Striking Immune Phenotypes in Gene-Targeted Mice Are Driven by a Copy-Number Variant Originating from a Commercially Available C57BL/6 Strain

Citation: Mahajan, Vinay S., Ezana Demissie, Hamid Mattoo, Vinay Viswanadham, Ajit Varki, Robert Morris, and Shiv Pillai. "Striking Immune Phenotypes in Gene-Targeted Mice Are Driven by a Copy-Number Variant Originating from a Commercially Available C57BL/6 Strain." Cell Reports 15:9 (May 2016), pp.1901-1909.

As Published: http://dx.doi.org/10.1016/j.celrep.2016.04.080

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/103644

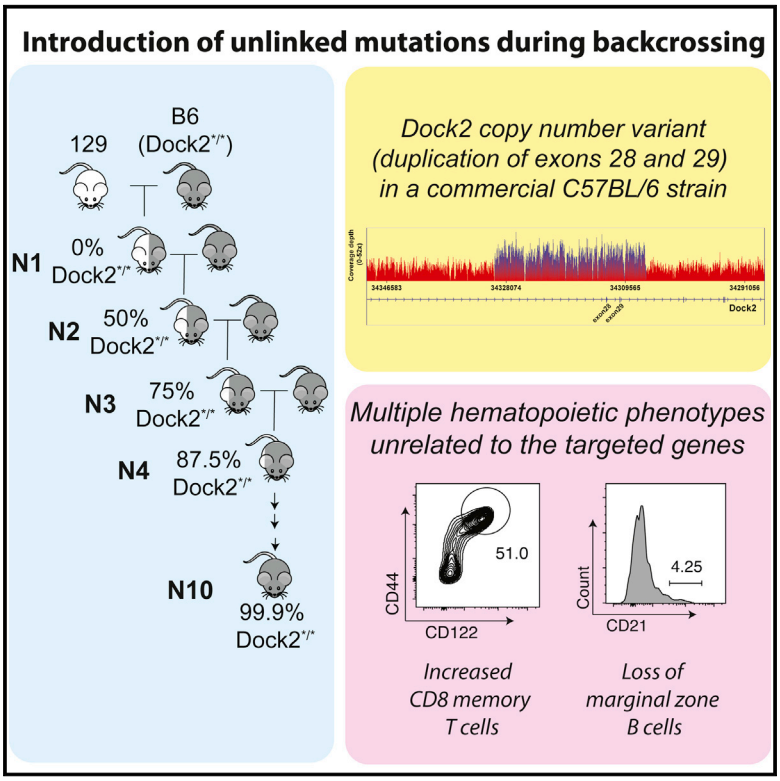
Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Report

Cell Reports

Striking Immune Phenotypes in Gene-Targeted Mice Are Driven by a Copy-Number Variant Originating from a Commercially Available C57BL/6 Strain

Graphical Abstract



Highlights

- A spontaneous *Dock2* mutation was found in a widely used C57BL/6 mouse strain
- The *Dock2*^{Hsd} allele has been inadvertently introduced into several gene-targeted mice
- The *Dock2*^{Hsd} allele may confound the interpretation of several gene-targeting studies
- Published studies using C57BL/6NHsd mice may need to be revisited

Authors

Vinay S. Mahajan, Ezana Demissie, Hamid Mattoo, Vinay Viswanadham, Ajit Varki, Robert Morris, Shiv Pillai

Correspondence

pillai@helix.mgh.harvard.edu

In Brief

Gene-targeted mice are often backcrossed into the C57BL/6 background. Mahajan et al. find that a homozygous copy-number variant disrupts the function of Dock2 in a specific commercially available C57BL/6 mouse strain (C57BL/6NHsd).

Mahajan et al., 2016, Cell Reports *15*, 1901–1909
May 31, 2016
http://dx.doi.org/10.1016/j.celrep.2016.04.080

Cell Reports
Report

Striking Immune Phenotypes in Gene-Targeted Mice Are Driven by a Copy-Number Variant Originating from a Commercially Available C57BL/6 Strain

Vinay S. Mahajan,[1,3] Ezana Demissie,[1,3] Hamid Mattoo,[1] Vinay Viswanadham,[1] Ajit Varki,[2] Robert Morris,[1] and Shiv Pillai[1,*]
[1]Ragon Institute of MGH, MIT and Harvard, 400 Technology Square, Cambridge, MA 02139, USA
[2]Departments of Medicine and Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[3]Co-first author
*Correspondence: pillai@helix.mgh.harvard.edu
http://dx.doi.org/10.1016/j.celrep.2016.04.080

SUMMARY

We describe a homozygous copy-number variant that disrupts the function of Dock2 in a commercially available C57BL/6 mouse strain that is widely used for backcrossing. This *Dock2* allele was presumed to have spontaneously arisen in a colony of *Irf5* knockout mice. We discovered that this allele has actually been inadvertently backcrossed into multiple mutant mouse lines, including two engineered to be deficient in Siae and Cmah. This particular commercially obtained subline of C57BL/6 mice also exhibits several striking immune phenotypes that have been previously described in the context of *Dock2* deficiency. Inadvertent backcrossing of a number of gene-targeted mice into this background has complicated the interpretation of several immunological studies. In light of these findings, published studies involving immune or hematopoietic phenotypes in which these C57BL/6 mice have been used as controls, as experimental animals, or for backcrossing will need to be reinterpreted.

INTRODUCTION

Over the last three decades, gene targeting has emerged as a powerful tool for functional analyses of immune genes in vivo. It has become a common practice to backcross gene-targeted mice for about ten generations into the C57BL/6 background to facilitate comparisons among gene-targeted mice, as well as for adoptive transfer experiments. However, numerous C57BL/6 sublines are in use around the world (Zurita et al., 2011), and the potential effect of variability among these C57BL/6 sublines on immune phenotypes is often not considered.

We had previously described defects in B cell development in two engineered mice strains with altered sialic acid physiology (Cariappa et al., 2009). Mice with a germline loss of either *Siae* (sialic acid acetyl esterase) or *Cmah* (cytidine monophosphate-N-acetylneuraminic acid hydroxylase) were found to lack marginal zone (MZ) B cells and exhibited hyperactive B cell receptor signaling (Cariappa et al., 2009). Given that these mice generate altered forms of sialic acid that are not recognized by key regulatory Siglecs expressed on B cells (such as CD22/Siglec-2 and Siglec-G), the defects in B cell development observed in these mice were presumed to arise from perturbations in Siglec function (Cariappa et al., 2009; Pillai et al., 2009). In addition, the observed phenotypes were largely compatible with previous studies of Siglec function (Mahajan and Pillai, 2016; Pillai et al., 2009). Both $Siae^{\Delta ex2/\Delta ex2}$ and *Cmah* knockout mice had been backcrossed into a specific commercially obtained C57BL/6 background for ten generations (Cariappa et al., 2009; Hedlund et al., 2007). We found that Siae-deficient mice unexpectedly lost their aberrant B cell development phenotype upon backcrossing for 13 additional generations into the C57BL/6J (Jackson Laboratory) background. We created an independent knockout line of Siae-deficient mice in the C57BL/6N background, and these mice exhibited no defects in B cell development.

Given these discrepant results, we re-examined the genetic basis of aberrant B cell development in $Siae^{\Delta ex2/\Delta ex2}$ mice using genetic crosses, SNP arrays, and whole-genome sequencing. These studies revealed that the defects in B cell development were not linked to *Siae*, which is present on chromosome 9 (chr9), but instead to a gene encoding a guanine nucleotide exchange factor, *Dock2*, on chromosome 11 (chr11).

Whole-genome sequencing revealed a duplication of exons 28 and 29 of *Dock2*. Surprisingly, an identical mutation in *Dock2* had been previously reported in two colonies of $Irf5^{-/-}$ mice (Purtha et al., 2012). Given that the same mutation in *Dock2* was identified in multiple gene-targeted mouse colonies despite different embryonic stem cell (ESC) lines being used to generate these mice, it appeared that it was most likely introduced during backcrossing into the C57BL/6 background. Furthermore, the presence of C57BL/6N SNPs in close linkage with the duplication suggests that it arose in a C57BL/6N subline. We were able to find this *Dock2* variant ($Dock2^{Hsd}$) in the homozygous state in the colony of C57BL/6NHsd mice maintained at Harlan Sprague Laboratories (now Envigo Biosciences). We also identified the $Dock2^{Hsd}$ variant in a colony of *Cmah* knockout mice that had been backcrossed into C57BL/6 mice obtained directly from Harlan. Examination of a range of other commercially available C57BL/6J and C57BL/6N mice revealed only wild-type *Dock2*. It is therefore likely that the $Dock2^{Hsd}$ variant originally arose

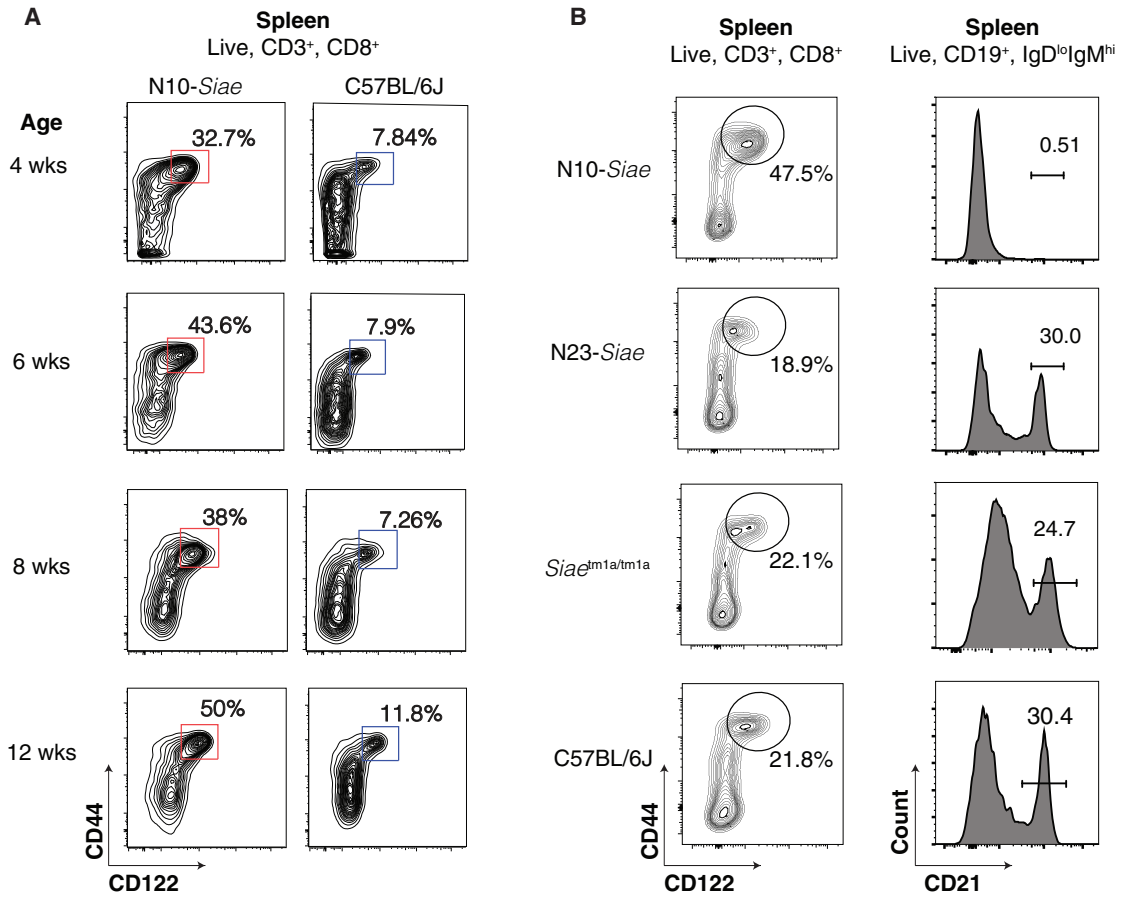


Figure 1. *Siae*-Deficient Mice Exhibit a Background-Dependent Increase in CD8$^+$ MP Cells and a Loss of MZ B Cells

(A) The proportion of CD8$^+$ MP cells in the spleens of 1- to 3-month-old N10-*Siae* mice with age-matched C57BL/6J controls.

(B) CD8$^+$ MP and MZ B cells in N10-*Siae*, N23-*Siae*, Siae$^{tm1a/tm1a}$, and C57BL/6J mice. The data shown are representative of five mice.

See also Figure S1.

within the C57BL/6NHsd mouse colony at Harlan. This study raises concerns that many other lines of gene-targeted mice bearing hematopoietic phenotypes may have been inadvertently compromised by backcrosses involving the use of C57BL/6NHsd mice.

RESULTS

The Loss of MZ B Cells and Enhanced CD8$^+$ Memory T Cell Phenotype Observed in *Siae*$^{\Delta ex2/\Delta ex2}$ Mice Is Not Linked to the Loss of Siae

We previously reported that *Siae*$^{\Delta ex2/\Delta ex2}$ mice, which were backcrossed into the C57BL/6 background for ten generations (henceforth referred to as N10-*Siae* mice), show a profound loss of MZ B cells (Cariappa et al., 2009). In subsequent studies, we noted that N10-*Siae* mice also exhibit a marked increase in CD8$^+$ CD44$^+$ CD122^{hi} memory-phenotype (MP) cells in the blood and spleen (Figure 1A). Surprisingly, both the defect in MZ B cell development and the increase in CD8$^+$ MP cells were lost upon further backcrossing of N10-*Siae* mice into the C57BL/6J background for an additional 13 generations (henceforth referred to as N23-*Siae* mice) (Figure 1B). To further determine whether the loss of Siae per se was responsible for any of the N10 phenotypes, we generated an independent line of Siae-deficient mice (*Siae*$^{tm1a/tm1a}$) using an ESC clone bearing a gene-targeted allele of Siae (*Siae*$^{tm1a(EUCOMM)Wtsi}$; MGI: 4842607) in the C57BL/6N background (Figure S1). These mice also exhibited normal numbers of MZ B cells and no increase in CD8$^+$ MP cells, confirming that the loss of Siae per se was not responsible for the phenotypes previously observed in N10-*Siae* mice (Figure 1B).

The Phenotypes Observed in N10-*Siae* Mice Are Inherited in a Mendelian Fashion and Linked to a Locus Distinct from *Siae*

To begin to identify the genetic locus responsible for the anomalous phenotypes observed in N10-*Siae* mice, we did a test cross to assess the inheritance pattern. N10-*Siae* × C57BL/6J mice (F1) were generated and backcrossed to N10-*Siae* mice. We found that the increase in CD8$^+$ MP cells was a Mendelian recessive trait that segregated independently of *Siae* (Figure 2A). In addition, analysis of these mice showed a 100% linkage between the loss of MZ B cells and an increase in CD8$^+$ MP

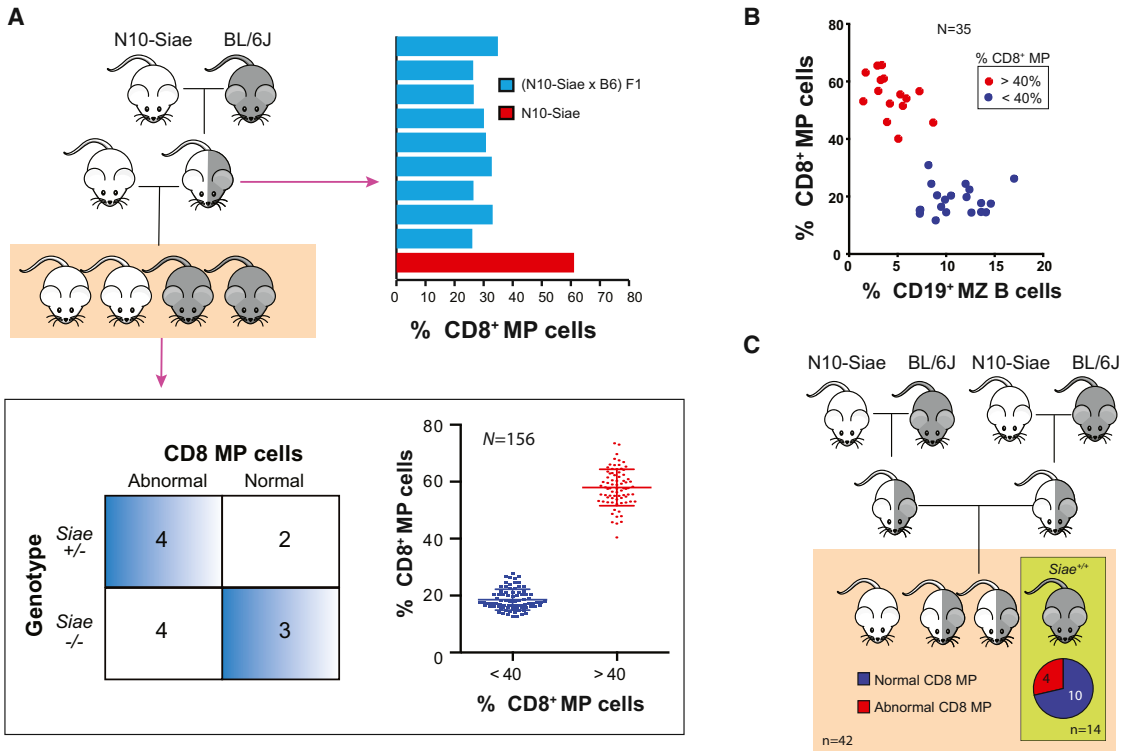


Figure 2. Hematopoietic Phenotypes Observed in N10-*Siae* Mice Are Mendelian, Recessive, and Unlinked to *Siae*

(A) (N10-Siae × C57BL/6J) F1 animals have a normal CD8$^+$ MP cell compartment. Upon subsequent backcrossing to the N10-Siae background, Siae does not segregate with the altered CD8$^+$ MP cells (n = 13). An approximately equal number of mice with (n = 79) and without (n = 77) the phenotype (<40% and >40% CD8 MP cells, respectively) are observed.

(B) A strong correlation between loss of MZ B cells and increase in CD8 MP is observed.

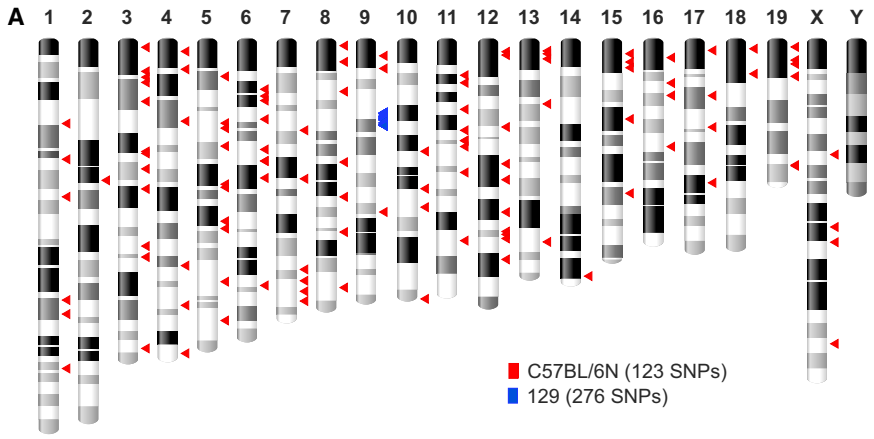
(C) About 25% of Siae$^{+/+}$ (wild-type) mice obtained from an intercross between heterozygous N10-*Siae* mice exhibit normal CD8$^+$ MP cells.

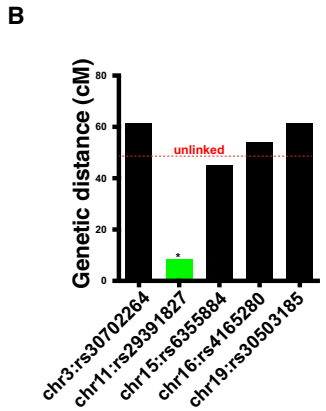
T cells, suggesting that a single genetic locus was likely responsible for both phenotypes (Figure 2B). Because the proportion of CD8$^+$ MP cells in blood could be easily measured, we used this trait for mapping the pathogenic locus. We also found that 4 of 18 mice that lacked both copies of the *Siae*$^{\Delta ex2}$ allele had an abnormal CD8$^+$ MP population, suggesting that *Siae* had no contribution to these observed phenotypes (Figure 2C).

A Genetic Marker on Chr11 Segregates with the N10-*Siae* Phenotype

In parallel, we performed a whole-genome SNP array (Affymetrix Mouse Diversity array) on two homozygous N10-*Siae* mice that exhibited a loss of MZ B cells and an increased proportion of CD8$^+$ MP cells. The N10-*Siae* whole-genome SNP arrays yielded an average call rate of 99%, with 98.7% homozygosity. Publicly available SNP array data from various 129 substrains and C57BL/6 sublines were used for comparison (Didion et al., 2012). The 538,667 SNPs showed a 100% call rate across all arrays. Given that the *Siae*$^{\Delta ex2}$ allele was generated in an R1 ESC that is derived from the (129X1/SvJ × 129S1/Sv)F1-Kitl$^+$ background (Nagy et al., 1993), it was not surprising that all SNPs (276 of 276) in a 6 Mbp region surrounding the *Siae* locus (chr9:35022155–41040054) that differed from the C57BL/6 reference genome were of 129 origin (Figure 3A) (Table S1). Surprisingly, all 129 origin SNPs were retained in the N23-*Siae* mice, suggesting that 13 additional generations of backcrossing into C57BL/6J mice had had no further effect on reducing the 129 contribution on chr9 (Figure S3). Considering that ~17.4% of the 538,667 SNPs calls analyzed in the whole-genome SNP array were different between the 129 and the C57BL/6 backgrounds and that we did not find any genomic region containing two or more contiguous SNPs of 129 origin outside of the chr9 locus containing *Siae*, we recognized that it was extremely unlikely that a mutation in the gene-targeted ESC clone of 129 origin had been retained in the N10-*Siae* mice.

All (123 of 123) SNPs on the remaining N10-*Siae* chromosomes that differed from the C57BL/6J allele matched the C57BL/6N consensus allele (shared between C57BL/6NTac, C57BL/6Nci, and C57BL/6NCrl), suggesting that the pathogenic locus could have arisen from a C57BL/6N strain used previously for backcrossing. We next tested the genetic linkage to a few randomly chosen candidate C57BL/6N SNPs in the (C57BL/6N × C57BL/6J) test crosses (Figures 2A and 2B). Fortuitously, a SNP of C57BL/6N origin, rs29391827, on chr11 (50962622) was linked to the increase in CD8$^+$ MP cells in the blood of 18 mice (Figure 3B). The linkage to rs29391827 was validated in a larger cohort of 156 mice and estimated to be at a distance of about 8.3 cM (Figure 3C).





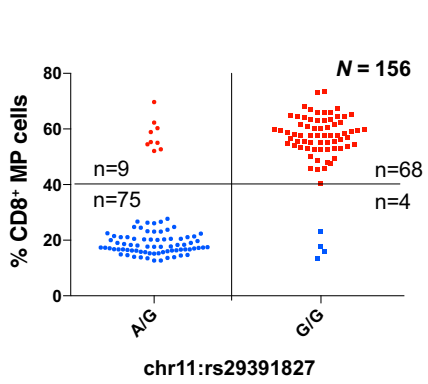


Figure 3. A SNP on Chr11 Segregates with Increase in CD8 MP T Cells

(A) Whole-genome SNP array revealed 399 SNPs that differ between N10-Siae and C57BL/6J. On chr9, 276 SNPs (blue triangles) in the *Siae* locus are shared by 129 strains; 123 SNPs (red triangles) appear to be of C57BL/6N origin.

(B) Linkage between candidate SNPs and $CD8^+$ MP expansion in the progeny of test crosses (n = 18) described in Figure 2A. A SNP on chr11 (rs29391827) was found to segregate with the CD8 MP (n = 18).

(C) The observation in (B) was validated in a larger cohort (n = 156).

See also Figure S2 and Table S1.

A Region on chr11:30552213–35421130 Is Tightly Associated with the N10-*Siae* Phenotype

We next performed whole-genome sequencing of an N10-*Siae* mouse that exhibited the phenotypic changes described. Single-end sequencing was performed for 85 cycles on a NextSeq 500 instrument, yielding a total of ~40 Gbp of sequence data. Of the reads, 93.9% mapped to the reference mouse genome (65.42% of the reads mapped to unique sites on the genome). A mean coverage depth of 15.33-fold was obtained. The Genome Analysis Toolkit (GATK) variant calling pipeline was used to identify positions on chr11 that differed from the reference genome (Figure S3) (McKenna et al., 2010). This analysis confirmed the results of the whole-genome SNP array (for SNPs with adequate sequencing coverage) and revealed a few additional SNPs that were not covered by the SNP array that differed between N10-*Siae* and C57BL/6J. All protein coding variants (3 of 3) that we observed within 25 Mbp of rs29391827 on chr11 were previously reported SNPs present in one or more inbred mouse strains (rs240617401, rs8261521, and rs6268547). Because structural variants could not be ruled out on the basis of this analysis, we used a panel of SNPs that differed between N10-*Siae* and C57BL/6J to analyze mice (Figure 3) that showed discrepant phenotypes with respect to rs29391827. This allowed us to progressively narrow the pathogenic genetic locus to a 5 Mbp region on chr11 (30552213–35421130) (Figure S4).

A *Dock2* Duplication on Chr11 Accounts for the Observed N10-*Siae* Phenotype

A focused examination of the whole-genome sequence in the chr11: 30552213–35421130 region revealed an ~2-fold increase in coverage in a 23.5 Kbp region encompassing exons 28 and 29 of the *Dock2* gene (Figure 4A). This was reminiscent of a previously published report identifying a duplication involving exons 28 and 29 of *Dock2* in $Irf5^{-/-}$ mice (Purtha et al., 2012). The presence of an identical duplication in *Dock2* was confirmed in N10-*Siae* mice using a previously reported PCR designed to identify this duplication (Figure 4B) (Yasuda et al., 2013). Sequencing of the PCR product revealed that the breakpoint in the *Dock2* duplication (chr11:34329863–34306424; GRCm38/mm10 assembly) was identical in $Irf5^{-/-}$ and N10-*Siae* mice (Figure S4). Because $Cmah^{-/-}$ mice also lack MZ B cells (Cariappa et al., 2009), we looked for the presence of the $Dock2^{Hsd}$ allele in three $Cmah^{-/-}$ mice. The $Dock2^{Hsd}$ allele was present in all three $Cmah^{-/-}$ mice analyzed (Figure 4B). This duplication has been demonstrated to be a loss-of-function allele, because it results in a frameshift mutation and nonsense-mediated decay of *Dock2* mRNA (Purtha et al., 2012).

The Mutant Dock2 Allele Was Introduced by Backcrossing into C57BL/6NHsd Mice

Given that the same duplication was seen in three independently gene-targeted mice, it confirmed our suspicion that the mutant *Dock2* allele was inadvertently introduced during backcrossing. As previously noted, the SNPs flanking the duplication were of C57BL6/N origin (rs29473246 at chr11:33548367 and rs29414108 at chr11:43358462), suggesting that this duplication arose in a C57BL/6N substrain. A survey of several commercially available C57BL6/N substrains showed the presence of a duplication in exons 28 and 29 of Dock2 with an identical breakpoint in 44 of 44 C57BL/6NHsd mice analyzed (Figures 5A and S5). We henceforth refer to this allele as $Dock2^{Hsd}$. The $Dock2^{Hsd}$ allele was not detected in any other C57BL/6N strains analyzed,

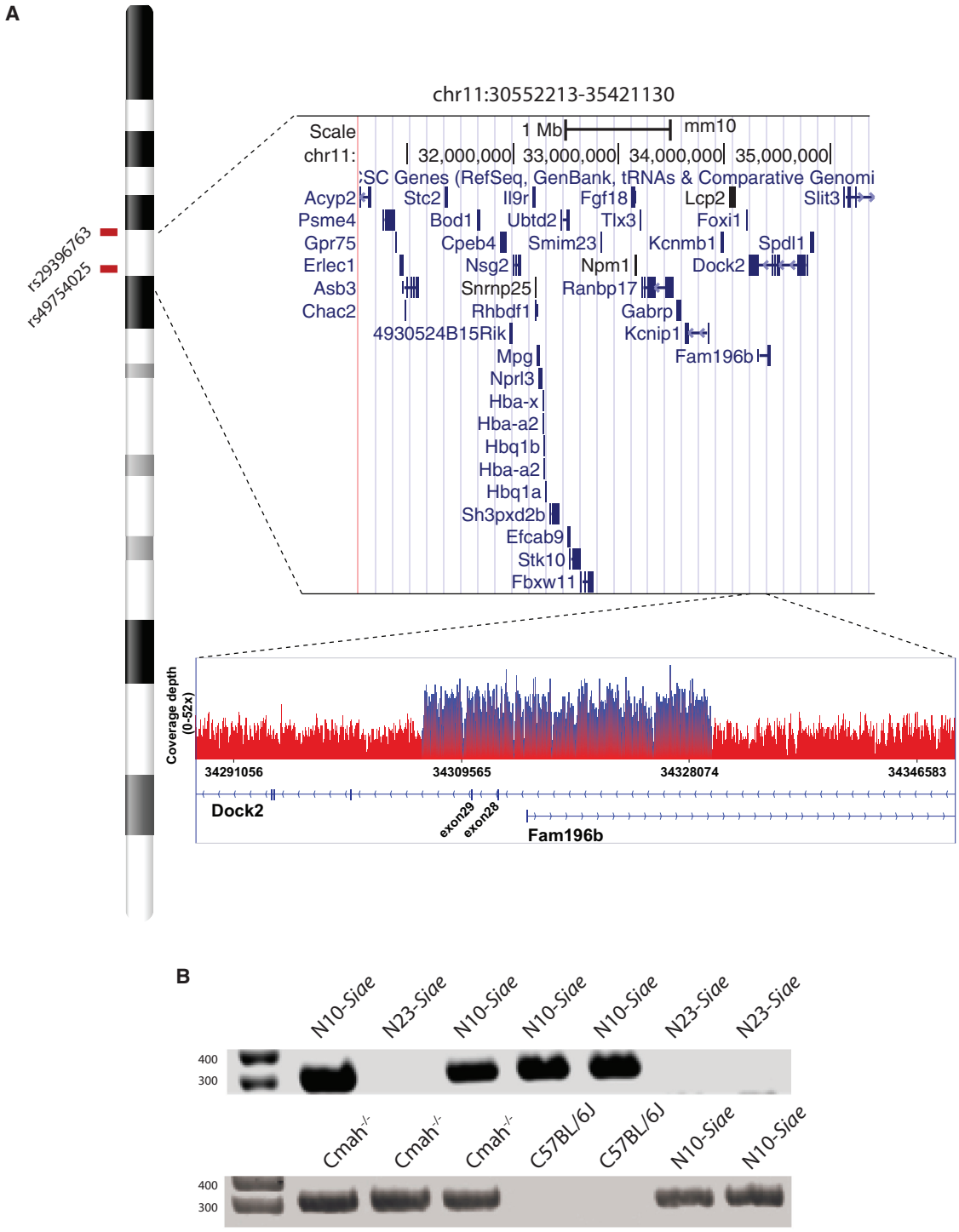


Figure 4. A Genomic Segment Encoding Exons 28 and 29 of *Dock2* Is Duplicated in N10-*Siae* and *Cmah*$^{-/-}$ Mice

(A) An abrupt increase in coverage depth within the putative pathogenic locus on chr11 mapped by SNPs encompasses exons 28 and 29 of *Dock2* in the N10-*Siae* genome.

(B) PCR detection of *Dock2* duplication in N10-*Siae*, Cmah$^{-/-}$, N23-*Siae*, and C57BL/6J mice.

See also Figures S3–S5.

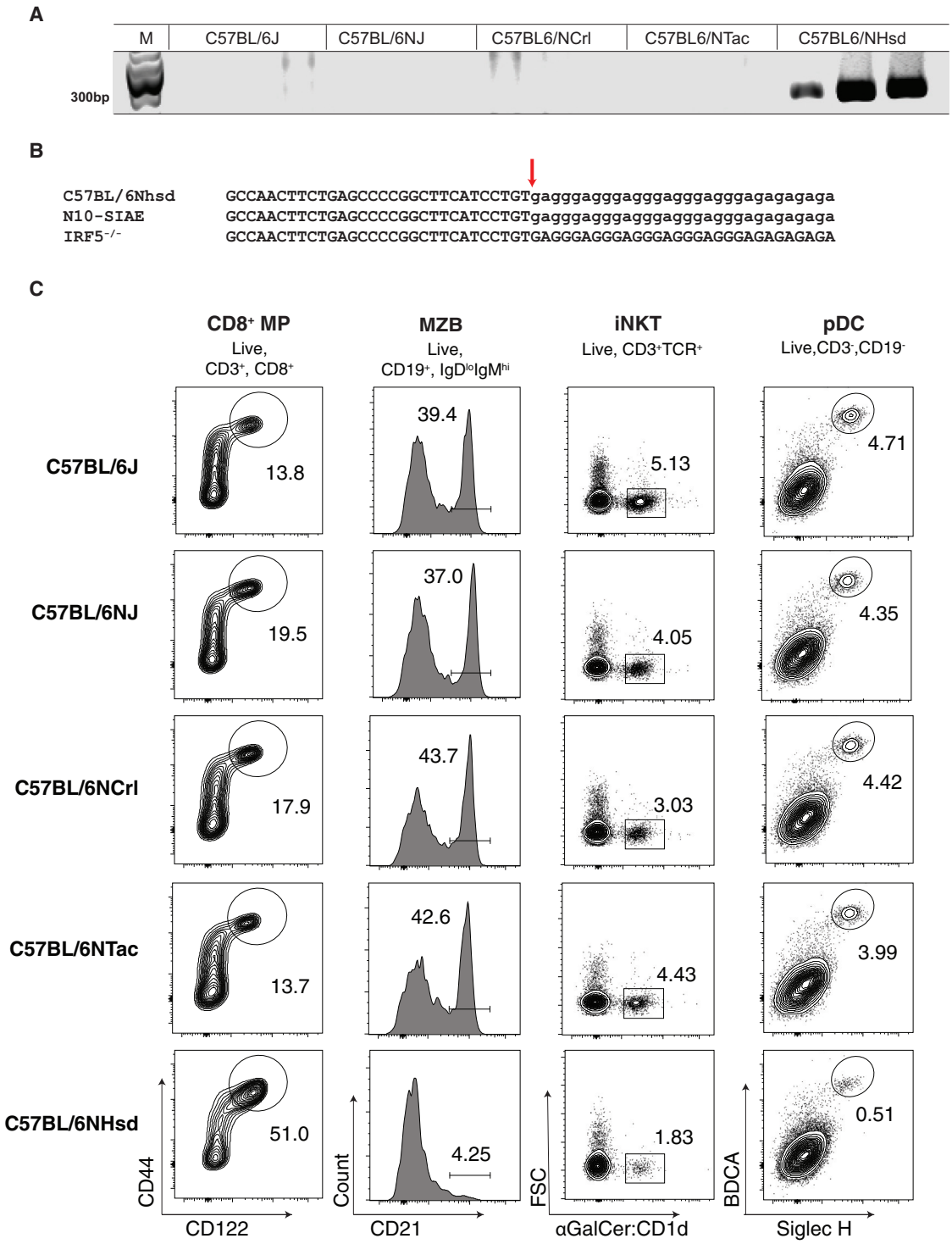


Figure 5. The Dock2 Duplication Is Present in the C57BL/6NHsd Strain, which Exhibits Multiple Characteristics of Dock2 Deficiency

(A) PCR detection of *Dock2* duplication in C57BL/6NHsd mice and other C57BL6/N sublines (C57BL/6NTac and C57BL6/NJ).

(B) Sequences of the Dock2 duplication breakpoint (arrow) from C57BL/6NHsd, N10-Siae, and IRF5$^{-/-}$ mice. A low-complexity sequence at the site of the breakpoint is shown in lowercase.

(C) Proportion of CD8^{+} MP cells, MZ B cells, invariant NK T cells, and plasmacytoid dendritic cells in C57BL/6NHsd, C57BL/6NTac, and C57BL/6J mice. The data shown are representative of five mice each.

See also Figure S5.

viz. C57BL/6NTac, C57BL/6NCrl, and C57BL/6NJ (Figure S5). This Dock2 duplication was also not observed in the C57BL/6NJ genome published by the Mouse Genomes Project (Wellcome Trust Sanger Institute) (Yalcin et al., 2012). Thus, the *Dock2*^{Hsd} allele appears to be a copy-number variant that has been fixed in the C57BL/6NHsd strain and may have been introduced into numerous other gene-targeted mouse lines during backcrossing. Consistent with this, both *Siae*$^{\Delta ex2/\Delta ex2}$ and *Cmah*$^{-/-}$ mice were generated from ESCs of 129 origin and backcrossed into the C57BL/6 background at UCSD using mice obtained from Harlan in the early 2000s.

The *Irf5*$^{-/-}$ mice were generated from an E14K ESC (129OlaHsd background) and backcrossed into the C57BL/6 background before being distributed to various investigators from a laboratory in Japan (Takaoka et al., 2005). However, the Dock2 duplication was found only in two of three colonies derived from the original *Irf5*$^{-/-}$ colony, which led the investigators to believe that the Dock2 duplication arose spontaneously within a subline of *Irf5*$^{-/-}$ mice (Purtha et al., 2012). Like the N10-*Siae* mice, the C57BL/6NHsd mice exhibit an expansion of CD8^{+} MP T cells (Figures 5C and S5). The C57BL/6NHsd mice also exhibit other phenotypes that have been previously observed in the *Dock2*$^{-/-}$ mice (Fukui et al., 2001). Both *Dock2*$^{-/-}$ and C57BL/6NHsd mice exhibit splenic cytopenia, as well as a profound loss of MZ B cells, invariant natural killer (NK) T cells, and plasmacytoid dendritic cells (Figure 5C) (Fukui et al., 2001; Gotoh et al., 2008; Kunisaki et al., 2006). The similarities between *Dock2*$^{-/-}$ and C57BL/6NHsd mice support the notion that the *Dock2*^{Hsd} allele inactivates the function of *Dock2* by creating a premature stop codon that contributes to non-sense-mediated decay of the resultant mRNA.

DISCUSSION

Striking phenotypes have been discovered in a number of gene-targeted mice that were previously assumed to be the result of the homozygous loss of a specific gene. To explain the loss of phenotypes in *Siae*$^{\Delta ex2/\Delta ex2}$ upon further backcrossing into the C57BL/6J background, we generated an additional mutant allele of *Siae*. This mutant mouse also lacked the phenotypes that we had previously reported.

To explain this loss of immune phenotypes, we undertook a series of studies involving conventional genetics, whole-genome genotyping, and whole-genome sequencing. This allowed us to identify a causal mutation on chr11 that was responsible for the previously observed immune phenotypes, such as the loss of MZ B cells and the increase in CD8+ MP T cells, in *Siae*$^{\Delta ex2/\Delta ex2}$ mice. We then found this same mutation, a homozygous duplication of two exons of *Dock2* that creates a non-functional product, in a subset of *Cmah* mutant mice. An identical mutation had been previously found in *Irf5*$^{-/-}$ mice, originally generated in Japan, and had been assumed to have arisen as a spontaneous mutation (Purtha et al., 2012; Takaoka et al., 2005). The most parsimonious explanation for a specific mutation being present in three different gene-targeted mice all derived from different ESCs, with two generated originally in California and a third generated in Japan, is that they had all been acquired by backcrossing into an unidentified mutant C57BL/6 strain. Although the loss of MZ B cells in *Cmah* mutant mice may be ascribed to the acquired *Dock2* mutation, other phenotypes have been reported in these mice that may relate to the absence of the N-glycolyl form of sialic acid observed in the absence of the Cmah protein (Naito et al., 2007). Thus, it remains possible that some functional phenotypes in such mice are related to altered interactions between sialic acids and Siglecs.

Because our SNP arrays showed the presence of specific C57BL/6N SNPs, we had a high degree of suspicion that this mutation had been acquired from a commercial C57BL/6N strain. We surveyed commercial C57BL/6N sublines and discovered that this *Dock2* gene duplication that was present only in C57BL/6N mice obtained from Harlan (C57BL/6NHsd). The set of 123 SNPs outside the *Siae* locus that differed between N10-*Siae* and C57BL/6J also comprised 97% of the SNP alleles (n = 128) from a pool of 538,667 SNPs analyzed in the whole-genome arrays that differ between C57BL/6J and C57BL/6N sublines but are identical to the consensus 129 alleles. Given that C57BL/6J and C57BL/6N strains were derived from a common stock of C57BL/6 mice, the differences between C57BL/6J and C57BL/6N substrains have been attributed so far to genetic drift. In contrast, our analysis indicates that C57BL/6N mice may have derived some genetic contribution from the 129 strain at a remote time during their history, before the divergence of various C57BL/6N sublines.

Siae$^{\Delta ex2/\Delta ex2}$ and *Cmah*-deficient mice were originally backcrossed into the C57BL/6 background using mice obtained from Harlan between 2005 and 2008. Currently available *Cmah* mutant mice that have been backcrossed into other C57BL/6 backgrounds lack the *Dock2* mutation. Initial reports of *Irf5*$^{-/-}$ mice backcrossed into the C57BL/6 background that bear phenotypes (e.g., lack of type I interferon) attributable to Dock2 deficiency were published in 2007 (Gotoh et al., 2010, 2008; Yasuda et al., 2007). Thus, we believe that the *Dock2* copy-number variant has been present in the C57BL/6NHsd colony at Harlan for at least a decade. The C57BL/6NHsd mice were derived by Harlan from the breeding nucleus of the C57BL/6N colony maintained at the NIH in 1988. Harlan was acquired by Envigo Biosciences in 2015. The C57BL/6N subline originated from the C57BL/6J colony at Jackson Laboratory in 1951 (Mekada et al., 2009). Given that other sublines derived from C57BL/6N and C57BL/6J strains lack the *Dock2*^{Hsd} variant, it is formally possible that a spontaneous mutation present in the C57BL/6N founders was expanded at Harlan. However, it is more likely that the *Dock2* duplication arose later in the C57BL/6NHsd colony after it diverged from the NIH breeding stock and was fixed by the breeding strategies based on continued brother-sister mating that are designed to minimize genetic drift in mouse sublines. Analysis of archived tissues from C57BL/6NHsd mice will help estimate when the *Dock2*^{Hsd} variant arose and help assess how many studies have been affected. Despite the relatively large size of the Dock2 gene, coding variants of *Dock2* have not been reported in other inbred mouse strains (Sherry et al., 2001).

The *Dock2*^{Hsd} variant is particularly significant because it causes wide-ranging hematopoietic phenotypes in a strain of mice that is widely used for immunological studies. Since 2003, Jackson Laboratory has implemented a Genetic Stability

Program in selected mouse strains to limit cumulative genetic drift, including that caused by copy-number variation, by regularly rebuilding foundation stocks from cryopreserved, pedigreed embryos every few generations (Taft et al., 2006). Commercial breeders of inbred mouse strains, as well as individual research groups maintaining knockout lines over long periods, should consider the implications of the choice of their breeding strategies on genetic drift within the colony.

The recent availability of ESCs from a C57BL/6 background obviates the need for backcrossing, but ESCs derived from the C57BL/6NHsd background should be tested for the presence of the *Dock2*^{Hsd} allele (Skarnes et al., 2011). Given that Dock2 is expressed primarily in the hematopoietic lineage, immune phenotypes that have been studied in the context of B6 mice from Harlan should be reviewed carefully. The *Dock2*^{Hsd} allele has likely affected the interpretation of a large number of studies in which C57BL/6NHsd mice obtained from Harlan were used as controls. We have identified a large number of studies in the literature that suggest to us that a plethora of assumed phenotypes should be revisited. A PCR screening protocol for the *Dock2*^{Hsd} allele is included in the Supplemental Experimental Procedures.

Loss-of-function variants of other genes, such as *Nnt*, *Mnrn1*, *Rd8*, and *Cyfip2*, have been described in one or more sublines of C57BL/6 and have been linked to glucose intolerance, impaired platelet function, retinal degeneration, and altered cocaine response, respectively (Freeman et al., 2006; Kumar et al., 2013; Mattapallil et al., 2012; Reheman et al., 2010). Loss-of-function variants of *Scna* are seen in some sublines of C57BL/6N, as well as in ESCs of C57BL/6N origin, but result in no obvious phenotype (Specht and Schoepfer, 2001). A copy-number variant that alters the expression of insulin-degrading enzyme (*Ide*) and fibroblast growth factor binding protein 3 (*Fgfbp3*) genes has been reported to have achieved a high allele frequency in C57BL/6J mice (Watkins-Chow and Pavan, 2008). While it is difficult to estimate the precise degree of genetic drift that results in deleterious or loss-of-function mutations, subline-specific variants can occur and should be considered as possible causes of phenotypic discrepancies in mouse sublines. Furthermore, the specific substrain of mice used for experiments or for backcrossing should be clearly documented.

EXPERIMENTAL PROCEDURES

Mice

An ESC clone (EPD0679_2_E06) bearing a targeted disruption of the Siae locus (*Siae*$^{tm1a(EUCOMM)Wtsi}$; MGI: 4842607) was obtained from the European Conditional Mouse Mutagenesis Program (EUCOMM) (Collins et al., 2007). The *Siae*$^{tm1a(EUCOMM)Wtsi}$ allele (*Siae*^{tm1a}) contains a beta-galactosidase reporter and a polyadenylation signal in intron 3 of *Siae*. Gene targeting had been initially confirmed using long-range PCR by EUCOMM and additionally validated by us with a Southern blot using the non-isotopic BrightStar Biodetect kit (Ambion) (Figure S1). The ESCs were then injected into C57BL/6 blastocysts at the Transgenic Core Facility at Brigham and Women's Hospital. The resulting chimeric mice were crossed into C57BL/6NTac. An *Siae*^{tm1a} founder was backcrossed for one generation into C57BL/6NTac and intercrossed to obtain *Siae*$^{tm1a/tm1a}$ homozygous mice. C57BL/6NHsd mice were purchased from Harlan Sprague Dawley (acquired by Envigo Biosciences in 2015). All mouse experimental protocols were reviewed and approved by the Institutional Animal Care and Use Committee.

Genetic Mapping

Whole-genome SNP arrays were performed at the Microarray Core Facility at the Dana Farber Cancer Institute using the Affymetrix Mouse Diversity Array platform. Analysis was performed using the Affymetrix Genotyping Console v.4.2.0.26. Publicly available SNP array data from various 129 substrains and C57BL/6 sublines were used for comparison (Didion et al., 2012). We focused on 538,667 SNPs that did not show any intra-strain variation among all C57BL/6N substrains, all 129 strains, and both N10-Siae mice analyzed. These were intended to represent the ancestral alleles of the 129 and C57BL6/N strains. Candidate SNPs were evaluated using PCR and Sanger sequencing. Sequencing chromatograms were analyzed using Mutation Surveyor v.3.24 (Softgenetics).

Whole-Genome Sequencing

A whole-genome library was constructed using the Kapa HyperPlus kit following manufacturer's recommendations. Briefly, 1 μg of genomic DNA was enzymatically fragmented, end repaired, deoxyadenosine (dA)-tailed, and ligated to Illumina Truseq adapters. No library amplification was performed to avoid introducing any coverage bias. The library was sequenced on the Illumina NextSeq 500 for 85 cycles using the NextSeq 500/550 High Output v.2 kit. A total of 567,104,840 single-end reads were aligned to the *Mus musculus* genome (GRCm38/mm10, December 2011 build) using Bowtie2 (Langmead and Salzberg, 2012). For 93.96% of reads mapping to the reference genome, 65.42% of reads mapped to unique sites on the genome and 28.54% of reads mapped to multiple regions. The GATK pipeline was used to identify variants (McKenna et al., 2010). Variants with quality scores < 30 or allele frequency < 100% were excluded, yielding 35,340 variants. Variant and read densities were calculated using BEDtools on Galaxy and visualized using the R statistical programming language (Quinlan and Hall, 2010).

Flow Cytometry

Mouse splenocytes or peripheral blood was hemolyzed with ammonium-chloride-potassium (ACK) buffer and stained with fluorescently conjugated antibodies analyzed on a BD LSR II flow cytometer. The following antibody clones were used in this study: anti-CD122 (TMB1), anti-CD44 (IM7), anti-CD3 (17A2), anti-CD8 (53-6.7), anti-CD19 (6D5), anti-B220 (RA3-6B2), anti-CD21 (7E9), anti-IgM (RMM-1), anti-CD1d (1B1), and anti-IgD (11-26c.2a) from BioLegend.

Statistical Analysis

Statistical analysis was performed using GraphPad Prism 6. The significance of differences between groups was evaluated using Student's t test.

ACCESSION NUMBERS

The accession number for the whole genome sequence of the N10-*Siae* mouse reported in this paper is NCBI Sequence Read Archive: SRA409984.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.04.080.

AUTHOR CONTRIBUTIONS

Conceptualization, V.S.M., E.D., and S.P.; Methodology, V.S.M., E.D., and H.M.; Software, V.V., V.S.M., and R.M.; Investigation, V.S.M., E.D., and H.M.; Resources, A.V.; Writing – Original Draft, V.S.M., E.D., and S.P.; Writing – Review & Editing, V.S.M., E.D., H.M., and S.P.; Visualization, V.S.M., E.D., and V.V.; Supervision, S.P. and V.S.M.

ACKNOWLEDGMENTS

We thank Patrick Secrest for the examination of mice currently at UCSD. This work was supported by grants from the NIH (AI064930 and GM032373).

Received: February 3, 2016
Revised: March 28, 2016
Accepted: April 21, 2016
Published: May 19, 2016

REFERENCES

Cariappa, A., Takematsu, H., Liu, H., Diaz, S., Haider, K., Boboila, C., Kalloo, G., Connole, M., Shi, H.N., Varki, N., et al. (2009). B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. J. Exp. Med. *206*, 125–138.

Collins, F.S., Rossant, J., and Wurst, W.; International Mouse Knockout Consortium (2007). A mouse for all reasons. Cell *128*, 9–13.

Didion, J.P., Yang, H., Sheppard, K., Fu, C.-P., McMillan, L., de Villena, F.P.-M., and Churchill, G.A. (2012). Discovery of novel variants in genotyping arrays improves genotype retention and reduces ascertainment bias. BMC Genomics *13*, 34.

Freeman, H.C., Hugill, A., Dear, N.T., Ashcroft, F.M., and Cox, R.D. (2006). Deletion of nicotinamide nucleotide transhydrogenase: a new quantitive trait locus accounting for glucose intolerance in C57BL/6J mice. Diabetes *55*, 2153–2156.

Fukui, Y., Hashimoto, O., Sanui, T., Oono, T., Koga, H., Abe, M., Inayoshi, A., Noda, M., Oike, M., Shirai, T., and Sasazuki, T. (2001). Haematopoietic cell-specific CDM family protein DOCK2 is essential for lymphocyte migration. Nature *412*, 826–831.

Gotoh, K., Tanaka, Y., Nishikimi, A., Inayoshi, A., Enjoji, M., Takayanagi, R., Sasazuki, T., and Fukui, Y. (2008). Differential requirement for DOCK2 in migration of plasmacytoid dendritic cells versus myeloid dendritic cells. Blood *111*, 2973–2976.

Gotoh, K., Tanaka, Y., Nishikimi, A., Nakamura, R., Yamada, H., Maeda, N., Ishikawa, T., Hoshino, K., Uruno, T., Cao, Q., et al. (2010). Selective control of type I IFN induction by the Rac activator DOCK2 during TLR-mediated plasmacytoid dendritic cell activation. J. Exp. Med. *207*, 721–730.

Hedlund, M., Tangvoranuntakul, P., Takematsu, H., Long, J.M., Housley, G.D., Kozutsumi, Y., Suzuki, A., Wynshaw-Boris, A., Ryan, A.F., Gallo, R.L., et al. (2007). N-glycolylneuraminic acid deficiency in mice: implications for human biology and evolution. Mol. Cell. Biol. *27*, 4340–4346.

Kumar, V., Kim, K., Joseph, C., Kourrich, S., Yoo, S.-H., Huang, H.C., Vitaterna, M.H., de Villena, F.P.-M., Churchill, G., Bonci, A., and Takahashi, J.S. (2013). C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates cocaine response. Science *342*, 1508–1512.

Kunisaki, Y., Tanaka, Y., Sanui, T., Inayoshi, A., Noda, M., Nakayama, T., Harada, M., Taniguchi, M., Sasazuki, T., and Fukui, Y. (2006). DOCK2 is required in T cell precursors for development of Valpha14 NK T cells. J. Immunol. *176*, 4640–4645.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Mahajan, V.S., and Pillai, S. (2016). Sialic acids and autoimmune disease. Immunol. Rev. *269*, 145–161.

Mattapallil, M.J., Wawrousek, E.F., Chan, C.-C., Zhao, H., Roychoudhury, J., Ferguson, T.A., and Caspi, R.R. (2012). The Rd8 mutation of the Crb1 gene is present in vendor lines of C57BL/6N mice and embryonic stem cells, and confounds ocular induced mutant phenotypes. Invest. Ophthalmol. Vis. Sci. *53*, 2921–2927.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Mekada, K., Abe, K., Murakami, A., Nakamura, S., Nakata, H., Moriwaki, K., Obata, Y., and Yoshiki, A. (2009). Genetic differences among C57BL/6 substrains. Exp. Anim. *58*, 141–149.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proc. Natl. Acad. Sci. USA *90*, 8424–8428.

Naito, Y., Takematsu, H., Koyama, S., Miyake, S., Yamamoto, H., Fujinawa, R., Sugai, M., Okuno, Y., Tsujimoto, G., Yamaji, T., et al. (2007). Germinal center marker GL7 probes activation-dependent repression of N-glycolylneuraminic acid, a sialic acid species involved in the negative modulation of B-cell activation. Mol. Cell. Biol. *27*, 3008–3022.

Pillai, S., Cariappa, A., and Pirnie, S.P. (2009). Esterases and autoimmunity: the sialic acid acetylesterase pathway and the regulation of peripheral B cell tolerance. Trends Immunol. *30*, 488–493.

Purtha, W.E., Swiecki, M., Colonna, M., Diamond, M.S., and Bhattacharya, D. (2012). Spontaneous mutation of the Dock2 gene in Irf5-/- mice complicates interpretation of type I interferon production and antibody responses. Proc. Natl. Acad. Sci. USA *109*, E898–E904. http://dx.doi.org/10.1073/pnas.1118155109.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Reheman, A., Tasneem, S., Ni, H., and Hayward, C.P.M. (2010). Mice with deleted multimerin 1 and alpha-synuclein genes have impaired platelet adhesion and impaired thrombus formation that is corrected by multimerin 1. Thromb. Res. *125*, e177–e183.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T., et al. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337–342.

Specht, C.G., and Schoepfer, R. (2001). Deletion of the alpha-synuclein locus in a subpopulation of C57BL/6J inbred mice. BMC Neurosci. *2*, 11.

Taft, R.A., Davisson, M., and Wiles, M.V. (2006). Know thy mouse. Trends Genet. *22*, 649–653.

Takaoka, A., Yanai, H., Kondo, S., Duncan, G., Negishi, H., Mizutani, T., Kano, S., Honda, K., Ohba, Y., Mak, T.W., and Taniguchi, T. (2005). Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. Nature *434*, 243–249.

Watkins-Chow, D.E., and Pavan, W.J. (2008). Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. Genome Res. *18*, 60–66.

Yalcin, B., Adams, D.J., Flint, J., and Keane, T.M. (2012). Next-generation sequencing of experimental mouse strains. Mamm. Genome *23*, 490–498.

Yasuda, K., Richez, C., Maciaszek, J.W., Agrawal, N., Akira, S., Marshak-Rothstein, A., and Rifkin, I.R. (2007). Murine dendritic cell type I IFN production induced by human IgG-RNA immune complexes is IFN regulatory factor (IRF)5 and IRF7 dependent and is required for IL-6 production. J. Immunol. *178*, 6876–6885.

Yasuda, K., Nündel, K., Watkins, A.A., Dhawan, T., Bonegio, R.G., Ubellacker, J.M., Marshak-Rothstein, A., and Rifkin, I.R. (2013). Phenotype and function of B cells and dendritic cells from interferon regulatory factor 5-deficient mice with and without a mutation in DOCK2. Int. Immunol. *25*, 295–306.

Zurita, E., Chagoyen, M., Cantero, M., Alonso, R., González-Neira, A., López-Jiménez, A., López-Moreno, J.A., Landel, C.P., Benítez, J., Pazos, F., and Montoliu, L. (2011). Genetic polymorphisms among C57BL/6 mouse inbred strains. Transgenic Res. *20*, 481–489.