The Journal of Experimental Medicine

Enteropathy-associated T cell lymphoma subtypes are characterized by loss of function of SETD2

Andrea B. Moffitt,[1,3] Sarah L. Ondrejka,[6] Matthew McKinney,[3] Rachel E. Rempel,[3] John R. Goodlad,[7] Chun Huat Teh,[8] Sirpa Leppa,[9] Susanna Mannisto,[9] Panu E. Kovanen,[10] Eric Tse,[11] Rex K.H. Au-Yeung,[11] Yok-Lam Kwong,[11] Gopesh Srivastava,[11] Javeed Iqbal,[12] Jiayu Yu,[12] Kikkeri Naresh,[13] Diego Villa,[14] Randy D. Gascoyne,[14] Jonathan Said,[15] Magdalena B. Czader,[16] Amy Chadburn,[17] Kristy L. Richards,[18] Deepthi Rajagopalan,[3] Nicholas S. Davis,[3] Eileen C. Smith,[3] Brooke C. Palus,[3] Tiffany J. Tzeng,[3] Jane A. Healy,[3] Patricia L. Lugar,[4] Jyotishka Datta,[2] Cassandra Love,[3] Shawn Levy,[19] David B. Dunson,[2] Yuan Zhuang,[5] Eric D. Hsi,[6] and Sandeep S. Dave[1,3]

[1]Duke Center for Genomics and Computational Biology and [2]Department of Statistical Science, Duke University, Durham, NC 27708
[3]Duke Cancer Institute, [4]Department of Medicine, and [5]Department of Immunology, Duke University School of Medicine, Durham, NC 27710
[6]Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH 44195
[7]Haematological Malignancy Diagnostic Service, St. James's University Hospital, Leeds LS9 7TF, England, UK
[8]Haematology Department, Western General Hospital, Edinburgh EH14 1TY, Scotland, UK
[9]Department of Oncology and Research Program Unit, Faculty of Medicine and [10]HUSLAB and Medicum, Helsinki University Hospital Cancer Center and University of Helsinki, 00014 Helsinki, Finland
[11]University of Hong Kong, Queen Mary Hospital, Hong Kong, China
[12]University of Nebraska Medical Center, Omaha, NE 68198
[13]Imperial College London, London SW7 2AZ, England, UK
[14]British Columbia Cancer Agency, University of British Columbia, Vancouver, BC V6R 1ZE, Canada
[15]University of California, Los Angeles, Los Angeles, CA 90095
[16]Indiana University, Indianapolis, IN 46202
[17]Presbyterian Hospital, Pathology and Cell Biology, Cornell University, New York, NY 10065
[18]University of North Carolina at Chapel Hill, Chapel Hill, NC 27599
[19]Hudson Alpha Institute for Biotechnology, Huntsville, AL 35806

Enteropathy-associated T cell lymphoma (EATL) is a lethal, and the most common, neoplastic complication of celiac disease. Here, we defined the genetic landscape of EATL through whole-exome sequencing of 69 EATL tumors. _SETD2_ was the most frequently silenced gene in EATL (32% of cases). The JAK-STAT pathway was the most frequently mutated pathway, with frequent mutations in _STAT5B_ as well as _JAK1_, _JAK3_, _STAT3_, and _SOCS1_. We also identified mutations in _KRAS_, _TP53_, and _TERT_. Type I EATL and type II EATL (monomorphic epitheliotropic intestinal T cell lymphoma) had highly overlapping genetic alterations indicating shared mechanisms underlying their pathogenesis. We modeled the effects of SETD2 loss in vivo by developing a T cell–specific knockout mouse. These mice manifested an expansion of γδ T cells, indicating novel roles for SETD2 in T cell development and lymphomagenesis. Our data render the most comprehensive genetic portrait yet of this uncommon but lethal disease and may inform future classification schemes.

INTRODUCTION

Enteropathy-associated T cell lymphoma (EATL) is an uncommon but lethal type of peripheral T cell lymphoma that appears to be rising in incidence (Swerdlow et al., 2008; Sharaiha et al., 2012). EATL is the most common oncologic complication of celiac disease, with a prevalence of ∼1% in those patients (Ferreri et al., 2011). The role of genetic mutations in the pathogenesis of EATL is largely unknown.

Past World Health Organization lymphoma classification systems (Swerdlow et al., 2008; Campo et al., 2011) have recognized two subtypes of EATL. Type I EATL has a more variable histology, is more prevalent in Northern Europe (Delabie et al., 2011), and is strongly associated with celiac disease. Type II EATL has a more uniform histology, occasional association with celiac disease (Delabie et al., 2011), and is more prevalent in Asia (Sun et al., 2011; Tse et al., 2012; Kikuma et al., 2014). The diagnostic distinction between these two types is not always clear. Cases are currently classified based on their morphology and immunophenotype, with both types sharing common T cell markers but type II cases

Correspondence to Sandeep S. Dave: sandeep.dave@duke.edu

Abbreviations used: DN, double negative; DP, double positive; EATL, enteropathy-associated T cell lymphoma; HSTL, hepatosplenic T cell lymphoma; IEL, intraepithelial lymphocyte; rRNA, ribosomal RNA; SNV, single nucleotide variant.

expressing CD56 more frequently (Delabie et al., 2011; Arps and Smith, 2013). Most type I patients carry the celiac disease–associated HLA genotype DQA1*0501, DQB1*0201, whereas that genotype occurs at a frequency similar to the general population in type II patients. Based on these distinctions, it has been recommended that type II EATL should be reclassified as a distinct entity, known as monomorphic epitheliotropic intestinal T cell lymphoma (Swerdlow et al., 2016). The genomic distinction between the two described types of EATL is not well understood. Compared with other peripheral T cell lymphomas, which are primarily derived from αβ T cells, both types of EATL frequently arise from γδ T cells, a feature they share with hepatosplenic T cell lymphoma (HSTL; Belhadj et al., 2003).

In this study, we applied whole-exome and transcriptome sequencing to define novel driver mutations, copy-number alterations, and expression patterns in both types of EATL. We found that both types of EATL are characterized by frequent silencing mutations in *SETD2*. We further modeled the effects of SETD2 loss in vivo by developing a T cell–specific knockout of the gene in mice. We found that mice with *Setd2* loss manifested an expansion of γδ T cells, suggesting novel roles for SETD2 in T cell development. Together, our data render the most comprehensive genetic portrait yet of EATL and may inform future classification schemes of this disease.

RESULTS

Whole–exome sequencing defines driver mutations and copy–number alterations in EATL

We performed whole-exome sequencing on 69 EATL tumors along with paired normal samples when available (*n* = 36), for a total of 105 exomes. Mean exome sequencing depth was greater than 70×. EATL cases were separated into the 36 cases with paired normal tissue available, deemed the "discovery set," and the remaining samples, a "validation set."

Fig. 1 a shows the mutation status of each significant EATL driver gene in each EATL case, with the number of cases summarized in the bar graphs of Fig. 1 b. Every gene shown had at least two confirmed somatic mutations in the discovery set with additional rare variants in the validation set that were comparable to those in the discovery set in terms of their frequency in the general population, distribution of amino acid alterations, location in protein domains, and evolutionary conservation, as we have described previously (Love et al., 2012; Zhang et al., 2014).

The most frequently mutated gene in EATL was found to be *SETD2* (32% of cases), with the mutations featuring a large proportion (41%) of loss-of-function frameshift indels or nonsense mutations. Other chromatin modifiers were also mutated in a significant number of cases, such as *TET2* and *YLPM1*, and typically exhibit loss-of-function type mutations.

The JAK-STAT pathway was the most commonly mutated signaling pathway, with frequent activating mutations in *STAT5B* (29%), followed by *JAK1* (23%), *JAK3* (23%), *STAT3* (16%), and frameshift mutations in the negative reg-

ulator of the pathway, *SOCS1* (7%). Interestingly, the RAS family of genes, which are rarely mutated in lymphomas, were also implicated in several EATL cases, with mutations at known activating hotspots (G12/G13) in *NRAS* and *KRAS*.

DNA damage response and repair genes were frequently mutated in EATL. *TP53* was mutated in 10% of the cases. Other DNA damage–related genes mutated include *BCL11B*, a known tumor suppressor in T cell lymphomas, and *BRIP1*, a DNA helicase known to be involved in the repair of double-strand breaks.

Other noteworthy genes mutated in EATL included apoptosis-associated *DAPK3* (10%), the cell cycle transcription factor *BBX* (16%), the interferon–related transcription factor *PRDM1* (9%), and the telomerase reverse transcription *TERT* (17%). Full details of somatic mutations detected in EATL cases are described in Table S1.

We analyzed the mutual exclusion patterns of the mutated driver genes using the weighted row exclusivity test (Leiserson et al., 2016). Examining sets of 3 genes, we found 11 significant sets, involving 12 different genes, which are listed in Table S2. The JAK-STAT pathway is prominent among the mutually exclusive genes, with all of the significant sets including at least one of the genes from *STAT5B*, *STAT3*, *JAK1*, and *JAK3*, as shown in Fig. S1 a. We also quantified how often each gene pair occurs among the significant sets of four genes, which showed *PRDM1* and *DAPK3*, followed by *STAT3* and *STAT5B*, as the most common mutually exclusive gene pairs (Fig. S1 b).

To further understand the genetic drivers of EATL, we investigated the distribution of allele frequencies of mutations in the driver genes and the fraction of mutations that appear to be clonal. Estimates of cancer cell fraction based on observed allele frequencies for each driver gene are shown in Fig. S1 (c and d). *STK10*, *DAPK3*, *TP53*, and *STAT5B* showed the highest proportion of clonal events among the driver genes.

We next examined copy-number alterations in these cases. Interestingly, most of the copy-number alterations in EATL cases occurred in very large regions, frequently comprising an entire arm of the chromosome. After correcting for these effects, there were no focal gene-level alterations that reached statistical significance.

Fig. 1 c depicts the significant copy number alteration status of each EATL case by chromosomal arm. Similar to previous studies investigating both type I and type II EATLs (deLeeuw et al., 2007; Ko et al., 2010; Tomita et al., 2015), we observed amplifications in chromosome 9q to be the most common in EATL (46%). We also noted gains in chromosomes 1q, 7q, and 8q, as well as losses in 8p, which were each present in 20–30% of cases. Many cases had more than one of the noted alterations. The total number of events per case out of the shown EATL driver events is depicted in Fig. 1 d. Additional details of copy-number alterations are in Table S3.

Clinical features of the EATL cohort

Median overall survival was 10 mo, with a 27% 2-yr survival rate (Fig. 1 e), highlighting the dismal prognosis in these pa-
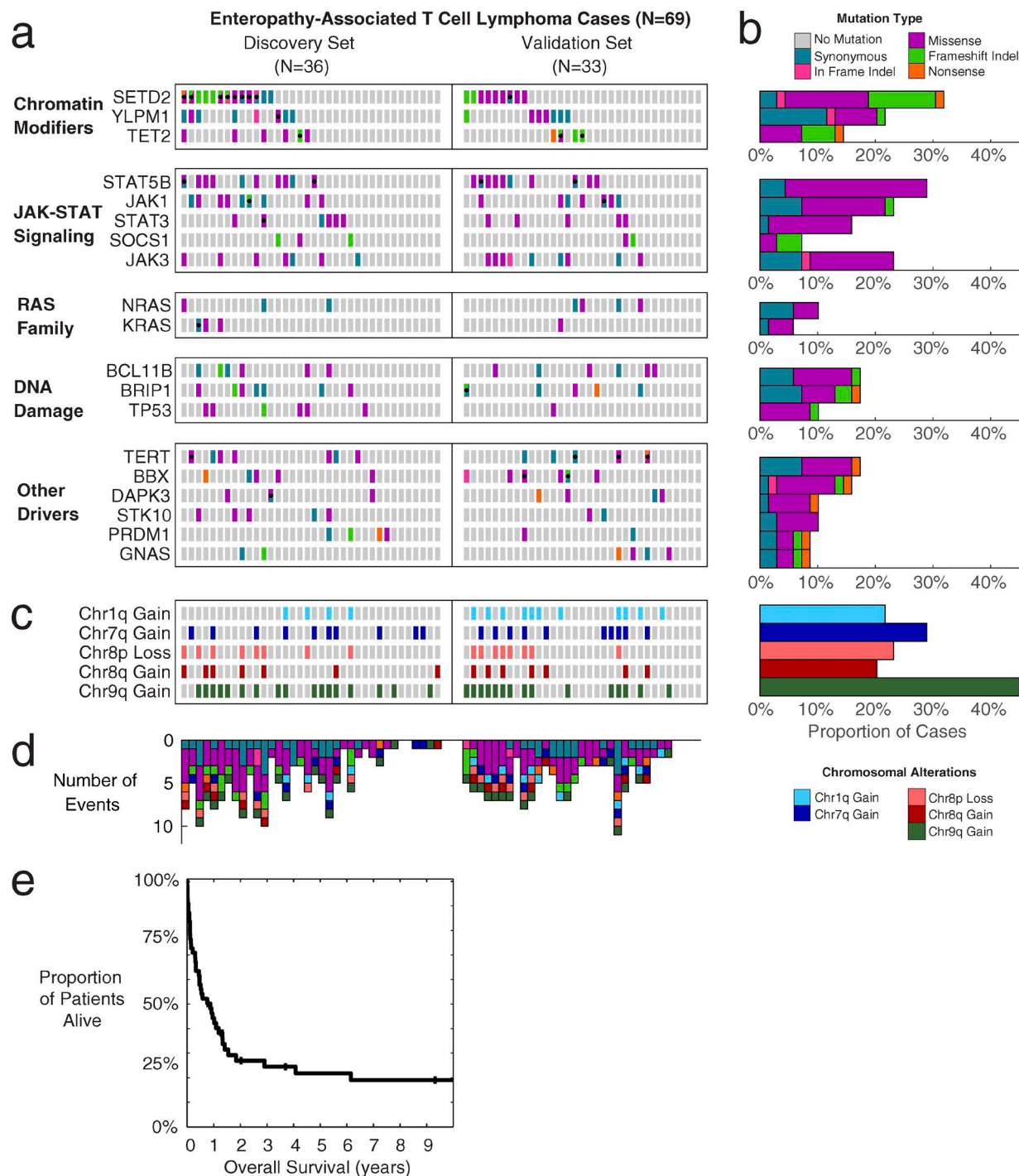
Figure 1. Somatic mutations, copy-number alterations, and survival of EATL patients. (a) Heat map of mutations in EATL cohort ($n = 69$). Every box represents the mutation status of a patient for a particular gene. Columns are split into discovery set with paired normals ($n = 36$) and validation set ($n = 33$). Gray, no mutation; teal, synonymous SNV; pink, in-frame indel; purple, missense SNV; green, frameshift indel; orange, nonsense SNV. Black dots indicate more than one mutation in that gene/patient, with boxes split diagonally to show different functions of multiple mutations. (b) Bar graph showing the percentage of cases in the cohort affected by each mutated gene. Bars are color coded by the most-damaging event type observed in each patient. (c) Heat map of arm-level copy-number alterations for each patient ($n = 69$). Light blue, chr1q gain; dark blue, chr7q gain; light red, chr8p loss; dark red, chr8q gain; dark green, chr9q gain. (d) Bar graph showing number of events from mutated genes and copy-number alterations shown in the heat maps above ($n = 69$). Bars are color-coded based on the type of alteration. (e) Kaplan-Meier curve showing overall survival of the EATL cohort ($n = 55$ with available outcome data). Median survival is 10 mo; 1-yr survival rate is 44%.

tients. Good performance status (P = 0.003), low stage (P = 0.006), and complete response to initial treatment (P < 0.001) were all associated with better outcome (Fig. S1, e–g).

Assignment of each case to the two types of EATL was determined based on World Health Organization criteria (Swerdlow et al., 2008). The two types of EATL were remarkably similar in terms of overall survival (median survival: 9 mo for type I, 11 mo for type II; Fig. S1 h), age of onset (64 yr for type I, 59 yr for type II), and gender (44% male for type I, 52% male for type II). As expected, the type I patients were primarily Caucasian (93%), whereas the type II patients were split (57% Caucasian and 39% Asian).

History or concurrent diagnosis of celiac disease was observed in 84% of type I EATL cases compared with 16% of type II cases. The DQ2 or DQ8 HLA genotypes were observed in 88% of type I EATL patients, and 43% of type II EATL patients (Table S4). Among patients with type I EATL, adherence to a gluten-free diet appeared to correlate with improved outcomes (P = 0.07; Fig. S1 i).

Additional clinical parameters are summarized in Table 1, with full clinical and pathological details in Table S5.

Genomic and transcriptomic differences between type I and type II EATL

Given the paucity of our understanding of the molecular distinction between the two described types of EATL, we further investigated the mutation frequencies in type I (*n* = 41) and type II (*n* = 23) EATLs (Fig. 2 a). Overall, we found that the overwhelming majority of the genes (90%) to be mutated in both types of EATL. A few selected genes appeared to be mutated more frequently in one type or the other. *SOCS1* and *DAPK3* had higher frequencies in type I, whereas *KRAS*, *STAT5B* (>50% in type II), and *SETD2* (>60% in type II) had higher frequencies in type II EATL. Fig. 2 b shows similar findings when we examined copy-number alterations separately in type I and II cases. Chromosome 9q gains were the most common alteration in both types. Diagnostic guidelines have attributed 1q gains primarily to type I EATL (Arps and Smith, 2013); however, we found chromosome 1q gains to be present in at least 20% of cases in both types of EATL. Chromosome 8 alterations, particularly gains in chromosome 8q, which includes *MYC*, were significantly more frequent in type II EATL but also occurred in type I.

These data indicate that type I and type II EATL cases have strongly overlapping genetic alterations and likely arise from shared genetic mechanisms.

We further investigated the gene expression profiles through transcriptome sequencing of 29 EATL tumors for which RNA was available, including type I (*n* = 16) and type II (*n* = 13) cases. Unbiased principal-component analysis revealed that a large amount of variation in the gene expression data are driven by the differences between type I and type II cases. Plotting each of the EATL samples in the principal component space (Fig. 2 c) shows that the majority of samples segregate into the two groups corresponding essentially

to the EATL subtypes, indicating robust differences in gene expression levels between the two subtypes. Based on these findings, we performed supervised differential expression analysis using the clinically derived labels of type I and type II. Fig. 2 d shows the genes that were at least 1.5-fold differentially expressed between the two types, with a false discovery rate of less than 0.1. Some notable genes with higher expression in the type I cases, compared with type II, included *STAT3*, *STAT5A*, *IRF1*, and *IRF4*. The genes *FASLG*, *SYK*, and *TGBR1* were found to be more highly expressed in type II EATL than in type I. Transglutaminase 2 (*TGM2*) is an important enzyme and autoantigen in celiac disease (Dieterich et al., 1997), which we found to be expressed at higher levels in the type I cases than in type II cases. *NCAM1*, also known as CD56, was expressed higher in type II EATL, consistent with the use of CD56 staining as a diagnostic feature of type II EATL. We did not find *MYC* to be differentially expressed, despite the higher proportion of chromosome 8q amplifications in type II. The full list of differentially expressed genes is included in Table S6.

Fig. 2 e depicts the heat map of the differentially expressed genes in the 29 EATL samples, along with HLA genotype and celiac disease status. With a false discovery rate (FDR) <0.1 and a fold change of at least 1.5× in either direction, 380 genes are higher in type II and 198 genes are higher in type I. The cluster assignment, based on consensus K-means clustering of the differentially expressed genes, was found to track very closely with the original clinical EATL subtype.

Gene set enrichment of the differentially expressed genes revealed significant overexpression of the interferon-γ signaling pathway (P < 0.001; Fig. 2, f and g) in type I EATLs. Genes expressed higher in type II EATL were found to be enriched for the natural killer–like cytotoxicity pathway (P = 0.02; Fig. 2, h and i). Both interferon-γ signaling and natural killer–like cytotoxicity (Meresse et al., 2006) are known functions of intraepithelial lymphocytes (IELs), which are the presumed cell of origin for EATLs (Ebert, 1990; Taguchi et al., 1991; Ishikawa et al., 1993; Kagnoff, 1998; Corazza et al., 2000). Increased interferon-γ production has been described in patients with untreated celiac disease, consistent with the higher expression of this function in celiac-associated type I EATL (Kagnoff, 1998; Nilsen et al., 1998). Other enriched gene sets of interest included MAPK signaling, up-regulated in type II, consistent with a previous study of high MAPK protein expression in type II (Kikuma et al., 2014).

Thus, the two EATL types are characterized by highly overlapping genetic alterations, as well as robust differences in gene expression.

To further investigate the profiles of the EATL tumors, we compared the gene expression of the EATL samples to published data on purified T cell subsets in mouse (Heng et al., 2008). When ranking the T cell samples by mean correlation to the EATL tumors, we find that the top ranked T cell subsets with most similarity to EATL samples include primarily γδ T cells (16/20), particularly activated IEL sam-

Table 1.   Summary of EATL clinical characteristics

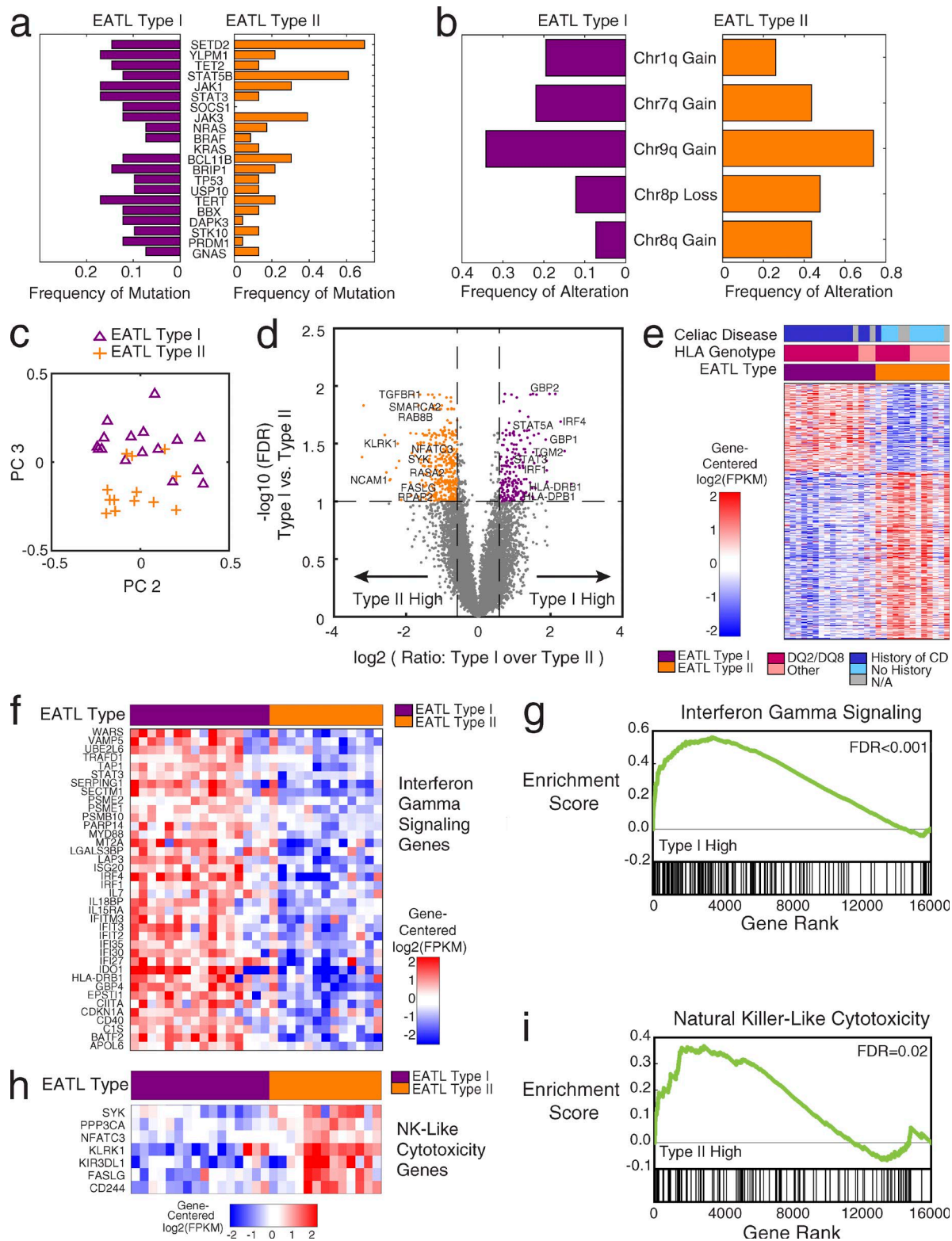| Parameter	All EATL (*n* = 63)	Type I (*n* = 41)	Type II (*n* = 23)
	Number (percentage) of cases		
Sex			
Male	33 (48)	18 (44)	12 (52)
Female	36 (52)	23 (56)	11 (48)
Race			
African-American	2 (3)	1 (2)	1 (4)
Caucasian	56 (81)	38 (93)	13 (57)
Asian	10 (14)	1 (2)	9 (39)
Age at diagnosis (yr)			
Median (range)	62.0 (36.0–92.0)	64.0 (36–92)	59.0 (39.0–89.0)
Celiac disease			
Yes	35 (61)	32 (84)	3 (16)
Adherence to gluten–free diet			
Yes	15 (36)	15 (44)	0 (0)
HLA type			
DQ2/DQ8	49 (71)	36 (88)	10 (43)
Tumor location			
Small intestine	61 (93)	40 (97)	20 (91)
Size of tumor (cm)			
Median (range)	7.0 (1.5–75.0)	8.0 (1.5–75.0)	6.0 (2.0–23.0)
Histopathology			
CD3$^+$	57 (85)	30 (75)	23 (100)
CD4$^-$	50 (91)	29 (91)	18 (95)
CD8$^+$	31 (46)	15 (37)	15 (68)
CD56$^+$	19 (33)	2 (5)	17 (85)
T cell receptor			
AB	16 (46)	12 (43)	4 (57)
GD	9 (26)	6 (21)	3 (43)
ECOG performance score			
0–1	20 (48)	14 (50)	6 (43)
2–4	22 (52)	14 (50	8 (57)
Ann Arbor stage			
I	17 (40)	10 (36)	7 (47)
II	18 (42)	14 (50)	4 (27)
III–IV	8 (19)	4 (14)	4 (27)
High LDH			
High LDH	31 (89)	21 (88)	10 (91)
B symptoms			
Yes	27 (59)	19 (63)	7 (47)
Small bowel perforation			
Yes	33 (57)	22 (58)	11 (55)
Number of nodal sites			
0	23 (48)	14 (48)	9 (56)
1 or more	15 (52)	15 (52)	7 (44)
Bone marrow involvement			
No	32 (89)	20 (91)	12 (92)
Treatment type			
Curative	32 (64)	19 (58)	12 (75)
Minimal	18 (36)	14 (42)	4 (25)
Response to initial treatment			
CR	22 (63)	15 (71)	7 (50)
PR	5 (14)	3 (14)	2 (14)
NR	8 (23)	3 (14)	5 (36)
Progression or relapse			
No	8 (24)	6 (32)	2 (14)
Yes	25 (76)	13 (68)	12 (86)
Survival			
Median overall survival	10 mo	9 mo	11 mo
1-yr survival rate	44%	41%	50%
2-yr survival rate	27%	31%	19%

Figure 2. Differences between type I and type II EATL: mutations, copy number, and gene expression. (a) Bar graph indicating frequency of mutation of each gene within the type I ($n = 41$) and type II ($n = 23$) EATL cohorts separately: type I frequency (purple; left); type II frequency (orange; right). (b) Bar graph indicating frequency of alteration of each arm-level copy-number alteration within the type I ($n = 41$) and type II ($n = 23$) EATL cohorts separately: type I frequency (purple; left); type II frequency (orange; right). (c) Principal-component (PC) analysis of EATL RNA-sequencing gene expression data ($n = 29$). The second principal component is on the x axis, and the third principal component is on the y-axis. One point is plotted for each sample in the analysis

ples. However, we did not find significant evidence within this dataset for T cell subsets that specifically resembled either type I or type II EATLs. More specific profiling data will be needed in future studies to fully answer the question of precise cell of origin for these EATL subtypes.

SETD2 is frequently mutated in EATL and plays a role in T cell development in vivo

Our mutation analysis in EATL tumors showed the histone lysine methyltransferase *SETD2* to be the most frequently mutated gene in the disease (22 of 69 cases). The recurrent nonsense mutations and frameshift insertions and deletions (41% of *SETD2* mutated cases) suggest that the role of mutations in this gene in EATL is likely to be loss of function. Further, nine EATL cases had multiple mutations in *SETD2*, further supporting its role as a tumor suppressor in EATL. *SETD2* has also been implicated as a driver gene with loss-of-function mutations in other malignancies, including clear cell renal carcinoma (Dalgliesh et al., 2010), DLBCL (Zhang et al., 2013), and acute leukemia (Zhu et al., 2014).

SETD2 is a histone lysine methyltransferase responsible for trimethylation of lysine 36 on histone 3 (H3K36me3), which is an epigenetic marker for active transcription. Its interactions with RNA polymerase II during transcription occur through the SRI domain (Kizer et al., 2005).

The distribution of mutations in *SETD2* in our EATL cases is shown in Fig. 3 a (top). In a separate study examining the genetics of HSTL, a PTCL that arises predominantly from γδ T cells (McKinney et al., 2017), we also found *SETD2* to be the most frequently silenced gene. For comparison, the distribution of mutations that we found in HSTL (McKinney et al., 2017) are also shown in Fig. 3 a (bottom). This distribution of mutations in both lymphomas appears similar. *SETD2* mutations in EATL were found to cluster within the SET domain, responsible for the methyltransferase activity of SETD2. In both EATL and HSTL, mutations cluster in the SRI domain, responsible for the interaction with RNA polymerase II. Both the SET and SRI domains have highly conserved amino acid sequences, as shown in Fig. 3 a. The missense mutations in *SETD2* may thus to serve to effec-

tively disrupt the essential functions of SETD2 by altering these protein domains.

Given that both HSTLs and EATLs arise disproportionately from γδ T cells and that γδ T cell expansion has been reported in patients with untreated celiac disease (Kagnoff, 1998; Nilsen et al., 1998), we further investigated a potential role for SETD2 in T cell development. We began by examining gene expression patterns throughout the mouse T cell lineage from early T cell precursors to early double negative (DN) to double positive (DP) and finally to the αβ or γδ T cell subsets (Heng et al., 2008; Fig. 3 b). The decision checkpoint for the αβ versus γδ split occurs at the DN3 stage (Carpenter and Bosselut, 2010; Kreslavsky et al., 2010). We found that *Setd2* is selectively up-regulated during the transition from early DN to DP T cells. *Setd2* expression is then down-regulated by DP CD69[pos] T cells on their way to becoming αβ T cells. These results indicate that *Setd2* up-regulation in the immature DP CD69[neg] stages is a critical step in confirming the αβ T cell lineage choice, and we hypothesized that the absence of *Setd2* up-regulation might favor the development of γδ T cells.

We further investigated the in vivo role of *SETD2* in T cell development by breeding *Setd2* conditional knockout mice with *Lck*-Cre transgenic mice, in which expression of Cre is restricted to the T cell lineage (Pan et al., 2002). The breeding scheme led to progeny whose T cells were either *Setd2* wild-type, heterozygous, or deleted (Fig. 3 c shows the final cross and the mice selected for analysis). We then compared T cell populations in young adult mice whose T cells were either wild-type or deleted for *Setd2*.

We first confirmed the loss of SETD2 protein in T cells from *Setd2[fl/fl];Lck*-Cre[+] mice. Protein lysates were made from splenic CD3e[pos] T cells and analyzed by immunoblot. As shown in Fig. 3 d, there is a significant reduction in SETD2 protein in T cells isolated from *Setd2[fl/fl];Lck*-Cre[+] mice compared with *Setd2[+/+];Lck*-Cre[+] mice. We also observed a corresponding decrease in H3K36me3, indicating a reduction of the histone methyltransferase activity in T cells lacking SETD2 protein.

We then assessed the effect of *Setd2* loss upon T cell populations by dissecting IELs and gut-associated lymphoid

in the principal-component space, labeled by the clinical assignment of type I or type II. (d) Volcano plot showing the selection of differentially expressed genes between EATL type I ($n = 16$) and type II ($n = 13$). Every point is one gene in the analysis (number of significant genes is 578). The x axis shows the log2 transformed ratio of mean expression in the type I samples over the mean expression in the type II samples. The y axis shows the log10 transformed q-value (after Benjamini–Hochberg correction) of a Student's *t* test between EATL type I and type II. Genes that pass the thresholds for significance (dotted lines) are colored in orange and purple (FDR <0.1 and fold change >1.5× in either direction). Notable genes are labeled with text. (e) Heat map showing the genes that are differentially expressed between EATL type I and type II (type I: $n = 16$; type II: $n = 13$; number of genes shown is 578). Genes are median centered per row. Red indicates higher expression; blue indicates lower expression. The color bar shows the twofold range of expression depicted. Genes shown have FDR <0.1 and fold change >1.5× in either direction. Bars above the heat map indicate the clinical subtype, HLA type, and history of celiac disease. (f) Differentially expressed genes in the interferon-γ signaling pathway, showing higher expression in the EATL type I samples (type I: $n = 16$; type II: $n = 13$). (g) Gene set enrichment plot showing the enrichment of interferon-γ genes at the top of the ranked list, with genes ordered by difference in type I versus type II (type I: $n = 16$; type 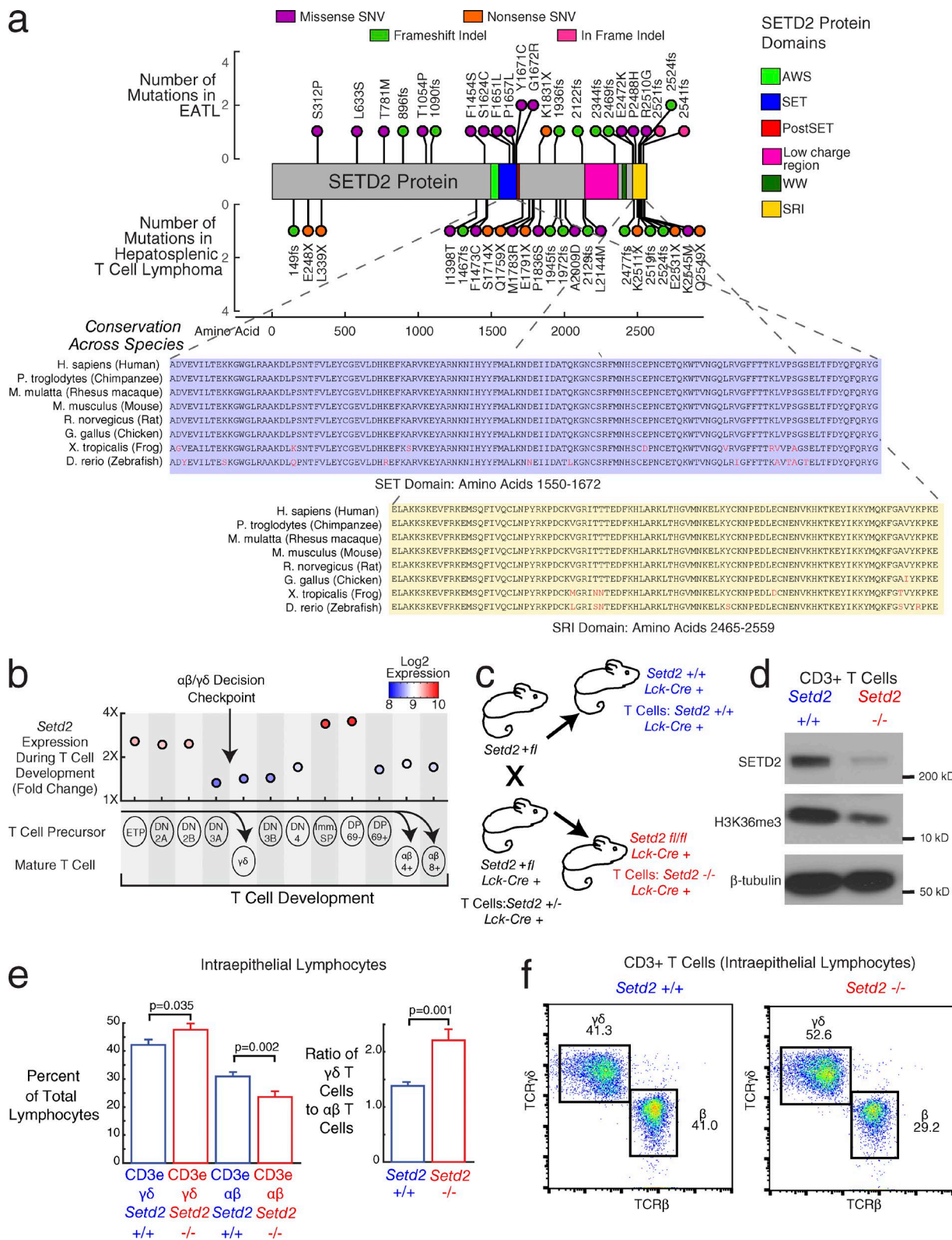II: $n = 13$). Gene set enrichment analysis KS-test FDR <0.001. (h) Differentially expressed genes in the natural killer (NK)–like cytotoxicity pathway, showing higher expression in the EATL type II samples (type I: $n = 16$; type II: $n = 13$). (i) Gene set enrichment plot showing the enrichment of natural killer–like cytotoxicity genes at the top of the ranked list, with genes ordered by difference in type II versus type I (type I: $n = 16$; type II: $n = 13$). Gene set enrichment analysis KS-test FDR = 0.02.

Figure 3. **SETD2 as a driver of increased γδ T cell proliferation.** (a) Stem plot showing distribution of mutations in SETD2. (Top) Mutations in EATL (*n* = 69). (Middle) Mutations in HSTL, a γδ-driven T cell lymphoma (*n* = 68). x axis indicates amino acid position of each mutation. y axis indicates number of patients with the same mutation. Circles are color coded by type of mutation (pink, in-frame indel; purple, missense SNV; green, frameshift indel; orange, nonsense SNV). Colored regions in protein diagram represent protein domains (lime green, AWS [associated with SET] domain; blue, SET domain; red,

tissue (Peyer's patches) from these mice and evaluating T cell markers by flow cytometry. The flow cytometry gating schemes for analysis and sorting are shown in Fig. S2 (a–g) for IELs. In IELs from *Setd2* wild type mice, CD3e^{pos} T cells ranged from 40% to 80% of the live cells within the lymphocyte gate. Typically, TCRγδ^{pos} were ~1.3 times more prevalent than TCRβ^{pos} cells and 90% of both subsets were CD8α^{pos}. In Peyer's patches from *Setd2* wild-type mice, CD3e^{pos} T cells were ~25% of the live cells within the lymphocyte gate with TCRβ^{pos} outnumbering TCRγδ^{pos} cells ~20 to 1. Although <25% of the TCRβ^{pos} cells were CD8α^{pos}, the majority (60%) of TCRγδ^{pos} cells were CD8α^{pos}.

We compared the proportion of IEL T cells that are CD3e^{pos} and TCRγδ^{pos} versus CD3e^{pos} and TCRβ^{pos} in the *Setd2* wild-type and deficient mice (Fig. 3 e) and found a significant increase in the γδ-positive T cell population (ANOVA F-test, P = 0.035), and a commensurate decrease in the αβ T cell population (ANOVA F-test, P = 0.002), resulting in an increase in the ratio of γδ T cells to αβ T cells from 1.4 to 2.2 (ANOVA F-test, P = 0.001). Fig. 3 f shows representative flow cytometry plots from *Setd2$^{+/+}$* and *Setd2$^{-/-}$* T cells, which show the shift toward a higher γδ T cell population in the *Setd2* knockout. An increase in the ratio of TCRγδ^{pos} to TCRβ^{pos} T cells was also evident in Peyer's patch T cells (Fig. S2, h–l).

To further investigate the phenotype of SETD2 loss in IELs, we profiled the gene expression of αβ and γδ sorted IELs, isolated from *Setd2$^{+/+}$* and *Setd2$^{-/-}$* T cells. Fig. S3 a shows the confirmation of the expected deletion of exon 2 of *Setd2* from the RNA-sequencing data. Some of the most highly expressed genes in the IELs included *Gzma, Gzmb, Ccl5, Txndr1, Rgs1,* and *B2m* (Table S7). We identified the top differentially expressed genes between the αβ and γδ T cells, which included known T cell marker genes such as *Cd4, Cd8b1, Icos,* and *Cd28* as more highly expressed in αβ T cells, and *Ctsh, Tgm2,* and *Ctsh* as more

highly expressed in γδ T cells (Fig. S3 b). Genes that differentiated *Setd2$^{-/-}$* and *Setd2$^{+/+}$* IELs included *Lst1, Tcea2, Bmp7, Klk1, Agr2,* and *Msrb1* (Fig. S3 c). Interestingly, we found that overall, the genes that had higher gene expression in γδ T cells (fold change >2 compared with αβ T cells), were highly enriched among the genes that increase in the *Setd2* knockout (Fig. S3, d and e), supporting the notion that *Setd2* loss may favor the development of T cells with a γδ-like phenotype.

No significant difference was found between the two genotypes in overall proportions of CD3e^{pos} or the proportion of the TCRβ^{pos} or TCRγδ^{pos} that are CD8α^{pos}. Table S8 lists the cell counts and proportions from IELs and Peyer's patches for αβ and γδ T cell populations, as well as a summary of the results for CD3e and CD8α. *Setd2* deficient mice did not thus far develop lymphoma at up to 10 mo of age, suggesting that additional collaborating genetic events are needed for oncogenic transformation.

Collectively, these data point to a novel role for *SETD2* in T cell development and suggest that SETD2 loss may favor the development of γδ T cells.

DISCUSSION

EATL is an uncommon but lethal disease. In this study, we have assembled one of the largest cohorts of patients with EATL to define the genetic underpinnings of the disease. Our work defines many previously unknown mutational drivers of the disease, including chromatin modifiers, DNA damage–related genes, and activated signaling pathways. Our results show that *SETD2* is the most frequently silenced gene in this disease, revealing the importance of loss of function of *SETD2*, as well as other epigenetic modifiers in general in the oncogenesis of EATL. Our data also suggest the JAK–STAT signaling pathway as an immediate therapeutic target in the disease, with over half of the EATL patients manifesting mutations in this pathway.

---

post-SET domain; pink, low-charge region; dark green, WW domain; yellow, SRI [Set2 Rpbl-interacting] domain). (Bottom) Conservation of SET (blue) and SRI (yellow) domain amino acid sequences. Nonconserved amino acids are colored in red. (b) *Setd2* expression during T cell development. Labeled circles indicate different stages of T cell development. Corresponding scatterplot shows mouse gene expression (fourfold range) during development, with circles colored by expression. Values represent mean expression from two to three samples per group (34 total samples). αβ T cells differentiate from the DP early T cells, whereas γδ T cell differentiate from the DN3 stage of early T cell. Stages: early T precursor (ETP), DN2A, DN2B, DN3A, DN3B, DN4, immature single positive (Imm. SP), DP CD69$^-$ (DP 69$^-$), DP CD69$^+$ (DP 69$^+$), γδ T cell (γδ), αβ CD4$^+$ T cell (αβ 4$^+$), αβ CD8$^+$ T cell (αβ 8$^+$). (c) Mouse breeding diagram. The final cross is between a heterozygous knockout *Setd2$^{+/fl}$* mouse and a heterozygous knockout *Setd2$^{+/fl}$* mouse with *Cre* transgene under the control of the T cell–specific *Lck* promoter (*Lck*-Cre$^+$). The progeny mice that are used for downstream experiments are the wild-type *Setd2* (+/+, blue) and null *Setd2* (−/− in T cells, red), both with one copy of the *Lck*-Cre transgene. (d) Western blot in CD3$^+$ splenic T cells from the *Setd2* wild-type mouse (left) and *Setd2*-null mouse (right). (Top) SETD2 protein, (middle) H3K36 trimethylation, (bottom) β-tubulin control. Blot is representative of three experimental replicates. (e) Bar graph showing the quantification of CD3e^{pos} and TCRγδ^{pos} or CD3e^{pos} and TCRβ^{pos} T cell populations within IELs. Blue bars show populations from the *Setd2$^{+/+}$* wild-type T cells. Red bars show populations from the *Setd2$^{-/-}$* null T cells. (Left) Left bars show γδ T cell receptor expressing populations, and right bars show αβ T cell receptor expressing populations. (Right) Ratio of γδ T cells to αβ T cells in *Setd2$^{+/+}$* T cells versus *Setd2$^{-/-}$* T cells. Error bars show standard error of the mean. Two-way ANOVA tests were done for comparisons across genotype, correcting also for experimental batch. P-values for genotype differences are 0.035 (γδ T cells), 0.002 (αβ T cells), and 0.001 (ratio of γδ T cells to αβ T cells). Data are shown from five different experimental batches with a mean of five mice per batch. Total number of *Setd2* wild-type mice: 11. Total number of *Setd2*-null mice: 15. (f) Flow cytometry representative plot showing the proportion of cells with TCR γδ and TCR β staining, corresponding to the αβ and γδ TCR-expressing T cell populations. (Left) *Setd2* wild-type mouse (+/+). (Right) *Setd2*-null mouse (−/−). Plots are representative of results found comparing 11 *Setd2* wild-type mice and 15 *Setd2*-null mice.

The classification and naming of EATL and its subtypes have been debated extensively (Chan et al., 2011; Tse et al., 2012; Nairismägi et al., 2016; Roberti et al., 2016). The classification schemes developed thus far largely rely on morphological distinctions and limited immunophenotyping and copy-number data. Next-generation sequencing provides a powerful tool for systematically defining the genetic differences between the subtypes. Interestingly, we observed that the types of EATL demonstrate highly convergent genetics in the face of divergent gene expression. These findings suggest that the two subtypes of EATL have potentially different cellular phenotypes that nevertheless follow similar genetic pathways to the development of EATL.

Knowing when to reclassify a disease is difficult, particularly in a rare entity such as EATL. There are many potential benefits of increasing the resolution of diagnosis, including the ability to define the basis of distinct clinical outcomes, as well as the ability to identify distinct aspects of biology. However, such fragmentation of the disease landscape can also pose challenges in trying to perform clinical trials and trying to improve clinical outcomes, particularly in uncommon diseases such as EATL, where the described subtypes have similar outcomes and distribution of therapeutically targetable mutations. DNA and RNA sequencing provides the basis for a data-driven approach to disease classification. It appears inevitable that such data will inform future classification schemes.

SETD2 is the most frequently silenced gene in EATL. T cell development involves the concerted action of many different genes at well-defined stages of development. Expression analysis in mouse data indicates that *Setd2* is selectively down-regulated in the steps that are critical precursors of γδ T cells and an up-regulation of *Setd2* is a feature of precursors of αβ T cells. SETD2 loss in mice is associated with an expansion of γδ T cells, suggesting a novel role for SETD2 in T cell development, in addition to its other described roles in oncogenesis (Dalgliesh et al., 2010; Zhu et al., 2014). Interestingly, a known hallmark of celiac disease is an increase in the proportion of γδ T cells (Kagnoff, 1998). The lack of lymphoma development in *Setd2* knockout mice suggests that the development of EATL requires the activation of other collaborating genetic events. More work is needed to fully define the role of SETD2 in the oncogenesis of EATL and other peripheral T cell lymphomas, such as HSTL, where it is also highly mutated.

This work defines the complete mutational, copy-number, and transcriptomic landscape of EATL genetics in one of the largest studies in this disease to date. Our in vivo experiments support a role for SETD2 loss in T cell development and lay the groundwork for future investigation into the role of this driver gene in T cell lymphomas.

MATERIALS AND METHODS
Study design and sample selection
EATL tumors and normal tissues were obtained from institutions that constitute the Hematologic Malignancies Research Consortium (Jima et al., 2010). Tumor samples were derived from FFPE tissues, and normal samples were from unaffected bone marrow, when available. Archival patient tumor and normal samples, as well as clinical data, were collected according to a protocol approved by the Duke University Institutional Review Board, which was exempt from informed consent, and in accordance with the Declaration of Helsinki.

Exome sequencing and derivation of genomic data
In brief, genomic DNA was sheared to 250 bp and size/concentration were verified using Bioanalyzer (Agilent Technologies). Sheared DNA was end-repaired, A-tailed, and ligated to Illumina paired-end adapters. The resulting libraries were amplified using Illumina PE specific primers. Samples were column purified and the size and quantity of the final libraries was determined using Bioanalyzer. The resulting libraries were then hybridized overnight to DNA baits provided in the SureSelect Human All Exon 50 MB kit (Agilent Technologies). The captured libraries were amplified and sequenced on Illumina Hiseq (Table S9).

RNA-sequencing library preparation for tumors
RNA was extracted and assessed for quality using the Bioanalyzer RNA Pico chip. Ribosomal RNA (rRNA) was depleted using a custom RNaseH method similar to other published methods (Morlan et al., 2012; Adiconis et al., 2013). 1 µg total RNA was hybridized with 1 µg of a custom DNA oligo mix tiling the 18s, 28s, 5.8s and mitochondrial rRNA sequences (Table S10). Each sample was then RNaseH treated to degrade complementary rRNA sequence and DNase treated to degrade the DNA oligo mix. The remaining rRNA-depleted samples were then purified using the RNeasy Mini kit as directed (QIAGEN). The Kapa Stranded RNA-Seq kit was used to generate stranded Illumina sequencing libraries (Kapa Biosystems). In brief, RNA was hybridized to random primers, followed by first-strand cDNA synthesis, second-strand cDNA synthesis with marking, A-tailing, ligation of Illumina paired-end adapters with 8-bp barcodes, and nine cycles of PCR amplification. Reactions were purified with Agencourt AMPure XP beads where necessary. Paired-end, library-prepared RNA was captured using SureSelect Exome V6+UTR capture baits from Agilent Technologies (Cieslik et al., 2015). The quality and quantity of the final captured libraries were assessed using the Bioanalyzer. Libraries were multiplexed in equimolar amounts and sequenced as paired-end 50-bp (100-bp) reads on an Illumina HiSeq Sequencing System.

Exome sequencing alignment
Reads in FASTQ format (Cock et al., 2010) were preprocessed with GATK (McKenna et al., 2010) version 3.2–2 to remove Illumina adapter sequences (analysis type −T ClipReads, -XF illumina.adapters.fa) and Phred-scaled base qualities of 10 and below (−QT 10) as described previously (Love et al., 2012; Zhang et al., 2013, 2014). After GATK processing, reads were mapped to the hg19 reference genome using Burrows–

Wheeler Aligner version 0.7.7 (BWA mem −M −p; Li and Durbin, 2010). Output was converted to BAM format with SAMtools (Li et al., 2009) version 0.1.19. All alignments were output as BAM files and merged using Picard version 1.96 (http://picard.sourceforge.net). PCR/optical duplicates were marked by Picard. Base quality recalibration was performed using GATK to generate a more accurate base quality score that takes into account its reported quality score in the original FASTQ file, position within the read, and sequence context, for example AC and TG dinucleotides. To improve accuracy and quality of alignment, localized indel realignments were performed using GATK, which infers the consensus indel call from multiple reads mapping to suspected indel genomic regions, rather than considering each read independently. Regions that should be realigned are determined by the GATK Realigner Target Creator.

Variant calling and annotation
SAMtools (Li et al., 2009) mpileup with settings "-C50 -m3 –F0.0002" was run on all cases and output to a single VCF file. The mpileup settings do the following: "-C50" limits the contribution of reads with many mismatches, and "-m3 − F0.0002" is used to maximize sensitivity to indel discovery by requiring three supporting reads at a minimum of 0.02% abundance, rather than the default abundance cutoff of 0.2%. Each variant was required to have an instance of genotype quality >30 and read depth >5. For all samples for which paired normal DNA was available, MuTect version 1.1.7 (Cibulskis et al., 2013) was used to detect truly somatic variants, which are absent in the normal germline DNA (tumor LOD score cutoff set to 4.5 for increased sensitivity). MuTect was also run on all human tumor exome samples to take advantage of the increased sensitivity for low allelic fraction variants. MuTect called variants were merged with mpileup variants for downstream processing. Individual single nucleotide variants (SNVs) and indels from human exomes were annotated with gene names, predicted function, population frequencies, and other variant annotations using Annovar (Wang et al., 2010), which takes as input genomic coordinates and variants, and outputs alteration (amino acid change, frameshift, stop gained, etc.), scores from various predictors, including SIFT (Sim et al., 2012), Polyphen-2 (Adzhubei et al., 2013), and CADD (Kircher et al., 2014), as well as any observed frequencies in the NHLBI 6500 Exomes Sequencing Project and dbSNP.

Variant filtering and the identification of EATL driver genes
The first round of variant filtering required variants to be absent from control populations (NHLBI 6500 Exomes Sequencing Project, in-house normal exomes, and EATL paired normals). We also required variants to have a frequency within the EATL dataset <30% for SNVs and <5% for indels (to eliminate obvious alignment artifacts) and to not have missing data in >50% of the samples. Highly recurrent SNVs required outside support from COSMIC and/or the literature. Furthermore, variants were filtered out that lie in

genomic regions with high duplication scores. Finally, the variant had to have an alternate allele read depth of at least 3, total read depth of at least 5, and an allele frequency of at least 5% in one sample. Next, we formed our discovery gene set using all genes implicated with at least two somatic events in our paired cases, as shown in the separation of discovery and validation sets in Fig. 1. Within these genes, we required these genes to be expressed in T cells (FPKM >1) and to not be overly variable in the normal population (restrictions on number of variants in the 6500 exome project and number of variants over gene size). Most importantly, to identify EATL driver genes, we required genes to have at least 3 not-synonymous events across at least 4 cases out of the highly filtered variants. Additional steps to identify drivers involved requiring a high proportion of identified variants in a gene to have high CADD scores and be completely absent in population databases and for the gene to have a high ratio of not-synonymous to synonymous events.

Mutual exclusion testing
Mutations in final driver gene list (excluding synonymous) were tested for mutual exclusion using the weighted row exclusivity test, using the Saddlepoint approximation for p-value (Leiserson et al., 2016). Sets of three and four genes were tested for mutual exclusion. Within the significant sets of four genes, the presence of all gene pairs in these sets was counted.

Arm–level copy-number calling
Copy numbers from exome sequencing data were generated using the EXCAVATOR (Magi et al., 2013) tool, in both the pooled and somatic mode. All tumor samples were run in pooled mode, against a pool of EATL normal samples, thus controlling for any population variation, and artifacts were generated in the exome sequencing process. Reported alterations were confirmed to occur somatically in the paired samples, via the paired mode of EXCAVATOR. Normals were also run against other normals to confirm no significant alterations in the paired normal samples and control for any remaining artifacts. This program evaluates read depths at each exon, based on the exome capture bait BED file, and normalizes based on the paired or pooled normals provided. Smoothing of the exon-level data provides the copy-number segmentation. Next, the program output was input into GISTIC 2.0 (Mermel et al., 2011) to determine significant arm-level copy-number changes in the EATL dataset. Significant arm level alterations in the EATL tumors from GISTIC were selected for further investigation and filtered against any alterations that were returned as significant in the normal dataset. Furthermore, to call sample-level arm-level amplification and deletions, we returned to the EXCAVATOR data and performed a rank-sum test between the exon level data of each arm of each sample against data from all other arms in that sample. To call a sample altered in that chromosome arm, we required that this test had a significant p-value (<0.01) after correcting for multiple testing with a Bonferroni correction, as well as a significant

increase/decrease in the tumor/normal ratio, equivalent to one copy gained or lost in a 10% pure tumor, as the minimum. Again, arms that are significant in any of the normal samples are removed. GISTIC runs were repeated with only type I or only type II samples separately, and alterations with significant q-values in those runs were also included.

Allele frequency and cancer cell fraction analysis

Purity estimates are taken as the maximum of the copy number based estimate from the driver copy-number alterations reported and the mutation allele frequency based estimate from the driver gene mutations reported. Allele frequencies are corrected for purity and assumed to be from heterozygous mutations to calculate cancer cell fraction. Only mutations with at least five reads are included in the analysis.

HLA genotyping

HLA genotyping was performed using EATL tumor and normal exome sequencing FASTQs and the PHLAT algorithm (Bai et al., 2014). The PHLAT python package was run using Python 2.7 and Bowtie2 (2.1.0). When possible, tumor and normal FASTQs were merged for analysis because the HLA genotype that we are interested in is the germline genotype. Tumor and normals were also assessed separately and were found to have a high degree of concordance. Table S4 shows the genotypes for DQA1 and DQB1, the two genes most relevant for establishing the celiac disease-associated HLA type, as well as the serotype groups for DQA1 and DQB1.

RNA–sequencing alignment and expression quantification

RNA-sequencing reads were first trimmed for removal of adapter sequences and low-quality bases with Trimmomatic (Bolger et al., 2014; version 0.32; adapters: 2:20:7:1:true, leading: 5, trailing: 5, slidingwindow: 4:10, minlen: 21). Next reads were aligned with Bowtie2 (Langmead and Salzberg, 2012; version 2.1.0, with local alignment) against a custom rRNA reference built by searching the NCBI nucleotide database for *Homo sapiens* rRNA, rRNA molecule type (or *Mus musculus*). Remaining reads were aligned with Tophat (Trapnell et al., 2012; version 2.0.9) to the hg19 transcriptome (Ensembl v74), with remaining reads aligned to the hg19 genome (or mm10 equivalents for mouse data). Expression quantification was done using Cuffquant and Cuffnorm (Trapnell et al., 2012; version 2.2.0, cuffquant −u −b). FPKM values from Cuffnorm were used in downstream analysis.

Differential expression and gene set enrichment analysis

Samples were filtered for quality, and genes that had a median expression greater than 0 in the dataset were used for further analysis. An additional step of quantile normalization was applied to the filtered dataset. Differentially expressed genes were identified using Student's *t* test and fold change (FDR <0.1, fold change > 1.5×). Principal-component analysis, volcano plots, and differential expression tests were all performed using Matlab (R2016a). Gene set enrichment was run as a hypergeometric test comparing each gene set to the list of differentially expressed genes between type I and type II EATL. The MSigDB Hallmark, KEGG, and REACTOME gene set lists were used for comparison. Ranked list gene set enrichment using GSEA (Subramanian et al., 2007) was also run using the same gene sets. Enrichment plots across the ranked list are generated using the GSEA software and gseapy (http://pydigger.com/pypi/gseapy). FDR correction is done for both differential expression and gene set enrichment according to the Benjamini–Hochberg method.

Comparison of EATL tumors to T cell subsets

EATL tumor RNA-sequencing samples were compared with the T cell subset gene expression data from the Immunological Genome Project (Heng et al., 2008). All T cells samples available from unmodified C57BL/6 mice were used for comparison. Datasets were filtered to common genes. Pearson correlations were computed for each EATL tumor against each T cell sample. T cell samples are then ranked by mean correlation across the EATL tumors.

Statistics

The log-rank test was used to test significance in survival associations with clinical covariates. Survival curves were drawn using the Kaplan–Meier method. Differentially expressed genes were selected as those with at least 1.5-fold difference between the two groups and an FDR (q-value) <0.1 (Benjamini–Hochberg FDR correction after Student's *t* test.) Principal-component analysis was run using the pca command in Matlab R2016a. Gene set enrichment analysis was run using the hallmark gene sets from MSigDB, Reactome, and KEGG gene sets. A two-way ANOVA model was used to compare T cell population percentages between wild-type and knockout mice while controlling for experimental batch effect.

Mouse strains

Mice were housed in a Duke University Medical Center Division of Laboratory Animal Resources facility and experiments approved by the Duke University and Medical Center Animal Care and Use Committee. The *Setd2* conditional knockout mouse line designated C57BL/6N-Setd2<tm1c(NCOM)Mfgc>/Tcp was made as part of the NorCOMM2 project with C57BL/6NSetd2<tm1a(NCOM)Mfgc>/Tcp made from NorCOMM ES cells at the Toronto Centre for Phenogenomics. The line was made available to us through the Canadian Mouse Mutant Repository. The *Lck-Cre* transgenic strain for T cell–specific CRE expression was provided by Y. Zhuang (Duke University, Durham, NC; Pan et al., 2002). For the evaluation of SETD2 function in T cells, *Setd2*[+/fl] mice were crossed to *Setd2*[+/fl]; *Lck-Cre*[+] mice. The progeny of such crosses included mice whose T cells were *Setd2* wild-type, heterozygous, or null.

Mouse genotyping

Genotyping was performed on tail-snip DNA using the QIA GEN HotStarTaq master mix. The *Setd2* alleles were identified using primers 5′-TAGCTGTGCTTGAGTTGCCAC AGAGC-3′ and 5′-TCCTGAAGCTGGTGCAAACCA CTAGG-3′ and an annealing temperature of 60°C. The wild-type allele produced a 571-bp product, whereas the floxed allele produced a 665-bp product. The *Lck–Cre* transgene was identified using primers 5′-GCAGGAAGTGGGTAACTA GACTAAC-3′ and 5′-TCTCCCACCGTCAGTACGTGA GATATC-3′ and an annealing temperature of 55°C. Presence of the transgene was indicated by a 750-bp product.

Mouse dissection and isolation of lymphocytes

Mouse Peyer's patches and IELs were harvested in RPMI-1640 medium containing 10% (vol/vol) FBS. Single-cell suspensions of Peyer's patch lymphocytes were made by squeezing dissected Peyer's patches between ground glass slides. IELs were isolated from dissected, washed, and inverted small intestinal fragments using a DTT-based extraction medium and sequential filtration steps. The detailed protocol is as follows: We modeled our isolation of IELs on the protocol by Couter and Surana (2016). Just before mouse dissection, IEL extraction media was prepared for each mouse (30 ml RPMI supplemented with 60 µl 0.5 M EDTA and 500 µl FBS). The extraction medium was placed into a 125-ml plastic culture flask and warmed to 37°C in a tissue culture incubator. Freshly prepared DTT (93 µl 5% wt/vol in dH_2O) was added immediately to each flask before addition of intestinal segments. The mouse was euthanized and an incision made to expose the gut contents. The small intestine was separated from the stomach by transection of the pyloric sphincter with scissors. The small intestine was removed from the peritoneum and mesenteric fat teased away. A second cut was made at the ileocecal junction. The isolated small intestine was rinsed and stored when not being manipulated in ice-cold RPMI/10% FBS to maximize cell viability. The small intestine was manipulated on paper towels moistened with RPMI/10% FBS. All the Peyer's patches were excised from the surface of the small intestine, placed in cold RPMI/10% FBS, and a single-cell suspension made by squeezing the tissue between ground glass slides (Micro Slides; Corning) and filtering the suspension through 70 µm cell strainers (Falcon). Red blood cells were lysed and remaining Peyer's patch cells were washed twice with PBS/2% FBS before flow cytometry. Upon removal of Peyer's patches, the central third of the small intestine was selected to be the source of IELs and remaining fat thoroughly and gently removed using curved forceps. The intestinal segment was then subdivided into three or four 1.5–2-in segments. Intestinal contents were removed by gently flushing the segments with 15–20 ml cold PBS using a butterfly needle affixed to a syringe. Each segment was inverted by cannulating the segment with curved forceps, grasping the distal end of the tissue, using a needle to pass the end of gut into the forceps and finally pulling the

forceps back out to invert the segment of gut. The segments were rinsed and stored in cold RPMI/10% FBS before being transferred into the culture flasks with prewarmed extraction medium. The flasks were shaken at 400 rpm for 15 min at 37°C. After incubation, a steel tea strainer was used to separate tissue pieces from the IELs contained in the cloudy filtrate. The filtrate was then put through a 100-µm cell strainer into a 50-ml conical tube, and for better recovery, the filter was rinsed with 20 ml RPMI/10% FBS. The solution was centrifuged at 500 *g* for 10 min at 4°C. The supernate was carefully decanted, and the pellet resuspended in 5 ml RPMI containing 10% FBS. The resuspended cells were filtered through a 40-µm filter, the filter rinsed with 15 ml RPMI/10% FBS, and the filtrate centrifuged again at 500 *g* for 10 min at 4°C. The supernate was carefully decanted and the pellet resuspended in 5 ml RPMI containing 10% FBS. The IELs were counted in Trypan blue with a yield of 15–50 million cells/animal with ∼70% viability.

Western blotting

Mouse splenic $CD3^{pos}$ T cells were isolated by filtering suspended spleen cells through a 70-µM cell strainer, exposed to RBC lysis buffer, and washed twice with PBS with 0.5% BSA. $CD3e^+$ spleen T cells were then positively selected for by cell surface staining with CD3e PE-Cy5.5 antibodies and MACs microbeads passed through LS columns (Miltenyi Biotech) according to the manufacturer's protocols. For each sample, $5 × 10^6$ cells were washed once in ice cold PBS and lysed in 150 microliters RIPA buffer (Sigma-Aldrich) with protease, phosphatase inhibitors (Roche), and 2 mM EDTA. Crude lysates were sonicated at 10% duty cycle, intensity 4, 200 cycles/burst on a Covaris S-series device then centrifuged at 13,000 *g* for 10 min, and supernatants were used for downstream experiments. 10–30 µg of protein were separated on 4–18% Bis-Tris gradient gels and transferred overnight to Immobilion PVDF membranes (EMD Millipore) in Tris-glycine buffer with 20% methanol at 30 V. Primary antibodies (anti-SETD2, anti-H3K36me3, and anti–β tubulin; Cell Signaling Technology) were diluted 1:1,000 in 5% BSA/TBS-T and bound overnight followed by washing and binding of HRP-conjugated anti–rabbit antibody (Santa Cruz Biotechnology, Inc.). Blots were developed using SuperSignal Pico chemiluminescent substrate (Thermo Fisher Scientific).

Flow cytometry

For the analysis of T cell populations, IEL preparations, and Peyer's patch cells were blocked using CD16/32 then incubated with antibodies to surface antigens for 20 min at 4°C. Cells were washed twice, resuspended in PBS/2% FBS, and analyzed by flow cytometry using a Sony LE-SH800 flow cytometer and cell sorter. Identification of TCR $\gamma\delta^{pos}$ and TCR β^{pos} cells was based on the staining and gating strategy illustrated in Fig. S2. Cells are within the lymphocyte gate, singlets, viable, and $CD3e^{pos}TCR\gamma\delta^{pos}$ or $CD3e^{pos}TCR\beta^{pos}$. Antibodies used in this study included BV421-conjugated anti-CD3e

(145-2C11; BioLegend), PE-conjugated anti-CD8a (53-6.7; BD Biosciences), APC-conjugated anti-TCR β (H57-597; BioLegend), and PE/Cy7 anti-TCR γδ (GL3; BioLegend) with Zombie Green (BioLegend) used to identify viable cells.

Flow cytometry analysis
Flow data from Peyer's patch samples were collected from nine experiments, and flow data from IEL preparation samples were collected from five experiments. Samples for each tissue were combined for analysis in FlowJo to standardize the gates. In the case of the Peyer's patches, samples were excluded if less than 70% of the total cells in the forward versus side scatter plot fell within the lymphocyte gate; this was to prevent incorporation of samples with significant IEL contamination caused by the incorporation of excessive gut wall tissue in the dissection. In the case of IEL preparations, samples were excluded if less than 5% of total cells in the forward versus side scatter plot fell within the lymphocyte gate; this effectively excluded samples with poor retrieval of viable lymphocytes.

RNA–sequencing library preparation for IELs
αβ and γδ IELs were isolated from the intestinal epithelium by sorting during flow cytometry as described. The purified IELs were then pelleted and lysed using RNA Lysis Buffer (Zymo). Total RNA was isolated using the Zymo Quick-RNA MicroPrep kit per manufacturer's instructions. RNA sequencing libraries were then created using the KAPA Stranded mRNA-Seq kit (Kapa Biosystems). In brief, mRNA was selected with poly-T beads, mRNA was fragmented and hybridized to random primers, followed by first-strand cDNA synthesis, second-strand cDNA synthesis with marking, A-tailing, ligation of Illumina paired-end adapters with 8-bp barcodes, and 12 cycles of PCR amplification. Reactions were purified with Agencourt AMPure XP beads where necessary. Samples were then sequenced as paired-end 75-bp (150-bp) libraries on the Illumina MiSeq Sequencing System.

Data availability
Primary sequence data are available at the European Genome-phenome Archive under accession no. EGAS00001001954.

Online supplemental material
Fig. S1 shows mutually exclusive gene sets, cancer cell fraction for driver genes, and exploratory Kaplan–Meier plots for clinical and molecular covariates (related to Fig. 1). Fig. S2 shows the flow cytometry scheme for sorting and quantifying T cell populations from the *Setd2* wild-type and knockout mice in IELs and the proportions of T cells in *Setd2* wild-type and knockout mice for Peyer's patches (related to Fig. 3). Fig. S3 shows the RNA-sequencing analysis of IELs (related to Fig. 3). Tables S1–S10 are available as tabs in an Excel file. Table S1 lists the variants in EATL significantly mutated genes. Table S2 lists the mutual exclusion sets. Table S3 lists the copy-number alterations in EATL. Table S4 lists the HLA genotypes for DQA1 and DQB1 genes. Table S5 lists the clinical and pathological features of the EATL cases. Table S6 lists the differentially expressed genes between type I and type II. Table S7 lists the RNA sequencing gene expression for the IELs. Table S8 lists the individual T cell proportions for IELs and Peyer's patches. Table S9 lists the summary statistics for exome and RNA sequencing. Table S10 lists the oligo sequences used for rRNA depletion.

REFERENCES

Adiconis, X., D. Borges-Rivera, R. Satija, D.S. DeLuca, M.A. Busby, A.M. Berlin, A. Sivachenko, D.A. Thompson, A. Wysoker, T. Fennell, et al. 2013. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods.* 10:623–629. http://dx.doi.org/10.1038/nmeth.2483

Adzhubei, I., D.M. Jordan, and S.R. Sunyaev. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7:20.

Arps, D.P., and L.B. Smith. 2013. Classic versus type II enteropathy-associated T-cell lymphoma: Diagnostic considerations. *Arch. Pathol. Lab. Med.* 137:1227–1231. http://dx.doi.org/10.5858/arpa.2013-0242-CR

Bai, Y., M. Ni, B. Cooper, Y. Wei, and W. Fury. 2014. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics.* 15:325. http://dx.doi.org/10.1186/1471-2164-15-325

Belhadj, K., F. Reyes, J.-P. Farcet, H. Tilly, C. Bastard, R. Angonin, E. Deconinck, F. Charlotte, V. Leblond, E. Labouyrie, et al. 2003. Hepatosplenic γδ T-cell lymphoma is a rare clinicopathologic entity with poor outcome: Report on a series of 21 patients. *Blood.* 102:4261–4269. http://dx.doi.org/10.1182/blood-2003-05-1675

Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170

Campo, E., S.H. Swerdlow, N.L. Harris, S. Pileri, H. Stein, and E.S. Jaffe. 2011. The 2008 WHO classification of lymphoid neoplasms and beyond: Evolving concepts and practical applications. *Blood.* 117:5019–5032. http://dx.doi.org/10.1182/blood-2011-01-293050

Carpenter, A.C., and R. Bosselut. 2010. Decision checkpoints in the thymus. *Nat. Immunol.* 11:666–673. http://dx.doi.org/10.1038/ni.1887

Chan, J.K., A.C. Chan, W. Cheuk, S.K. Wan, W.K. Lee, Y.H. Lui, and W.K. Chan. 2011. Type II enteropathy-associated T-cell lymphoma: A distinct aggressive lymphoma with frequent γδ T-cell receptor expression. *Am. J. Surg. Pathol.* 35:1557–1569. http://dx.doi.org/10.1097/PAS.0b013e318222dfcd

Cibulskis, K., M.S. Lawrence, S.L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E.S. Lander, and G. Getz. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31:213–219. http://dx.doi.org/10.1038/nbt.2514

Cieslik, M., R. Chugh, Y.-M. Wu, M. Wu, C. Brennan, R. Lonigro, F. Su, R. Wang, J. Siddiqui, R. Mehra, et al. 2015. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* 25:1372–1381. http://dx.doi.org/10.1101/gr.189621.115

Cock, P.J., C.J. Fields, N. Goto, M.L. Heuer, and P.M. Rice. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38:1767–1771. http://dx.doi.org/10.1093/nar/gkp1137

Corazza, N., S. Müller, T. Brunner, D. Kägi, and C. Mueller. 2000. Differential contribution of Fas- and perforin-mediated mechanisms to the cell-mediated cytotoxic activity of naive and in vivo–primed intestinal intraepithelial lymphocytes. *J. Immunol.* 164:398–403. http://dx.doi.org/10.4049/jimmunol.164.1.398

Couter, C.J., and N.K. Surana. 2016. Isolation and flow cytometric characterization of murine small intestinal lymphocytes. *J. Vis. Exp.* e54114. http://dx.doi.org/10.3791/54114

Dalgliesh, G.L., K. Furge, C. Greenman, L. Chen, G. Bignell, A. Butler, H. Davies, S. Edkins, C. Hardy, C. Latimer, et al. 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature.* 463:360–363. http://dx.doi.org/10.1038/nature08672

Delabie, J., H. Holte, J.M.Vose, F. Ullrich, E.S. Jaffe, K.J. Savage, J.M. Connors, L. Rimsza, N.L. Harris, K. Müller-Hermelink, et al. 2011. Enteropathy-associated T-cell lymphoma: Clinical and histological findings from the international peripheral T-cell lymphoma project. *Blood.* 118:148–155. http://dx.doi.org/10.1182/blood-2011-02-335216

deLeeuw, R.J., A. Zettl, E. Klinker, E. Haralambieva, M. Trottier, R. Chari, Y. Ge, R.D. Gascoyne, A. Chott, H.K. Müller-Hermelink, and W.L. Lam. 2007. Whole-genome analysis and HLA genotyping of enteropathy-type T-cell lymphoma reveals 2 distinct lymphoma subtypes. *Gastroenterology.* 132:1902–1911. http://dx.doi.org/10.1053/j.gastro.2007.03.036

Dieterich, W., T. Ehnis, M. Bauer, P. Donner, U. Volta, E.O. Riecken, and D. Schuppan. 1997. Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat. Med.* 3:797–801. http://dx.doi.org/10.1038/nm0797-797

Ebert, E.C. 1990. Intra-epithelial lymphocytes: Interferon-gamma production and suppressor/cytotoxic activities. *Clin. Exp. Immunol.* 82:81–85. http://dx.doi.org/10.1111/j.1365-2249.1990.tb05407.x

Ferreri, A.J., P.L. Zinzani, S. Govi, and S.A. Pileri. 2011. Enteropathy-associated T-cell lymphoma. *Crit. Rev. Oncol. Hematol.* 79:84–90. http://dx.doi.org/10.1016/j.critrevonc.2010.06.006

Heng, T.S.P., M.W. Painter, K. Elpek, V. Lukacs-Kornek, N. Mauermann, S.J. Turley, D. Koller, F.S. Kim, A.J. Wagers, N. Asinovski, et al. Immunological Genome Project Consortium. 2008. The Immunological Genome Project: Networks of gene expression in immune cells. *Nat. Immunol.* 9:1091–1094. http://dx.doi.org/10.1038/ni1008-1091

Ishikawa, H., Y. Li, A. Abeliovich, S. Yamamoto, S.H. Kaufmann, and S. Tonegawa. 1993. Cytotoxic and interferon gamma–producing activities of gamma delta T cells in the mouse intestinal epithelium are strain dependent. *Proc. Natl. Acad. Sci. USA.* 90:8204–8208. http://dx.doi.org/10.1073/pnas.90.17.8204

Jima, D.D., J. Zhang, C. Jacobs, K.L. Richards, C.H. Dunphy, W.W. Choi, W.Y. Au, G. Srivastava, M.B. Czader, D.A. Rizzieri, et al. Hematologic Malignancies Research Consortium. 2010. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood.* 116:e118–e127. http://dx.doi.org/10.1182/blood-2010-05-285403

Kagnoff, M.F. 1998. Current concepts in mucosal immunity. III. Ontogeny and function of γδ T cells in the intestine. *Am. J. Physiol.* 274:G455–G458.

Kikuma, K., K. Yamada, S. Nakamura, A. Ogami, S. Nimura, M. Hirahashi, H. Yonemasu, S. Urabe, S. Naito, Y. Matsuki, et al. 2014. Detailed clinicopathological characteristics and possible lymphomagenesis of type II intestinal enteropathy-associated T-cell lymphoma in Japan. *Hum. Pathol.* 45:1276–1284. http://dx.doi.org/10.1016/j.humpath.2013.10.038

Kircher, M., D.M.Witten, P. Jain, B.J. O'Roak, G.M. Cooper, and J. Shendure. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46:310–315. http://dx.doi.org/10.1038/ng.2892

Kizer, K.O., H.P. Phatnani, Y. Shibata, H. Hall, A.L. Greenleaf, and B.D. Strahl. 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.* 25:3305–3316. http://dx.doi.org/10.1128/MCB.25.8.3305-3316.2005

Ko, Y.H., S. Karnan, K.M. Kim, C.K. Park, E.S. Kang, Y.H. Kim, W.K. Kang, S.J. Kim, W.S. Kim, W.Y. Lee, et al. 2010. Enteropathy-associated T-cell lymphoma—A clinicopathologic and array comparative genomic hybridization study. *Hum. Pathol.* 41:1231–1237. http://dx.doi.org/10.1016/j.humpath.2009.11.020

Kreslavsky, T., M. Gleimer, A.I. Garbe, and H. von Boehmer. 2010. αβ versus γδ fate choice: counting the T-cell lineages at the branch point. *Immunol. Rev.* 238:169–181. http://dx.doi.org/10.1111/j.1600-065X.2010.00947.x

Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9:357–359. http://dx.doi.org/10.1038/nmeth.1923

Leiserson, M.D., M.A. Reyna, and B.J. Raphael. 2016. A weighted exact test for mutually exclusive mutations in cancer. *Bioinformatics.* 32:i736–i745. http://dx.doi.org/10.1093/bioinformatics/btw462

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 26:589–595. http://dx.doi.org/10.1093/bioinformatics/btp698

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 25:2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352

Love, C., Z. Sun, D. Jima, G. Li, J. Zhang, R. Miles, K.L. Richards, C.H. Dunphy, W.W. Choi, G. Srivastava, et al. 2012. The genetic landscape of mutations in Burkitt lymphoma. *Nat. Genet.* 44:1321–1325. http://dx.doi.org/10.1038/ng.2468

Magi, A., L. Tattini, I. Cifola, R. D'Aurizio, M. Benelli, E. Mangano, C. Battaglia, E. Bonora, A. Kurg, M. Seri, et al. 2013. EXCAVATOR: Detecting copy number variants from whole-exome sequencing data. *Genome Biol.* 14:R120. http://dx.doi.org/10.1186/gb-2013-14-10-r120

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303. http://dx.doi.org/10.1101/gr.107524.110

McKinney, M., A.B. Moffitt, P. Gaulard, M. Travert, L. De Leval, A. Nicolae, M. Raffeld, E.S. Jaffe, S. Pittaluga, L. Xi, et al. 2017. The genetic basis of hepatosplenic T-cell lymphoma. *Cancer Discov.* 7:369–379. http://dx.doi.org/10.1158/2159-8290.CD-16-0330

Meresse, B., S.A. Curran, C. Ciszewski, G. Orbelyan, M. Setty, G. Bhagat, L. Lee, M. Tretiakova, C. Semrad, E. Kistner, et al. 2006. Reprogramming of CTLs into natural killer-like cells in celiac disease. *J. Exp. Med.* 203:1343–1355. http://dx.doi.org/10.1084/jem.20060028

Mermel, C.H., S.E. Schumacher, B. Hill, M.L. Meyerson, R. Beroukhim, and G. Getz. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12:R41. http://dx.doi.org/10.1186/gb-2011-12-4-r41

Morlan, J.D., K. Qu, and D.V. Sinicropi. 2012. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One.* 7:e42882. http://dx.doi.org/10.1371/journal.pone.0042882

Nairismägi, M.L., J. Tan, J.Q. Lim, S. Nagarajan, C.C.Y. Ng, V. Rajasegaran, D. Huang, W.K. Lim, Y. Laurensia, G.C. Wijaya, et al. 2016. JAK-STAT and G-protein-coupled receptor signaling pathways are frequently altered in epitheliotropic intestinal T-cell lymphoma. *Leukemia.* 30:1311–1319. http://dx.doi.org/10.1038/leu.2016.13

Nilsen, E.M., F.L. Jahnsen, K.E. Lundin, F.E. Johansen, O. Fausa, L.M. Sollid, J. Jahnsen, H. Scott, and P. Brandtzaeg. 1998. Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease. *Gastroenterology.* 115:551–563. http://dx.doi.org/10.1016/S0016-5085(98)70134-9

Pan, L., J. Hanrahan, J. Li, L.P. Hale, and Y. Zhuang. 2002. An analysis of T cell intrinsic roles of E2A by conditional gene disruption in the thymus. *J. Immunol.* 168:3923–3932. http://dx.doi.org/10.4049/jimmunol.168.8.3923

Roberti, A., M.P. Dobay, B. Bisig, D. Vallois, C. Boéchat, E. Lanitis, B. Bouchindhomme, M.-C. Parrens, C. Bossard, L. Quintanilla-Martinez, et al. 2016. Type II enteropathy-associated T-cell lymphoma features a unique genomic profile with highly recurrent SETD2 alterations. *Nat. Commun.* 7:12602. http://dx.doi.org/10.1038/ncomms12602

Sharaiha, R.Z., B. Lebwohl, L. Reimers, G. Bhagat, P.H. Green, and A.I. Neugut. 2012. Increasing incidence of enteropathy-associated T-cell lymphoma in the United States, 1973-2008. *Cancer.* 118:3786–3792. http://dx.doi.org/10.1002/cncr.26700

Sim, N.-L., P. Kumar, J. Hu, S. Henikoff, G. Schneider, and P.C. Ng. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40:W452-W457. http://dx.doi.org/10.1093/nar/gks539

Subramanian, A., H. Kuehn, J. Gould, P. Tamayo, and J.P. Mesirov. 2007. GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics.* 23:3251–3253. http://dx.doi.org/10.1093/bioinformatics/btm369

Sun, J., Z. Lu, D. Yang, and J. Chen. 2011. Primary intestinal T-cell and NK-cell lymphomas: A clinicopathological and molecular study from China focused on type II enteropathy-associated T-cell lymphoma and primary intestinal NK-cell lymphoma. *Mod. Pathol.* 24:983–992. http://dx.doi.org/10.1038/modpathol.2011.45

Swerdlow, S., E. Campo, N.L. Harris, N.L. Harris, E.S. Jaffe, S.A. Pileri, H. Stein, J. Thiele, and J.W. Vardiman. 2008. WHO classification of tumours of haematopoietic and lymphoid tissues. IARC Press, France. 439 pp.

Swerdlow, S.H., E. Campo, S.A. Pileri, N.L. Harris, H. Stein, R. Siebert, R. Advani, M. Ghielmini, G.A. Salles, A.D. Zelenetz, and E.S. Jaffe. 2016. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 127:2375–2390. http://dx.doi.org/10.1182/blood-2016-01-643569

Taguchi, T., W.K. Aicher, K. Fujihashi, M. Yamamoto, J.R. McGhee, J.A. Bluestone, and H. Kiyono. 1991. Novel function for intestinal intraepithelial lymphocytes. Murine CD3+, gamma/delta TCR+ T cells produce IFN-gamma and IL-5. *J. Immunol.* 147:3736–3744.

Tomita, S., Y.Y. Kikuti, J. Carreras, M. Kojima, K. Ando, H. Takasaki, R. Sakai, K. Takata, T. Yoshino, S. Bea, et al. 2015. Genomic and immunohistochemical profiles of enteropathy-associated T-cell lymphoma in Japan. *Mod. Pathol.* 28:1286–1296. http://dx.doi.org/10.1038/modpathol.2015.85

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, and L. Pachter. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7:562–578. http://dx.doi.org/10.1038/nprot.2012.016

Tse, E., H. Gill, F. Loong, S.J. Kim, S.B. Ng, T. Tang, Y.H. Ko, W.J. Chng, S.T. Lim, W.S. Kim, and Y.L. Kwong. 2012. Type II enteropathy-associated T-cell lymphoma: A multicenter analysis from the Asia Lymphoma Study Group. *Am. J. Hematol.* 87:663–668. http://dx.doi.org/10.1002/ajh.23213

Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. http://dx.doi.org/10.1093/nar/gkq603

Zhang, J., V. Grubor, C.L. Love, A. Banerjee, K.L. Richards, P.A. Mieczkowski, C. Dunphy, W. Choi, W.Y. Au, G. Srivastava, et al. 2013. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. USA.* 110:1398–1403. http://dx.doi.org/10.1073/pnas.1205299110

Zhang, J., D. Jima, A.B. Moffitt, Q. Liu, M. Czader, E.D. Hsi, Y. Fedoriw, C.H. Dunphy, K.L. Richards, J.I. Gill, et al. 2014. The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells. *Blood.* 123:2988–2996. http://dx.doi.org/10.1182/blood-2013-07-517177

Zhu, X., F. He, H. Zeng, S. Ling, A. Chen, Y. Wang, X. Yan, W. Wei, Y. Pang, H. Cheng, et al. 2014. Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat. Genet.* 46:287–293. http://dx.doi.org/10.1038/ng.2894