,334 mouse mutations, of which 1,088 are deleterious and 2,246 are neutral. In order to evaluate the performance of the constructed model in predicting the membership of the new mutation category, the input dataset was randomly divided into two sets: one set consisting of 2,668 mutations (80% of original dataset, 871 deleterious mutations and 1,797 neutral mutations) was used to train and validate the logistic regression model, and a second set of the remaining 666 mutations was used to test the performance of the established model. The 80/20 splits for training and testing were conducted 10 times randomly; the ROC curve shown in *SI Appendix*, Fig. S3 yielded an AUC close to the average AUC value of 0.853 ± 0.014.

Quartile based correspondence between raw damage scores and probability of protein damage to phenovariance is shown in *SI Appendix*, Table S1.

E-Score. E-score is used to estimate the likelihood of lethality in mice when the gene is knocked out. Our approach is based on the assumption that essential and nonessential genes in mice can be distinguished by various independent features of genes. The logistic regression method is used to fit the features of known essential and nonessential genes in mice to obtain a trained model for predicting the unknown essentiality of genes.

The model uses the following gene features: 1) From the OGEE database (15): gene conservation, connectivity in protein–protein interaction network, expression stage during development, evolutionary age, GO terms, copy number of genes, and length of gene product. These features have been suggested to be associated with gene essentiality of many species, including mouse. 2) The essentiality of human orthologous genes: the genes required for cell proliferation and viability in tested cell lines are defined as essential genes under specific conditions. Frequency of being essential in tested human cell lines was used as a feature in our model. 3) pLI score from the ExAC (probability of loss-of-function

intolerance): the closer the score is to 1, the more likely the gene is essential to human survival. 4) Minimum *P* values for an ENU-targeted mouse gene obtained from the lethal model by the Linkage Analyzer program.

The phenotypic description of the 8,032 genes in MGI, which may be knocked out in mice, was carefully reviewed and a set of genes designated as "essential" or "nonessential" were manually curated according to the following criterion: 1) If the homozygous knockout allele is explicitly described as causing embryonic lethality, neonatal lethality, prenatal lethality, perinatal lethality, or preweaning lethality, the gene was considered to be required to survive before weaning and was classified as an essential gene. An E-score of 1 was assigned to the gene. 2) If homozygous knockout alleles are compatible with viability, normal growth, no obvious phenotype, or some phenotype, but not apparent effect on viability, then it was classified as a nonessential gene. An E-score of 0 was assigned to the gene. In addition, an E-score of 1 was assigned to those genes verified in our CRISPR knockout experiments as causing significant lethality before weaning; an E-score of 0 was assigned to genes verified in our CRISPR knockout experiments as resulting in normal Mendelian ratios in crosses of heterozygous mutants.

A set of 7,009 genes, in which 2,587 were labeled as essential genes and 4,422 as nonessential genes, was integrated with the above-mentioned gene features. The resulting dataset was used to train and optimize a logistic regression model using the train function of the R caret package with 10-fold cross-validation (https://topepo.github.io/caret/model-training-and-tuning.html); this was repeated three times. The scaling of the data was performed by the preProcess function. The constructed model was then used to predict the essentiality of remaining mouse genes. The predicted score is between 0 and 1. The closer the score is to 1, the more likely the gene is essential.

To assess the performance of constructed model in predicting unknown essentiality of genes, the dataset used to construct the model was randomly divided into two sets: one set consisting of 5,608 genes (80% of original dataset, 3,538 nonessential genes and 2,070 essential genes) was used to train

Xu et al.
Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled

and validate the logistic regression model, and the remaining 1,401 genes were used to test the performance of the established model in the training dataset. The 80/20 splits for training and testing were conducted 10 times randomly; the ROC curve shown in *SI Appendix*, Fig. S4 yielded an AUC close to the average AUC value of 0.891 ± 0.0087.

Algorithmic Score. Each mutation–phenotype association starts with an algorithmic score of zero that is adjusted according to the rules in Table 5.

GO Analysis. Summaries of GO annotations in Dataset S4 were generated using the Alliance of Genome Resources SimpleMine tool (tazendra.caltech.edu/~azurebrd/cgi-bin/forms/agr_simplemine.cgi). Enriched GO annotations associated with a gene list (Datasets S5 and S6) were determined using GO TermFinder (16) (https://go.princeton.edu/cgi-bin/GOTermFinder) set to use the *Mus musculus* annotations (MGI) and exclude evidence code "IEA" (inferred from electronic annotation). GO TermMapper (16) was used to assign genes to 70 static GO parent annotations (17) (https://go.princeton.edu/cgi-bin/GOTermMapper) (Dataset S7).

Gene-Expression Data. The mouse Gene Expression Database (4) was queried by batch submission of the 386 gene symbols. Tissue × Gene Matrix results, all of which were from RNA-sequencing experiments, were filtered by "anatomical system: immune system" and by "TPM level: High AND Medium."

Data Availability. The CE data are part of the Mutagenetix database and are publicly accessible at https://mutagenetix.utsouthwestern.edu/linksplorer/candidate.cfm. Raw phenotype data are available through Candidate Explorer by clicking the screen name for any mutation. Sequences of small base-pairing guide RNA used for CRISPR/Cas9 targeting are available by request from the corresponding author. All other data are available in the main text or the supporting information.

1. T. Wang *et al.*, Real-time resolution of point mutations that cause phenovariance in mice. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E440–E449 (2015).
2. M. M. Simon *et al.*, Current strategies for mutation detection in phenotype-driven screens utilising next generation sequencing. *Mamm. Genome* **26**, 486–500 (2015).
3. T. Wang *et al.*, Probability of phenotypically detectable protein damage by ENU-induced mutations in the Mutagenetix database. *Nat. Commun.* **9**, 441 (2018).
4. R. M. Baldarelli *et al.*, The mouse Gene Expression Database (GXD): 2021 update. *Nucleic Acids Res.* **49**, D924–D931 (2021).
5. H. Suzuki *et al.*, Xid-like immunodeficiency in mice with disruption of the p85alpha subunit of phosphoinositide 3-kinase. *Science* **283**, 390–392 (1999).
6. M. E. Conley *et al.*, Agammaglobulinemia and absent B lineage cells in a patient lacking the p85α subunit of PI3K. *J. Exp. Med.* **209**, 463–470 (2012).
7. N. Sakaguchi *et al.*, Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature* **426**, 454–460 (2003).
8. M. E. Elder *et al.*, Human severe combined immunodeficiency due to a defect in ZAP-70, a T cell tyrosine kinase. *Science* **264**, 1596–1599 (1994).
9. T. A. Chatila *et al.*, JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. *J. Clin. Invest.* **106**, R75–R81 (2000).
10. M. E. Brunkow *et al.*, Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat. Genet.* **27**, 68–73 (2001).
11. C. M. Roifman *et al.*, Depletion of CD8+ cells in human thymic medulla results in selective immune deficiency. *J. Exp. Med.* **170**, 2177–2182 (1989).
12. Y. Minegishi *et al.*, Mutations in Igalpha (CD79a) result in a complete block in B-cell development. *J. Clin. Invest.* **104**, 1115–1121 (1999).
13. P. Georgel, X. Du, K. Hoebe, B. Beutler, ENU mutagenesis in mice. *Methods Mol. Biol.* **415**, 1–16 (2008).
14. National Research Council (US) Committee for the Update of the Guide for the Care and Use of Laboratory Animals, *Guide for the Care and Use of Laboratory Animals* [National Academies Press (US), Washington, DC, ed. 8, 2011].
15. S. Gurumayum *et al.*, OGEE v3: Online GEne Essentiality database with increased coverage of organisms and human cell lines. *Nucleic Acids Res.* **49** (D1), D998–D1003 (2021).
16. E. I. Boyle *et al.*, GO:TermFinder—Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715 (2004).
17. M. A. Harris *et al.*; Gene Ontology Consortium, The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).

GENETICS

Xu et al.
Thousands of induced germline mutations affecting immune cells identified by automated meiotic mapping coupled with machine learning

PNAS | 11 of 11
https://doi.org/10.1073/pnas.2106786118
www.manaraa.com