UNIVERSITY OF CALIFORNIA SAN DIEGO

Role of TET proteins in B cells and Treg cells

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Daniela Samaniego Castruita

Committee in charge:

Professor Anjana Rao, Chair
Professor Cornelis Murre, Co-Chair
Professor Steven Briggs
Professor Christopher Glass
Professor Lorraine Pillus

2022

The Dissertation of Daniela Samaniego Castruita is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To the memory of Juan A. Arias del Angel, my friend, neighbor and former labmate; your happy spirit will always live in my heart and memory.

To my parents Ana Maria Castruita Marquez and Jose Alfredo Samaniego Gaxiola: I can't stress enough how much I love you and how grateful I am for all your support. I wouldn't be here if it weren't because of you. Thank you for never letting me down.

To my siblings Jose A. Samaniego Castruita and Patricia Samaniego Castruita: thank you for always cheering me up and checking on me, regadless of the distance.

To my daughters and love of my life: Ana Victoria Gonzalez Samaniego and Amanda Gonzalez Samaniego, you are the motor of my life and the most precious event that could ever happen to me.

Last but not least, to my husband and long-term collaborator, Edahi Gonzalez Avalos, thank you for always believing in me and for being by my side in all these years.

EPIGRAPH

“Life is what happens while you are busy making other plans” –John Lennon-

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SUPPLEMENTAL FILES

Supplemental File Chapter 2. Table S2.1, Table S2.2 and Table S2.3.

Supplemental File Chapter 3.1. Table S3.1, Table S3.2 and Table S3.3.

Supplemental File Chapter 3.2. Table S3.4.

Supplemental File Chapter 3.3. Table S3.5, Table S3.6 and Table S3.7.

Supplemental File Chapter 3.4. Table S3.8, Table S3.9 and Table S3.10.

Supplemental File Chapter 3.5. Table S3.11 and Table S3.12.

ACKNOWLEDGEMENTS

I am deeply grateful with all the people that have supported me during my academic and personal life. Without them, this dissertation wouldn't be possible:

First, I would like to thank my advisor Dr. Anjana Rao for giving me the opportunity of joining her lab and for her invaluable support during these years. I am very grateful to have learned from her and her lab.

Second, I would like to thank my committee meeting advisors, Dr. Cornelis Murre, Dr. Lorraine Pillus, Dr. Steven Briggs and Dr. Christopher Glass for all their guidance during these years.

Next, I would like to thank two of my close friends, collegues and mentors, Dr. Xiaojing Yue and Dr. Vipul Shukla, for all their unconditional support and guidance. I am deeply grateful to have coincide with such amazing souls.

I would also like to thank all my collaborators and friends Dr. Edahi Gonzalez Avalos, Dr. Benjamin Barwick, Dr. Zhen Dong, Dr. Hyunseok Seo, Dr. Jerry Lio for believing in me and inviting me to work in amazing projects.

I am thankful with all the amazing people I have met in Dr. Rao's lab that have helped me during these years. Thank you, Dr. Atsushi Onodera, Dr. Hugo Sepulveda, Roberta Nowak, Nupura Hirve, Carlos Angel, Wade Zhang, Payal Ramchandani. Also, thanks to Dr. Patrick Hogan for all his feedback during the lab meetings.

I would like to thank the people from the Biological Sciences graduate program at UCSD that have always supported me and helped me with all the burocratics. Thank you, Melody Bazyar, Suzi Harlow, Lien Ngo, Donna Aizuss, Marifel Alfaro, Irish Fontan-Basinga. I would like to thank

CONACyT and UC-MEXUS for the finantial support that I received from them in the first 5 years of my doctoral program.

Thanks a lot, to the amazing people I have met during these years who have made this experience even more enjoyable. Special thanks to the friends I have made in graduate housing while raising my kids. Thank you to my dear Mexican friends, Esmeralda Garcia, Sayra Serrano, Diana Vera, Aurora Labastida and Paulina Diaz.

I am very thankful for the unconditional support I got from my two ESAs, Babychami and Cotton Truffle, thank you for making my life much happier and for teaching me how to live a much easier life. I don't exaggerate when I say you have been life changing to me.

I would like to thank my family who has always supported me and without them I wouldn't be here: my mom and my dad, Ana Maria Castruita Marquez and Jose Alfredo Samaniego Gaxiola, for ther invaluable support through all my life, thank you for taking care of my daughters and myself while I was pursuing this degree. Thank you to my siblings, Pepe and Paty for cheering me up and for always trusting in me. Thank you for never letting me down.

Special thanks to my husband, Edahi Gonzalez Avalos, for all his support during these years. Thank you for being by my side and for helping me when I got stuck.

Thanks a lot, to my two daughters and love of my life: Ana Victoria and Amanda. I am honored to be your mother and I would do everything for you. Thank you for helping me doing experiments while being in the belly.

Chapter 2, in full, is a reprint with modifications as it appears in "TET deficiency perturbs mature B cell homeostasis and promotes oncogenesis associated with accumulation of G-quadruplex and R-loop structures", *Nature Immunology* (2022). DOI: https://doi.org/10.1038/s41590-021-01087-w. The dissertation author was the co-primary

investigator and co-first author of this paper. Other authors include Vipul Shukla, Zhen Dong, Edahí González-Avalos, Qingqing Yan, Kavitha Sarma, and Anjana Rao.

Chapter 3, in full, is a reprint with modifications as it appears in "Whole-genome analysis of TET dioxygenase function in regulatory T cells", *EMBO reports* (2021). DOI: https://doi.org/10.15252/embr.202152716. The dissertation author was the co-primary investigator and co-first author of this paper. Other authors include Xiaojing Yue, Edahí González-Avalos, Xiang Li, Benjamin Barwick, and Anjana Rao.

VITA

2015 National Autonomous University of México
Bachelor of Genomic Sciences.

2014-2015 University of Colorado
Undergraduate Research Technician.

2015-2016 La Jolla Institute for Immunology
Research Technician.

2016-2022 University of California San Diego
Doctor of Philosophy, Biology.

PUBLICATIONS

Shukla, V. *, **Samaniego-Castruita D.***, Dong Z., Gonzalez-Avalos E., Yan Q, Sarma K, Rao A. TET deficiency perturbs mature B cell homeostasis and promotes oncogenesis associated with accumulation of G-quadruplex and R-loop structures. Nature immunology 2022.

Yue, X. *, **Samaniego-Castruita D.** *, González-Avalos, E., Li, X., Barwick, B. G., & Rao, A. Whole-genome analysis of TET dioxygenase function in regulatory T cells. EMBO reports, e52716.

Shukla, V., Halabelian, L., Balagere, S., **Samaniego-Castruita D.**, Feldman, D. E., Arrowsmith, C. H., Rao A., Aravind, L. HMCES Functions in the Alternative End-Joining Pathway of the DNA DSB Repair during Class Switch Recombination in B Cells. Molecular Cell. 2019.

Seo H., Chen J., Gonzalez-Avalos E., **Samaniego-Castruita D.**, Das A., Wang Y., Lopez-Moyado I., Georges R., Zhang W., Onodera A., Wu C., Lu L., Hogan P., Bhandoola A., Rao A. TOX and TOX2 cooperate with NR4A transcription factors to impose CD8+ T cell exhaustion. Proceedings of the National Academy of Sciences, 116(25), 12410-12415.

Yue X., Lio C-W. J., **Samaniego-Castruita D.**, Li X. Rao A. Loss of TET proteins in regulatory T cells unleashes effector function. Nature communications. 2019 May 10:2011

Lio C.-W. J.*, Shukla V.*, **Samaniego-Castruita D.**, Gonzalez-Avalos E., Chakraborthy A., Yue X., Schatz D., Ay F., Rao A. TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer Science immunology. 2019 April

Ganesan A.P., Clarke J., Wood O., Garrido-Martin E.M., Chee S.J., Mellows T., **Samaniego-Castruita D.**, Singh D., Seumois G., Alzetani A., Woo E.. Tissue-resident memory features are

linked to the magnitude of cytotoxic T cell responses in human lung cancer. Nature immunology. 2017 Aug;18(8):940.

Schmiedel B.J., Seumois G., **Samaniego-Castruita D.**, Cayford J., Schulten V., Chavez L., Ay F., Sette A., Peters B., Vijayanand P. 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. Nature communications. 2016 Nov 16;7:13426.

Engel I., Seumois G., Chavez L., **Samaniego-Castruita D.**, White B., Chawla A., Mock D., Vijayanand P., Kronenberg M. Innate-like functions of natural killer T cell subsets result from highly divergent gene programs. Nature immunology. 2016 Jun;17(6):728.

*These authors contributed equally to this work.

ABSTRACT OF THE DISSERTATION

Role of TET proteins in B cells and Treg cells

by

Daniela Samaniego Castruita

Doctor of Philosophy in Biology

University of California San Diego, 2022

Professor Anjana Rao, Chair
Professor Cornelis Murre, Co-Chair

DNA demethylation is promoted by TET enzymes, a family of three proteins -TET1, TET2 and TET3- which oxidize methyl group of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). TET proteins have been previously associated to cancer. In particular, Tet2 loss of function is strongly associated to haematopoietic malignancies in humans, such as B cell and T

cell lymphoma. In this dissertation we explore the role of TET proteins in B and Treg cells. To do this we rely in the use of *in vivo* and *in vitro* models.

We generated a mouse model with *Tet2/3* deletion, *CD19 Tet2/3 DKO* mice. The TET deficient mice showed an expansion of Germinal Center B cells and developed B cell lymphoma, as early as 6 weeks of age; coupled to this, we observed an increase of G-quaruplexes and R-loops and increase of DNA damage. Mice generated with the adiddtional deletion of DNMT1, the enzyme responsible to maintain DNA methylation and also a G-quadruplex binding protein, on top of the *Tet2/3* deletion, showed an inhibition of Germinal Center B cells and a significant delay on B cell lymphoma.

We also explored the role of TET proteins in Treg cells using genomic and epigenomic approaches in *in vivo* and *in vitro* models. The Tet2/3 deficient Treg cells showed a loss of Treg-specific molecular features, and an impairement of IL-2/STAT5 signaling pathway. In our *in vitro* system, we induced Treg (iTreg) cells with TGFβ and treated them with Vitamin C, a potent activator of TET proteins and other dioxygenases. Oposite to what we observed *in vivo*, boosting TET activity with Vitamin C enhanced IL-2 responsiveness in iTreg cells by increasing IL2Rα expression and STAT5 activity.

CHAPTER 1: Introduction

1.1 TET enzymes

5-methylcytosine (5mC) is a covalent DNA modification catalyzed de novo by DNA Methyltransferase 3A (DNMT3A) and 3B (DNMT3B) proteins, and maintained during DNA replication by the DNMT1/UHRF1 protein complex (Moore *et al.* 2012). Enzymes of the TET (Ten-Eleven-Translocation; TET1, TET2, TET3) family catalyze the sequential oxidation of the methyl group of 5mC in DNA to 5-hydroxymethylcytosine (5hmC) and the further oxidized products 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (He *et al.* 2011; Ito *et al.* 2011; Iyer *et al.* 2009; Tahiliani *et al.* 2009).

TET proteins contain a conserved double-stranded β-helix (DSBH) domain, a cysteine-rich domain, and binding sites for the cofactors Fe (II) and 2-oxoglutarate (2-OG) that together form the core of their catalytic region. Like all Fe (II) and 2-oxoglutarate-dependent dioxygenases, TET enzymes utilize 2OG, Fe (II) and molecular oxygen to generate their oxidized products, with CO2 and succinate formed as byproducts (Hausinger, 2004). TET1 and TET3 proteins have an N-terminal CXXC zinc finger domain that can bind DNA (Rasmussen & Helin, 2016). While TET2 protein lacks this CXXC domain, the IDAX gene, linearly adjacent to the Tet2 gene, encodes a CXXC4 protein and is thought to play a role in regulating TET2 activity by facilitating its recruitment to unmethylated CpGs (Rasmussen & Helin, 2016).

The 5hmC, 5fC and 5caC modified bases are intermediates in DNA demethylation as well as potential epigenetic marks capable of recruiting reader proteins that exert specific functions (Pastor & Rao, 2013; Tsagaratou *et al.* 2017b; Wu & Zhang, 2017). 5hmC, referred as the "sixth

DNA base", is the most abundant oxidized form of 5mC and has been shown to be correlated with active gene expression, its presence at enhancers correlates with chromatin accessibility (Tsagaratou *et al.* 2017a; Lio *et al.* 2016) and has been consistently associated with active genomic regions or "epigenetically dynamic loci" (Szulwach *et al.* 2011).

1.2 TET proteins in hematological cancers

TET proteins regulate several fundamental biological processes including lineage commitment, and play important roles in embryonic, neuronal and hematopoietic development (Lio *et al.* 2016). From the three mammalian TET proteins (TET1, TET2 and TET3), TET2 and TET3 are most highly expressed in cells of the immune and hematopoietic systems (Tsagaratou *et al.* 2013). It has been observed that TET genes mutations and decreased TET activity, through impaired regulation of metabolic enzymes that affect TET activity, leads to hematological malignances and solid cancers (Lio *et al.* 2020; Cimmino *et al.* 2011; Huang *et al.* 2014; Ko *et al.* 2015; Lio *et al.* 2019).

Of the three TET proteins, TET2 is the most commonly mutated in hematopoietic cancers (Ko *et al.* 2015). TET2 mutations have been identified in a several myeloid malignancies, such as CMML, AML, MDS, MPN (Ko *et al.* 2015). TET2 mutations are also commonly found in lymphoid cancers, such as T cell and B cell lymphomas (Ko *et al.* 2015). For example, in humans, TET2 is recurrently mutated in ~10% of Diffuse Large B-cell Lymphoma (DLBCL), a heterogeneous malignancy originating in mature B cells undergoing activation and differentiation in germinal centers (GCs) (Schmitz *et al.* 2018; Reddy *et al.* 2017; Chapuy *et al.* 2018). Moreover, TET2 mutations represent an early driver event in DLBCL (Dominguez *et al.* 2018).

To better understand the function of TET proteins in hematopoietic development and cellular transformation, we and others have generated TET deficient mouse models (Ko *et al.*

2015). For example, Tet2 deletion in mouse models in hematopoietic lineages disrupted GC B cell homeostasis and promoted development of more aggressive lymphomas when the transcription factor BCL6 was constitutively overexpressed (Dominguez *et al.* 2018). Also, effects of Tet2 and Tet3 deletion (using Mb1-Cre) at early stages of mouse B cell development are impaired light chain rearrangement and developmental blockade, that led to development of an acute precursor-B-cell-derived leukemia with 100% penetrance (Orlanski *et al.* 2016). Compared to Tet2 and Tet3, inducible deletion (Mx1-Cre) of both Tet1 and Tet2 promoted the development of acute lymphoblastic leukemia derived from precursor B cells. Single deletion of Tet1 caused only B cell lymphomas with an extended latency (Quivoron *et al.* 2011). Due to the TET proteins having shared functions among its family members, studies of TET-mediated gene regulation are best performed in systems where more than one TET gene is deleted acutely, rather than during development.

1.3 TET proteins and Vitamin C

The enzymatic activity of TET proteins can be modulated by diverse factors, including the ambient levels of oxygen (hypoxia versus normoxia) and diverse metabolic enzymes that regulate the intracellular levels of either 2OG itself or its competitive inhibitor 2-hydroxyglutarate (2HG) (Dang & Su, 2017; Lio & Rao, 2019; Lio *et al.* 2020). Vitamin C is a potent activator of TET proteins. Vitamin C functions as a cofactor for TET enzymes by facilitating the reduction of Fe(III) back to Fe(II) at the dioxygenase active site in each catalytic cycle (Blaschke *et al.* 2013; Yue *et al.* 2016).

When vitamin C is added to cultures of mouse ES cells, it promotes TET activity, quickly leading to a global increase in 5hmC, followed by DNA demethylation of many gene promoters and upregulation of demethylated germline genes, which was not observed in ES cells lacking

Tet1 and Tet2. Similar observations were reported for mouse fibroblast, where addition of vitamin C increased 5hmC generation in a dose- and time-dependent manner (Minor *et al.* 2013). Further evidence of vitamin C interaction with TET proteins was shared by Yin and colleagues (Yin *et al.* 2013) where they reported a direct interaction between vitamin C and the catalytic domain of Tet2, as shown by the ability of vitamin C to quench the intrinsic fluorescence of recombinant Tet2 catalytic domain.

TET proteins are required to maintain the stability and function of regulatory T (Treg) cells. (Josefowicz *et al.* 2012; Li & Zheng, 2015; Sakaguchi *et al.* 2008; Yang *et al.* 2015; Yue *et al.* 2016). Treg cells are defined by expression of the X-chromosome-encoded transcription factor FOXP3, a lineage-determining transcription factor essential for Treg development and function (Josefowicz *et al.* 2012; Li & Zheng, 2015; Sakaguchi *et al.* 2008). TET proteins also control the stability of FOXP3 expression in "induced" regulatory T cells (iTreg cells), generated in vitro by activation of naïve CD4+ T cells through the T cell receptor (TCR) and the costimulatory receptor CD28 in the presence of TGFβ and/or retinoic acid (RA) (Yue *et al.* 2016).

The stability of FOXP3 expression is regulated by the DNA methylation status of an intronic enhancer, CNS2 (conserved non-coding sequence 2) located in the first intron of the Foxp3 gene (Feng *et al.* 2014b; Li *et al.* 2014; Zheng *et al.* 2010). The 12 CpG sites in the mouse Foxp3 CNS2 enhancer are predominantly "methylated" in iTreg cells generated with TGFβ or TGFβ and RA but predominantly "unmethylated" in endogenous Treg cells (Floess *et al.* 2007; Huehn *et al.* 2009; Toker & Huehn, 2011). Notably, deletion of the Tet2 and Tet3 genes with either CD4Cre or Foxp3Cre during Treg development, results in increased methylation of Foxp3 CNS2 in Tet2/3 DKO Treg cells compared to WT Treg cells and a marked defect in Foxp3 stability (Yue *et al.* 2019; Yue *et al.* 2016). In a parallel gain-of-function approach, we have shown that addition of

the TET activator Vitamin C during in vitro iTreg differentiation leads to TET-dependent demethylation of Foxp3 CNS2 and a substantial increase in Foxp3 stability after adoptive transfer in vivo or cell division in vitro (Sasidharan Nair *et al.* 2016; Yue *et al.* 2016) (**Fig. 1.1**).

On top of the benefits vitamin C has in promoting TET's enzymatic activity, a body of work studies the importance of TET enzymes and vitamin C in regulating functions of the immune system. For example, TET2 mutations in humans cooperate with FLT3ITD mutations to induce acute myeloid leukemia (AML); however, vitamin C depletion accelerated the development of AML from Flt3ITDTet2+/− and Flt3ITDTet2−/− cells, where this phenomenon was reversed by dietary repletion of vitamin C (Agathocleous *et al.* 2017). Another example illustrated that loss of Tet2 in hematopoietic stem/precursor cells (HSPCs) resulted in an aberrant increase in self-renewal, increased DNA methylation, aberrant myeloid lineage differentiation, and increased cell death (Cimmino *et al.* 2017). Restoration of Tet2 reversed these phenomena, and treatment with vitamin C pharmacologically mimicked the effects of Tet2 restoration in HSPCs and blocked the progression of myeloid disease (Cimmino *et al.* 2017).

1.4 Figures

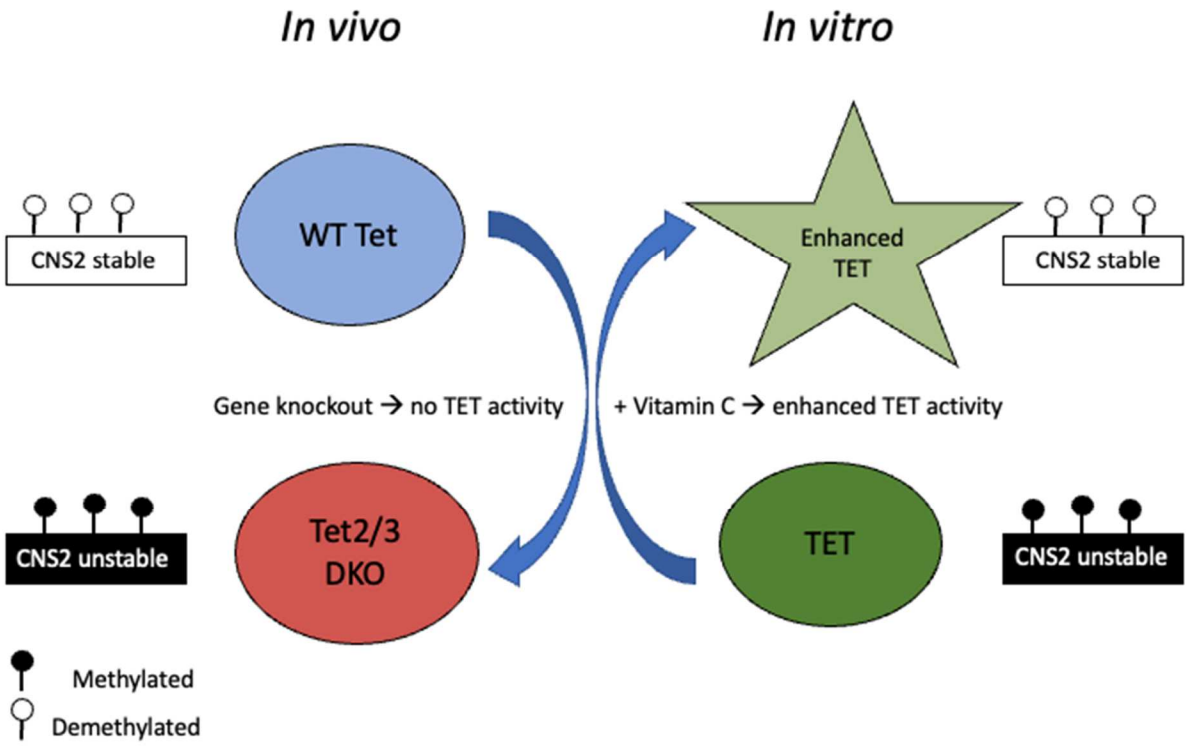


Figure 1.1 Schematic representation of *CNS2* in *Foxp3* locus.
TET proteins regulate the stability of *Foxp3* in Treg cells through regulating the methylation levels in *CNS2* locus. *In vivo*, Tet deletion causes DNA methylation in the *CNS2* and impairs Tregs stability. Addition of vitamin C into iTregs, an *in vitro* model, increases TET activity which results in decreased DNA methylation in the *CNS2* locus, ultimately causing enhancement of Treg stability.

1.5 References

Agathocleous, M., Meacham, C.E., Burgess, R.J., Piskounova, E., Zhao, Z., Crane, G.M., Cowin, B.L., Bruner, E., Murphy, M.M., Chen, W., Spangrude, G.J., Hu, Z., DeBerardinis, R.J. and Morrison, S.J. (2017). Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. Nature, 549(7673), pp.476–481.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., Lorincz, M.C. and Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature 500: 222-226.

Chapuy, B., Stewart, C., Dunford, A.J., Kim, J., Kamburov, A., Redd, R.A., Lawrence, M.S., Roemer, M.G., Li, A.J., Ziepert, M. and Staiger, A.M., 2018. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature medicine, 24(5), pp.679-690.

Chédin, F., 2016. Nascent connections: R-loops and chromatin patterning. Trends in Genetics, 32(12), pp.828-838.

Cimmino, L., Dolgalev, I., Wang, Y., Yoshimi, A., Martin, G.H., Wang, J., Ng, V., Xia, B., Witkowski, M.T., Mitchell-Flack, M., Grillo, I., Bakogianni, S., Ndiaye-Lobry, D., Martín, M.T., Guillamot, M., Banh, R.S., Xu, M., Figueroa, M.E., Dickins, R.A. and Abdel-Wahab, O. (2017). Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. Cell, 170(6), pp.1079-1095.e20.

Crossley, M.P., Bocek, M. and Cimprich, K.A., 2019. R-loops as cellular regulators and genomic threats. Molecular cell, 73(3), pp.398-411.

Dang L, Su SM (2017) Isocitrate Dehydrogenase Mutation and (R)-2-Hydroxyglutarate: From Basic Discovery to Therapeutics Development. Annu Rev Biochem 86: 305-331

De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O. and Capranico, G., 2019. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. Proceedings of the National Academy of Sciences, 116(3), pp.816-825.

Dominguez, P.M., Ghamlouch, H., Rosikiewicz, W., Kumar, P., Béguelin, W., Fontan, L., Rivas, M.A., Pawlikowska, P., Armand, M., Mouly, E. and Torres-Martin, M., 2018. TET2 deficiency causes germinal center hyperplasia, impairs plasma cell differentiation, and promotes B-cell lymphomagenesis. Cancer discovery, 8(12), pp.1632-1653.

Gray, L.T., Vallur, A.C., Eddy, J. and Maizels, N., 2014. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. Nature chemical biology, 10(4), pp.313-318.

Hänsel-Hertsch, R., Di Antonio, M. and Balasubramanian, S., 2017. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nature reviews Molecular cell biology, 18(5), pp.279-284.

Hausinger R.P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. Crit Rev Biochem Mol Biol 39: 21-68

He Y.F., Li B.Z., Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science 333: 1303-1307

Ito S, Shen L, Dai Q, Wu S.C., Collins L.B., Swenberg J.A., He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science 333: 1300-1303

Iyer L.M., Tahiliani M, Rao A, Aravind L (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle 8: 1698-1710

Kafer, G.R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I. and Carlton, P.M., 2016. 5-Hydroxymethylcytosine marks sites of DNA damage and promotes genome stability. Cell reports, 14(6), pp.1283-1292.

Kharat, S.S., Ding, X., Swaminathan, D., Suresh, A., Singh, M., Sengodan, S.K., Burkett, S., Marks, H., Pamala, C., He, Y. and Fox, S.D., 2020. Degradation of 5hmC-marked stalled replication forks by APE1 causes genomic instability. Science signaling, 13(645).

Ko, M., An, J., Pastor, W.A., Koralov, S.B., Rajewsky, K. and Rao, A., 2015. TET proteins and 5-methylcytosine oxidation in hematological cancers. Immunological reviews, 263(1), pp.6-21.

Lio C.W.J., Zhang J, González-Avalos E, Hogan P.G., Chang X, Rao A (2016) Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. Elife 5

Lio C.W.J., Rao A (2019) TET Enzymes and 5hmC in Adaptive and Innate Immune Systems. Front Immunol 10: 210

Lio, C.W.J., Yue, X., López-Moyado, I.F., Tahiliani, M., Aravind, L. and Rao, A., 2020. TET methylcytosine oxidases: new insights from a decade of research. Journal of biosciences, 45(1), pp.1-14.

Miglietta, G., Russo, M. and Capranico, G., 2020. G-quadruplex–R-loop interactions and the mechanism of anticancer G-quadruplex binders. Nucleic acids research, 48(21), pp.11942-11957.

Minor, E.A., Court, B.L., Young, J.I. and Wang, G. (2013). Ascorbate Induces Ten-Eleven Translocation (Tet) Methylcytosine Dioxygenase-mediated Generation of 5-Hydroxymethylcytosine. The Journal of Biological Chemistry, 288(19), pp.13669–13674.

Moore, L.D., Le, T. and Fan, G. (2012). DNA Methylation and Its Basic Function. Neuropsychopharmacology 38, 23–38.

Niehrs, C. and Luke, B., 2020. Regulatory R-loops as facilitators of gene expression and genome stability. Nature Reviews Molecular Cell Biology, 21(3), pp.167-178.

Pastor W.A., Rao A (2013) TETonic shift: Biological roles of TET proteins in DNA demethylation and transcription. Nature Reviews Molecular Cell Biology.

Orlanski, S., Labi, V., Reizel, Y., Spiro, A., Lichtenstein, M., Levin-Klein, R., Koralov, S.B., Skversky, Y., Rajewsky, K., Cedar, H. and Bergman, Y., (2016). Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. Proceedings of the National Academy of Sciences 113, 5018-5023.

Quivoron, C., Couronné, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H. and Godley, L., (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer cell 20, 25-38.

Rasmussen, K.D. and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. Genes & Development 30(7), pp.733–750.

Reddy, A., Zhang, J., Davis, N.S., Moffitt, A.B., Love, C.L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L., Karjalainen-Lindsberg, M.L. and Nørgaard, P., 2017. Genetic and functional drivers of diffuse large B cell lymphoma. Cell, 171(2), pp.481-494.

Rhodes, D. and Lipps, H.J., 2015. G-quadruplexes and their regulatory roles in biology. Nucleic acids research, 43(18), pp.8627-8637.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X. and Chédin, F., 2016. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Molecular cell, 63(1), pp.167-178.

Sauer, M. and Paeschke, K., 2017. G-quadruplex unwinding helicases and their function in vivo. Biochemical Society Transactions, 45(5), pp.1173-1182.

Schmitz, R., Wright, G.W., Huang, D.W., Johnson, C.A., Phelan, J.D., Wang, J.Q., Roulland, S., Kasbekar, M., Young, R.M., Shaffer, A.L. and Hodson, D.J., (2018). Genetics and pathogenesis of diffuse large B-cell lymphoma. New England Journal of Medicine 378, 1396-1407.

Scott-Browne, J.P., Lio, C.W.J. and Rao, A., (2017). TET proteins in natural and induced differentiation. Current opinion in genetics & development 46, 202-208.

Skourti-Stathaki, K. and Proudfoot, N.J., 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & development, 28(13), pp.1384-1396.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., Yoon, Y.-S., Ren, B., He, C. and Jin, P. (2011). Integrating 5-Hydroxymethylcytosine into the Epigenomic Landscape of Human Embryonic Stem Cells. PLoS Genetics 7, e1002154.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324: 930-935

Tsagaratou A, González-Avalos E, Rautio S, Scott-Browne J.P., Togher S, Pastor W.A., Rothenberg E.V., Chavez L, Lähdesmäki H, Rao A (2017a) TET proteins regulate the lineage specification and TCR-mediated expansion of iNKT cells. Nat Immunol 18: 45-53

Tsagaratou A, Lio C.W.J., Yue X, Rao A (2017b) TET Methylcytosine Oxidases in T Cell and B Cell Development and Function. Front Immunol 8: 220

Tsagaratou A, Rao A (2013) TET proteins and 5-methylcytosine oxidation in the immune system. Cold Spring Harb Symp Quant Biol 78: 1-10

Wu X, Zhang Y (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet 18: 517-534

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., Xiao, M.-T., Liu, L., Jiang, W., Liu, J., Zhang, J., Wang, B., Frye, S., Zhang, Y., Xu, Y., Lei, Q. and Guan, K.-L. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell 19: 17-30

Ye D, Guan K.L., Xiong Y (2018) Metabolism, Activity, and Targeting of D- and L-2-Hydroxyglutarates. Trends Cancer 4: 151-165

Yin, R., Mao, S.-Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., Jiao, Y., Li, C., Liu, S., Wu, D., Gu, W., Yang, Y.-G., Xu, G.-L. and Wang, H. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. Journal of the American Chemical Society, 135(28), pp.10396–403.

Yue, X., Trifari, S., Äijö, T., Tsagaratou, A., Pastor, W.A., Zepeda-Martínez, J.A., Lio, C.-W.J., Li, X., Huang, Y., Vijayanand, P., Lähdesmäki, H. and Rao, A. (2016) Control of Foxp3 stability through modulation of TET activity. J Exp Med 213: 377-397.

CHAPTER 2: TET deficiency perturbs mature B cell homeostasis and promotes oncogenesis associated with accumulation of G-quadruplex and R-loop structures.

2.1 Abstract

Enzymes of the TET family are methylcytosine dioxygenases that undergo frequent mutational or functional inactivation in human cancers. Recurrent loss-of-function mutations in TET proteins are frequent in human Diffuse Large B-Cell Lymphoma (DLBCL). Here we investigate the role of TET proteins in B-cell homeostasis and development of B cell lymphomas with features of DLBCL. We show that deletion of *Tet2* and *Tet3* genes in mature B cells in mice perturbs B-cell homeostasis and results in spontaneous development of germinal center-derived B cell lymphomas with increased G-quadruplexes and R-loops. At a genome-wide level, G-quadruplexes and R-loops were associated with increased DNA double strand breaks at immunoglobulin switch regions. Deletion of the DNA methyltransferase DNMT1 in TET-deficient B cells prevented expansion of germinal center B cells, diminished the accumulation of G-quadruplexes and R-loops, and delayed B lymphoma development, consistent with the opposing functions of DNMT and TET enzymes in DNA methylation and demethylation. CRISPR-mediated depletion of nucleases and helicases that regulate G-quadruplexes and R-loops decreased the viability of TET-deficient B cells. Our studies suggest a molecular mechanism by which TET loss-of-function might predispose to development of B cell malignancies.

2.2 Introduction

The three mammalian TET enzymes (TET1, TET2 and TET3) are Fe(II), O_2 and α-ketoglutarate dependent dioxygenases that sequentially oxidize 5-methylcytosine (5mC) to 5-hydroxymethyl- (5hmC), 5-formyl- (5fC) and 5-carboxyl- (5caC) cytosine (Tahiliani *et al.* 2009; Ko *et al.* 2010; Ito el at., 2011). TET enzymes regulate enhancer activity and DNA methylation dynamics during development (including B cell development) (Lio *et al.* 2016; Orlasnski *et al.* 2016; Dominguez *et al.* 2018; Lio *et al.* 2019; Rosikiewicz *et al.* 2020), cell differentiation and cell lineage specification (reviewed in Lio *et al.* 2020; Wu *et al.* 2014; Pastor *et al.* 2013; Rasmussen *et al.* 2016). *TET2* gene mutations and/or decreased TET activity have been observed in many hematological malignancies and solid cancers, often through impaired regulation of metabolic enzymes that affect TET activity (reviewed in Lio *et al.* 2020; Cimmino *et al.* 2011; Huang *et al.* 2014; Ko *et al.* 2015; Lio *et al.* 2019). For instance, *TET2* is recurrently mutated in ~10% of Diffuse Large B-cell Lymphoma (DLBCL) (Reddy *et al.* 2017; Schmitz *et al.* 2018; Chapuy *et al.* 2018), a heterogeneous malignancy originating in mature B cells undergoing activation and differentiation in germinal centers (GCs). *TET2* mutations represent an early driver event in DLBCL (Dominguez *et al.* 2018); in mouse models, deletion of *Tet2* in hematopoietic lineages disrupted GC B cell homeostasis and promoted development of more aggressive lymphomas when the transcription factor BCL6 was constitutively overexpressed (Dominguez *et al.* 2018). 5hmC deposition has been observed at sites of DNA double strand breaks (DSBs) in HeLa cells (Kafer *et al.* 2016), and TET2 is associated with degradation of stalled replication forks in BRCA2-deficient mouse cells (Kharat *et al.* 2020), suggesting that TET proteins regulate genomic integrity.

Two non-canonical DNA structures, R-loops and G-quadruplexes, can act as physical impediments to DNA and RNA polymerases during transcription and DNA replication, and have been linked to replication fork stalling and genome instability (reviewed in Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014; Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017). R-loops form, mostly at genomic regions with high GC content, when RNA binds to the transcribed strand of DNA, displacing the non-transcribed DNA strand (Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014). G-quadruplexes often form on the displaced G-rich strand when four guanines, one from each of four tracts of two or more guanine bases interspersed with variable numbers of random nucleotides ($G_{>2}N_{1\text{-}n}G_{>2}N_{1\text{-}n}G_{>2}N_{1\text{-}n}G_{>2}$), form square planar structures known as G-quartets, that are stabilized by Hoogsteen hydrogen bonding (Skourti-Stathaki *et al.* 2014; Rhodes *et al.* 2015; De Magis *et al.* 2019; Gray *et al.* 2014; Miglietta el al., 2020). G-quartets stack above each other and the resulting G-quadruplex structure is further stabilized by monovalent cations (Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017). G-quadruplexes and R-loops are commonly associated with gene promoters, 5' untranslated regions, DNA replication origins, telomeres and other regulatory elements in the mammalian genome (Crossley *et al.* 2019; Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017; Chedin *et al.* 2016; Niehrs *et al.* 2020; Sanz *et al.* 2016). The pathological effects of these structures can be mitigated by the concerted actions of diverse nucleases and helicases, among them RNases H1 and H2 which destroy RNA in RNA:DNA hybrids (Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014), and several helicases (e.g. ATRX, FANCD2, and Werner's (WRN) and Bloom (BLM) syndrome Recq-like DNA helicases) that can bind and resolve G-quadruplexes (Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017).

Here we show that profound TET loss-of-function, induced by deletion of the *Tet2* and *Tet3* genes in mature B cells in mice, is associated with the rapid development of DLBCL-like tumors from GC B cells. Like other malignancies associated with TET loss-of-function in primary mouse cells (An *et al.* 2015; Cimmino *et al.* 2015; Tsagaratou *et al.* 2017), Tet2/3-deficient B cells exhibit increased DNA damage, based on increased staining with phosphorylated H2AX (γH2AX). 5hmC deposition has been observed at sites of DNA double strand breaks (DSBs) in HeLa cells (Kafer *et al.* 2016), and While exploring possible mechanisms for increased γH2AX staining, we noticed a marked accumulation of G-quadruplex (G4) structures and R-loops in expanded TET-deficient B, T and myeloid cells. CRISPR-mediated depletion of RNASEH1 or ATRX, FANCD2 and BLM helicases led to a slight increase in DNA damage and apoptosis in TET-deficient compared to control B cells. Genome-wide mapping and high-throughput genome-wide translocation sequencing (HTGTS) showed a strong correlation of increased G-quadruplex and R-loop structures with increased DNA DSBs in switch regions of immunoglobulin genes in TET-deficient B cells. TET-deficient B cells also showed upregulation of the maintenance DNA methyltransferase DNMT1, as well as a slight but significant increase in DNA methylation at regions associated with G-quadruplexes and R-loops. DNMT1 was recently shown to bind G-quadruplex structures (Mao *et al.* 2018), and deletion of DNMT1 in TET-deficient B cells was associated with a striking reduction in GC B cells, decreased levels of G-quadruplex and R-loop structures in the surviving B cells, and a marked delay in B lymphoma development. Together, our findings suggest molecular mechanisms through which TET loss-of-function could promote oncogenesis and genome instability in multiple TET-deficient cell types.

2.3 Results

TET deficiency in mature B cells causes B cell lymphoma. Tet2 and Tet3 are the major TET paralogs expressed in mature B cells (Lio *et al.* 2016; Orlasnski *et al.* 2016; Lio *et al.* 2019). To disrupt Tet2 and Tet3 activity specifically in mature B cells, we generated compound transgenic mice bearing floxed alleles of the *Tet2* and *Tet3* genes (*Tet2*$^{fl/fl}$, *Tet3*$^{fl/fl}$, here termed *double-floxed (Dfl*), the Cre recombinase under control of the CD19 locus (*CD19Cre*), and the yellow fluorescent protein (YFP) reporter preceded by the *loxP-stop-loxP* cassette in the *Rosa26* locus (*Rosa26-YFP*^{LSL}) (**Fig. 2.1A**). Whereas *CD19Cre* and *Dfl* control mice, and mice with individual deletions of *Tet2* or *Tet3* (*CD19 Tet2 KO* and *CD19 Tet3 KO* mice), survived normally for more than a year of age, *Tet2*$^{fl/fl}$, *Tet3*$^{fl/fl}$, *CD19Cre* (*CD19 DKO*) mice with profound TET deficiency showed spontaneous development of lymphoma with complete penetrance and a median survival age of ~20 weeks (**Fig. 2.1B**). The disease was marked by lymphadenopathy and splenomegaly as early as 6 weeks of age (**Fig. 2.1C**), with disruption of normal splenic architecture (**Fig. S2.1A**), early expansion of activated CD4 and CD8 T cells ($CD44^+CD62L^{low}$) as well as T follicular helper cells ($CD4^+PD1^+CXCR5^+$) and progressive expansion of TET-deficient $CD19^+YFP^+$ B cells (**Fig. 2.1D**, and **Fig. S2.1B-K**). Transfer of the expanded YFP^+ B cells from 8-week-old *CD19 DKO* ($CD45.2^+$) mice into immunocompetent, $CD45.1^+$ congenic host mice recapitulated the lymphoma, whereas transfer of B cells from *Dfl* control mice did not result in long-term engraftment (**Fig. S2.1L-O**).

Transcriptional profiling of *CD19 DKO* and control *Dfl* B cells from 8-week-old mice revealed 2,678 differentially expressed genes (DEGs) in *CD19 DKO* B cells compared with *Dfl* control B cells. 1630 genes were upregulated and 1048 genes were downregulated (log2 fold change ≥2; FDR ≤0.05; **Fig. 2.1G** and **Table S2.1**). The transcriptional profile of *CD19 DKO* B cells resembled that of early germinal center (GC) B cells (**Fig. S2.2A-B**). Compared to control

Dfl or *CD19Cre* B cells, a subset of *CD19 DKO* B cells expressed low IgD and high Ephrin B1 (EFNB1), characteristic of GC B cells (**Fig. S2.2C**). The cells also expressed higher levels of BCL6, AID (activation-induced deaminase, a GC B cell-specific enzyme), IgG1, IgG2b and IgG3, indicating that they were undergoing class switch recombination, albeit without any major changes in germline transcription at the *IghM* (*μ*) and *IgG1* (*γ1*) loci (**Fig. S2.2D-G**). Analysis of variable gene segments of immunoglobulin heavy and kappa (κ) light chains (clonotypes) from *Dfl* and *CD19 DKO* B cells showed that the *CD19 DKO* B cells had undergone oligoclonal expansion compared to *Dfl* B cells (**Fig. S2.2H-I**). Furthermore, *CD19 DKO* mice showed expansion of GC B cells in Peyer's patches (**Fig. S2.2J-K**). These findings point to a GC origin for expanded B cells from *CD19 DKO* mice.

Consistent with previous reports in other TET-deficient cell types in mice (An *et al.* 2015; Cimmino *et al.* 2015; Tsagaratou *et al.* 2017), *CD19 DKO* B cells showed increased staining with antibodies to phosphorylated histone H2AX (γH2AX), which marks sites of DNA damage in the genome, compared to B cells isolated from *CD19Cre* and *Dfl* control mice (**Fig. 2.1E-F**). To further characterize the potential mechanisms by which TET deficiency predisposes to increased genomic instability, we scored differentially expressed genes from *CD19 DKO* B cells (**Fig. 2.1G**) for enrichment of annotated molecular pathways. The most downregulated pathways included signaling and scaffolding components of the Ig B cell receptor, the MHC complex and the RAS GTPase complex, which relay signals essential for B cell differentiation, but interestingly, the major upregulated pathways were linked to alterations in secondary DNA structures and DNA damage signaling (**Fig. 2.1H**). Compared to control B cells, TET-deficient B cells showed upregulation of mRNAs encoding several helicases and other proteins – FANCD2, BLM, WRN, PIF1, RECQL4, RNASEH2B, nucleolin (Ncl), Ataxin (Atxn1) – that recognize or regulate G-

quadruplex (G4) structures and R-loops (**Fig. 2.1G**). At the protein level, we confirmed that four known G4-binding helicases ATRX, FANCD2, BLM and WRN (Rhodes *et al.* 2015; Sauer *et al.* 2017), DNMT1 (which also binds G-quadruplex structures (Mao *et al.* 2018), and RNASEH1 (which like the RNASEH2 complex, specifically degrades the RNA strand of RNA:DNA hybrids in R-loops (Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014) showed increased expression in TET-deficient compared to control *Dfl* B cells (**Fig. 2.1I**).

TET deficiency is associated with increased G4s and R-loops. We found that TET deficiency resulted in increased levels of G-quadruplex structures and R-loops in *CD19 DKO* B cells. To detect G-quadruplex structures, we used BG4-Ig, a fusion protein of BG4, a single chain variable fragment (scFv) that recognizes G-quadruplex structures (Biffi *et al.* 2013) fused to the mouse IgG1 constant region to improve its valency and sensitivity (**Fig. 2.2A**). Treatment of activated primary B cells with a G4-stabilizing ligand, pyridostatin (PDS) (Muller *et al.* 2012) led to a significant increase in BG4-Ig fluorescence signal, authenticating this approach for G-quadruplex detection (**Fig. S2.3A-B**). *CD19 DKO* B cells from 8 to 10 weeks old mice showed an approximately 1.5- to 2-fold increase in staining with BG4-Ig antibody compared with *Dfl* or *CD19Cre* control B cells (**Fig. 2.2B-C**). We confirmed this result using N-methyl mesoporphyrin IX (NMM), a G4-binding compound that shows a strong increase in fluorescence when it binds G-quadruplex structures (Sabharwal *et al.* 2014) (**Fig. S2.3C-E**). (Muller *et al.* 2012; Sabharwal *et al.* 2014; Nicoludis *et al.* 2012) *CD19 DKO* B cells showed a marked increase (~3 fold) in NMM fluorescence signal compared with *Dfl* and *CD19Cre* control B cells (**Fig. 2.2D-E**). Amnis imaging flow cytometry further confirmed increased staining of *CD19 DKO* B cells with both BG4-Ig and NMM, compared with *CD19Cre* control B cells (**Fig. S2.3H-K**).

To detect R-loops, we used a recombinant, V5-epitope-tagged version of a catalytically inactive, mutant (D210N) RNASEH1 protein (rRNASEH1) (Chen *et al.* 2017; Yan etal., 2019; Crossley *et al.* 2021) (**Fig. 2.2A**) in preference to the S9.6 antibody against RNA:DNA hybrids (Boguslawski *et al.* 1986) which gave high background staining. CH12 B cells stained with rRNASEH1 displayed a strong fluorescence signal over background that was significantly diminished upon treatment with catalytically active, RNASE H enzyme (**Fig. S2.3F-G**). There was a clear increase (~2-fold) in R-loop levels in *CD19 DKO* B cells stained with rRNASEH1 compared with *Dfl* or *CD19Cre* control B cells (**Fig. 2.2F-G**; and **Fig. S2.3L-M**). Together, these studies provide compelling evidence for upregulation of G-quadruplex and R-loop structures in TET-deficient B cells.

The increase in G-quadruplex structures and R-loops occurred early (within 12 days) after deletion of the *Tet2 and Tet3* genes in B cells from *Cγ1CreTet2$^{fl/fl}$Tet3$^{fl/fl}$Rosa26-YFP^{LSL}* (*Cγ1 DKO*) mice, in which Cre recombinase expression and TET deletion are induced primarily in GC B cells after immunization. Unlike *CD19 DKO* mice, unimmunized *Cγ1 DKO* mice did not show early signs of morbidity and mortality. Immunization with a model antigen, sheep red blood cells (SRBCs), led to a slight increase in spleen cellularity but no apparent change in the frequency of GC B cells in *Cγ1 DKO* compared with *Cγ1Cre* or *Dfl* control mice (**Fig. S2.4A-C**), but resulted in a significant increase in G4 and R-loop levels in *Cγ1 DKO* GC B cells compared with *Cγ1Cre* or *Dfl* GC B cells (**Fig. 2.3A-D**).

We used the 40LB cell culture system, in which B cells are cultured on a mouse fibroblast cell line expressing CD40 ligand and B cell activation factor (BAFF) to induce *Cγ1-Cre* recombinase activity (Nojima *et al.* 2011), to explore the kinetic relationship between TET loss-of-function and alterations in G-quadruplex and R-loop structures. ~95 to 98% of *Cγ1 DKO* and

Cγ1Cre B cells cultured on 40LB cells underwent genetic recombination (deletion of the *LoxP-stop-LoxP* cassette), as judged by an increase in YFP expression within 3 days (**Fig. 2.3E**). Compared with *Cγ1Cre* control B cells, *Cγ1 DKO* B cells displayed a significant increase in G-quadruplex and R-loop levels by 4 days (**Fig. 2.3F-I**). In addition, we deleted *Tet* genes acutely by tamoxifen treatment of *ERT2-Cre Tet2$^{fl/fl}$ Tet3$^{fl/fl}$ Rosa26-YFP^{LSL} (ERT2-Cre DKO)* mice; compared to control *Dfl* mice, these mice show similar B cell development *in vivo* and similar proliferation of naïve B cells *in vitro* upon activation with lipopolysaccharide (LPS) and interleukin-4 (IL-4) (Lio *et al.* 2019). Naïve *ERT2-Cre DKO* B cells, as well as *ERT2-Cre DKO* B cells activated in vitro with LPS and IL-4 for 3 days in the presence of tamoxifen (**Fig. S2.4D**), exhibited a significant increase in G-quadruplex levels compared to control *Dfl* B cells (**Fig. S2.4E-F**). Increased levels of G-quadruplexes were also observed in expanded TET-deficient myeloid cells (*ERT2-Cre Tet1$^{fl/fl}$ Tet2$^{fl/fl}$ Tet3$^{fl/fl}$, ERT2creTKO*) and iNKT cells (*CD4Cre Tet2$^{-/-}$ Tet3$^{fl/fl}$, CD4Cre DKO*) relative to their TET-sufficient counterparts (**Fig. S2.4G-J**) (Tsagaratou *et al.* 2017; Lopez-Moyado *et al.* 2019). In summary, increased levels of G-quadruplexes and R-loops are early events associated with TET deficiency in B cells and other hematopoietic cells.

TET-deficient B cells are sensitive to G4 and R-loop targeting. Since mRNAs encoding G4-resolving helicases and the R-loop resolving enzyme RNase H were upregulated in TET-deficient B cells (**Fig. 2.1G**), we asked whether depletion of ATRX, BLM, FANCD2 and RNASEH1 proteins would affect the growth or survival of TET-deficient B cells. Cells were nucleofected with Cas9 ribonucleoprotein complexes (Cas9 RNPs) loaded with the appropriate CRISPR guide RNAs (**Fig. S2.5A**), then cultured in the presence of LPS for 48 hours. This procedure led to efficient depletion of the targeted proteins in *CD19 DKO* B cells, compared to their levels in cells nucleofected with a control Cas9 RNP directed to the *Cd4* locus (*Ctrl*) (**Fig.**

S2.5B). Individual depletion of ATRX, BLM, FANCD2 and RNASEH1 caused a significant increase in apoptosis of *CD19 DKO* B cells but not *Dfl* control B cells (measured by flow cytometry for activated caspase 3; **Fig. S2.5C-D**), concomitantly with increased DNA damage and increased G-quadruplex levels (assessed by flow cytometry for γH2AX and NMM respectively; **Figs. 2.5E-F**). We also used a G4 ligand, pyridostatin (PDS) (**Fig. S2.5G**), to stabilize G-quadruplexes in *Dfl* and *CD19 DKO* B cells cultured with LPS with or without PDS for 48 hours. PDS treatment led to an ~2.5-fold increase in apoptosis in *CD19 DKO* B cells but not in *Dfl* control B cells (**Fig. S2.5H-K**). These studies show that depletion of enzymes that regulate G-quadruplexes and R-loops increases apoptosis in TET-deficient B cells.

Relation of G4s and R-loops to DNA modifications. We mapped the genome-wide distribution of G-quadruplexes and R-loops in control-*Dfl* and *CD19 DKO* B cells by chromatin immunoprecipitation followed by sequencing (ChIP-seq) using BG4-Ig (Hansel-Hertsch *et al.* 2016) and MapR (Yan *et al.* 2019) (a modified CUT&RUN-based method for mapping R-loops which utilizes a recombinant catalytically inactive RNASEH1 protein fused to micrococcal nuclease) respectively. We defined a total of 9722 regions that showed enrichment for G4 and/or R-loop signals over the local background in replicate experiments (**Fig. 2.4A**). Consistent with the flow cytometry results, *CD19 DKO* B cells showed increased G4 and R-loop signals compared with *Dfl* control B cells (**Fig. 2.4B-C**). A majority of these (6212 regions; 64%) were present at annotated promoters and 5'UTRs (within +/- 1 kb of transcription start sites (TSS), and showed enrichment for G-rich DNA sequences predicted to be capable of forming G-quadruplex structures compared to the control regions selected randomly from the genome (**Fig. S2.6A-B**). The remaining 3510 regions (36%) were located distant from promoters (median distance ~14 kb from TSS). The G-quadruplex signal was increased at both promoter and non-promoter regions, but

only promoter regions showed a significant increase in R-loops signal in *CD19 DKO* B cells (**Fig. S2.6C-D**). The distributions of R-loops and G-quadruplexes at promoter regions were both shifted slightly 5' of the TSS in *CD19 DKO* B cells, consistent with the known propensity for G-quadruplex formation on the displaced single DNA strand of R-loops (Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014; Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017; Miglietta *et al.* 2020). The actual shift in R-loop signal upstream of the TSS might suggest transcriptional pausing at promoter proximal regions, however, the precise reason for the change in the R-loop signal profile in *CD19 DKO* B cells remains to be determined.

To relate the distribution of TET-regulated DNA modifications to the distribution of G-quadruplexes and R-loops, we mapped 5hmC by HMCP (5hmC Pull-down), a method similar to hME-Seal (Song *et al.* 2011) in which 5hmC is glucosylated and biotinylated prior to DNA precipitation using streptavidin beads. As expected, *CD19 DKO* B cells showed a strong depletion of 5hmC across the genome (**Fig. S2.6E**). In *Dfl* B cells, we observed a significant enrichment of 5hmC signal at and near (+/- 1kb) G-quadruplex and R-loop forming regions compared to control random regions in euchromatin (**Fig. S2.6F**). Whole Genome Bisulfite Sequencing (WGBS) analysis identified 6948 differentially-methylated regions (DMRs) genome-wide, of which 5934 (~85%) were hypermethylated and 1014 (~15%) were hypomethylated in *CD19 DKO* B cells compared with control cells (**Fig. S2.6G**). G4- and R-loop-enriched promoter and non-promoter regions displayed a slight but significant increase in DNA methylation in their flanking regions (+/- 1 to 2 kb) in *CD19 DKO* compared to *Dfl* B cells (**Fig. S2.6C-D,H** and **Fig. S2.7A-C**). In contrast, random genomic regions in *CD19 DKO* B cells were hypomethylated compared with those in control B cells (**Fig. S2.6I**), as we have previously reported for other TET-deficient genomes (An *et al.* 2015; Tsagaratou *et al.* 2017; Lopez-Moyado *et al.* 2019).

In general, more highly expressed genes in both *Dfl* and *CD19 DKO* B cells showed greater enrichment for G-quadruplexes and R-loops near the TSS (**Fig. S2.6J**). We also asked whether the presence of G-quadruplex structures and R-loops correlated with differential gene expression. Among all differentially expressed genes (DEGs) in *CD19 DKO* B cells, a significant proportion (~36%) harbored a G4- and R-loop-enriched region within 1 kb of their TSSs, compared to the proportion (64%) of DEGs without any G4- and R-loop-region and the proportion (22%) of non-DEGs containing a G4- and R-loop enriched region in their promoters (**Fig. S2.6K**, scatter plot and pie chart). However, both up- and down-regulated genes had G-quadruplexes and R-loops at their promoters (**Fig. S2.6K** and **Fig. S2.7A-C**), consistent with the proposed roles of G4 and R-loop structures in both up- and down-regulation of gene expression (Crossley *et al.* 2019; Rhodes *et al.* 2015).

TET-deficient B cells show genome-wide increase in Ig-translocations. To investigate the potential link between increased G-quadruplexes/R-loop levels and increased DNA breaks in *CD19 DKO* compared to control *Dfl* B cells (**Fig. 2.1E-F**), we performed locus-directed high-throughput genome-wide translocation sequencing (HTGTS (Hu *et al.* 2016). Using a biotinylated capture oligonucleotide bait located at a well-characterized G4- and R-loop-forming region present within the switch μ region in the IgH locus (**Fig. 2.4D**), we observed a striking increase in the absolute numbers of translocations arising from DNA breaks within or up to 500 bases 5' of this region (**Fig. 2.4D-E,G** and **Fig. S2.7D**), as well as an increase in the number of genomic loci harboring translocations in *CD19 DKO* compared with *Dfl* B cells (**Fig. 2.4F-G**; **Fig. S2.7D** and **Table S2.2**). The regions captured as translocation partners of the switch μ region by HTGTS showed significant enrichment of G-quadruplex and R-loop signals compared to control regions in both *Dfl* and *CD19 DKO* B cells (**Fig. 2.4H**). The absolute number of DNA breaks

(translocations) that directly overlap G4- and R-loop-enriched regions were also significantly increased in *CD19 DKO* B cells compared with *Dfl* B cells (**Fig. 2.4I**). The vast majority (~98%) of DNA breaks were located in the IgH locus associated with IgM as well as other Ig isotypes (**Fig. 2.4D**). Compared to control euchromatic regions, a substantial proportion (~85%) of HTGTS hits harbored canonical motifs predicted to form G-quadruplexes (**Fig. S2.7E**) and showed significant enrichment of DNA motifs targeted by the cytidine deaminase AID in B cells (**Fig. S2.7F**). Together, these studies show increased G-quadruplexes and R-loops in *CD19 DKO* B cells, correlating with an increase in the levels of DNA double-strand breaks at Ig switch regions.

DNMT1 deletion delays oncogenesis in TET-deficient mice. DNMT and TET enzymes have opposing biochemical activities in DNA methylation and demethylation respectively. A recent study showed that binding of DNMT1 to G-quadruplex structures inhibits its catalytic activity (Mao *et al.* 2018). We confirmed the DNMT1-G-quadruplex interaction (**Fig. S2.8A,B**), then asked whether (given the strong upregulation of DNMT1 in TET-deficient B cells (**Fig. 2.1I**), depleting DNMT1 in TET-deficient B cells affected the pathogenesis of B cell lymphoma. We generated *Dnmt1 fl/fl Rosa26-YFP*^{LSL} mice expressing *Cre* recombinase under control of the CD19 locus (*CD19Cre Dnmt1 knock out (KO)* and bred them with *Tet2fl/fl Tet3fl/fl* mice (Triple floxed-*Tfl*) (**Fig. 2.5A**). The *CD19 DKO* mice generated from these crosses succumbed to B cell lymphoma with a median survival of 20 weeks, whereas *CD19 TKO* mice (with additional deletion of *Dnmt1*) showed a substantial increase in survival (median survival 98 weeks) (**Fig. 2.5B**). Compared with *CD19 DKO* mice, *Tfl*, *CD19 Dnmt1 KO* and *CD19 TKO* mice displayed no apparent signs of splenomegaly at 10 weeks of age (**Fig. 2.5C** and **Fig. S2.8C-D**); a small proportion (4 out of 14 mice) of *CD19 TKO* mice did develop splenomegaly, albeit with a very long latency (**Fig. S2.8E**).

Since GC B cells are the cell-of-origin of the expanded B cells in *CD19 DKO* mice, we examined the numbers of GC B cells in Peyer's patches of *CD19 DKO* versus *Dnmt1-Tet2/3 CD19 TKO* mice. Deletion of *Dnmt1* in B cells completely abrogated GC B cells (**Fig. 2.5D-E**), similar to previous reports using DNMT1 germline hypomorphs (Shaknovich *et al.* 2011). *CD19 TKO* mice displayed GC B cell frequencies considerably lower than the expanded GC B cell population in *CD19 DKO* mice, but only slightly lower than those in *Tfl* control mice when gated on YFP+ B cell population (**Fig. 2.5D-E**). Together, these findings point to a functional interplay between TET and DNMT1 activities in GC B cell development and oncogenic transformation of TET-deficient B cells.

We assessed the levels of G-quadruplexes and R-loops in total splenic B cells and Peyer's patch GC B cells of 10-week-old mice by flow cytometry. *Tet2/3, Dnmt1*-deficient *CD19 TKO* B cells demonstrated a notable decrease in the levels of G4 (using both NMM and BG4-Ig staining) and R-loop (rRNASE H1 staining) structures compared with *Tet2/3*-deficient *CD19 DKO* B cells, almost to the levels of control *Tfl* B cells (**Fig. 2.5F-I**; and **Fig. S2.8F-G**). The decrease in the levels of G4 and R-loop structures in *CD19 TKO* B cells was also accompanied by decreased levels of γH2AX compared with *CD19 DKO* B cells (**Fig. S2.8H**). Similar decreases in G-quadruplex and R-loop levels were observed when *CD19 TKO* GC B cells isolated from Peyer's patches were compared to *CD19 DKO* GC B cells (**Fig. S2.8I-L**). Altogether, these results show that DNMT1 deletion inhibits the development of GC B cells to delay oncogenesis, and is associated with decreased levels of G-quadruplex and R-loop structures (**Fig. S2.8M**).

2.4 Discussion

Our studies establish a causal relationship between TET deficiency and the development of mature B cell neoplasms. The fully penetrant progression of B cell lymphomas in CD19 Tet2/3 DKO mice is consistent with the frequent occurrence of TET gene mutations or dysregulation of TET activity in human DLBCL (Reddy *et al.* 2017; Schmitz *et al.* 2018; Chapuy *et al.* 2018). We show that profound inactivation of TET function, through deletion of both the Tet2 and Tet3 genes, perturbs normal B cell homeostasis, leading to spontaneous expansion of B cells with a GC phenotype. We also document a clear association between TET deficiency, increased G-quadruplexes and R-loops, and increased DNA double strand breaks, particularly at Ig-switch regions in B cells; indeed, accumulation of G-quadruplex and R-loop structures is a novel feature of TET deficiency in B cells and cells of other hematopoietic lineages. TET-deficient B cells upregulated mRNAs encoding several proteins that recognize and regulate R-loops and G-quadruplex structures, including RNase H1, Rnaseh2b, DNMT1, Fancd2/FANCD2, ATRX, BLM, WRN, Recql4, and Pif1 (Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014; Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017). Genetic deletion of one of the G4 binders, DNMT1, in CD19 Tet2/3 DKO B cells was associated with decreased levels of the precursor GC B cells, decreased G-quadruplex and R-loop structures and a notable increase in the survival of Tet2/3-deficient mice. Together with the de novo DNA methyltransferases Dnmt3a and 3b, Dnmt1 is expected to maintain the increased DNA methylation resulting from Tet gene deletion in cells; thus, Dnmt1 deletion might be expected to counter, at least partially, any deleterious effects resulting from altered DNA methylation in Tet2/3-deficient B cells.

A recent study showed that TET proteins limit the activation of self-reactive B cells in the periphery and serve as critical regulators of B cell tolerance (Tanaka *et al.* 2020). While the identity

of the antigen(s) driving the expansion of CD19 DKO B cells in our system is not known, it is possible that TET-deficient B cells expand because they are self-reactive. This is consistent with the well-documented relationship between the expression of self-reactive B cell receptors and two mature B cell-derived malignancies, chronic lymphocytic leukemia (CLL) and the activated B cell (ABC) subtype of DLBCL, in humans (Young *et al.* 2015). The oligoclonal (rather than polyclonal) nature of the BCRs observed in expanded CD19 DKO B cells implies antigen-driven selection (Young *et al.* 2015). Two possibilities, not mutually exclusive, warrant further investigation. The intracellular and transcriptional signals emanating from BCRs may be regulated by TET enzymes to limit cellular expansion; alternatively, loss of TET proteins may inhibit a physiological brake on antigen-driven expansion through a mechanism that does not involve transcription.

G-quadruplex structures have been implicated in initiation of the frequent IgH-BCL2 translocations in follicular lymphoma (Nambiar *et al.* 2011; Rabkin *et al.* 2008). We observed an association between increased G-quadruplex and R-loop structures and DNA breaks at the switch regions (primarily switch μ) of Ig isotypes in Tet2/3-deficient B cells. Due to the lack of robust methods for genome-wide mapping of DNA breaks, we relied on HTGTS, a locus-directed method for DNA break mapping at single base resolution (Hu *et al.* 2016). Since HTGTS maps DNA breaks relative to DNA junctions formed by break ligation (with possible errors due to indels), it is likely that we have underestimated the total number of DNA breaks. Nevertheless, we observed a striking increase in the absolute numbers of DNA DSBs in CD19 DKO B cells when compared with control B cells. The DNA breaks could arise either from conflicts of G-quadruplex and R-loop structures with transcription or DNA replication machineries, or from specific targeting of AID cytidine deaminase, the B cell mutator, to G4 structures and R-loops. In fact, AID possesses a G4-binding activity that is important for its genome-wide targeting (Qiao *et al.* 2017; Xu *et al.*

2020; Yewdell *et al.* 2020). A recent study reported an association between G4 structures and AID activity at commonly mutated genes in B cell lymphoma (Xu *et al.* 2020); consistent with these findings, we observed an enrichment of AID hotspot motifs within translocation sites with increased levels of G-quadruplexes and R-loops. The detailed mechanisms through which G-quadruplexes and R-loops recruit AID and/or promote genomic instability in B cells remain to be addressed.

TET-dependent methylation changes do not necessarily play a direct biochemical role in the observed increase of R-loop and G quadruplex structures in Tet2/3 deficient B cells. We based this conclusion on the fact that although we observed a slight but significant link between loss of TET activity and increased DNA methylation in the vicinity of R-loops and G-quadruplex structures, we observed no change in the levels of DNA methylation at the R-loop/G-quadruplex-containing regions (median size ~750 bp) themselves. Thus, further studies are needed to fully understand how TET deficiency in multiple cell types results in increased levels of G-quadruplex and R-loop structures.

DNMT1 has been previously shown to be induced in GC B cells (Shaknovich *et al.* 2011), where it may be needed to maintain the DNA methylation landscape during rapid GC B cell proliferation. Thus, the apparent upregulation of DNMT1 in TET-deficient B cells is likely due to their prominent GC phenotype compared to WT B cells. Conversely, Dnmt1 deletion results in a dramatic decrease of GC B cells, potentially the primary reason for the delayed oncogenesis observed in triple Tet2/3, Dnmt1-deficient B cells. In fact, the late-onset cancers may reflect delayed outgrowth of cells that escaped deletion of one or both Dnmt1 alleles, since in one experiment (unpublished data), PCR analysis of YFP+ total B cells and GC B cells in Peyer's patches in much older (7-month-old) mice showed almost complete deletion of Tet2 and Tet3 but

only ~50% deletion of Dnmt1. The decrease in R-loop and G-quadruplex structures observed upon Dnmt1 deletion in total Tet2/3-deficient B cells suggests a possible role for DNMT1 activity in stabilization of these structures. DNMT1 is known to be overexpressed in several different hematological and solid cancers (Zhang & Xu *et al.* 2017). The potential interplay between TET and DNMT activities in regulating oncogenesis as well as R-loop and G-quadruplex levels in B cells and other cell types remains to be investigated.

Our studies suggest that G-quadruplexes and R-loops could be therapeutic vulnerabilities in cancers with TET loss-of-function. G4-stabilizing ligands were recently shown to decrease cell viability in ATRX-deficient gliomas and BRCA1/2-deficient tumor cells (Wang, Y. *et al.* 2019; Xu, H. *et al.* 2017). In our hands, the use of a G4-stabilizing ligand, or depletion of proteins known to regulate G4 or R-loop structures, was associated with increased DNA DSBs and a slight increase in apoptosis in TET-deficient B cells. Furthermore, deletion of DNMT1 in TET-deficient B cells prevented the accumulation of R-loop and G-quadruplex structures in splenic B cells and Peyer's patch GC B cells and rescued the survival of TET-deficient mice. Follow-up studies in pre-clinical models could test whether a combination of G4-stabilizing agents and DNA methyltransferase inhibitors might synergize to delay the onset and/or progression of B cell lymphomas and other malignancies with TET loss-of function.

2.5 Materials and Methods

Mice. Tet2fl/fl and Tet3fl/fl mice were generated as previously described (Kang *et al.* 2015; Ko *et al.* 2011). C57BL/6J (000664), CD19 cre (006785), Rosa26-LSL-EYFP (006148), Cγ1-Cre (010611), Ubc-CreERT2 (008085; described as ERT2cre herein) and CD45.1 mice (002014, ptprca) were obtained from Jackson Laboratory. To induce ERT2cre-mediated deletion, Cre-expressing and control mice were intra-peritoneally injected with 2 mg tamoxifen (Sigma) dissolved in 100 uL corn oil (Sigma) daily for 5 days. For transplantation studies, CD45.1 mice were sub-lethally irradiated with 600 rads of X-rays 24 hours prior to transfer of 2 million CD19 DKO or Dfl B cells through the retro-orbital sinus. All mice used were 8 to 16 weeks of age (unless otherwise indicated) and were on a C57BL/6 genetic background, housed in specific-pathogen free animal facility at La Jolla Institute for Immunology. All studies were performed according to protocols approved by the Institutional Animal Care and Use Committee.

Immunization. Cγ1-Cre, Dfl and Cγ1-DKO mice were immunized with sheep red blood cells-SRBCs (#31102, Colorado Serum company, CO, USA) washed two times with PBS and injected in 2 doses, first primed with 200X106 SRBCs followed by a boost at day 5 with 109 SRBCs before analysis of splenocytes at day 12 post first immunization.

B cell isolation and cultures. Primary B cells were isolated using flow sorting of YFP+ CD19 DKO B cells for RNA-Seq analysis, EasySep Mouse Pan B cell isolation kit (#19844 Stem Cell Technology, Canada) for in vitro culture of B cells and CD19 positive selection (for G4 mapping, R-loop mapping, WGBS and HTGTS studies) from splenocytes. Primary B cells and CH12F3 (CH12) cells were cultured at 37°C, 5% CO2 in RPMI 1640 media supplemented with 10% FBS, 1x MEM non-essential amino acids, 10mM HEPES, 2mM Glutamax, 1mM sodium pyruvate, 55μM 2-mercaptoethanol (all from Life technologies). B cells (5x105-1x106 cells/ mL)

were activated with 10 ug/mL LPS (for Cas9 RNP targeting studies) from E. coli O55:B5 (Sigma, St. Louis, MO), 25 ug/mL LPS and 10 ng/mL rmIL-4 for stimulation of Dfl and ERT2cre B cells in the presence of 1 uM 4-hydroxytamoxifen (Tocris). For pyridostatin (PDS) (#SML2690, Sigma-Aldrich) treatment B cells were activated with 10 ug/mL LPS for 48 hours in presence of 10 uM PDS before analysis. All cytokines used above were from Peprotech (Rocky Hill, NJ).

Flow cytometry. Primary cells and in vitro cultured cells were stained in FACS buffer (1% bovine serum albumin, 1mM EDTA, and 0.05% sodium azide in PBS) with indicated antibodies for 30 mins on ice. Cells were washed and then fixed with 1% paraformaldehyde (diluted from 4% with PBS; Affymetrix) for 10 minutes at room temperature before FACS analysis using FACS Celesta and FACS LSR II (BD Biosciences). Antibodies and dyes were from BioLegend, eBioscience, and BD Pharmingen. Data were analyzed with FlowJo (FlowJo LLC, Ashland, OR).

Immunoblotting. Proteins isolated from cells with NP-40 lysis buffer were resolved using NuPAGE 4-12% Bis-Tris gel (ThermoFisher) and transferred from gel to PVDF membrane using Wet/Tank Blotting Systems (Bio-Rad). Membrane was blocked with 5% non-fat milk in TBSTE buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 0.05% Tween-20, 1mM EDTA), incubated with indicated primary antibodies, followed by secondary antibodies conjugated with horse-radish peroxidase (HRP) and signal was detected with enhanced chemiluminescence reagents (Invitrogen) and X-ray film. Antibodies against ATRX (clone D5), FANCD2 (clone Fl17) and BLM (clone B4) were purchased from Santa Cruz Biotechnology. Antibodies against RNASE H1 (NBP2-20171) from Novus biologicals, WRN (clone 8H3) and β-ACTIN HRP (clone 13E5) from Cell signaling and DNMT1 (ab19905) from Abcam.

Cas9 RNP targeting. Alt-R crRNA and Alt-R tracrRNA (from IDT) were reconstituted to concentration of 100 uM in Nuclease-Free Duplex buffer (IDT). RNA duplexes were prepared by

mixing oligonucleotides (Alt-R crRNA and Alt-R tracrRNA) at equimolar concentrations in a sterile PCR tube (e.g. 4 uL Alt-R crRNA and 4 uL Alt-R-tracrRNA). Mixed oligonucleotides were annealed by heating at 95C for 5 minutes in a PCR machine, followed by incubation at room temperature for at least 1 hour. 5.9 uL of crRNA-tracrRNA duplexes (180 pmol) were then mixed with 2.1 uL IDT Cas9 v3 (80 pmol) (together Cas9 RNP) by gentle pipetting and incubated at room temperature for at least 10 min. 35 uL of cell media was aliquoted in a 24 well plate. 2 million B cells were then washed with PBS and resuspended in 15 uL resuspension buffer (4D-Nuclcofector X Kit S, #V4XP-4032; Lonza). Cells in resuspension buffer (12.5ul) were then mixed with Cas9-RNPs (7.5 ul) and transferred to the nucleofection cuvette strips and electroporated using the CM-137 program with 4D nucleofector. After electroporation, cells were transferred into 35 uL media in 24 well plate, and incubated for 20 min at 37C before transferring cells to activation media with 10ug/ml LPS. The crRNA sequences used are listed in **Table S2.3**.

G4 detection and mapping. For flow cytometry-based detection of G4s, cells were stained for cell surface markers according to the flow cytometry staining protocol described above. Following cell surface staining, cells were fixed with 4% paraformaldehyde in PBS for 12 minutes at room temperature. Using the intracellular transcription factor staining kit (Invitrogen), cells were then permeabilized at 4C overnight according to manufacturer's instructions. Cells were then washed with wash buffer and treated with RNASE A (1:50, AM2269; Ambion) for 30 minutes at room temperature followed by washing and incubation with 1:100 dilutions of BG4-Ig antibody (Ab00174-1.1, 1mg/ml; Absolute Antibody) for 30 minutes at room temperature or 4C. This was followed by washing and subsequent staining with anti-mouse IgG1 fluorophore conjugated antibody (Biolegend) to reveal the signal.

The excitation and emission spectra of NMM in the presence of G-quadruplex DNA, but not in the presence of non-G-quadruplex control DNA, closely mimicked that of the fluorophore Brilliant Violet 605 (BV605). Therefore, we also used NMM to stain for G-quadruplexes. Following RNase A, treatment cells were incubated with 10 μM NMM (NMM580, Frontier Scientific) for 30 min at 25 °C or 4 °C before reading the fluorescence signal in the BV605 channel. Mouse CH12 B cells treated with PDS, which binds G-quadruplex structures in a manner distinct from NMM (Muller *et al.* 2012; Sabharwal *et al.* 2014) showed increased NMM fluorescence in flow cytometry (**Fig. S2.3E**), validating NMM as a flow cytometric probe for G-quadruplex structures in cells.

G4 mapping was performed as previously described with minor modifications (Hansel-Hertsch *et al.* 2016). Briefly, ~10 million cells were fixed with 1% formaldehyde at room temperature with nutation for 10 minutes mins at 1x106 cell/mL in media, quenched with 125 mM glycine, washed twice with ice cold PBS. Cells were pelleted, snap-frozen in liquid nitrogen, and stored at -80°C until use. Cell pellets were then lyzed using 800 ul of ChIP hypotonic solution (Chromatrap) and 150 ul lysis buffer (Chromatrap) supplemented with protease inhibitor cocktail (invitrogen) on ice according to manufacturer's instructions. The lyzed nuclei suspension was then sonicated using the bioruptor pico sonicator (Diagenode), 16 cycles 30 seconds on / 30 seconds off and the lysates were cleared by centrifugation at >12000g for 10 minutes. The lysates were diluted 1:10 times in intracellular buffer (25 mM HEPES pH 7.5, 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl2, 1 mM MgCl2) (total 200 ul) then treated with 1mg/ml RNASE A (AM2269; Ambion, 1:100 dilution) for 20 minutes at 37C. Chromatin lysates were then incubated with BG4-Ig antibody (Ab00174-1.1, 1mg/ml; Absolute Antibody) (1:100) for 1.5 hours at 16 C followed by incubation with protein G magnetic dynabeads (Invitrogen). The beads were then washed 5 times

with 1 ml wash buffer (intracellular buffer with 0.05% tween 20). The beads were re-suspended in elution buffer (100 mM NaHCO3, 1% SDS, 1 mg/mL RNaseA; Qiagen) treated with RNASE A (1:100, 100mg/ml) at 37 C for 30 minutes and proteinase K (1:40, 20mg/ml; Ambion) at 65 C for 2 hours to overnight. DNA was purified with Zymo ChIP DNA Clean & Concentrator-Capped Column (Zymo Research, Irvine, CA). Library was prepared with NEB Ultra II library prep kit (NEB) following manufacturer's instruction and was sequenced on an Illumina NovaSeq 6000 (50 bp paired-end reads). 2 independent experiments were combined for analysis.

NMM oligonucleotide fluorescence enhancement assay. G4 forming oligonucleotides and control oligonucleotides with mutated G-stretches were ordered from Integrated DNA technologies using standard desalting purification (**Table S2.3**). The oligonucleotides were reconstituted at 100μM concentration in 10mM Tris and 1mM EDTA (TE buffer) and then diluted to a final concentration of 10μM in buffer containing 20mM HEPES, 250mM KCl and 1mM DTT. The diluted oligonucleotides were then heated at 95C for 10 minutes followed by gradual cooling to room temperature (cooled down at 0.1C per second ramp rate) in a PCR machine. 50 μl of 10μM was then mixed with 50 μl of 20μM NMM incubated at room temperature for 1 hour and the fluorescence was measured in the Spectramax M2 (Spectral labs) plate reader with excitation at 400nm and emission at 610nm.

Recombinant RNASE H1 purification. Coding sequence of N-terminal V5-tagged mutant human RNASE H1 (D210N) (Yan *et al.* 2019) was sub-cloned in frame with Glutathione-S-Transferase (GST) protein separated by a Precission protease cleavage site in the pGEX6p1 vector. BL21 (DE3) cells were transformed with the pGEX6p1 vector with GST-RNASE H1 fusion. Single colonies were picked and grown overnight in 5 ml LB at 37C before further expanding in 500ml LB (1:100) with shaking at 37C for 2-3 hours until the OD at 600 nm reached

0.5-0.6. 1mM IPTG was then added to induce the protein expression for 3 hours at 37C with shaking. Expression of protein was confirmed by coomassie staining on a SDS-PAGE gel by comparing with an uninduced control culture. Cultures were then pelleted at 4500g for 15 minutes and cell pellets were re-suspended in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl), 15ml for 200ml culture. The cell suspension was then lysed by sonication for a total of 4 minutes (5 seconds on, 30 seconds off cycle) with a probe sonicator (Missonix) followed by centrifuge at 10,000g for 20min, at 4C. The supernatant was then mixed with 150ul washed and equilibrated glutathione agarose beads (#16100; Pierce) and rotated for 4 hours to overnight. The lysate beads mixture was then washed 3 times with lysis buffer and then re-suspended in 1ml elution buffer (25mM Tris-HCl (pH 7.6), 10% Glycerol, 0.2mM EDTA, 100mM KCl, 1mM DTT) and treated overnight with precission protease (#Z02799; Genescript). The supernatant was then collected and the protein was quantified by absorbance at 280 nm using Nanodrop. The purity of the protein was additionally confirmed by coomassie staining of samples on an SDS-PAGE gel.

R-loop detection and mapping (MapR). For flow cytometry-based detection of R-loops, cells were stained for the cell surface marker according to the flow cytometry staining protocol described above. Following cell surface staining, cells were fixed with 4% paraformaldehyde in PBS for 12 min at 25 °C. Using the intracellular transcription factor staining kit (Invitrogen), cells were permeabilized at 4 °C overnight according to manufacturer's instructions. Cells were then washed and incubated with 1:50 dilutions of V5-tagged recombinant mutant RNase H1 for 30 min at 4 °C followed by washing and subsequent staining with V5 tag antibody (1:200, clone D3H8Q; Cell Signaling Technologies) and rabbit secondary antibody (1:1,000; A-21245, Invitrogen) to reveal the signal. For RNase H digestion and R-loop quantification, CH12 cells were fixed, permeabilized and treated with 20 U of RNase H (NEB, M0297S) in 100 µl of digestion buffer

diluted in water for 2 h at 37 °C before proceeding to R-loop staining. MapR and immunoprecipitation with S9.6 antibody identify a set of R-loops in common as well as sets of unique R-loops that depend on differential sequence specifities and/or preferences of S9.6 and RNase H; thus, R-loops identified by MapR may represent a slightly different subset than those identified by other methods (Chedin el al., 2021; Konig *et al.* 2017).

R-loop mapping was performed by the MapR method, as previously described (Yan *et al.* 2019). The libraries were sequenced on an Illumina NovaSeq 6000 (50-bp paired-end reads). Three independent experiments were combined for analysis

Detection of G-quadruplex binding proteins. Purified naïve B cells were activated in presence of 25 ug/ml LPS and 10ng/ml IL-4 and nuclear extracts from 20 million cells were prepared using Dignam extraction. Cells were resuspended in 10 ml dignam buffer A (10 mM KCl, 1.5 mM MgCl2) supplemented with protease inhibitor cocktail (Thermofisher) and incubated on ice for at least 10 minutes. The cells were then homogenized and lyzed with a dounce homogenizer (10 strokes) and nuclei were collected by centrifugation at 600 g for 5 minutes and washed again with buffer A. Cells were then resuspended in 1ml of dignam buffer C (0.2 mM EDTA, 25% glycerol (v/v), 20 mM HEPES-KOH (pH 7.9), 0.42 M NaCl, 1.5 mM MgCl2) supplemented with protease inhibitor cocktail and rotated at 4C for 1 hour. The lysates were collected and passed through a PD-10 buffer exchange column (GE healthcare) equilibrated with intracellular buffer (25 mM HEPES pH 7.5, 10.5 mM NaCl, 110 mM KCl, 130 nM CaCl2, 1 mM MgCl2). The nuclear lysates collected in intracellular buffer were spike-in with 1 ug flag-tagged BG4 scFv antibody (Millipore), pre-cleared with myone T1 streptavidin-conjugated dynabeads (Life technologies) for one hour and incubated with streptavidin beads conjugated with a mixture of G4 or non-G4 forming oligonucleotides for 4 hours (**Table S2.3**). Samples were then washed with intracellular

buffer with 0.05% tween 20 for a total of 5 washes and denatured using the SDS-PAGE sample buffer for immunoblotting.

RNA-Seq. RNA from cells was isolated with Trizol (ThermoFisher, Waltham, MA) following manufacturer's instructions. Isolated RNA was further purified using RNA clean up and concentrator kit (R1013, Zymo research) according to manufacturer's instructions. 40 ng of purified RNA was then used for preparation of RNA-sequencing libraries with the Nugen ovation RNA-Seq V2 system (now Tecan Genomics) according to the manufacturer recommended protocol. Samples were sequenced on Illumina Hiseq 2500 (single-end 50 bp reads). 2 independent experiments were combined for analysis.

Whole genome bisulfite sequencing (WGBS). 1 μg of genomic DNA was spiked-in with unmethylated lambda-phage DNA (1:200) and sonicated using Bioruptor (9 cycles 30 seconds on, 30 seconds off). The sonicated DNA fragments were end repaired, A-tailed and ligated with methylated adaptors (all using NEB kits). The adaptor ligated fragments were then treated with sodium bisulfite using the Methylcode kit (Invitrogen) for a total of 4 hours according to manufacturer's protocol. The bisulfite converted, adaptor ligated fragments were then amplified with Kapa Hifi Hotstart Uracil+ PCR mix with NEB universal dual indexing primers. The libraries were then sequenced on an Illumina NovaSeq 6000 (150 bp paired-end reads). 2 independent experiments were combined for analysis.

5hmC pull down (HMCP). 5hmC mapping was performed on purified genomic DNA using the HMCP method in collaboration with Cambridge Epigenetix according to the manufacturer's recommended protocol. The libraries were then sequenced on an Illumina Hiseq 2500 (single-end 50 bp reads). 2 independent experiments were combined for analysis.

Linear Amplification Mediated High Throughput Genome-wide Translocation Capture Sequencing (LAM-HTGTS or HTGTS). HTGTS protocol was adapted from previously published method (Hu *et al.* 2016). Briefly, genomic DNA was sonicated with 2 pulses of 5 secs ON and 1 minute off using the Bioruptor pico sonicator (Diagenode). LAM-PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo Fisher) and Sμ 5' biotinylated bait (TAGTAAGCGAGGCTCTAAAAAGCA). 4 independent PCRs were run with ~2.5 μg of sonicated DNA each, with 90 cycles of primer elongation. The biotinylated DNA fragments were purified using Dynabeads T1 streptavidin beads (Invitrogen) at room temperature, on a rotisserie, overnight. On bead ligation was performed with T4 DNA ligase (NEB) and annealed bridge adaptors (GCGACTATAGGGCACGCGTGGNNNNNN-NH2 and 5'P-CCACGCGTGCCCTATAGTCGC-NH2) for 1h at 25°C, 2h at 22°C, and overnight at 16°C. 15 cycles of nested PCR were performed on the DNA-bead complexes using an Sμ nested primer (ACACTCTTTCCCTACACGACGCTCTTCCGATCT-GGTAAGCAAAGCTGGGCTTG) and a reverse universal primer I7 (GACTGGAGTTCAGACGTGTGCTCTTCCGATCT-GACTATAGGGCACGCGTGG). The bold regions represent the complementary regions to the genomic sequence in the Sμ nested primer and the bridge adaptor for I7 primer. PCR products were cleaned with the QIAquick Gel Extraction Kit (Qiagen) and the final barcoding/indexing PCR was performed using the P5-I5 and P7-I7 universal dual indexing primers (NEB) for 15 cycles and purified using Ampure beads purification with a cutoff of 200 bp. Library quality was evaluated by Bioanalyzer (Agilent) and sequencing performed using the Illumina Miseq Reagent Kit v2 (600 cycles) following the manufacturer's instructions. DNA junctions from 2 independent experiments were combined for analyses. The complete list of primers and sequences are included in **Table S2.3**.

Bioinformatics analyses. The reference genome used was mm10. Heatmaps and profile plots were generated using DeepTools (Ramirez *et al.* 2016).

HMCP analysis. Single-end reads were mapped to the mouse genome mm10 GRCm38 (Dec. 2011) indexed with spiked-in phage lambda, using Bowtie (V 1.1.2) (-S -p 3) (Langmed *et al.* 2009). Reads that mapped to spike-ins were filtered out with samtools view; remaining mapped reads in sam files were sorted and PCR duplicates were removed with Picard (V 2.7.1). Peaks were called with MACS2 (callpeak --keep-dup all -g mm) (Xi *et al.* 2009). Peaks from the replicates from same condition were merged with mergePeaks from HOMER and kept only those that intersected between the replicates. Intersected peaks from the replicates were merged between the conditions to create the total number of peaks. Tag directories were created and replicates were merged with makeTagDirectory from HOMER (-genome mm10) (Heinz *et al.* 2010); makeMultiWigHub.pl was used to generate tracks.

WGBS analysis. WGBS reads were mapped with BSMAP (v2.90) (Xi *et al.* 2009) to both the mouse genome mm10 GRCm38 (Dec. 2011) from UCSC and the lambda genome. Bisulfite conversion efficiency was estimated based on cytosine methylation in all contexts. For all the samples the bisulfite conversion efficiency was higher than 0.9996. Duplicated reads caused by PCR amplification were removed by PICARD's MarkDuplicates. CpG DNA methylation at both DNA strands was called by methratio.py script, from BSMAP (v2.90) (-g - i "correct" -x CG,CHG,CHH). To identify differentially methylated cytosines and regions (DMCs and DMRs) we used RADmeth methpipe-3.4.2 (adjust -bins 1:100:1; merge -p 0.05).

RNA-seq and BCR analysis. Single-end (50bp) reads were mapped to the mouse genome mm10/GRCm38 using STAR (Dobin *et al.* 2013) (--runThreadN 8 --genomeLoad LoadAndRemove --outFilterMultimapNmax 1 --outFilterType BySJout --alignSJoverhangMin 8

--alignSJDBoverhangMin 1 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 100000 --outFilterMismatchNmax 0). Counts were obtained with featureCounts (subread-1.4.3-p1) (-g gene_name -s 1). Differentially expressed genes were calculated with DESeq2 (Love *et al.* 2014), filtering out genes that did not have any count in any condition; cut off to define DEGs was an adjusted pvalue 0.05 and a log2 fold change >= +/-1

BCR sequences were retrieved from RNA-Seq data sets, and the frequency of IG chain clonotypes was determined using MiXCR (Bolotin *et al.* 2017) (mixcr-1.7–2.1) package with the default parameters "align -c IG -s mmu -p rna-seq -OallowPartialAlignments=true". Two rounds of contig assembly were performed by employing the "assemblePartial" function; extension of incomplete BCR was done with "extendAlignments" function; assembly and export of the clonotypes was performed using the "assemble" and the "exportClones" --preset min -fraction -targets -vHits -dHits -jHits -vAlignments -dAlignments -jAlignments) functions, respectively.

G4 analysis. Paired-end (50bp) reads were mapped to the mouse genome mm10 GRCm38 (Dec. 2011) indexed with spiked-in G4 sequences, using Bowtie (V 1.1.2) (Xi *et al.* 2009). Reads that mapped to spike-ins were filtered out with samtools view; remaining reads in sam format were sorted and PCR duplicates were marked and removed with Picard (V 2.7.1). Peaks were calculated with MACS (Zhang *et al.* 2008) (callpeak –keep-dup all -g mm -p 0.0001), using input as control. Peaks from the replicates from same condition were merged with mergePeaks from HOMER and kept only those that intersected between the replicates. Intersected peaks from the replicates were merged between the conditions to create the total number of peaks. Tag directories were created and replicates were merged with makeTagDirectory from HOMER (Heinz *et al.* 2010) (-genome mm10); makeMultiWigHub.pl was used to generate tracks.

MapR. Paired-end (50bp) reads were mapped to the mouse genome mm10 GRCm38 (Dec. 2011, using Bowtie (V 1.1.2). Reads that mapped to spike-ins were filtered out with samtools view; remaining reads Mapped reads were sorted in sam format were sorted and PCR duplicates were marked and kept with Picard (V 2.7.1). Peaks were calculated with MACS2 (callpeak –keep-dup all -g mm --broad --broad-cutoff 0.1), using input as control. Peaks from the replicates from same condition were merged with mergePeaks from HOMER (Heinz *et al.* 2010) and kept only those that intersected between the replicates. Intersected peaks from the replicates were merged between the conditions to create the total number of peaks. Tag directories were created and replicates were merged with makeTagDirectory from HOMER (-genome mm10); makeMultiWigHub.pl was used to generate tracks. For heatmaps and profiles plots generated with Deeptools, MNase control signal was removed from the rRNASEH1-MNase signal.

G-quadruplexes and R-loops peaks. The 9722 G-quadruplex and R-loop regions were obtained from the union of G-quadruplex and R-loop peaks identified by using 'intersectBed' from bedtools (Quinlan *et al.* 2010).

HTGTS. Analysis was done following HTGTS (Hu *et al.* 2016) pipeline (https://github.com/robinmeyers/transloc_pipeline). Briefly, paired-end reads were trimmed with TranslocPreprocess.pl and full analysis was performed with TranslocWrapper.pl. Post filter analysis was done using TranslocFilter.pl with default parameters. Circos plots were generated using the Circos software (Krzywinski *et al.* 2009). HTGTS hits were identified in a 10kb window and color scale was assigned based on the number of hits per window.

Potential to form G-quadruplexes analysis. Regular expressions in 'awk' commands were used to categorize the potential to form G-quadruplexes and R-loops peaks from fasta sequences. Loop1-7 is defined as with the following expression: G3+N1-7 G3+N1-7 G3+N1-7G3.

The long-loop is defined as sequences with a G4 with any loop of length >7 (up to 12 for any loop and 21 for the middle loop). The simple-bulge is defined as sequences with a G4 with a bulge of 1–7 bases in one G-run or multiple 1-base bulges. And the Two Tetrad/Complex bulge is defined as sequences with G4s with two G-bases per G-run with several bulges of 1–5 bases.

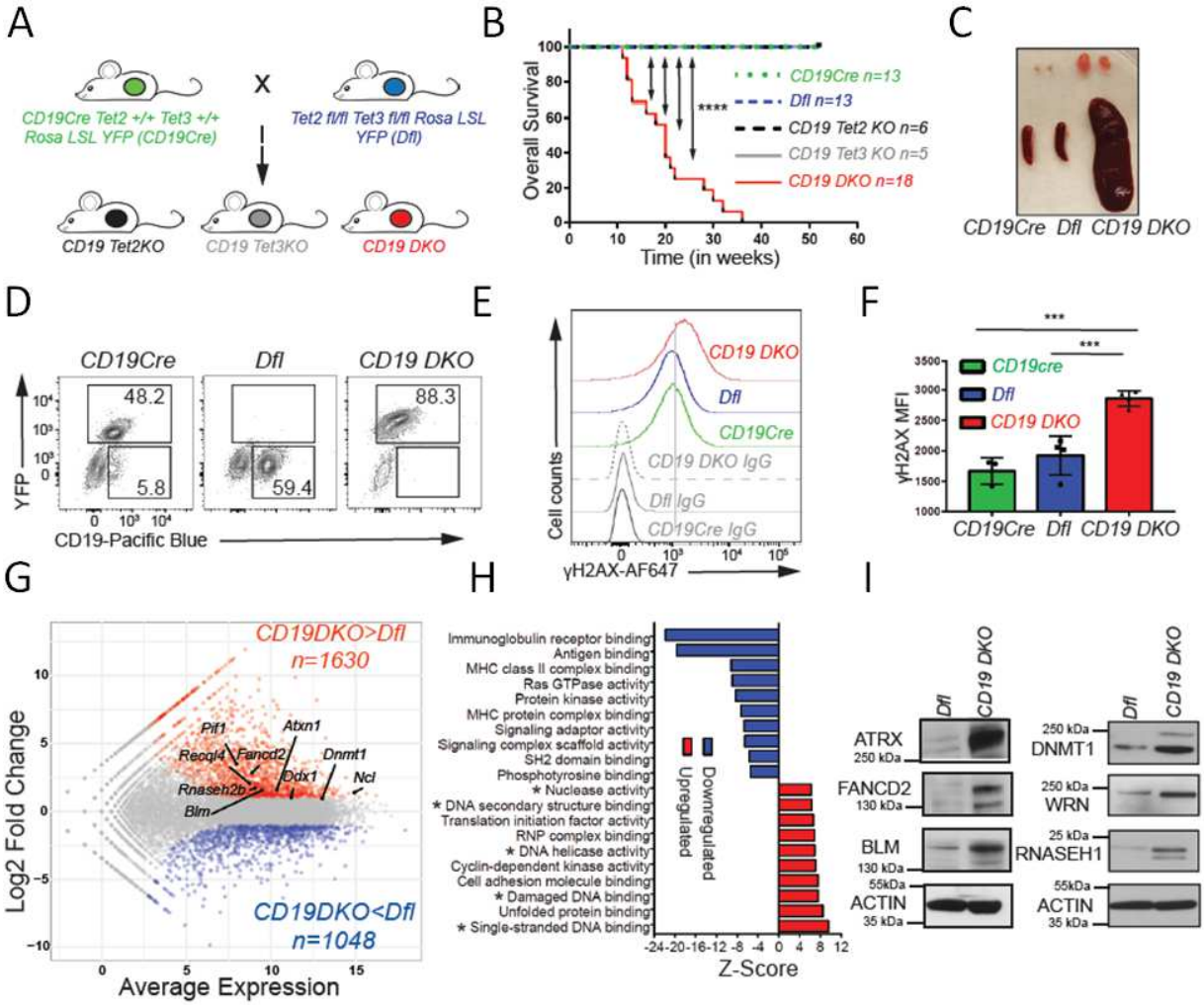
AID motifs hotspots analysis. Fasta sequences were retrieved from bed files from the G-quadruplexes and R-loops peaks. Library "stringr" and function str_locate_all in R was used to identify the "[A|T][A|G]C[C|T]" or "[A|G]G[C|T][A|T]" motifs.

Random regions. Random regions were retrieved with the function "shuffleBed" from the bedtools (Quinlan *et al.* 2010) using as the reference the whole mm10 genome. In particular cases, random regions were retrieved from euchromatin regions defined by the A compartment from Hi-C analysis in naïve and activated B cells (Kieffer-Kwon *et al.* 2017).

Quantification and statistical analysis. Statistical analyses were performed with Graph Prism 8 and R version 3.4. The statistical tests used to determine significance in each analysis are described in the figure legends of the corresponding figures. Parametric tests (t-tests, ANOVA) were used in experiments where normal distribution could be assumed, whereas in other cases, where normal distribution could not be assumed, non-parametric tests (Wilcoxon signed-rank test or Kruskal–Wallis test and the ad hoc Dunn's test) were used. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications (Lio *et al.* 2019; Tsagaratou *et al.* 2017). No data points were excluded from the analysis, and appropriate animals/samples for each experiment were chosen randomly. Because our analysis required genotyping of the experimental mouse groups, we were not able to blind ourselves during data collection.

2.6 Figures

Figure 2.1. TET deficiency in mature B cells causes B cell lymphoma.
(A) Breeding strategy used to generate single Tet2 or Tet3-deficient mice (CD19 Tet2KO or CD19 Tet3KO) and double Tet2, Tet3-deficient mice (CD19 DKO). **(B)** Kaplan-Meir curves displaying overall survival of control Dfl (blue) and CD19 Cre (green) mice, CD19 Tet2KO (black), CD19 Tet3KO (grey) and CD19 DKO (red) mice. Y-axis denotes percent survival and X-axis shows time in weeks. Only the CD19 DKO mice develop B cell lymphoma in the 52-week time period shown. **(C)** Enlarged spleen and lymph nodes in 18-week-old CD19 DKO and control mice. **(D)** Flow cytometry plots of splenocytes from 18-week-old CD19Cre, Dfl and CD19 DKO mice. Numbers show frequencies of B cells among splenocytes. **(E)** Flow cytometry histograms showing γH2AX staining compared to isotype IgG controls in B cells from 8-week-old CD19cre (YFP+), Dfl and CD19 DKO (YFP+) mice. **(F)** Quantification of the median fluorescence intensity (MFI) of γH2AX in CD19Cre (YFP+), Dfl and CD19 DKO (YFP+) B cells from 4 independent experiments. **(G)** MA plot of RNA-sequencing data, displaying changes in gene expression in CD19 DKO B cells compared to Dfl B cells. The highlighted genes represent known G4 and R-loop binders. **(H)** Upregulated (red) and downregulated pathways (blue) in CD19 DKO compared to Dfl B cells. X-axis, Z-score. * Asterisks highlight pathways related to DNA structures and DNA damage. **(I)** Protein levels of RNASE H1 and selected G4-binding helicases and proteins in Dfl and CD19 DKO B cells, assessed by immunoblotting. The data is representative of at least 2 independent experiments. Statistical significance is calculated using the log-rank test **(B)** or one-way ANOVA **(F).** Error bars represent mean +/- standard deviation, *** p value ≤0.001, **** p value ≤0.0001.


A
CD19Cre Tet2 +/+ Tet3 +/+ Rosa LSL YFP (CD19Cre)
x
Tet2 fl/fl Tet3 fl/fl Rosa LSL YFP (Dfl)
CD19 Tet2KO
CD19 Tet3KO
CD19 DKO
B
Overall Survival
CD19Cre n=13
Dfl n=13
CD19 Tet2 KO n=6
CD19 Tet3 KO n=5
CD19 DKO n=18
****
Time (in weeks)
C
CD19Cre Dfl CD19 DKO
D
CD19Cre
Dfl
CD19 DKO
48.2
5.8
59.4
88.3
YFP
CD19-Pacific Blue
E
CD19 DKO
Dfl
CD19Cre
CD19 DKO IgG
Dfl IgG
CD19Cre IgG
Cell counts
γH2AX-AF647
F
***
***
CD19cre
Dfl
CD19 DKO
γH2AX MFI
CD19Cre Dfl CD19 DKO
G
CD19DKO>Dfl
n=1630
CD19DKO<Dfl
n=1048
Pif1
Atxn1
Recql4
Fancd2
Dnmt1
Ddx1
Ncl
Rnaseh2b
Blm
Log2 Fold Change
Average Expression
H
Immunoglobulin receptor binding
Antigen binding
MHC class II complex binding
Ras GTPase activity
Protein kinase activity
MHC protein complex binding
Signaling adaptor activity
Signaling complex scaffold activity
SH2 domain binding
Phosphotyrosine binding
* Nuclease activity
* DNA secondary structure binding
Translation initiation factor activity
RNP complex binding
* DNA helicase activity
Cyclin-dependent kinase activity
Cell adhesion molecule binding
* Damaged DNA binding
Unfolded protein binding
* Single-stranded DNA binding
Upregulated
Downregulated
Z-Score
I
Dfl
CD19 DKO
ATRX
250 kDa
FANCD2
130 kDa
BLM
130 kDa
55kDa
ACTIN
35 kDa
DNMT1
250 kDa
WRN
250 kDa
RNASEH1
25 kDa
55kDa
ACTIN
35 kDa

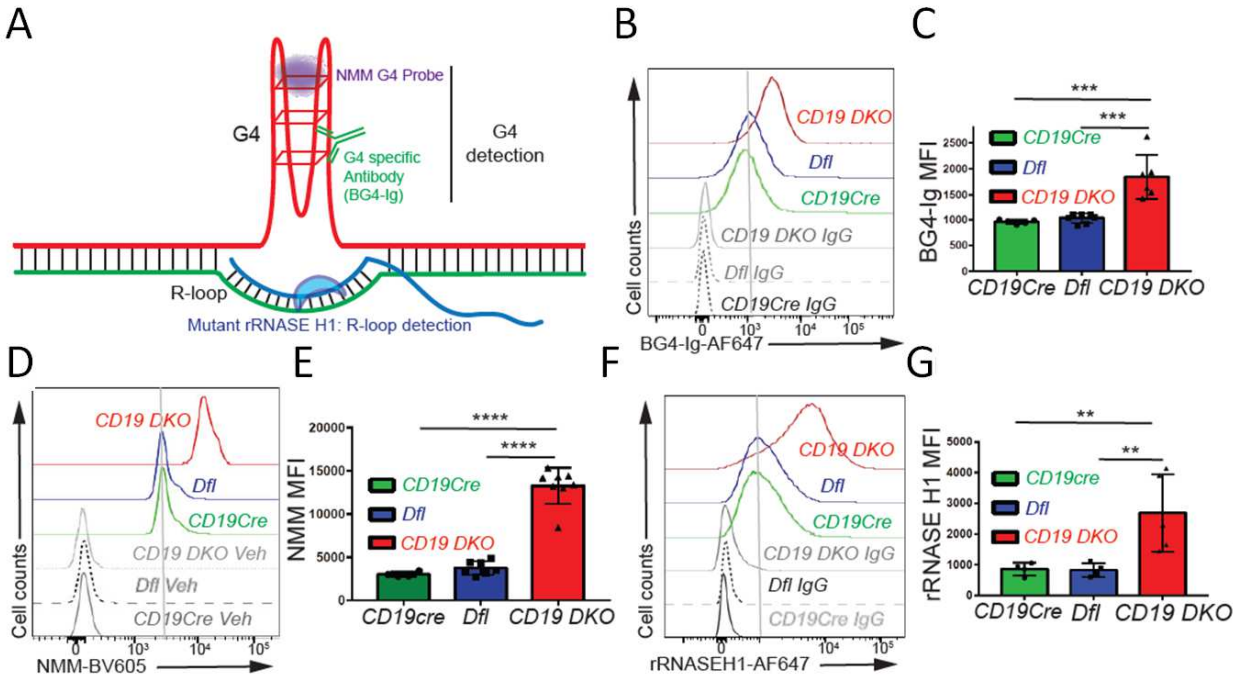


Figure 2.2. TET deficiency is associated with increased levels of G-quadruplexes and R-loops. **(A)** Diagrammatic representation of a G-quadruplex with an associated R-loop structure, illustrating the reagents used for detection of G-quadruplexes and R-loops. All experiments were performed in 8 to 10-week-old mice. **(B)** Flow cytometric detection of G-quadruplexes by staining of permeabilized cells with BG4-Ig antibody or isotype IgG controls in B cells from CD19cre (YFP+), Dfl and CD19 DKO (YFP+) mice. **(C)** Quantification of median fluorescence intensity (MFI) of BG4-Ig signal from CD19cre (YFP+), Dfl and CD19 DKO (YFP+) B cells from 4 independent experiments. **(D)** Flow cytometric detection of G-quadruplexes after incubation of cells with NMM or DMSO vehicle controls (Veh) in B cells from CD19Cre (YFP+), Dfl and CD19 DKO (YFP+) mice. **(E)** Quantification of median fluorescence intensity (MFI) of NMM signal from CD19Cre, Dfl and CD19 DKO B cells from 6 independent experiments. **(F)** Flow cytometric detection of R-loops using V5-epitope-tagged recombinant RNASE H1 (rRNASE H1) in B cells from CD19Cre (YFP+), Dfl and CD19 DKO (YFP+) mice. Samples stained with anti-V5 and anti-rabbit secondary antibodies were used as controls (IgG). **(G)** Quantification of median fluorescence intensity (MFI) of R-loops (rRNASE H1) signal from CD19cre (YFP+), Dfl and CD19 DKO (YFP+) B cells from 4 independent experiments. Statistical significance is calculated using one-way ANOVA in **(C)**, **(E)** and **(G)**. Error bars represent mean +/- standard deviation, ** p value ≤0.01, *** p value≤0.002, **** p value ≤0.0001.

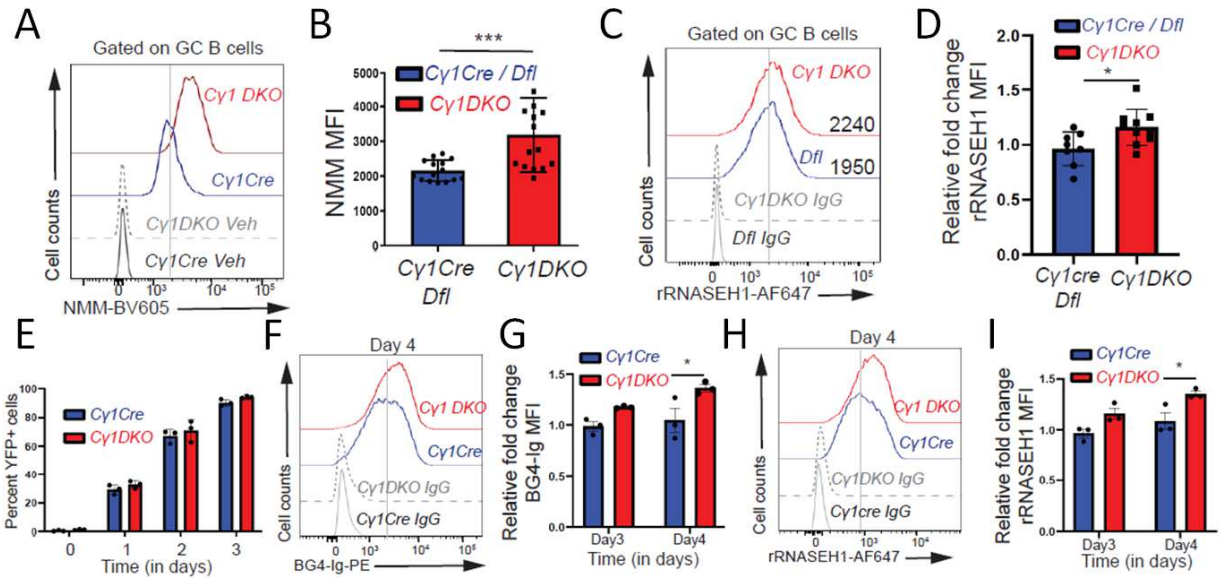


Figure 2.3. Acute TET deletion is associated with increased levels of G-quadruplexes and R-loops.

(A) Flow cytometric detection of G-quadruplexes with NMM versus DMSO vehicle staining controls (Veh) in GC B cells from Cγ1Cre or Dfl control and Cγ1 DKO mice. **(B)** Quantification of median fluorescence intensity (MFI) of NMM signal from Cγ1Cre or Dfl control and Cγ1 DKO GC B cells from 3 independent experiments. **(C)** Flow cytometric detection of R-loops (rRNASE H1) signal versus IgG staining controls (IgG) in GC B cells from Cγ1cre or Dfl control and Cγ1 DKO mice. **(D)** Fold change in median fluorescence intensity (MFI) signal of R-loops (rRNASE H1) Cγ1 DKO GC B cells relative to Cγ1Cre or Dfl control GC B cells from 2 independent experiments. **(E)** Percent YFP+ B cells from Cγ1Cre control and Cγ1 DKO mice cultured on 40LB media, X-axis denotes time (in days) of culture. **(F)** Flow cytometry detection of G-quadruplexes with BG4-Ig antibody or isotype IgG controls in Cγ1Cre and Cγ1 DKO B cells at day 4 of culture on 40LB feeder cells. (g) Quantification of median fluorescence intensity (MFI) of G-quadruplexes (BG4-Ig) signal from Cγ1cre and Cγ1 DKO B cells at day 3 and 4 of culture on 40LB cells from 3 biological replicates. **(H)** Flow cytometry detection of R-loops (rRNASE H1) signal versus anti-rabbit secondary antibody controls (IgG) in Cγ1cre and Cγ1 DKO B cells at day 4 of culture on 40LB cells. **(I)** Quantification of median fluorescence intensity (MFI) of R-loops (rRNASE H1) signal from Cγ1cre and Cγ1 DKO B cells at day 3 and 4 of culture on 40LB cells from 3 biological replicates. Statistical significance is calculated using two-tailed student t-test **(B)** and **(D)** and two-way ANOVA in **(G)** and **(I)**. Error bars represent mean +/- standard deviation, * p value ≤0.05, *** p value≤0.002.

Figure 2.4. TET-deficient B cells show a genome-wide increase in G-quadruplexes and R-loops and increased translocations to Ig switch regions.
(A) Heatmap showing enrichment of G-quadruplex (G4) structures (average of 2 replicates) and R-loops (average of 3 replicates) in *Dfl* and *CD19 DKO* B cells. Reads per million (RPM) values in 9722 regions with overlapping high intensity signals for both G-quadruplexes and R-loops are plotted in +/- 2 kb windows from the center of the region. **(B)**, **(C)** Box and whisker plots quantifying enrichment (RPKM, reads per kilobase per million) of **(B)** G-quadruplex structures (2 biological replicates) and **(C)** R-loops (3 biological replicates) in the 9722 regions from *Dfl* and *CD19 DKO* B cells. **(D)** Genome browser view showing data from *Dfl* (blue tracks) and *CD19 DKO* (red tracks) B cells. Green and blue bars at the bottom show the location of Sμ and G4 regions. The red arrow indicates the bait sequence located 5' of the switch μ (Sμ) region used to capture translocations by HTGTS. The zoomed in panel on the right shows the distribution of signals around the *IghM* region. **(E)- (F)** Quantification of **(E)** total number of translocations (hits), and **(F)** the number of genomic loci to which translocations occur in *CD19 DKO* and *Dfl* B cells from 2 biological replicates. **(G)** Circos plots to visually depict all translocations identified by HTGTS in *Dfl* and *CD19 DKO*. Colored lines connect the Sμ bait with the translocation partner regions. Color scale represents the number of translocations identified in a 10 kb window. Translocations from two *Dfl* and *CD19 DKO* replicates were concatenated for this representation. **(H)** Box and whisker plots quantifying G-quadruplexes (*left panel*) and R-loops (*right panel*) RPKM signal in translocation partner regions of translocations identified in *CD19 DKO* and *Dfl* B cells. Regions were extended +/-1 kb from the center of the junctions in the translocation partner regions from 2 biological replicates. **(I)** Quantification of the number of translocations (hits) overlapping G-quadruplexes and R-loops regions in *CD19 DKO* and *Dfl* B cells from 2 biological replicates. Statistical significance is calculated using Kruskal-Wallis test and the ad hoc Dunn's test in **(B)**, **(C)** and **(H)**. Boxes in box and whisker plots represent median (center) with 25th to 75 th percentile and whiskers represent maxima/minima. Error bars represent mean +/- standard deviation, ***** p value <0.000001.

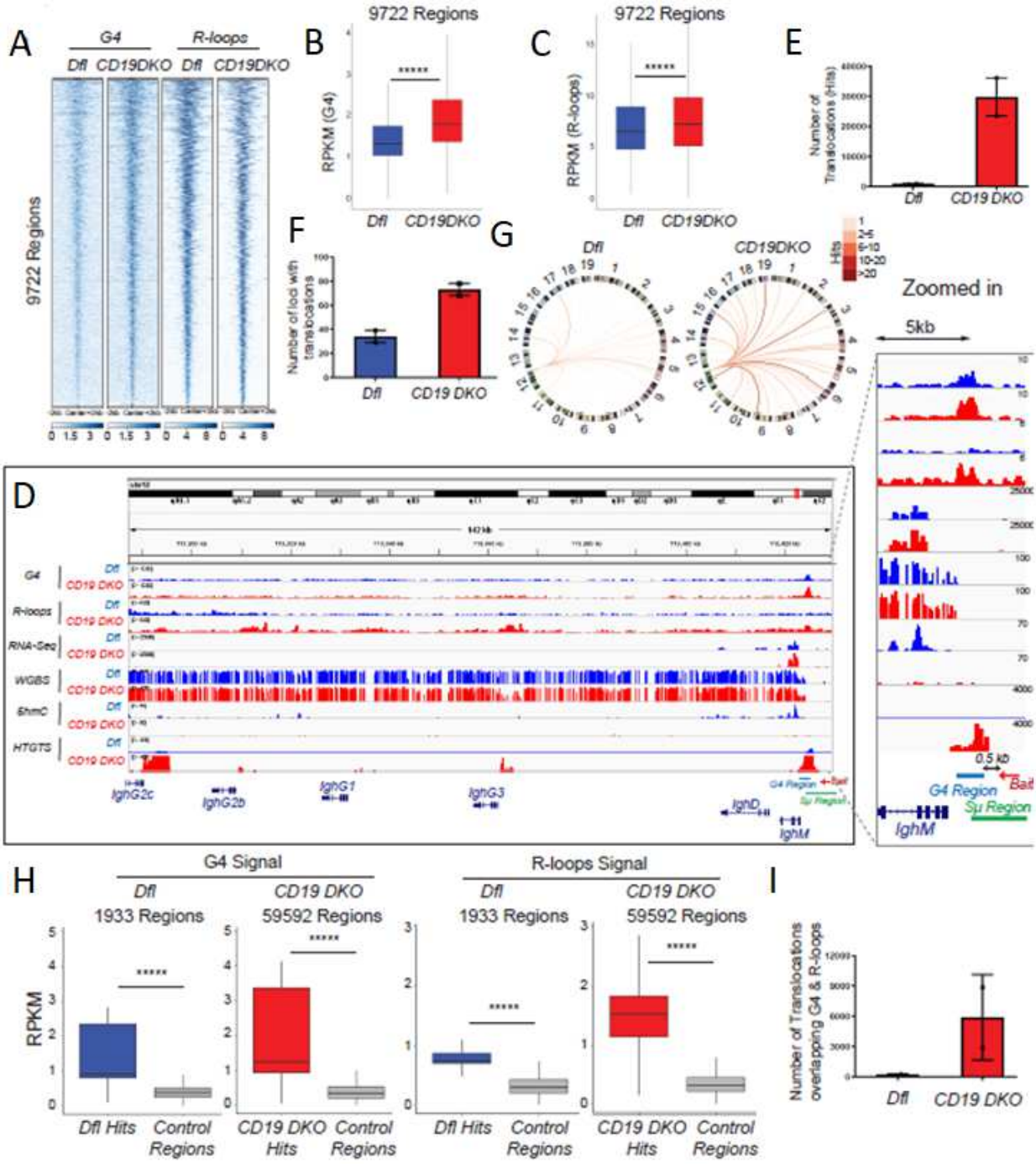
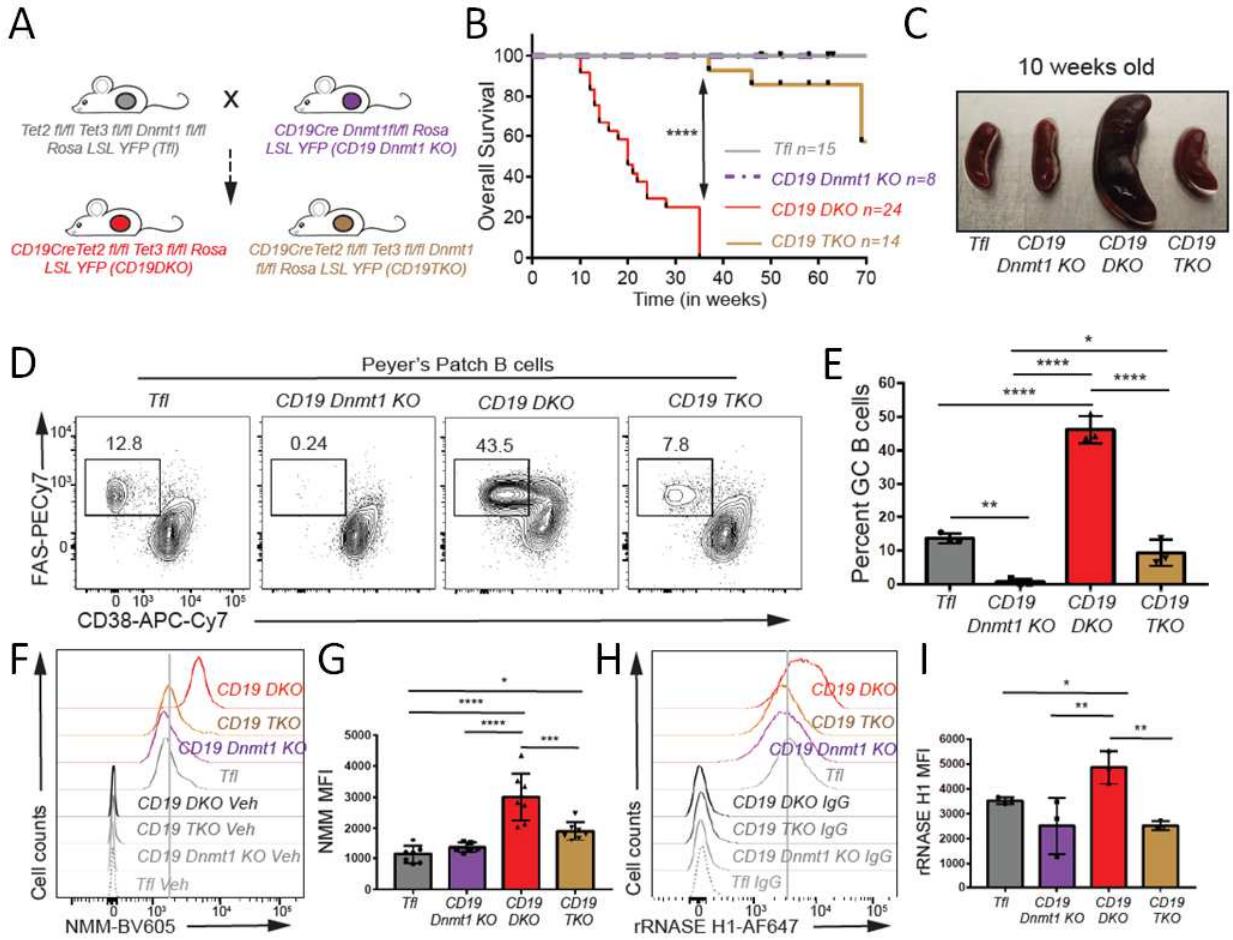

A
G4
R-loops
Dfl
CD19DKO
Dfl
CD19DKO
9722 Regions
B
9722 Regions
RPKM (G4)
*****
Dfl
CD19DKO
C
9722 Regions
RPKM (R-loops)
*****
Dfl
CD19DKO
E
Number of Translocations (Hits)
Dfl
CD19 DKO
F
Number of loci with translocations
Dfl
CD19 DKO
G
Dfl
CD19DKO
Hits
1
2-5
6-10
10-20
>20
Zoomed in
5kb
D
G4
R-loops
RNA-Seq
WGBS
5hmC
HTGTS
Dfl
CD19 DKO
IghG2c
IghG2b
IghG1
IghG3
IghD
IghM
0.5 kb
G4 Region
Bait
Sμ Region
IghM
H
G4 Signal
R-loops Signal
Dfl
CD19 DKO
1933 Regions
59592 Regions
RPKM
*****
Dfl Hits
Control Regions
CD19 DKO Hits
Control Regions
I
Number of Translocations overlapping G4 & R-loops
Dfl
CD19 DKO

Figure 2.5. DNMT1 deletion delays oncogenesis in TET-deficient mice.
(A) Breeding strategy used to generate the triple *Dnmt1*, *Tet2*, *Tet3* deficient mice (*CD19TKO*). **(B)** Kaplan-Meir curves displaying the overall survival of *Tfl* (*grey*), *CD19 Dnmt1 KO* (*purple*), *CD19 DKO* (*red*) and *CD19 TKO* (*brown*) mice. Y-axis denotes percent survival and X-axis shows time in weeks. **(C)** Spleens from 10-week-old *Tfl, CD19 Dnmt1, CD19 DKO* and *CD19 TKO* mice. **(D)** Representative flow cytometry data gated on Peyer's patch B cells from 10-week-old *Tfl, CD19 Dnmt1 KO, CD19 DKO and CD19 TKO* mice. Numbers represent frequency of GC B cells, identified as FAS^+ (Y-axis) and $CD38^-$ (X-axis). **(E)** Quantification of GC B cell frequency in Peyer's patches of *Tfl* (YFP^-, *grey*), *CD19 Dnmt1 KO* (YFP^+, *purple*), *CD19 DKO* (YFP^+, *red*) *and CD19 TKO* (YFP^+, *brown*) mice from 3 biological replicates and 2 independent experiments. **(F)** Flow cytometric detection of G-quadruplexes by staining of permeabilized cells with NMM or DMSO vehicle controls (Veh) in B cells from *Tfl* (YFP^-), *CD19 Dnmt1 KO* (YFP^+), *CD19 DKO* (YFP^+) and *CD19 TKO* (YFP^+) mice. **(G)** Quantification of median fluorescence intensity (MFI) of NMM signal from *Tfl*, *CD19 Dnmt1 KO*, *CD19 DKO* and *CD19 TKO* B cells from 7 biological replicates and 5 independent experiments. **(H)** Flow cytometric detection of R-loops using V5-epitope-tagged recombinant RNASE H1 (rRNASE H1) in B cells from *Tfl* (YFP^-), *CD19 Dnmt1 KO* (YFP^+), *CD19 DKO* (YFP^+) and *CD19 TKO* (YFP^+) mice. Samples stained with anti-V5 and anti-rabbit secondary antibodies were used as controls (IgG). **(I)** Quantification of median fluorescence intensity (MFI) of R-loops (rRNASE H1) signal from *Tfl, CD19 Dnmt1KO, CD19 TKO* and *CD19 DKO* B cells from 3 biological replicates and 2 independent experiments. Statistical significance is calculated using one-way ANOVA. Error bars represent mean +/- standard deviation, * p value ≤ 0.05, ** p value ≤ 0.01, *** p value ≤ 0.0005 **** p value < 0.0001.


A
Tet2 fl/fl Tet3 fl/fl Dnmt1 fl/fl Rosa LSL YFP (Tfl)
x
CD19Cre Dnmt1fl/fl Rosa LSL YFP (CD19 Dnmt1 KO)
CD19CreTet2 fl/fl Tet3 fl/fl Rosa LSL YFP (CD19DKO)
CD19CreTet2 fl/fl Tet3 fl/fl Dnmt1 fl/fl Rosa LSL YFP (CD19TKO)
B
Overall Survival
Time (in weeks)
****
Tfl n=15
CD19 Dnmt1 KO n=8
CD19 DKO n=24
CD19 TKO n=14
C
10 weeks old
Tfl
CD19 Dnmt1 KO
CD19 DKO
CD19 TKO
D
Peyer's Patch B cells
Tfl
CD19 Dnmt1 KO
CD19 DKO
CD19 TKO
12.8
0.24
43.5
7.8
FAS-PECy7
CD38-APC-Cy7
E
Percent GC B cells
F
CD19 DKO
CD19 TKO
CD19 Dnmt1 KO
Tfl
CD19 DKO Veh
CD19 TKO Veh
CD19 Dnmt1 KO Veh
Tfl Veh
Cell counts
NMM-BV605
G
NMM MFI
H
CD19 DKO IgG
CD19 TKO IgG
CD19 Dnmt1 KO IgG
Tfl IgG
rRNASE H1-AF647
I
rRNASE H1 MFI

2.7 Supplemental Figures

Figure S2.1. TET deficiency leads to development of mature B cell lymphoma.
(A) H&E staining in spleens from 8-week-old Dfl and CD19 DKO mice; the data is representative of 2 independent experiments. **(B)** Kaplan-Meir curves displaying lymphoma-free survival of CD19 DKO (red) and control Dfl (blue) and CD19 Cre (green) mice. Y-axis, percent mice without lymphoma, defined by > 3-fold increase in spleen weight and > 2-fold increase in cellularity. **(C) – (F)** Quantification of **(C)** cell numbers, **(D)** spleen weights, e) B cell numbers, f) percent of B cells in spleens of 9-week-old CD19 DKO (red) and control Dfl (blue) and, CD19 Cre (green) mice from at least 5 biological replicates and 4 independent experiments. **(G)** Percent of B cells in spleens of CD19 DKO (red) and control Dfl (blue) and CD19 Cre (green) mice at different ages from 3 independent experiments. **(H)** Quantification of cell numbers of CD4+ T, CD8+ T and CD4+ Tfh cells from spleens of 9-week-old CD19 DKO (red) and control Dfl (blue) mice from 3 biological replicates and 2 independent experiments. **(I)** Percent of activated CD4+ and, CD8+ T cells, and CD4 Tfh cells from spleens of 9-week-old CD19 DKO (red) and control Dfl (blue) mice from 3 biological replicates and 2 independent experiments. **(J)** and **(K)** Flow cytometry plots of **(J)** Activated CD62Llow (X-axis) and CD44+ (Y-axis) CD4+ and CD8+ T cells and **(K)** CD4+ Tfh cells PD1+ (Y-axis) CXCR5+ (X-axis) from spleen of 9-week-old CD19 DKO (red) and control Dfl (blue) mice. Numbers represent frequency of gated populations. **(L)** Scheme of adoptive transfer experiment. B cells from Dfl and CD19 DKO CD45.2 mice were transferred retro-orbitally into sub-lethally irradiated CD45.1 immunocompetent host mice. **(M)** Kaplan-Meir curves displaying overall survival of CD45.1 host mice transplanted with B cells from Dfl or CD19 DKO mice. X-axis denotes time in weeks after transplantation. **(N)** Enlarged spleens in CD45.1 host mice transplanted with CD19 DKO compared to Dfl B cells 8 weeks after adoptive transfer. **(O)** Representative flow cytometry data from the spleens of CD45.1 host mice 8 weeks after transplantation with Dfl and CD19 DKO B cells. Frequencies of the CD45.1+ and CD45.2+ cell populations are shown. Statistical significance is calculated using the log-rank test for **(B)** and **(M)**, one-way ANOVA for **(C)-(F)** and two-way ANOVA for **(G)-(I)**. Error bars represent mean +/- standard deviation, ** p value ≤0.01 *** p value ≤0.001, **** p value ≤0.0001.

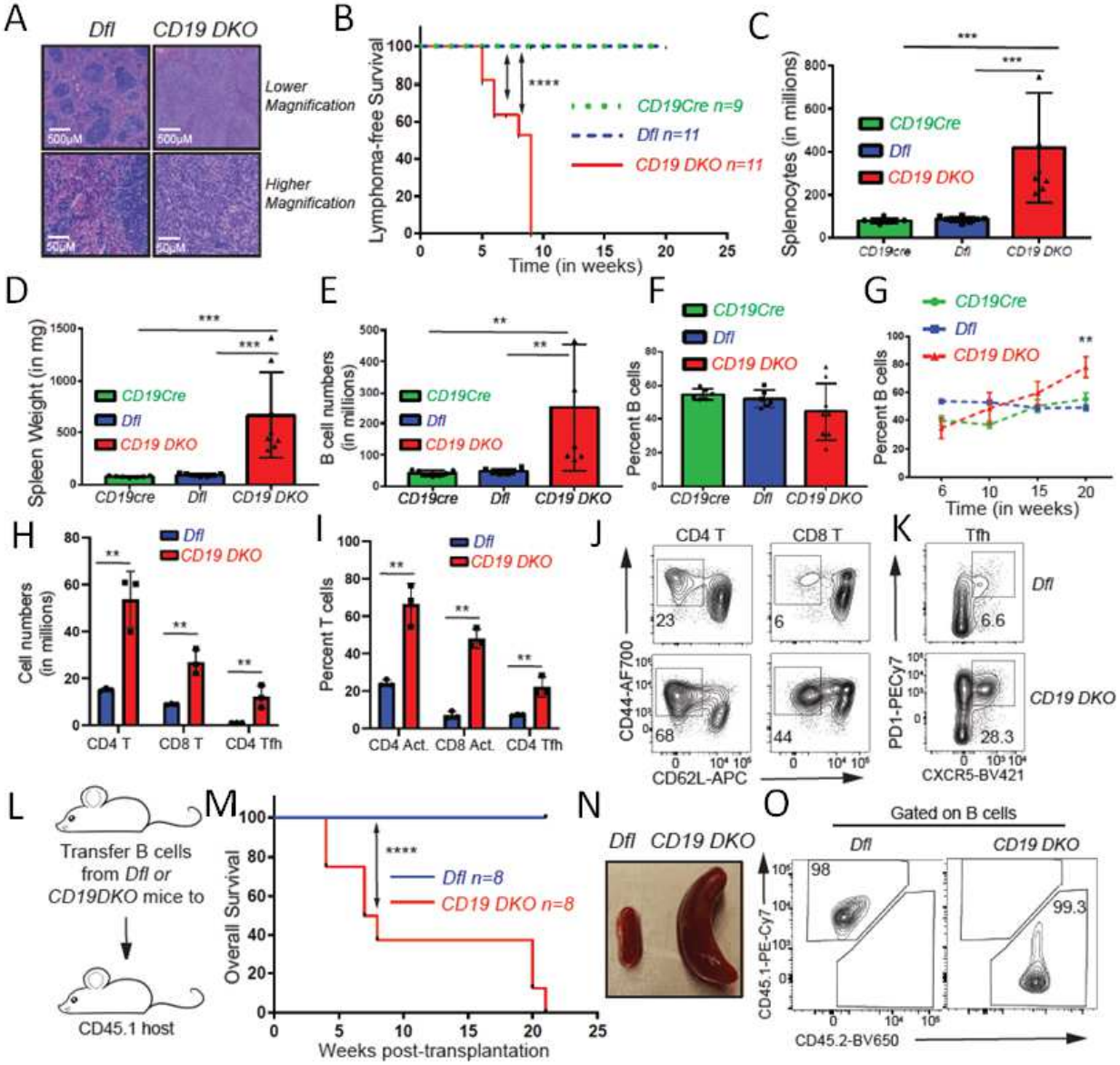

A
Dfl
CD19 DKO
Lower Magnification
Higher Magnification
500μM
50μM
B
Lymphoma-free Survival
CD19Cre n=9
Dfl n=11
CD19 DKO n=11
****
Time (in weeks)
C
Splenocytes (in millions)
CD19Cre
Dfl
CD19 DKO
***
D
Spleen Weight (in mg)
E
B cell numbers (in millions)
F
Percent B cells
G
Percent B cells
Time (in weeks)
H
Cell numbers (in millions)
CD4 T
CD8 T
CD4 Tfh
I
Percent T cells
CD4 Act.
CD8 Act.
J
CD4 T
CD8 T
CD44-AF700
CD62L-APC
23
6
68
44
K
Tfh
PD1-PECy7
CXCR5-BV421
6.6
28.3
L
Transfer B cells from Dfl or CD19DKO mice to
CD45.1 host
M
Overall Survival
Dfl n=8
CD19 DKO n=8
Weeks post-transplantation
N
O
Gated on B cells
CD45.1-PE-Cy7
CD45.2-BV650
98
99.3

Figure S2.2. Expanded B cells from CD19 DKO mice have a germinal center (GC) origin. **(A)**, **(B)** Gene set enrichment analysis (GSEA) plots showing enrichment for a GC B cell transcriptional signature in the transcriptional profile of CD19 DKO compared to Dfl B cells, using gene sets from **(A)** GC versus follicular B cells and **(B)** early GC versus late GC B cells. Y-axis denotes enrichment score. NES, Normalized enrichment score, FDR, False discovery rate. **(C)** Representative flow cytometry data gated on splenic B cells from 8-week-old Dfl, CD19Cre (YFP+) and CD19 DKO (YFP+) mice. Numbers in the rectangles represent frequencies of GC B cells, identified as EFNB1+ (Y-axis) and IgDlow (X-axis). **(D)** Representative flow cytometry data showing Ig isotype expression, gated on splenic B cells from 8-week-old Dfl, CD19Cre (YFP+) and CD19 DKO (YFP+) mice. Top, IgG1; middle, IgG2b; bottom, IgG3 X-axis shows expression of the default IgD isotype. Numbers represent frequencies of gated cell populations. **(E)** Immunoblot showing AID expression in Dfl and CD19 DKO B cells (2 replicate experiments). Actin is used as a loading control. **(F)** Relative fold-change in expression of μ and γ1 germline, Tet3, Irf4 and Myc transcripts measured by qRT-PCR in Dfl and CD19 DKO B cells from 2 biological replicates. **(G)** Histogram (left panel) and bar-graph (right panel) showing staining with BCL6 antibody compared to isotype IgG controls in B cell from 8-week-old CD19Cre, Dfl and CD19 DKO mice from 3 independent experiments. **(H)** and **(I)** Bar plots displaying the proportions of **(H)** IgVH and **(I)** Igκ clonotypes (rearranged variable gene segments) from Dfl (blue) and CD19 DKO (red) B cells, identified from BCR repertoire analysis of RNA-Seq data. Y-axis represents the proportion of each clonotype. Each individual IgVH and Igκ clonotype is displayed using a different color in the bar plots. Numbers at the bottom represent the number of clonotypes identified in two independent replicates of Dfl (blue) and CD19 DKO (red) B cells. **(J)** Representative flow cytometry data gated on Peyer's patch B cells from 8-week-old Dfl and CD19 DKO mice. Numbers represent frequency of GC B cells, identified as FAS+ (Y-axis) and CD38- (X-axis). **(K)** Quantification of GC B cell frequency in Peyer's patches of Dfl (blue) and CD19 DKO (red) mice from 3 independent experiments. Statistical significance is calculated using two-tailed student t-test in **(F)**, **(K)** and one-way ANOVA in **(G)**. Error bars represent mean +/- standard deviation in **(F), (G)** and **(K)**. ** p value ≤ 0.01.

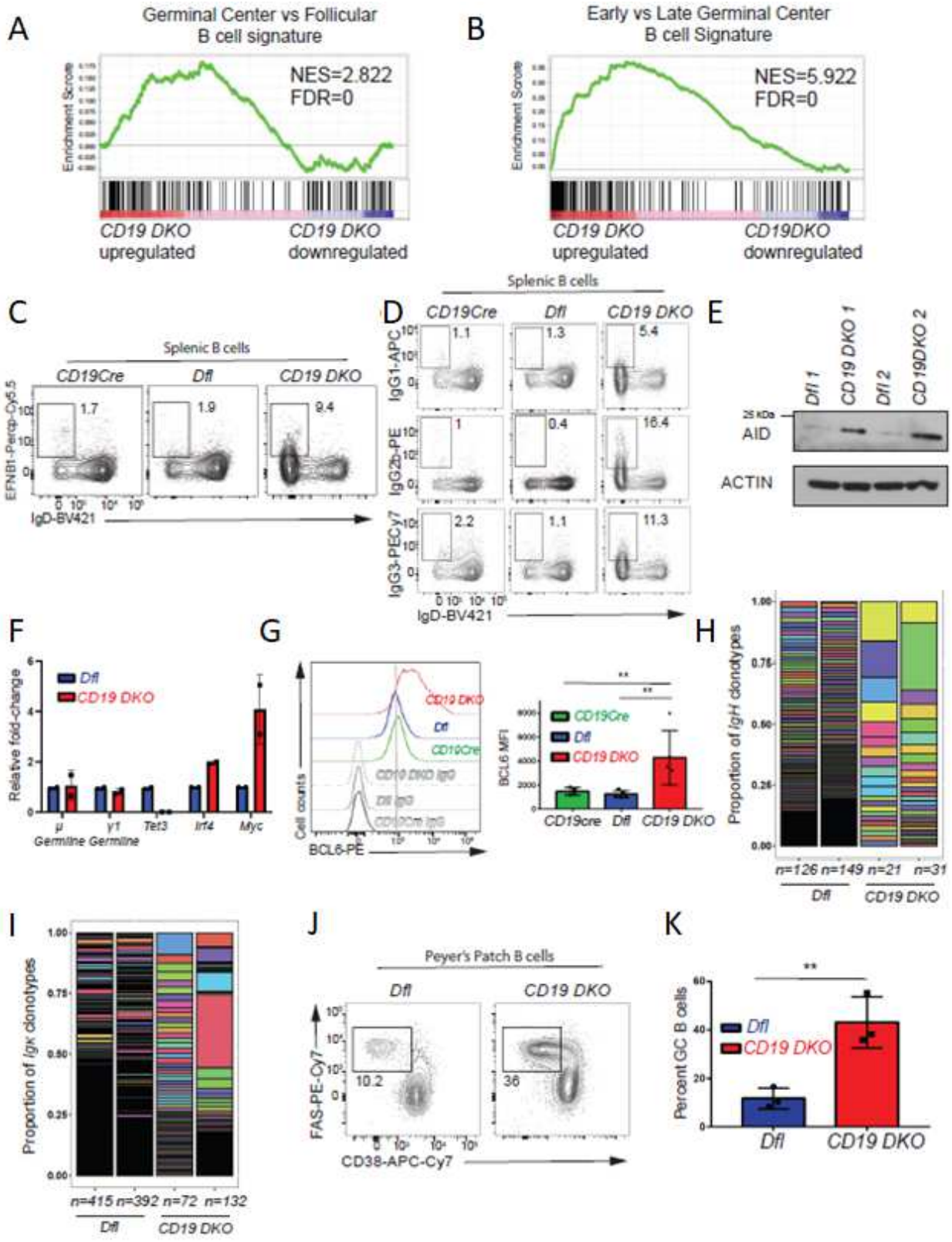

A
Germinal Center vs Follicular B cell signature
NES=2.822
FDR=0
Enrichment Score
CD19 DKO upregulated
CD19 DKO downregulated
B
Early vs Late Germinal Center B cell Signature
NES=5.922
FDR=0
Enrichment Score
CD19 DKO upregulated
CD19DKO downregulated
C
Splenic B cells
CD19Cre
Dfl
CD19 DKO
EFNB1-Percp-Cy5.5
IgD-BV421
1.7
1.9
9.4
D
Splenic B cells
CD19Cre
Dfl
CD19 DKO
IgG1-APC
IgG2b-PE
IgG3-PECy7
IgD-BV421
1.1
1.3
5.4
1
0.4
16.4
2.2
1.1
11.3
E
Dfl 1
CD19 DKO 1
Dfl 2
CD19DKO 2
35 kDa
AID
ACTIN
F
Dfl
CD19 DKO
Relative fold-change
μ Germline
γ1 Germline
Tet3
Irf4
Myc
G
CD19 DKO
Dfl
CD19Cre
CD19 DKO IgG
Dfl IgG
CD19Cre IgG
Cell counts
BCL6-PE
BCL6 MFI
CD19Cre
Dfl
CD19 DKO
H
Proportion of IgH clonotypes
n=126 n=149 n=21 n=31
Dfl
CD19 DKO
I
Proportion of Igκ clonotypes
n=415 n=392 n=72 n=132
Dfl
CD19 DKO
J
Peyer's Patch B cells
Dfl
CD19 DKO
FAS-PE-Cy7
CD38-APC-Cy7
10.2
36
K
Percent GC B cells
Dfl
CD19 DKO

Figure S2.3. TET deficiency is associated with increased levels of G-quadruplexes and R-loops.
(A) Flow cytometric detection of G-quadruplexes with BG4-Ig antibody or isotype IgG controls in primary B cells stimulated with 25 μg/ml LPS for 48 hours and treated with 10 μM pyridostatin (PDS, G4 ligand) for 24 hours. Numbers represent median fluorescence intensity. **(B)** Quantification of median fluorescence intensity (MFI) of BG4-Ig signal from primary B cells treated with (red) or without (blue) PDS. Lines connect paired samples treated with or without PDS from 3 independent experiments. **(C)** Fluorescence emission spectrum of NMM in the presence of a G4-forming oligonucleotide (oligo) from the human c-Kit gene promoter or a control oligo in which guanines in G4-forming regions (G-tracts) were mutated. **(D)** Fluorescence enhancement over background (no oligos) for NMM at 610 nm in presence of known G4-forming oligos (Kit1, Kit2, Spb1) from the c-Kit gene locus or the telomeric repeat (Telo) or their respective mutated versions. **(E)** G-quadruplex levels assessed by NMM or DMSO vehicle control (Veh) staining in untreated CH12 B cells or cells treated with 5 μM pyridostatin (PDS, G4 ligand) for 24 hours. Numbers represent median fluorescence intensity from 3 independent experiments. **(F)** Flow cytometric detection of R-loops using V5-epitope-tagged recombinant RNASE H1 (rRNASE H1) in CH12 cells with or without RNASE H enzyme digestion during staining. Numbers represent median fluorescence intensity. **(G)** Quantification of median fluorescence intensity (MFI) of R-loops (rRNASE H1) signal from CH12 cells with (red) or without (blue) RNASE H enzyme digestion. Lines connect paired samples with or without RNASE H digestion from 3 independent experiments. **(H) – (M)** Representative images **(H, J, L)** and quantification of mean fluorescence signal (i, k, m) of CD19cre and CD19 DKO YFP+ B cells stained with DAPI or propidium iodide and CD19, BG4-Ig **(H, I)**, NMM **(J, K)** and rRNASE H1 **(L, M)** or respective controls using the AMNIS imagestream. Data are from two independent experiments. Statistical significance is calculated using paired student t-test in **(B)** and **(G)**, two-tailed student t-test **(I)**, **(K)** and **(M)**. Error bars represent mean +/- standard deviation, ** p value ≤0.01, **** p value ≤0.0001.

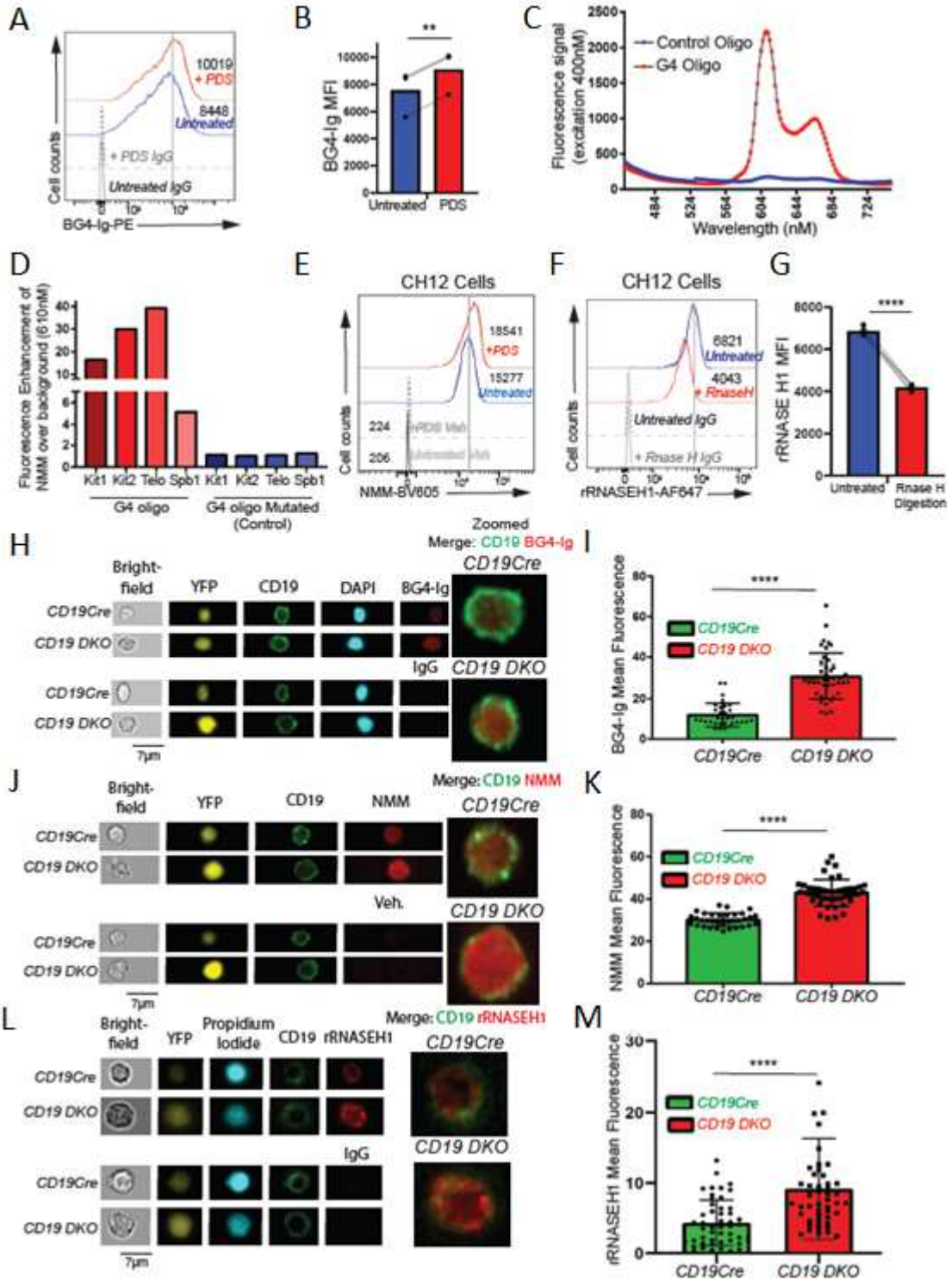

A
10019
+ PDS
8448
Untreated
+ PDS IgG
Untreated IgG
Cell counts
BG4-Ig-PE
B
BG4-Ig MFI
Untreated
PDS
C
Control Oligo
G4 Oligo
Fluorescence signal (excitation 400nM)
Wavelength (nM)
D
Fluorescence Enhancement of NMM over background (610nM)
Kit1 Kit2 Telo Spb1 Kit1 Kit2 Telo Spb1
G4 oligo
G4 oligo Mutated (Control)
E
CH12 Cells
18541
+PDS
15277
Untreated
224
206
Cell counts
NMM-BV605
F
CH12 Cells
6821
Untreated
4043
+ RnaseH
Untreated IgG
+ Rnase H IgG
Cell counts
rRNASEH1-AF647
G
rRNASE H1 MFI
Untreated
Rnase H Digestion
H
Zoomed
Merge: CD19 BG4-Ig
Bright-field
YFP
CD19
DAPI
BG4-Ig
CD19Cre
CD19 DKO
IgG
7μm
I
BG4-Ig Mean Fluorescence
CD19Cre
CD19 DKO
J
Merge: CD19 NMM
Bright-field
YFP
CD19
NMM
Veh.
K
NMM Mean Fluorescence
L
Merge: CD19 rRNASEH1
Bright-field
YFP
Propidium Iodide
CD19
rRNASEH1
IgG
M
rRNASEH1 Mean Fluorescence

Figure S2.4. TET deficiency in multiple primary cell types is associated with increased DNA G-quadruplex structures.
(A) Representative flow cytometry data gated on splenic B cells from 8-week-old *Cγ1Cre*, *Dfl* and *Cγ1 DKO* mice 12 days after immunization with SRBCs. GC B cells are identified as FAS^+ (Y-axis) and $CD38^-$ (X-axis). Numbers represent frequency of GC B cells. **(B)**, **(C)** Quantification of **(B)** GC B cell frequencies and **(C)** absolute numbers of splenocytes from 8-week-old *Cγ1Cre*, *Dfl* and *Cγ1 DKO* mice 12 days after immunization with SRBCs from 3 independent experiments. **(D)** Experimental design. *ERT2Cre DKO* or control *Dfl* mice were injected for 5 consecutive days with tamoxifen to induce Cre expression and *TET* deletion, then rested for 2 days. Splenic B cells were activated for 72 hours *in vitro* with LPS and IL-4 in the presence of 4-hydroxytamoxifen (4-OHT). **(E)** G-quadruplex levels in naïve (*left panel*) and activated (right *panel*) B cells from tamoxifen-treated *ERT2Cre DKO (YFP^+)* or control *Dfl* mice. Numbers represent median fluorescence intensity. **(F)** Quantification of median fluorescence intensity (MFI) of NMM signal from naïve and activated B cells from *ERT2Cre DKO (YFP^+)* or control *Dfl* mice from 3 independent experiments. **(G)**, **(I)** G-quadruplex levels assessed by NMM or DMSO vehicle staining (Con) in **(G)** transferred $CD45.2^+$ myeloid cells from *ERT2Cre TKO (YFP^+)* or control *Tfl* mice, and **(I)** transferred $CD45.2^+$ T cells from *CD4Cre DKO (YFP^+)* or control *Dfl* mice. **(H)**, **(J)** Quantification of median fluorescence intensity (MFI) of NMM signal in **(H)** transferred $CD45.2^+$ myeloid cells from *ERT2Cre TKO (YFP^+)* or control *Tfl* mice from 2 biological replicates and **(J)** transferred $CD45.2^+$ T cells from *CD4Cre DKO (YFP+)* or control *Dfl* mice from 3 biological replicates. Statistical significance is calculated using one-way ANOVA in **(C)**, two-way **(F)** and two-tailed student t-test in **(H)** and **(J)**. Error bars represent mean +/- standard deviation, * p value ≤ 0.01, ** p value ≤ 0.005.

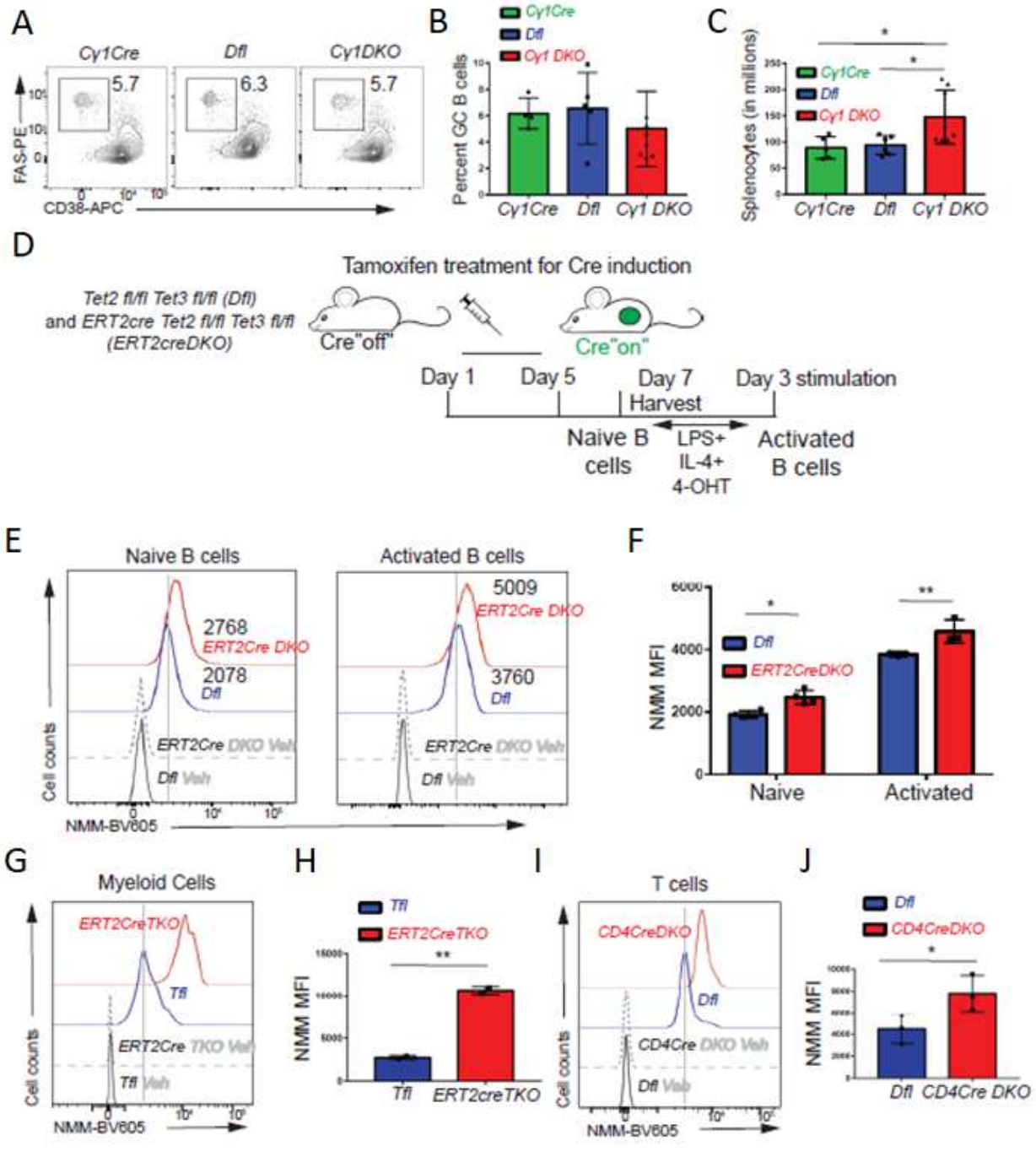

A
Cγ1Cre
Dfl
Cγ1DKO
5.7
6.3
5.7
FAS-PE
CD38-APC
B
Cγ1Cre
Dfl
Cγ1 DKO
Percent GC B cells
C
Splenocytes (in millions)
D
Tamoxifen treatment for Cre induction
Tet2 fl/fl Tet3 fl/fl (Dfl)
and ERT2cre Tet2 fl/fl Tet3 fl/fl
(ERT2creDKO)
Cre"off"
Cre"on"
Day 1
Day 5
Day 7
Harvest
Day 3 stimulation
Naive B cells
LPS+ IL-4+ 4-OHT
Activated B cells
E
Naive B cells
Activated B cells
2768
ERT2Cre DKO
2078
Dfl
5009
ERT2Cre DKO
3760
Dfl
ERT2Cre DKO Veh
Dfl Veh
Cell counts
NMM-BV605
F
NMM MFI
Dfl
ERT2CreDKO
Naive
Activated
G
Myeloid Cells
ERT2CreTKO
Tfl
ERT2Cre TKO Veh
Tfl Veh
H
Tfl
ERT2CreTKO
I
T cells
CD4CreDKO
Dfl
CD4Cre DKO Veh
J
CD4CreDKO
Dfl CD4Cre DKO

Figure S2.5. Increased apoptosis and DNA DSBs in TET-deficient B lymphoma cells depleted of enzymes that resolve G-quadruplexes and R-loops.
(A) Experimental design. Primary B cells from *Dfl* and *CD19 DKO* B cells were nucleofected with Cas9 RNPs loaded with sgRNAs against *Rnase H1* or the known G4-binding helicases *Atrx, Blm* and *Fancd2*, then stimulated with 10 μg/ml LPS for 48 hours before assessing the frequency of apoptotic cells by flow cytometry for cleaved Caspase 3. **(B)** Representative immunoblots showing decreased protein levels of ATRX, BLM, FANCD2 and RNASE H1 in *CD19 DKO* B cells nucleofected 48 hours earlier with the corresponding or *CD4* Cas9 RNPs (*Ctrl*). The data is representative of at least 2 independent experiments. **(C)** Representative flow cytometry plots quantifying percent apoptotic cells in *Dfl* and *CD19 DKO* B cells nucleofected with Cas9 RNPs. Y-axis, staining for cleaved Caspase 3; X-axis, forward scatter (FSC). **(D)** Quantification of apoptosis, measured as percent of cells showing staining for cleaved Caspase 3, in cells nucleofected with Cas9 RNPs to *Atrx, Blm*, *Fancd2, Rnase H* from 3 biological replicates. **(E)** Quantification of G-quadruplexes as NMM median fluorescence intensity (MFI) in *Dfl* and *CD19 DKO* B cells 48 hours after nucleofection with Cas9 RNPs. The signal is normalized to the NMM MFI of the same biological sample nucleofected with *CD4* Cas9 RNPs (*Ctrl*) from 3 biological replicates. **(F)** Quantification of DNA DSBs, assessed by γH2AX median fluorescence intensity (MFI) in *Dfl* and *CD19 DKO* B cells 48 hours after nucleofection with the indicated Cas9 RNPs from 3 biological replicates. The signal is normalized to the γH2AX MFI of the same biological sample nucleofected with control Cas9 RNP loaded with sgRNA against *CD4 (Ctrl)*. **(G)** Experimental design. *Dfl* and *CD19 DKO* B cells were treated for 2 days with the G-quadruplex stabilizing compound pyridostatin (PDS) prior to activation for 48 hours with LPS. **(H)** Quantification of apoptosis, measured as percent of cells showing staining for cleaved caspase 3, in *Dfl* and *CD19 DKO* B cells cultured without (untreated) or with 10 μM PDS from 5 biological replicates and 3 independent experiments. **(I)** Representative flow-cytometry plots quantifying percent apoptotic cells in *Dfl* and *CD19 DKO* B cells with or without PDS treatment. Y-axis, staining for cleaved caspase 3; X-axis, forward scatter (FSC). Statistical significance is calculated using two-way ANOVA. Error bars represent mean +/- standard error, * p value ≤0.05, ** p value ≤0.01, *** p value ≤0.0005 **** p value <0.0001.

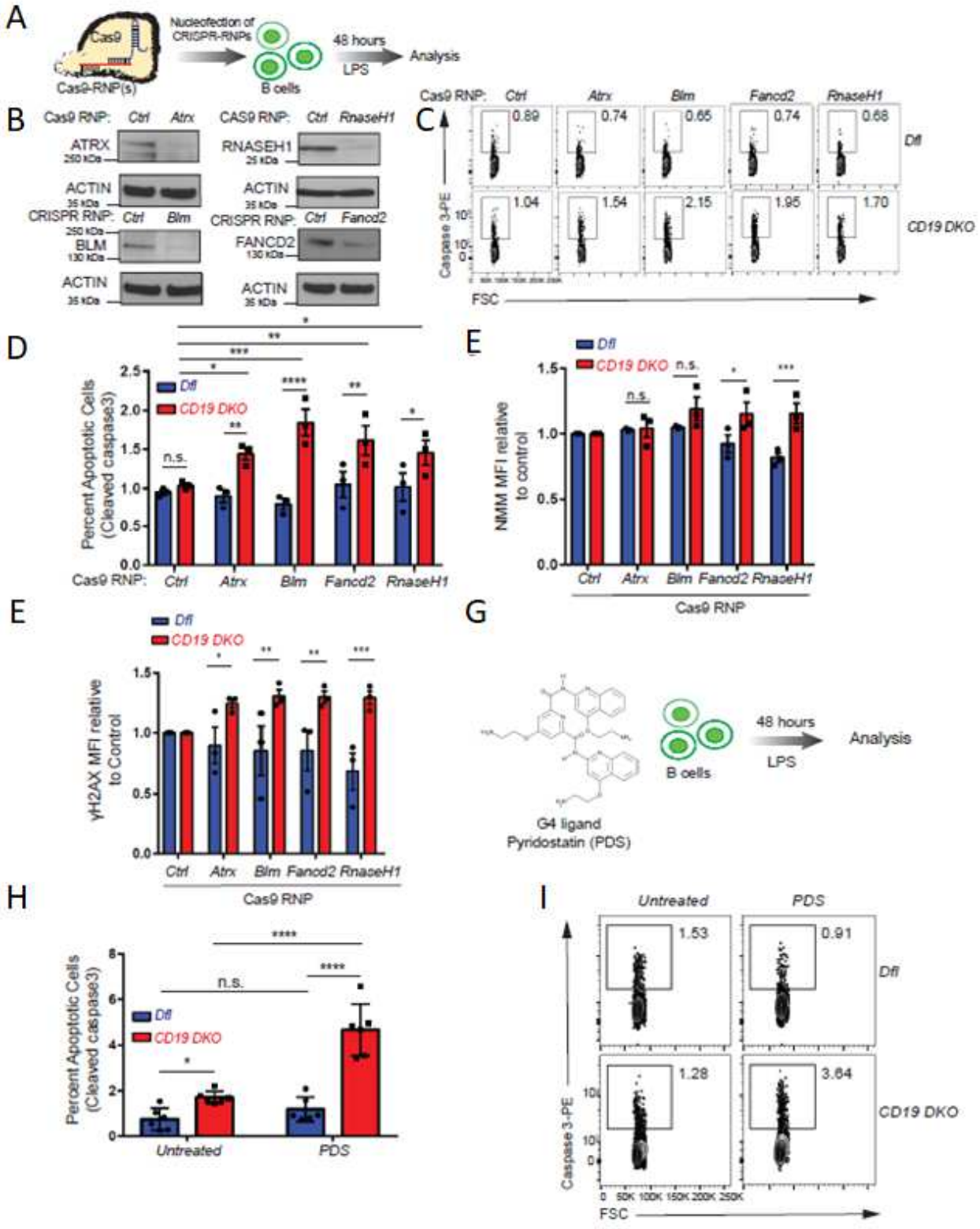

A
Cas9
Cas9-RNP(s)
Nucleofection of CRISPR-RNPs
B cells
48 hours
LPS
Analysis
B
Cas9 RNP: Ctrl Atrx
ATRX
ACTIN
CRISPR RNP: Ctrl Blm
BLM
ACTIN
CAS9 RNP: Ctrl RnaseH1
RNASEH1
ACTIN
CRISPR RNP: Ctrl Fancd2
FANCD2
ACTIN
C
Cas9 RNP: Ctrl Atrx Blm Fancd2 RnaseH1
0.89 0.74 0.65 0.74 0.68
Dfl
1.04 1.54 2.15 1.95 1.70
CD19 DKO
Caspase 3-PE
FSC
D
Percent Apoptotic Cells (Cleaved caspase3)
Dfl
CD19 DKO
n.s.
Cas9 RNP: Ctrl Atrx Blm Fancd2 RnaseH1
E
NMM MFI relative to control
Dfl
CD19 DKO
n.s.
Ctrl Atrx Blm Fancd2 RnaseH1
Cas9 RNP
E
γH2AX MFI relative to Control
Dfl
CD19 DKO
Ctrl Atrx Blm Fancd2 RnaseH1
Cas9 RNP
G
G4 ligand
Pyridostatin (PDS)
B cells
48 hours
LPS
Analysis
H
Percent Apoptotic Cells (Cleaved caspase3)
Dfl
CD19 DKO
n.s.
Untreated
PDS
I
Untreated
PDS
1.53 0.91
Dfl
1.28 3.64
CD19 DKO
Caspase 3-PE
FSC

Figure S2.6. TET deficiency is associated with genome-wide accumulation of G-quadruplexes and R-loops.
(A) Genome annotations of regions enriched for G-quadruplexes (G4) and R-loops (*right bar*) compared to their representation in the mouse genome (mm10) (*left bar*). **(B)** Relative representation of different classes of motifs predicted to form G-quadruplexes (pG4) in control regions selected randomly from the genome (*left bar*) and regions enriched for G-quadruplexes and R-loops (*right bar*). **(C)** Heat maps showing enrichment (RPM) for G-quadruplexes and R-loops in *CD19 DKO* and control B cells. The signal is plotted in a +/- 2 kb window from the center of the regions ordered based on decreasing intensity from top to bottom in the entire 4 kb window. R-loop signal is plotted after background subtraction of MNase-alone control. **(D)** Profile histograms showing the signals for G-quadruplexes (G4) (RPM, reads per million), R-loops (RPM), WGBS (percent of 5mC+5hmC/unmodified C) and 5hmC (RPM). The 9722 regions enriched for both G-quadruplexes and R-loops are divided into two categories – 6212 regions overlapping promoters (*left panels*) and 3510 regions not at promoters (*right panels*). Dashed grey lines indicate the center of the region and the 1 kb boundaries located on either side of the center. Blue and red lines show data from *Dfl* and CD19 DKO B cells, respectively. Asterisks represent statistical significance calculated by comparing the signals between *Dfl* and *CD19 DKO* B cells, either within the G-quadruplex and R-loop forming regions, the region to +/-1kb window or +/-1kb to 2kb window for respective datasets. **(E)** Profile histograms showing the 5hmC signal in *Dfl* (blue) and *CD19 DKO* (red) B cells in 23,467 regions identified as enriched for 5hmC signal. **(F)** Violin plots quantifying enrichment (RPKM) of 5hmC signal in *Dfl* B cells in the +/-1 kb from G-quadruplex and R-loop forming regions at promoters, non-promoter regions and control regions randomly located in euchromatin (Hi-C A genomic compartment) from 2 biological replicates. **(G)** Pie chart showing the differentially methylated regions (DMRs) in *CD19 DKO* compared to control *Dfl* B cells. Of a total of 6948 DMRs identified by WGBS, 1014 (15%) showed reduced DNA methylation (hypomethylation) and 5934 (85%) showed increased DNA methylation (hypermethylation). **(H)** Box and whisker plots quantifying percent of 5mC+5hmC/unmodified C (from WGBS) at and near the G4 and R-loop forming regions overlapping promoters and regions not overlapping promoters in *Dfl* (blue) and *CD19 DKO* (red) B cells from 2 biological replicates. The signal is plotted in three windows; window 1, within the G4 and R-loop regions; window 2, from region to +/- 1kb on either side and; window 3, +/-1kb to +/-2kb on either side. **(I)** Percent of 5mC+5hmC/unmodified C (from WGBS) in random genomic regions of *Dfl* (blue) and *CD19 DKO* (red) B cells. **(J)** Heatmaps of enrichment (RPM) for G-quadruplexes (*left*) and R-loops (*right*) in *Dfl* and *CD19 DKO* B cells, ordered in descending order of gene expression of *CD19 DKO* B cells. **(K)** MA plot (*left*) showing differentially expressed genes (DEGs) in *CD19 DKO* B cells. *Red dots*, upregulated DEGs; *blue dots*, downregulated DEGs; black dots, DEGs with G-quadruplexes and R-loops at their promoters (+/-1kb of TSS); *yellow dots*, non-DEGs with G-quadruplexes and R-loops at their promoters; *grey dots*, non-DEGs without G-quadruplexes and R-loops. The pie-chart (*right*) shows the percent of DEGs with (*green*) and without (*brown*) G-quadruplexes and R-loops at their promoters. Asterisks indicate statistical significance. Statistical significance is calculated using Kruskal-Wallis test and the ad hoc Dunn's test in **(D)**, **(F)** and **(H)**, Chi-square test in **(J)**. Boxes in box and whisker plots represent median (center) with 25th to 75 th percentile and whiskers represent maxima/minima. **** p value ≤0.0001 and ***** p value <0.000001.

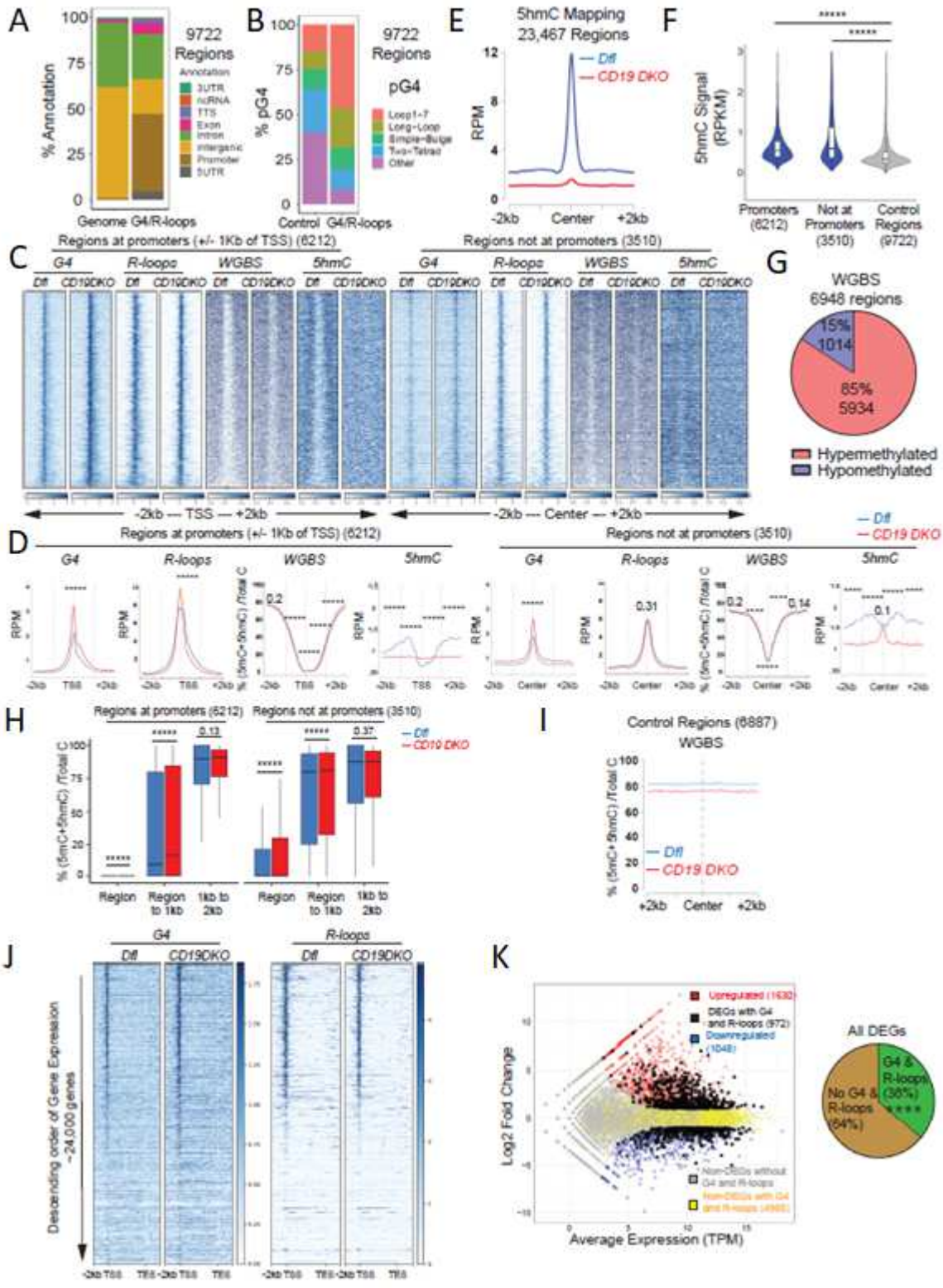

A
% Annotation
9722 Regions
Annotation
3UTR
ncRNA
TTS
Exon
Intron
Intergenic
Promoter
5UTR
Genome G4/R-loops
B
% pG4
9722 Regions
pG4
Loop1-7
Long-Loop
Simple-Bulge
Two-Tetrad
Other
Control G4/R-loops
E
5hmC Mapping
23,467 Regions
Dif
CD19 DKO
RPM
-2kb Center +2kb
F
5hmC Signal (RPKM)
Promoters (6212)
Not at Promoters (3510)
Control Regions (9722)
C
Regions at promoters (+/- 1Kb of TSS) (6212)
Regions not at promoters (3510)
G4
R-loops
WGBS
5hmC
Dif
CD19DKO
-2kb — TSS — +2kb
-2kb — Center — +2kb
G
WGBS
6948 regions
15%
1014
85%
5934
Hypermethylated
Hypomethylated
D
Regions at promoters (+/- 1Kb of TSS) (6212)
Regions not at promoters (3510)
Dif
CD19 DKO
RPM
% (5mC+5hmC) /Total C
0.2
0.31
0.14
0.1
H
Regions at promoters (6212)
Regions not at promoters (3510)
Dif
CD19 DKO
0.13
0.37
% (5mC+5hmC) /Total C
Region
Region to 1kb
1kb to 2kb
I
Control Regions (6887)
WGBS
% (5mC+5hmC) /Total C
Dif
CD19 DKO
+2kb Center +2kb
J
G4
R-loops
Dif
CD19DKO
Descending order of Gene Expression
~24,000 genes
-2kb TSS
TES -2kb TSS
TES
K
Upregulated (1632)
DEGs with G4 and R-loops (972)
Downregulated (1048)
Log2 Fold Change
Average Expression (TPM)
All DEGs
G4 & R-loops (36%)
No G4 & R-loops (64%)

Figure 2.7. Genome-wide analysis of TET deficient B cells.
(A)-(C) Genome browser tracks showing the distribution of G-quadruplexes (G4), R-loops, RNA-Seq, WGBS, and 5hmC datasets for *Dfl* (blue tracks) and *CD19 DKO* (red tracks) B cells. Grey boxes indicate regions of interest. The blue arrows at the bottom show the location of the TSS and the direction of transcription. **(D)** Circos plots to visually depict all translocations identified by HTGTS in *Dfl* and *CD19 DKO* replicates. Colored lines connect the Sμ bait with the translocation partner regions. Color scale represents the number of translocation partner regions identified in a 10 kb window. Translocations from two *Dfl* and *CD19 DKO* replicates are represented separately. **(E)** Relative representation of different classes of motifs predicted to form G-quadruplexes (pG4) in control genomic regions selected randomly (*right bar*) and +/-300 bp from the center of translocation partner junctions identified from translocations in *Dfl* and *CD19 DKO* B cells (*right bar*). The numbers (n) of *Dfl* and *CD19* DKO hits, and control regions are included in the plots. **(F)** Density of AID motifs (WRCY/RGYW) in +/-300bp from the center of translocation partner junctions identified from translocations in *Dfl* and *CD19 DKO* B cells compared to control random regions in euchromatin (Hi-C A compartment) from 2 biological replicates. Statistical significance is calculated using the Wilcoxon signed rank test **(F)**, ***** p value <0.000001.

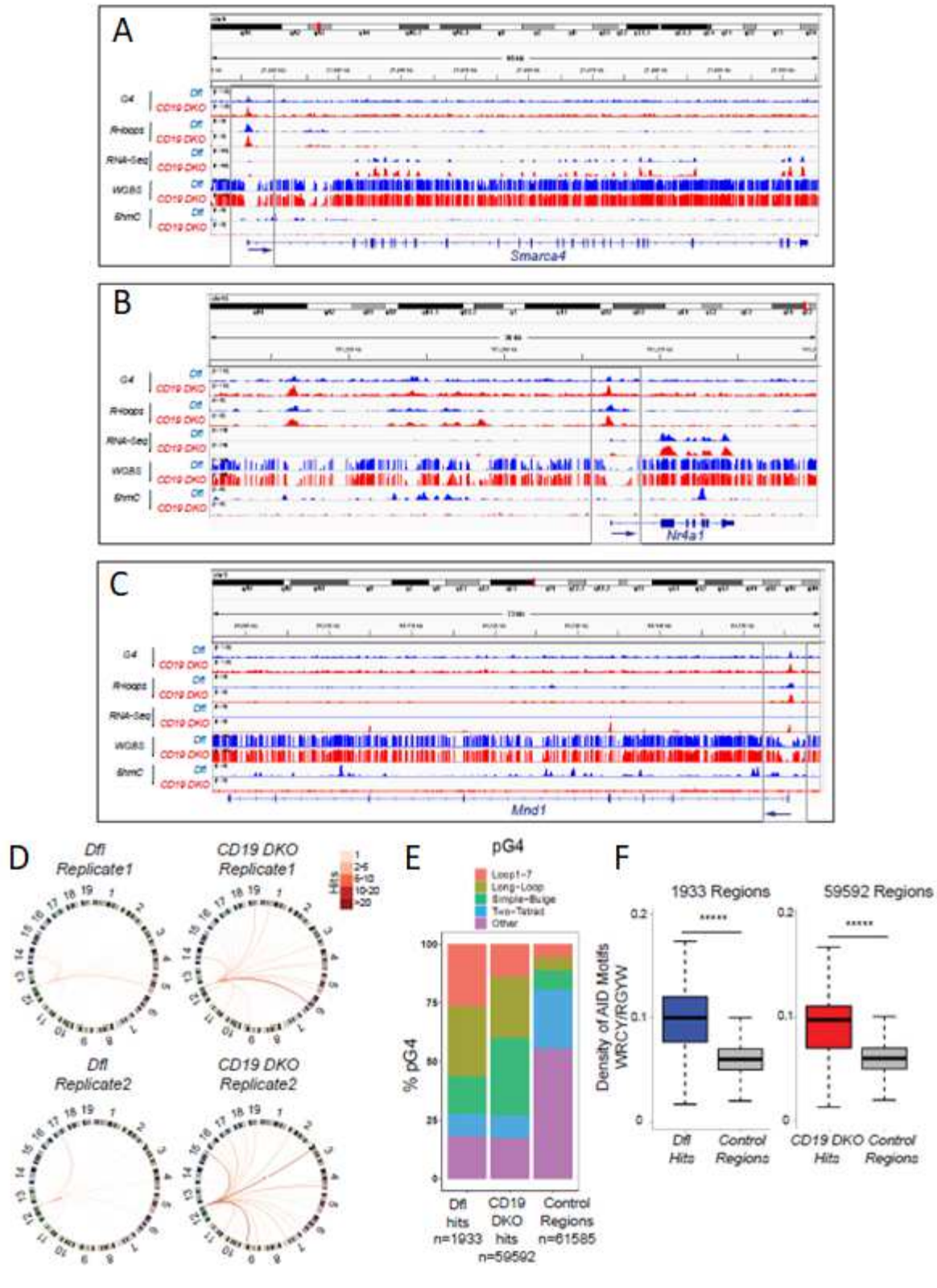

A
G4
R-loops
RNA-Seq
WGBS
5hmC
Dfl
CD19 DKO
Smarca4
B
Nr4a1
C
Mnd1
D
Dfl Replicate1
CD19 DKO Replicate1
Dfl Replicate2
CD19 DKO Replicate2
Hits
E
pG4
Loop1-7
Long-Loop
Simple-Bulge
Two-Tetrad
Other
% pG4
Dfl hits n=1933
CD19 DKO hits n=59592
Control Regions n=61585
F
1933 Regions
59592 Regions
Density of AID Motifs WRCY/RGYW
Dfl Hits
Control Regions
CD19 DKO Hits
Control Regions

Figure S2.8. DNMT1 deletion delays oncogenesis in TET-deficient mice.
(A) Diagrammatic representation of the strategy used to confirm G-quadruplex binding. Nuclear lysates of activated B cells were incubated with biotin-conjugated single stranded G4- or non-G4-forming control oligonucleotides (Oligos) captured using streptavidin beads. **(B)** Immunoblots showing flag-tagged BG4 (positive control), ATRX, BLM and DNMT1 proteins. *Left lane*, $1/10^{th}$ input from nuclear lysates ($1/10^{th}$ Input); *middle lane,* proteins pulled down with G4 forming oligonucleotides; and *right lane*, proteins pulled down with non-G4 control oligonucleotides. The data is representative of at least 2 independent experiments. **(C) – (D)** Quantification of **(C)** cell numbers, **(D)** spleen weights, of 10-week-old *Tfl* (grey), *CD19 Dnmt1 KO* (purple), *CD19 DKO* (red) and *CD19 TKO* (brown) mice from 5 independent experiments. **(E)** Enlarged spleen of 75-week-old *CD19 TKO* mice compared with *Tfl* control mice. **(F)** Flow cytometric detection of G-quadruplexes with BG4-Ig antibody or isotype IgG controls in B cells from *Tfl* (YFP^-, *grey*), *CD19 Dnmt1 KO* (YFP^+, *purple*), *CD19 DKO* (YFP^+, *red*) and *CD19 TKO* (YFP^+, *brown*) mice. **(G) – (H)** Quantification of median fluorescence intensity (MFI) of **(G)** BG4-Ig signal, **(H)** γH2AX signal from *Tfl* (YFP^-), *CD19 Dnmt1 KO* (YFP^+), *CD19 DKO* (YFP^+) and *CD19 TKO* (YFP^+) B cells from 4 independent experiments. **(I), (K)** Flow cytometric detection of **(I)** G-quadruplexes with BG4-Ig antibody or isotype IgG controls and **(K)** R-loops using V5-epitope-tagged recombinant RNASE H1 (rRNASE H1) or IgG controls in GC B cells (Fas^+) from *Tfl* (YFP^-, *grey*), *CD19 DKO* (YFP^+, *red*) and *CD19 TKO* (YFP^+, *brown*) mice. **(J), (L)** Quantification of median fluorescence intensity (MFI) of **(J)** BG4-Ig signal and **(L)** R-loops (rRNASE H1) signal in GC B cells (Fas^+) from *Tfl* (YFP^-, *grey*), *CD19 DKO* (YFP^+, *red*) and *CD19 TKO* (YFP^+, *brown*) mice from 3 independent experiments. **(M)** Model proposing functional interplay between TET and DNMT activities to limit GC B cell expansion. TET deficiency in B cells leads to increased G4 and R-loop structures and is associated with altered gene expression, DNA damage and development of B cell lymphoma. Statistical significance is calculated using one-way ANOVA in **(C), (D), (G), (H), (J)** and **(L)**. Error bars represent mean +/- standard deviation, * p value ≤0.05, ** p value ≤0.01, *** p value ≤0.0005 **** p value <0.0001.

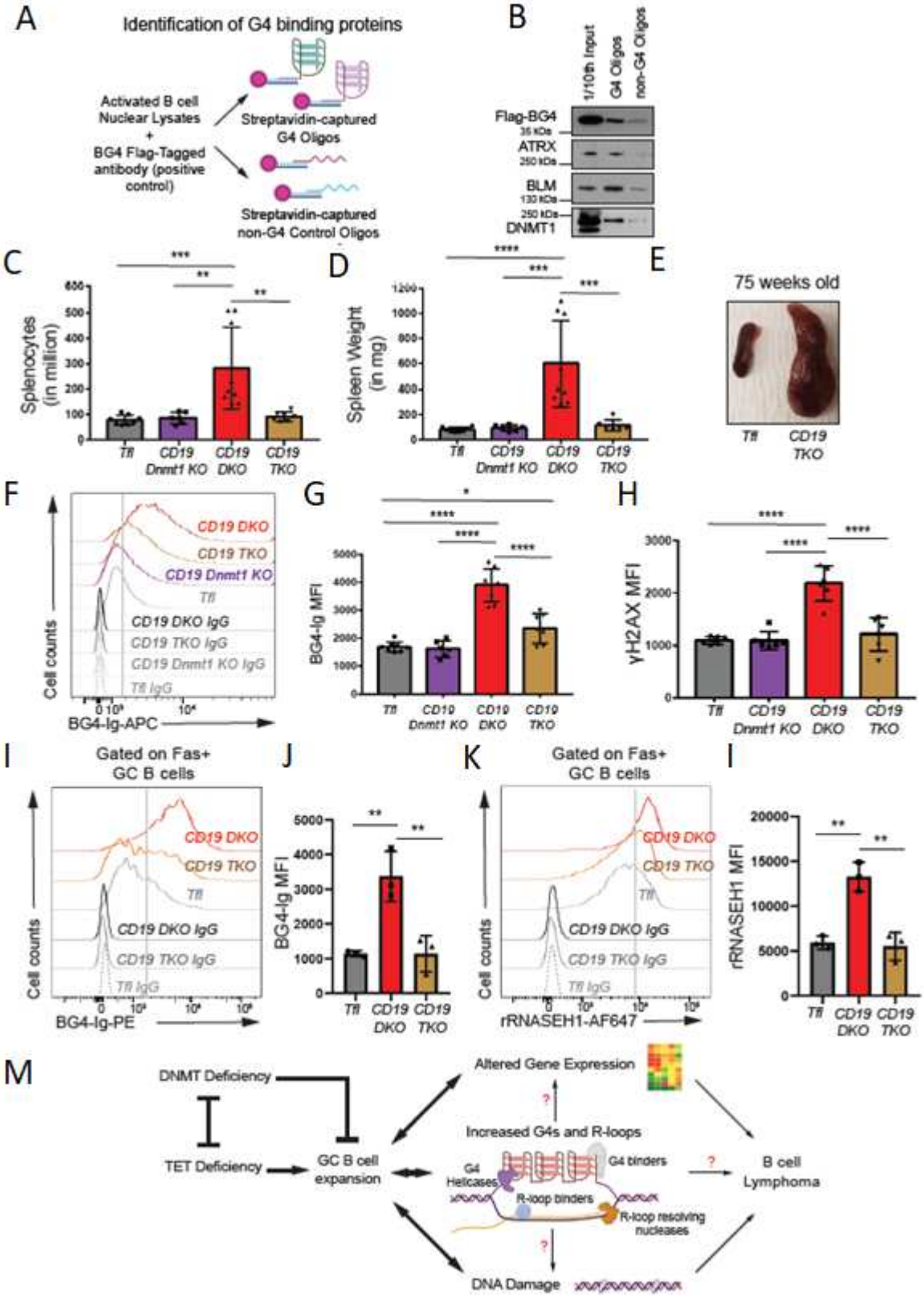
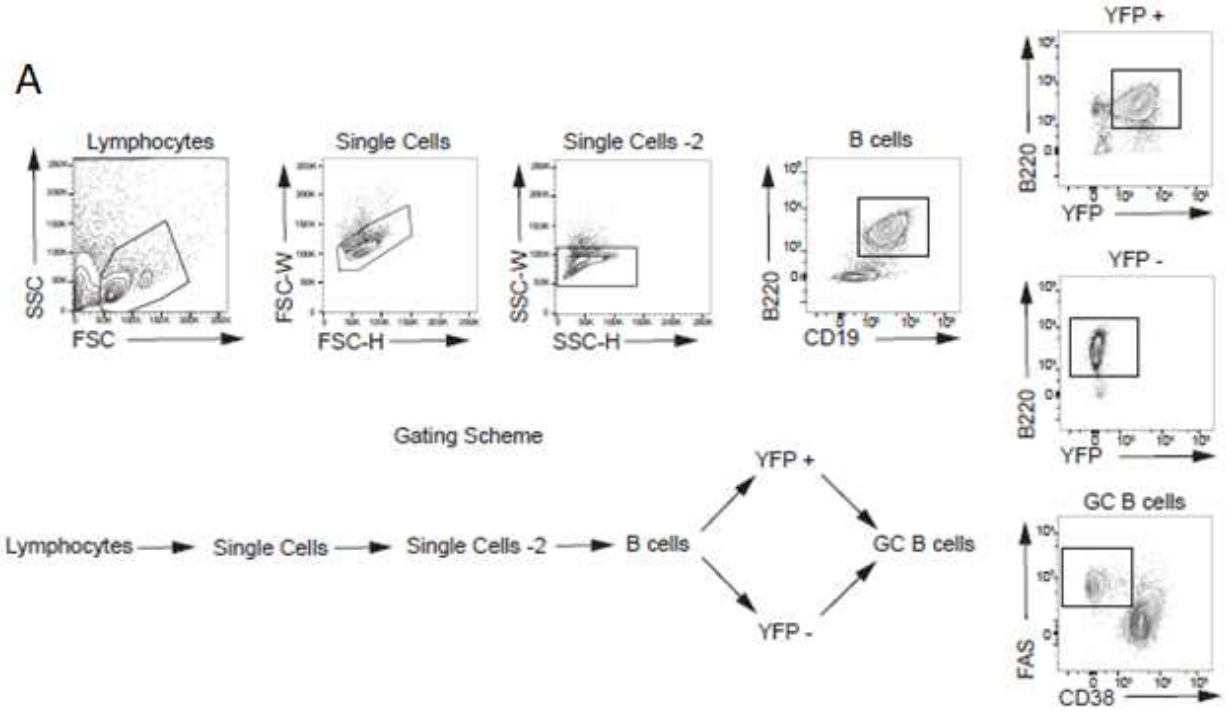
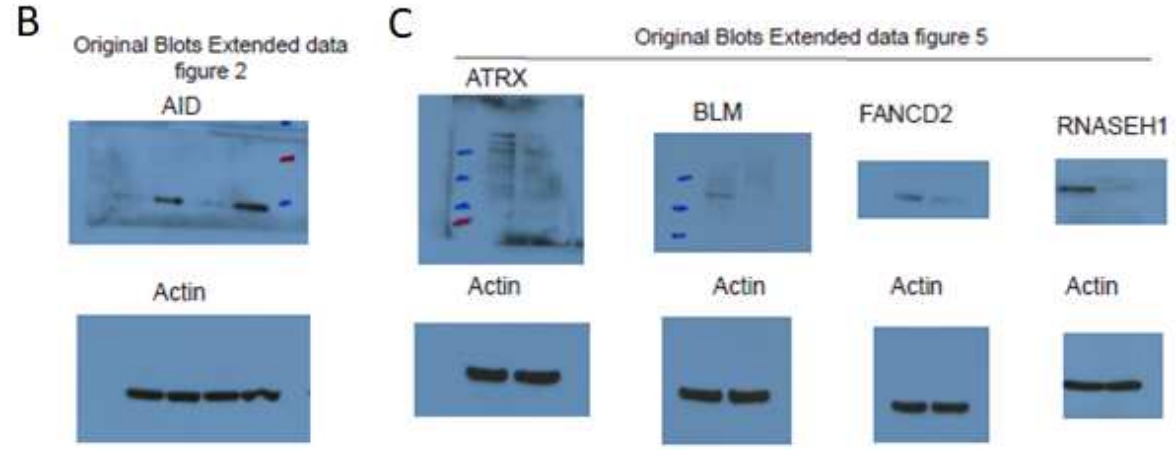
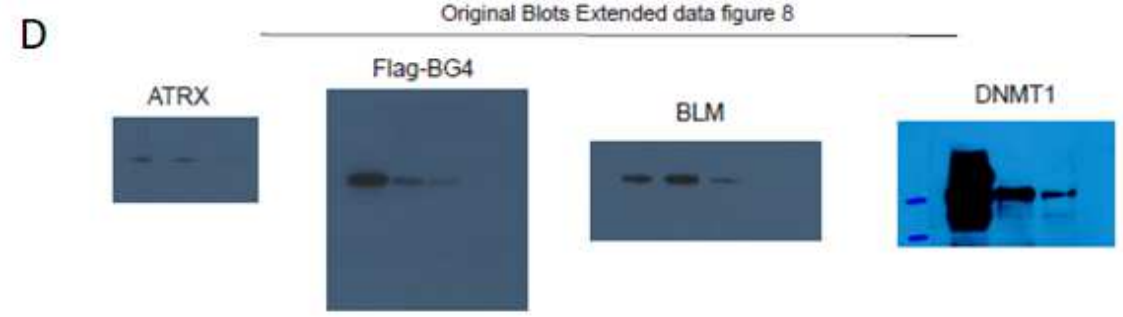

A
Identification of G4 binding proteins
Activated B cell Nuclear Lysates + BG4 Flag-Tagged antibody (positive control)
Streptavidin-captured G4 Oligos
Streptavidin-captured non-G4 Control Oligos
B
1/10th Input
G4 Oligos
non-G4 Oligos
Flag-BG4
35 kDa
ATRX
250 kDa
BLM
130 kDa
250 kDa
DNMT1
C
Splenocytes (in million)
T/I
CD19 Dnmt1 KO
CD19 DKO
CD19 TKO
D
Spleen Weight (in mg)
E
75 weeks old
F
CD19 DKO
CD19 TKO
CD19 Dnmt1 KO
T/I
CD19 DKO IgG
CD19 TKO IgG
CD19 Dnmt1 KO IgG
T/I IgG
Cell counts
BG4-Ig-APC
G
BG4-Ig MFI
H
γH2AX MFI
I
Gated on Fas+ GC B cells
BG4-Ig-PE
J
K
rRNASEH1-AF647
L
rRNASEH1 MFI
M
DNMT Deficiency
TET Deficiency
GC B cell expansion
Altered Gene Expression
Increased G4s and R-loops
G4 Helicases
G4 binders
R-loop binders
R-loop resolving nucleases
B cell Lymphoma
DNA Damage

Figure S2.9. FACS gating strategy and original blots.
(A) Sequential gating strategy used for the flow cytometry analysis. The respective gate names are mentioned in the corresponding figures. **(B)-(E)** scanned immunoblots for **Fig. S2.2E**, **Fig. S2.5B** and **Fig. S2.8B**.


A
Lymphocytes
SSC
FSC
Single Cells
FSC-W
FSC-H
Single Cells -2
SSC-W
SSC-H
B cells
B220
CD19
YFP +
B220
YFP
YFP -
B220
YFP
Gating Scheme
Lymphocytes
Single Cells
Single Cells -2
B cells
YFP +
YFP -
GC B cells
GC B cells
FAS
CD38


B
Original Blots Extended data figure 2
AID
Actin
C
Original Blots Extended data figure 5
ATRX
BLM
FANCD2
RNASEH1
Actin
Actin
Actin
Actin


D
Original Blots Extended data figure 8
ATRX
Flag-BG4
BLM
DNMT1

2.8 Author Contributions

V.S. and D.S.C. acquired, analyzed and interpreted the data. Z.D. performed R-loop mapping experiments and helped with the interpretation of results. E.D.A and D.S.C. performed the bioinformatics analysis. Q.Y. and K.S. provided the plasmids and suggestions for MapR. A.R. and V.S. supervised the studies, conceptualized the experiments and helped with the interpretation of data. V.S., D.S.C. and A.R. wrote the manuscript. All authors were involved in reviewing and editing the manuscript.

2.9 Acknowledgements

We thank Drs. H. Yuita and I. Lopez-Moyado for generating the *ERT2-Cre Tet1$^{fl/fl}$ Tet2$^{fl/fl}$ Tet3$^{fl/fl}$, ERT2creTKO* mice; Dr. D. Kitamura at the Tokyo University of Science for sharing the 40LB cells; Drs. U. Basu and B. Laffleur at Columbia University for help with the HTGTS protocol; our collaborators at Cambridge Epigenetix (UK) for providing the 5hmC mapping kits; the LJI Flow Cytometry Core team: C. Kim, D. Hinz, C. Dillingham, M Haynes, S. Ellis for help with cell sorting; and the LJI Next generation sequencing core members: J. Day, S. Alarcon, H. Dose, K. Tanaguay and A. Hernandez for help with sequencing. BD FACSAria II is supported by NIH (NIH S10OD016262, NIH S10RR027366) and our research used resources of the Advanced Light Source, which is a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231. The NovaSeq 6000 and the HiSeq 2500 were acquired through the Shared Instrumentation Grant (SIG) Program (S10); NovaSeq 6000 S10OD025052 and HiSeq 2500 S10OD016262. K.S. acknowledges support from National Institutes of Health Grants DP2-NS105576. V.S. is supported by Leukemia and Lymphoma Society Postdoctoral Fellowship (grant ID: 5463-18) and a K99/R00 award from National Cancer Institute (grant ID: CA248835). D.S.C. and E.G.A are supported by CONACYT/UCMEXUS Fellowship. This work is supported by the National Institutes of Health (NIH) grants R35 CA210043, R01 AI109842 and AI128589 to A.R and, K99/R00 CA248835, research funds from LLS grant 5463-18 and the Tullie and Rickey families SPARK award from LJI to V.S.

Chapter 2, in full, is a reprint with modifications as it appears in "TET deficiency perturbs mature B cell homeostasis and promotes oncogenesis associated with accumulation of G-quadruplex and R-loop structures", *Nature Immunology* (2022). DOI: https://doi.org/10.1038/s41590-021-01087-w. The dissertation author was the co-primary

investigator and co-first author of this paper. Other authors include Vipul Shukla, Zhen Dong, Edahí González-Avalos, Qingqing Yan, Kavitha Sarma, and Anjana Rao.

2.10 References

An, J., González-Avalos, E., Chawla, A., Jeong, M., López-Moyado, I.F., Li, W., Goodell, M.A., Chavez, L., Ko, M. and Rao, A., 2015. Acute loss of TET function results in aggressive myeloid cancer in mice. Nature communications, 6(1), pp.1-14.

Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S., 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. Nature chemistry, 5(3), pp.182-186.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L. and Carrico, R.J., 1986. Characterization of monoclonal antibody to DNA· RNA and its application to immunodetection of hybrids. Journal of immunological methods, 89(1), pp.123-130.

Bolotin, D.A., Poslavsky, S., Davydov, A.N., Frenkel, F.E., Fanchi, L., Zolotareva, O.I., Hemmers, S., Putintseva, E.V., Obraztsova, A.S., Shugay, M. and Ataullakhanov, R.I., 2017. Antigen receptor repertoire profiling from RNA-seq data. Nature biotechnology, 35(10), pp.908-911.

Chapuy, B., Stewart, C., Dunford, A.J., Kim, J., Kamburov, A., Redd, R.A., Lawrence, M.S., Roemer, M.G., Li, A.J., Ziepert, M. and Staiger, A.M., 2018. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature medicine, 24(5), pp.679-690.

Chédin, F., 2016. Nascent connections: R-loops and chromatin patterning. Trends in Genetics, 32(12), pp.828-838.

Chédin, Frédéric, Stella R. Hartono, Lionel A. Sanz, and Vincent Vanoosthuyse. "Best practices for the visualization, mapping, and manipulation of R-loops." The EMBO Journal 40, no. 4 (2021): e106394.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H. and Zhou, Y., 2017. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. Molecular cell, 68(4), pp.745-757.

Cimmino, L., Abdel-Wahab, O., Levine, R.L. and Aifantis, I., 2011. TET family proteins and their role in stem cell differentiation and transformation. Cell stem cell, 9(3), pp.193-204.

Cimmino, L., Dawlaty, M.M., Ndiaye-Lobry, D., Yap, Y.S., Bakogianni, S., Yu, Y., Bhattacharyya, S., Shaknovich, R., Geng, H., Lobry, C. and Mullenders, J., 2015. TET1 is a tumor suppressor of hematopoietic malignancy. Nature immunology, 16(6), pp.653-662.

Crossley, M.P., Bocek, M. and Cimprich, K.A., 2019. R-loops as cellular regulators and genomic threats. Molecular cell, 73(3), pp.398-411.

Crossley, M.P., Brickner, J.R., Song, C., Zar, S.M.T., Maw, S.S., Chédin, F., Tsai, M.S. and Cimprich, K.A., 2021. Catalytically inactive, purified RNase H1: A specific and sensitive probe for RNA–DNA hybrid imaging. Journal of Cell Biology, 220(9), p.e202101092.

De Magis, A., Manzo, S.G., Russo, M., Marinello, J., Morigi, R., Sordet, O. and Capranico, G., 2019. DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. Proceedings of the National Academy of Sciences, 116(3), pp.816-825.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), pp.15-21.

Dominguez, P.M., Ghamlouch, H., Rosikiewicz, W., Kumar, P., Béguelin, W., Fontan, L., Rivas, M.A., Pawlikowska, P., Armand, M., Mouly, E. and Torres-Martin, M., 2018. TET2 deficiency causes germinal center hyperplasia, impairs plasma cell differentiation, and promotes B-cell lymphomagenesis. Cancer discovery, 8(12), pp.1632-1653.

Gray, L.T., Vallur, A.C., Eddy, J. and Maizels, N., 2014. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. Nature chemical biology, 10(4), pp.313-318.

Hänsel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M. and Tannahill, D., 2016. G-quadruplex structures mark human regulatory chromatin. Nature genetics, 48(10), pp.1267-1272.

Hänsel-Hertsch, R., Di Antonio, M. and Balasubramanian, S., 2017. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nature reviews Molecular cell biology, 18(5), pp.279-284.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell, 38(4), pp.576-589.

Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W. and Frock, R.L., 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. Nature protocols, 11(5), pp.853-871.

Huang, Y. and Rao, A., 2014. Connections between TET proteins and aberrant DNA modification in cancer. Trends in Genetics, 30(10), pp.464-474.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science, 333(6047), pp.1300-1303.

Kafer, G.R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I. and Carlton, P.M., 2016. 5-Hydroxymethylcytosine marks sites of DNA damage and promotes genome stability. Cell reports, 14(6), pp.1283-1292.

Kang, J., Lienhard, M., Pastor, W.A., Chawla, A., Novotny, M., Tsagaratou, A., Lasken, R.S., Thompson, E.C., Surani, M.A., Koralov, S.B. and Kalantry, S., 2015. Simultaneous deletion of the

methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. Proceedings of the National Academy of Sciences, 112(31), pp.E4236-E4245.

Kharat, S.S., Ding, X., Swaminathan, D., Suresh, A., Singh, M., Sengodan, S.K., Burkett, S., Marks, H., Pamala, C., He, Y. and Fox, S.D., 2020. Degradation of 5hmC-marked stalled replication forks by APE1 causes genomic instability. Science signaling, 13(645).

Kieffer-Kwon, K.R., Nimura, K., Rao, S.S., Xu, J., Jung, S., Pekowska, A., Dose, M., Stevens, E., Mathe, E., Dong, P. and Huang, S.C., 2017. Myc regulates chromatin decompaction and nuclear architecture during B cell activation. Molecular cell, 67(4), pp.566-578.

Ko, M., An, J. and Rao, A., 2015. DNA methylation and hydroxymethylation in hematologic differentiation and transformation. Current opinion in cell biology, 37, pp.91-101.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B. and Rao, A., 2011. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proceedings of the National Academy of Sciences, 108(35), pp.14566-14571.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R. and Liu, X.S., 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature, 468(7325), pp.839-843.

König, F., Schubert, T. and Längst, G., 2017. The monoclonal S9. 6 antibody exhibits highly variable binding affinities towards different R-loop sequences. PLoS One, 12(6), p.e0178875.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome research, 19(9), pp.1639-1645.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology, 10(3), pp.1-10.

Lio, C.W., Zhang, J., González-Avalos, E., Hogan, P.G., Chang, X. and Rao, A., 2016. Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. Elife, 5, p.e18290.

Lio, C.W.J., Shukla, V., Samaniego-Castruita, D., González-Avalos, E., Chakraborty, A., Yue, X., Schatz, D.G., Ay, F. and Rao, A., 2019. TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. Science immunology, 4(34).

Lio, C.W.J., Yue, X., López-Moyado, I.F., Tahiliani, M., Aravind, L. and Rao, A., 2020. TET methylcytosine oxidases: new insights from a decade of research. Journal of biosciences, 45(1), pp.1-14.

Lio, C.W.J., Yuita, H. and Rao, A., 2019. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. Blood, 134(18), pp.1487-1497.

López-Moyado, I.F., Tsagaratou, A., Yuita, H., Seo, H., Delatte, B., Heinz, S., Benner, C. and Rao, A., 2019. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. Proceedings of the National Academy of Sciences, 116(34), pp.16933-16942.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15(12), pp.1-21.

Mao, S.Q., Ghanbarian, A.T., Spiegel, J., Cuesta, S.M., Beraldi, D., Di Antonio, M., Marsico, G., Hänsel-Hertsch, R., Tannahill, D. and Balasubramanian, S., 2018. DNA G-quadruplex structures mold the DNA methylome. Nature structural & molecular biology, 25(10), pp.951-957.

Miglietta, G., Russo, M. and Capranico, G., 2020. G-quadruplex–R-loop interactions and the mechanism of anticancer G-quadruplex binders. Nucleic acids research, 48(21), pp.11942-11957.

Müller, S., Sanders, D.A., Di Antonio, M., Matsis, S., Riou, J.F., Rodriguez, R. and Balasubramanian, S., 2012. Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. Organic & biomolecular chemistry, 10(32), pp.6537-6546.

Nambiar, M., Goldsmith, G., Moorthy, B.T., Lieber, M.R., Joshi, M.V., Choudhary, B., Hosur, R.V. and Raghavan, S.C., 2011. Formation of a G-quadruplex at the BCL2 major breakpoint region of the t (14; 18) translocation in follicular lymphoma. Nucleic acids research, 39(3), pp.936-948.

Nicoludis, J.M., Miller, S.T., Jeffrey, P.D., Barrett, S.P., Rablen, P.R., Lawton, T.J. and Yatsunyk, L.A., 2012. Optimized end-stacking provides specificity of N-methyl mesoporphyrin IX for human telomeric G-quadruplex DNA. Journal of the American Chemical Society, 134(50), pp.20446-20456.

Niehrs, C. and Luke, B., 2020. Regulatory R-loops as facilitators of gene expression and genome stability. Nature Reviews Molecular Cell Biology, 21(3), pp.167-178.

Nojima, T., Haniuda, K., Moutai, T., Matsudaira, M., Mizokawa, S., Shiratori, I., Azuma, T. and Kitamura, D., 2011. In-vitro derived germinal centre B cells differentially generate memory B or plasma cells in vivo. Nature communications, 2(1), pp.1-11.

Orlanski, S., Labi, V., Reizel, Y., Spiro, A., Lichtenstein, M., Levin-Klein, R., Koralov, S.B., Skversky, Y., Rajewsky, K., Cedar, H. and Bergman, Y., 2016. Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. Proceedings of the National Academy of Sciences, 113(18), pp.5018-5023.

Pastor, W.A., Aravind, L. and Rao, A., 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nature reviews Molecular cell biology, 14(6), pp.341-356.

Qiao, Q., Wang, L., Meng, F.L., Hwang, J.K., Alt, F.W. and Wu, H., 2017. AID recognizes structured DNA for class switch recombination. Molecular cell, 67(3), pp.361-373.

Quinlan, A.R. and Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), pp.841-842.

Rabkin, C.S., Hirt, C., Janz, S. and Dölken, G., 2008. t (14; 18) Translocations and risk of follicular lymphoma. Journal of the National Cancer Institute Monographs, 2008(39), pp.48-51.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F. and Manke, T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research, 44(W1), pp.W160-W165.

Rasmussen, K.D. and Helin, K., 2016. Role of TET enzymes in DNA methylation, development, and cancer. Genes & development, 30(7), pp.733-750.

Reddy, A., Zhang, J., Davis, N.S., Moffitt, A.B., Love, C.L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L., Karjalainen-Lindsberg, M.L. and Nørgaard, P., 2017. Genetic and functional drivers of diffuse large B cell lymphoma. Cell, 171(2), pp.481-494.

Rhodes, D. and Lipps, H.J., 2015. G-quadruplexes and their regulatory roles in biology. Nucleic acids research, 43(18), pp.8627-8637.

Rosikiewicz, W., Chen, X., Dominguez, P.M., Ghamlouch, H., Aoufouchi, S., Bernard, O.A., Melnick, A. and Li, S., 2020. TET2 deficiency reprograms the germinal center B cell epigenome and silences genes linked to lymphomagenesis. Science advances, 6(25), p.eaay5872.

Sabharwal, N.C., Savikhin, V., Turek-Herman, J.R., Nicoludis, J.M., Szalai, V.A. and Yatsunyk, L.A., 2014. N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. The FEBS journal, 281(7), pp.1726-1737.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X. and Chédin, F., 2016. Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. Molecular cell, 63(1), pp.167-178.

Sauer, M. and Paeschke, K., 2017. G-quadruplex unwinding helicases and their function in vivo. Biochemical Society Transactions, 45(5), pp.1173-1182.

Schmitz, R., Wright, G.W., Huang, D.W., Johnson, C.A., Phelan, J.D., Wang, J.Q., Roulland, S., Kasbekar, M., Young, R.M., Shaffer, A.L. and Hodson, D.J., 2018. Genetics and pathogenesis of diffuse large B-cell lymphoma. New England Journal of Medicine, 378(15), pp.1396-1407.

Shaknovich, R., Cerchietti, L., Tsikitas, L., Kormaksson, M., De, S., Figueroa, M.E., Ballon, G., Yang, S.N., Weinhold, N., Reimers, M. and Clozel, T., 2011. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood, The Journal of the American Society of Hematology, 118(13), pp.3559-3569.

Skourti-Stathaki, K. and Proudfoot, N.J., 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & development, 28(13), pp.1384-1396.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X. and Wang, J., 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nature biotechnology, 29(1), pp.68-72.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 324(5929), pp.930-935.

Tanaka, S., Ise, W., Inoue, T., Ito, A., Ono, C., Shima, Y., Sakakibara, S., Nakayama, M., Fujii, K., Miura, I. and Sharif, J., 2020. Tet2 and Tet3 in B cells are required to repress CD86 and prevent autoimmunity. Nature immunology, 21(8), pp.950-961.

Tsagaratou, A., González-Avalos, E., Rautio, S., Scott-Browne, J.P., Togher, S., Pastor, W.A., Rothenberg, E.V., Chavez, L., Lähdesmäki, H. and Rao, A., 2017. TET proteins regulate the lineage specification and TCR-mediated expansion of i NKT cells. Nature immunology, 18(1), pp.45-53.

Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A. and Huse, J.T., 2019. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nature communications, 10(1), pp.1-14.

Wu, H. and Zhang, Y., 2014. Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell, 156(1-2), pp.45-68.

Xi, Y. and Li, W., 2009. BSMAP: whole genome bisulfite sequence MAPping program. BMC bioinformatics, 10(1), pp.1-9.

Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Dos Santos, N., Silvester, J., Wei, V., Garcia, J. and Kabeer, F., 2017. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nature communications, 8(1), pp.1-18.

Xu, Y.Z., Jenjaroenpun, P., Wongsurawat, T., Byrum, S.D., Shponka, V., Tannahill, D., Chavez, E.A., Hung, S.S., Steidl, C., Balasubramanian, S. and Rimsza, L.M., 2020. Activation-induced cytidine deaminase localizes to G-quadruplex motifs at mutation hotspots in lymphoma. NAR cancer, 2(4), p.zcaa029.

Yan, Q., Shields, E.J., Bonasio, R. and Sarma, K., 2019. Mapping native R-loops genome-wide using a targeted nuclease approach. Cell reports, 29(5), pp.1369-1380.

Yewdell, W.T., Kim, Y., Chowdhury, P., Lau, C.M., Smolkin, R.M., Belcheva, K.T., Fernandez, K.C., Cols, M., Yen, W.F., Vaidyanathan, B. and Angeletti, D., 2020. A Hyper-IgM Syndrome Mutation in Activation-Induced Cytidine Deaminase Disrupts G-Quadruplex Binding and Genome-wide Chromatin Localization. Immunity, 53(5), pp.952-970.

Young, R.M., Wu, T., Schmitz, R., Dawood, M., Xiao, W., Phelan, J.D., Xu, W., Menard, L., Meffre, E., Chan, W.C.C. and Jaffe, E.S., 2015. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proceedings of the National Academy of Sciences, 112(44), pp.13447-13454.

Zhang, W. and Xu, J., 2017. DNA methyltransferases and their roles in tumorigenesis. Biomarker research, 5(1), pp.1-8.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome biology, 9(9), pp.1-9.

CHAPTER 3: Whole-genome analysis of TET dioxygenase function in regulatory T cells

3.1 Abstract

TET methylcytosine dioxygenases are essential for the stability and function of regulatory T cells (Treg cells), which maintain immune homeostasis and self-tolerance and express the lineage-determining transcription factor Foxp3. Here we use whole genome analyses to show that the transcriptional program and epigenetic features (DNA modification, chromatin accessibility) of Treg cells are attenuated in the absence of Tet2 and Tet3. Conversely, addition of the TET activator Vitamin C during TGFβ-induced iTreg cell differentiation in vitro potentiates the expression of Treg signature genes and alters the epigenetic landscape to better resemble that of Treg cells generated in vivo. Vitamin C enhances IL-2 responsiveness in iTreg cells by increasing IL2Rα expression, STAT5 phosphorylation, and STAT5 binding, mimicking the IL-2/STAT5 dependence of Treg cells generated in vivo. In summary, TET proteins play essential roles in maintaining Treg molecular features and promoting their dependence on IL-2. TET activity during endogenous Treg development, and potentiation of TET activity by Vitamin C during iTreg differentiation, are necessary to maintain the transcriptional and epigenetic features of Treg cells.

3.2 Introduction

Enzymes of the TET family are Fe(II) and 2-oxoglutarate-dependent dioxygenases that catalyze sequential oxidation of the methyl group of 5-methylcytosine (5mC) in DNA to 5-hydroxymethylcytosine (5hmC) and the further oxidized products 5-formalcytosine (5fC) and 5-carboxylcytosine (5caC) (He *et al.* 2011; Ito *et al.* 2011; Iyer *et al.* 2009; Tahiliani *et al.* 2009). These modified bases are intermediates in DNA demethylation as well as potential epigenetic marks capable of recruiting reader proteins that exert specific functions (Pastor & Rao, 2013; Tsagaratou *et al.* 2017b; Wu & Zhang, 2017). There are three mammalian TET proteins, TET1, TET2 and TET3, of which TET2 and TET3 are most highly expressed in cells of the immune and hematopoietic systems (Tsagaratou & Rao, 2013). Like all Fe(II) and 2-oxoglutarate-dependent dioxygenases, TET enzymes utilize 2-oxoglutarate (2OG), Fe (II) and molecular oxygen to generate their oxidized products, with CO_2 and succinate formed as byproducts (Hausinger, 2004). The enzymatic activity of TET proteins can be modulated by diverse factors, including the ambient levels of oxygen (hypoxia versus normoxia) and diverse metabolic enzymes that regulate the intracellular levels of either 2OG itself or its competitive inhibitor 2-hydroxyglutarate (2HG) (Dang & Su, 2017; Lio & Rao, 2019; Lio *et al.* 2020). Both stereoisomers of 2HG are inhibitors of TET proteins and other dioxygenases, with the D (or *S(+)* stereoisomer being significantly more potent than the L (or *R(-)* form (Xu *et al.* 2011; Ye *et al.* 2018). In contrast, Vitamin C (ascorbate) is a potent activator of TET proteins and other dioxygenases; it is believed to act by restoring the reduced state of Fe(II) at the dioxygenase active site in each catalytic cycle in which Fe(II) is converted to Fe(III) (Blaschke *et al.* 2013; Yue *et al.* 2016).

TET proteins are required to maintain the stability and function of regulatory T (Treg) cells, a distinct lineage of $CD4^+$ T cells that maintain immune self-tolerance and homeostasis

(Josefowicz *et al.* 2012; Li & Zheng, 2015; Sakaguchi *et al.* 2008; Yang *et al.* 2015; Yue *et al.* 2016). T regulatory cells are defined by expression of the X-chromosome-encoded transcription factor FOXP3, a lineage-determining transcription factor essential for Treg development and function (Josefowicz *et al.* 2012; Li & Zheng, 2015; Sakaguchi *et al.* 2008). TET proteins also control the stability of FOXP3 expression in "induced" regulatory T cells (iTreg cells), generated *in vitro* by activation of naïve $CD4^+$ T cells through the T cell receptor (TCR) and the costimulatory receptor CD28 in the presence of TGFβ and/or retinoic acid (RA) (Yue *et al.* 2016). Both during thymic development and during iTreg differentiation, TET proteins are dispensable for the initial acquisition of FOXP3 expression, but are critical for maintaining the stability of FOXP3 expression through cell division *in vitro* and *in vivo* (Yang *et al.* 2015; Yue *et al.* 2016).

The stability of FOXP3 expression is regulated by the DNA methylation status of an intronic enhancer, *CNS2* (conserved non-coding sequence 2) located in the first intron of the *Foxp3* gene (Feng *et al.* 2014b; Li *et al.* 2014; Zheng *et al.* 2010). The 12 CpG sites in the mouse *Foxp3 CNS2* enhancer are predominantly "methylated" in iTreg cells generated with TGFβ or TGFβ and RA but predominantly "unmethylated" in endogenous Treg cells (Floess *et al.* 2007; Huehn *et al.* 2009; Toker & Huehn, 2011); the words "methylated" and "unmethylated" are in quotation marks because bisulfite sequencing does not distinguish 5mC and 5hmC, or unmodified cytosine from 5fC and 5caC (Huang *et al.* 2010). Notably, deletion of the *Tet2* and *Tet3* genes with either *CD4Cre* or *Foxp3Cre* during Treg development, results in increased methylation of *Foxp3 CNS2* in *Tet2/3* DKO Treg cells compared to WT Treg cells and a marked defect in Foxp3 stability (Yue *et al.* 2019; Yue *et al.* 2016). In a parallel gain-of-function approach, we have shown that addition of the TET activator Vitamin C during *in vitro* iTreg differentiation leads to TET-dependent

demethylation of *Foxp3 CNS2* and a substantial increase in Foxp3 stability after adoptive transfer *in vivo* or cell division *in vitro* (Sasidharan Nair *et al.* 2016; Yue *et al.* 2016).

To gain a better understanding of the role of TET dioxygenases in Treg function, we embarked on genome-wide analyses to compare wildtype (WT) $CD4^+$ naïve T cells with endogenous WT Treg cells and *Tet2/3*-deficient Treg cells. In parallel, we used the gain-of-function approach to document the genome-wide effects of Vitamin C on iTreg differentiation, comparing iTreg cells generated by activation in the presence of TGFβ alone or TGFβ alone plus retinoic acid (RA) with cells generated under the same conditions but with Vitamin C added to the culture medium. We assessed (i) transcriptional profiles (by RNA-sequencing, RNA-seq); (ii) 5hmC distribution (by CMS-IP, DNA immunoprecipitation with antibodies to cytosine 5-methylenesulphonate (CMS), the product of 5hmC after bisulfite treatment); (iii) genome-wide DNA methylation patterns (by whole-genome bisulfite sequencing, WGBS); and (iv) the chromatin accessibility landscape (by assay for transposase-accessible chromatin using sequencing, ATAC-seq). We find that in the absence of Tet2 and Tet3, Treg cells fail to maintain Treg-specific epigenetic features and the expression of Treg signature genes. Moreover, the changes induced by Vitamin C during TGFβ-induced iTreg differentiation *in vitro* resemble those occurring in a TET-dependent manner in endogenous Tregs *in vivo*, although the in vitro effects of Vitamin C on iTreg differentiation are, as expected, less striking than the differences between WT and TET-deficient Tregs generated *in vivo*. TET deletion in vivo resulted in impaired IL2/STAT5 signaling, while boosting TET activity with Vitamin C enhanced IL-2 responsiveness in iTreg cells by increasing IL2Rα expression and STAT5 activity. Our study provides compelling evidence for manipulation of TET activity in future Treg-mediated treatment of autoimmune disease.

3.3 Results

Dysregulation of gene expression patterns in Treg cells in the absence of Tet2 and Tet3. To examine the transcriptional changes in Treg cells upon *Tet2* and *Tet3* deletion, we performed RNA-seq on WT Treg cells, *Tet2/Tet3*-deficient Treg cells, and WT $CD4^+$ naïve T cells. Floxed exons 8-10 of *Tet2* and exon 2 of *Tet3* were efficiently deleted in Treg cells isolated from *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$*CD4Cre* (*Tet2/3* DKO) mice (Tsagaratou *et al.* 2017a; Yue *et al.* 2016) compared to WT Treg cells, as shown by RNA-seq (**Fig. A3.1A**). Principal component analysis of gene expression data showed that samples clustered by cell type: $CD4^+$ naïve T cell replicates clustered together, as did replicates for WT and *Tet2/3* DKO Treg cells (**Fig. 3.1A**). Principal component 1 (PC1) distinguished $CD4^+$ naïve T cells from both WT Treg and *Tet2/3* DKO Treg cells, whereas principal component 2 (PC2) separated WT Treg cells from *Tet2/3* DKO Treg cells (**Fig. 3.1A**). Differential analysis showed that 2,245 genes were differentially expressed in WT Treg cells compared to $CD4^+$ naïve T cells, while 614 genes were differentially expressed in WT Treg compared to *Tet2/3* DKO Treg cells (**Fig. 3.1B, Table S3.1**).

Gene Set Enrichment Analysis (GSEA) of Hallmark gene sets (Liberzon *et al.* 2015; Subramanian *et al.* 2005) showed that the IL-2/STAT5 signaling gene set was highly enriched in WT Treg cells compared to $CD4^+$ naïve T cells as expected (**Fig. 3.1C**, *upper panel*), since the IL-2/STAT5 signaling axis is essential for the functional programming of Treg cells and the maintenance of immune homeostasis (Ross & Cantrell, 2018). The IL2/STAT5 signaling gene set was also highly enriched in WT Treg cells compared to *Tet2/3* DKO Treg cells (**Fig. 3.1C**, *lower panel*). Consistent with the gene expression changes, the expression of IL2Rα (CD25) was reduced in *Tet2/3* DKO Treg cells compared to WT Treg cells (**Fig. A3.2A**), as well as the phospho-STAT5 levels upon IL-2 restimulation (**Fig A3.2B**). These data suggested that TET deletion led to

impaired IL-2/STAT5 signaling. We also found that the E2F Target gene set, which includes genes essential for DNA replication and cell cycle and is one of the most frequently altered pathway categories upon TET deletion (Tsagaratou *et al.* 2017a; Yue *et al.* 2019), was highly enriched in *Tet2/3* DKO Treg cells compared to WT Treg cells (**Fig 3.1D**, *lower panel*).

We also examined the representation of Treg signature genes from Hill et al (Hill *et al.* 2007) using GSEA. Treg signature genes upregulated in Treg cells compared to conventional T cells (Hill *et al.* 2007) were significantly enriched in WT Treg cells compared to $CD4^+$ naïve T cells as expected (**Fig. 3.1E**, *upper panel*) but were significantly depleted in *Tet2/3* DKO Treg compared to WT Treg cells (**Fig. 3.1E**, *lower panel*). Conversely, Treg signature genes downregulated in Treg cells compared to conventional T cells (Hill *et al.* 2007) were significantly depleted in WT Treg cells compared to $CD4^+$ naïve T cells (**Fig. A3.1B**, *upper panel*) but were evenly distributed between WT Treg and *Tet2/3* DKO Treg cells (**Fig. A3.1B**, *lower panel*). These data suggested that TET proteins are more involved in activation of Treg signature genes than in their suppression. Selected Treg signature genes – *Foxp3*, *Il10ra*, *Nrp1*, *Ikzf4* and *Ccr6* – are shown in **Fig. 3.1F**; their expression levels were lower in *Tet2/3* DKO Treg than in WT Treg cells, except for *Foxp3*, which was used as a marker for Treg cell isolation. The expression levels of genes related to the effector program, cell cycle and cell adhesion – *Eomes*, *Chek1, Ccl4* and *Fn1* – were dramatically increased upon TET deletion (**Fig. 3.1F**). Together, these data show clearly that the transcriptional program of Treg cells is altered upon TET deletion; specifically, *Tet2/3* DKO Treg cells show notably impaired expression of genes in the Treg signature program.

5hmC distribution correlates strongly with gene expression levels. WT Treg cells and $CD4^+$ naïve T cells have similar amounts of genomic 5hmC, as assessed by dot blot of sodium bisulfite-treated DNA with antibodies recognizing cytosine 5-methylenesulphonate (CMS), the

product of 5hmC after bisulfite treatment (Huang *et al.* 2012; Pastor *et al.* 2011) (**Fig. 3.2A**). This was also reflected in the similar numbers of 5hmC-marked regions enriched over input in $CD4^+$ naïve T cells and WT Treg cells based on immunoprecipitation (CMS-IP) of genomic DNA with anti-CMS (**Fig. A3.3A-B**). 12,279 5hmC-marked regions were differentially hydroxymethylated between these two cell types (Treg > naïve, 9,058 regions, naïve > Treg, 3,221 regions; **Fig. 3.2B, Table S3.2**). Compared to the mm9 genome, 5hmC was depleted in intergenic regions but significantly enriched at promoter regions (±1 kb of TSS) and in transcribed regions of genes – introns, exons, 5' and 3' UTRs, and transcription termination sites (**Fig. A3.3C-D**). 5hmC levels were also low at transcription start sites (TSS) (**Fig. A3.3E**) because these regions have lower levels of the substrate 5mC ((Tsagaratou *et al.* 2017a); also see **Fig. S3.1B** below). However, 5hmC levels in the latter two-thirds of gene bodies showed a very good correlation with gene expression: as shown for other cell types (Tsagaratou *et al.* 2014; Tsagaratou *et al.* 2017a), the greatest amounts of 5hmC were present in the gene bodies of the highest expressed genes in both $CD4^+$ naïve T cells and WT Treg cells (**Fig. A3.3E**).

We examined the relation of differentially hydroxymethylated regions (DhmRs) to differentially expressed genes (DEGs), and found that DhmRs that were more hydroxymethylated in WT Treg cells (Treg-specific DhmRs) were significantly associated with DEGs expressed at higher levels in WT Treg cells compared to $CD4^+$ naïve T cells (**Fig. 3.2C***, blue dots*). Conversely, DhmRs that were more hydroxymethylated in $CD4^+$ naïve T cells (naïve T cell-specific DhmRs) were significantly associated with DEGs expressed at higher levels in $CD4^+$ naïve T cells compared to WT Treg cells (**Fig. 3.2C***, black dots*). This point is illustrated by the genome browser tracks and quantified 5hmC signal for the *Foxp3* and *Il2ra* genes (**Fig. 3.2D-E**); again, integrated

5hmC levels in the latter part of gene bodies correlated well with levels of gene expression (**Fig. 3.2F**).

These data highlight the strong correlation of 5hmC levels with gene expression, cell differentiation and lineage specification, as previously noted for other immune and non-immune cell types (Neri *et al.* 2013; Tsagaratou *et al.* 2014; Tsagaratou *et al.* 2017a). We note, however, that despite showing the expected profound loss of 5hmC (Yue *et al.* 2016), *Tet2/3* DKO Treg cells show equivalent expression of many genes compared to WT Treg cells (**Fig. 3.1B**), indicating that 5hmC modulates gene expression rather than driving gene transcription.

Treg-specific DNA demethylation patterns are impaired in the absence of Tet2 and Tet3. Because our data suggested that TET proteins were important for correct specification of the full Treg signature program during Treg differentiation, we interrogated the effects of TET deletion on DNA methylation by whole-genome bisulfite sequencing (WGBS) to identify CpGs that were differentially modified in naïve $CD4^+$ T cells, WT and *Tet2/3* DKO Treg cells (**Fig. 3.3**). Bisulfite sequencing cannot distinguish 5mC and 5hmC, and therefore reports on the sum of 5mC+5hmC; nevertheless, since 5hmC is typically <5-10% of 5mC, the data are often termed "methylation" for convenience. Principal component analysis of DNA methylation data showed that $CD4^+$ naïve T cell replicates clustered together, as did replicates for WT and *Tet2/3* DKO Treg cells; principal component 1, which distinguished naïve $CD4^+$ T cells from WT Treg cells as well as WT Treg cells from *Tet2/3* DKO Treg cells, represented the majority of the variation (97.6%) whereas principal component 2 only separated WT Treg cells from *Tet2/3* DKO Treg cells (**Fig. 3.3A**). Global CpG methylation levels were very similar in all three cell types (**Fig. S3.1A**). As previously noted by our own and other groups (Laurent *et al.* 2010; Lee *et al.* 2002; Tsagaratou *et*

al. 2017a), the most highly expressed genes in each cell type showed prominent demethylation at promoter/TSS regions, and this demethylation extended well into the gene body (**Fig. S3.1B**).

We then analyzed the differentially methylated CpGs (DM-CpGs) (**Table S3.3**). Comparing $CD4^+$ naïve T cells and WT Treg cells, the majority of the DM-CpGs (11,370/15,450; 73.6%) were more methylated in naïve $CD4^+$ T cells (**Fig. 3.3B**, *left panel*). Similarly, comparing *Tet2/3* DKO with WT Treg cells, most of the 52,805 DM-CpGs (76.9%) were more methylated in the *Tet2/3* DKO cells as expected (**Fig. 3.3B**, *right panel*). K-means clustering of DM-CpGs in selected pairwise comparisons ($CD4^+$ naïve T vs WT Treg and WT Treg vs *Tet2/3* DKO Treg) identified clusters with distinct DNA methylation patterns among the cell types (**Fig. 3.3C**, *left panel*). Cluster 1 contained ~23,000 DM-CpGs whose demethylated status, at least in mature Treg cells, was maintained by Tet2 and/or Tet3; these CpGs were demethylated in both $CD4^+$ naïve T cells and WT Treg cells, but showed increased DNA methylation in *Tet2/3* DKO Treg cells. Cluster 2 contained ~24,000 DM-CpGs that showed Treg-specific, TET-regulated DNA methylation patterns: the majority of these CpGs were highly methylated in $CD4^+$ naïve T cells and became demethylated in WT Treg cells, but remained highly methylated in *Tet2/3* DKO Treg cells (**Fig 3.3C**, *left panel*). Notably, both CpG clusters 1 and 2 were marked by a significant amount of 5hmC in both $CD4^+$ naïve T cells and WT Treg cells, confirming that TET proteins act directly to oxidize 5mC at these CpG sites (**Fig. 3.3C**, *right panel*). In contrast, cluster 3 contained ~14,000 DM-CpGs that showed increased DNA methylation in WT Treg cells compared to $CD4^+$ naïve T cells and *Tet2/3* DKO Treg cells (**Fig. 3.3C,** *left panel*). Notably, this cluster was depleted for 5hmC in both $CD4^+$ naïve T cells and WT Treg cells (**Fig. 3.3C,** *right panel*). We speculate that the CpGs in this cluster reside in heterochromatin, perhaps in genes or regions that are repressed during the course of Treg differentiation, since they are highly methylated in WT Treg

cells but undergo a paradoxical loss (rather than the expected gain) of DNA methylation in TET-deficient Treg cells (Lopez-Moyado *et al.* 2019). We then annotated the CpGs in each K-means cluster using Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean *et al.* 2010). The results showed that CpGs in cluster 1 and 2 are proximal to genes regulating immune responses, lymphocyte cell activation and cell adhesion (**Table S3.4**; CpGs proximal to the selected genes were annotated in **Fig. 3.3C**, *middle panel*). In contrast, CpGs in cluster 3, which were aberrantly demethylated in Tet2/3 DKO Treg cells, were associated with genes regulating amino acid transport (**Table S3.4**).

We assessed DNA methylation (5mC+5hmC) levels at selected consensus transcription factor (TF) binding sites (±100bp, centered on the sites; **Table S3.5**). DNA methylation levels surrounding consensus ETS-binding motifs were lower in both $CD4^+$ naïve T cells and WT Treg cells compared to *Tet2/3* DKO Treg cells, consistent with the well-established roles of ETS family members in T cell activation and Treg development and function (Luo *et al.* 2017; Mouly *et al.* 2010; Muthusamy *et al.* 1995) (**Fig. 3.3D,** *left*). DNA methylation levels surrounding consensus binding motifs for members of the basic region-leucine zipper (bZIP), Rel-homology domain (RHD), and nuclear receptor (NR) transcription factor families were predominantly located in CpG cluster 2, i.e. Treg-specific (methylation levels lower in WT Treg cells compared to $CD4^+$ naïve T cells) as well as TET-regulated (methylation levels higher in *Tet2/3* DKO Treg cells compared to WT Treg cells) (**Fig. 3.3D**, *middle*). These transcription factor families include JunB, BATF, BACH2 (bZIP), NFAT and NFκB RelA and c-Rel (RHD) and NR4A (NR), are all important regulators of Treg development and function (Grinberg-Bleyer *et al.* 2017; Hayatsu *et al.* 2017; Koizumi *et al.* 2018; Ronin *et al.* 2019; Roychoudhuri *et al.* 2013; Sekiya *et al.* 2013; Wu *et al.* 2006). Consistent with these data, motif enrichment analysis of DhmRs showed that consensus

binding sequences of transcription factors ETS, HMG (High-Mobility Group) and ETS:RUNT composite motifs were enriched in DhmRs specific to naïve T cells, whereas consensus binding sequences of NR, RHD and bZIP families were enriched in Treg-specific DhmRs (**Fig. S3.1C, Table S3.6**). In contrast, DNA methylation levels at consensus zinc finger (Znf) transcription factor binding sites displayed a different pattern: methylation levels were comparable or even lower in *Tet2/3* DKO Treg cells compared to WT Treg cells (**Fig. 3.3D**, *right*).

Genome browser views of DNA methylation levels at the *Hopx* and *Il10* loci showed Treg-specific and TET-regulated DM-CpGs (**Fig. 3.3E**). Many of these overlapped or were adjacent to Treg-specific DhmRs, as especially apparent in the *Il10* locus (**Fig. 3.3E**, and **Fig. S3.1D**, *grey shaded regions*). The *Cd28/Ctla4/Icos, Il2ra* and *Bach2* loci also displayed stretches of CpGs with Treg-specific and TET-regulated methylation patterns (**Fig. S3.1E**), again mainly located in CpG cluster 2 (**Fig. 3.3C**).

Together these data suggest that TET proteins have a critical role in maintaining Treg-specific demethylated CpGs. Moreover, motifs displaying Treg-specific TET-regulated methylation patterns correlated well with motifs enriched at Treg-specific DhmRs.

Chromatin accessibility patterns of Treg cells in the absence of Tet2 and Tet3. Analysis of chromatin accessibility by ATAC-seq (assay for transposase-accessible chromatin using sequencing) (Buenrostro *et al.* 2013), followed by principal component analysis, again showed that samples from naïve $CD4^+$ T cells, WT Treg cells and *Tet2/3* DKO Treg cells clustered by cell type (a published ATAC-seq dataset was used for the naïve $CD4^+$ T cells (Yoshida *et al.* 2019); again, the differences between $CD4^+$ naïve T cells and Treg cells, either WT or *Tet2/3* DKO, were more striking than the differences between WT and *Tet2/3* DKO Treg cells, accounting for 47% and 18% of the variability respectively (shown as principal components 1 and 2, PC1 and

PC2, in **Fig. 3.4A**). Of 3,471 regions showing altered accessibility, 2,270 regions were more accessible and 1,201 regions were less accessible in WT compared to *Tet2/3* DKO Treg cells (**Fig. 3.4B, Table S3.7**). K-means clustering of the 3,471 regions identified clusters with distinct patterns of accessibility changes (**Fig. 3.4C**): regions that lost accessibility in *Tet2/3* DKO Treg cells compared to both $CD4^+$ naïve T cells and WT Treg cells (cluster 1); regions that were more accessible in WT Treg cells compared to $CD4^+$ naïve T cells (Treg-specific) and whose accessibility decreased upon TET deletion (TET-regulated) (cluster 2); regions that were more accessible in *Tet2/3* DKO Treg cells compared to both naïve $CD4^+$ T cells and WT Treg cells (cluster 3); and regions that were similarly accessible in naïve $CD4^+$ T cells and *Tet2/3* DKO Treg cells but which lost accessibility in WT Treg cells (cluster 4).

Motif analysis of regions in each cluster (**Fig. 3.4D, Table S3.8**) showed enrichment for consensus ETS binding motifs in cluster 1, corresponding with the previously noted increase in DNA methylation levels at ETS binding sites in *Tet2/3* DKO Treg cells compared to naïve $CD4^+$ T cells and WT Treg cells (**Fig. 3.3D**, *left*). Similarly, RHD, bZIP and NR binding motifs were enriched in cluster 2, which contained Treg-specific TET-regulated accessible regions, again consistent with the Treg-specific TET-regulated DNA methylation patterns at these transcription factor-binding sites (**Fig. 3.3D**, *middle*). These patterns are illustrated in genome browser views of chromatin accessibilities at the *Hopx* and *Il10* loci (**Fig. 3.4E**), which in many cases overlapped with Treg-specific TET-regulated methylation patterns (**Fig. 3.3E**), and corresponded well with gene expression changes among the three cell types (**Fig. 3.4F**).

The *Tiam1, Tox* and *Rpl24* loci also displayed Treg-specific and TET-regulated accessible regions (**Fig. A3.4A-C***, left panels*). The accessibility changes corresponded well with gene expression changes for Tiam1 and Tox, however, as expected for a component of the ribosome

60S subunit, Rpl24 was constitutively expressed in all three cell types (**Fig. A3.4A-C***, right panels*). As previously reported, the *CNS2* enhancer in the *Foxp3* locus is already accessible in Treg precursor cells (Samstein *et al.* 2012), and its accessibility was only slightly diminished in *Tet2/3* DKO compared to WT Treg cells (**Fig. A3.4D,** *left panels*). *Foxp3* expression was only slightly reduced in *Tet2/3* DKO compared to WT Treg cells (**Fig. A3.4D,** *right panels*; also see **Fig 1F**).

Overall, our results indicate that as with the transcriptional, 5hmC and WGBS changes described in previous sections, Treg-specific chromatin accessibility was significantly diminished in the absence of Tet2 and Tet3.

Vitamin C enhances Treg molecular features in iTreg cells differentiating *in vitro*.

Vitamin C is a well-established activator of TET proteins and other iron- and α-ketoglutarate-dependent dioxygenases (Blaschke *et al.* 2013; Yue & Rao, 2020). We previously showed that compared with iTreg cells differentiated with TGFβ alone, iTreg cells differentiated with TGFβ and Vitamin C showed demethylation of *CNS1* and *CNS2*, two intronic enhancers within the *Foxp3* locus, and increased the mean fluorescent intensity (MFI) and stability of Foxp3 expression (Yue *et al.* 2016). Tet2 and Tet3 were major targets of Vitamin C in iTreg cells since both *CNS2* demethylation and Foxp3 MFI increase were abolished in Vitamin C treated *Tet2/3 DKO* iTreg cells using an ERT2-Cre inducible system (Yue *et al.* 2016). We observed similar effects in *Tet2/3 DKO* iTreg cells mediated by CD4Cre: Vitamin C treatment increased the MFI of Foxp3 expression in WT iTreg cells but not in *Tet2/3* DKO iTreg cells (**Fig. A3.5**). To determine whether Vitamin C could potentiate Treg signatures in differentiating iTreg cells, we analyzed Th0 cells activated with anti-CD3 and anti-CD28 in the absence of any polarizing cytokines or neutralizing anti-cytokine antibodies, and iTreg cells activated with anti-CD3 and anti-CD28 and differentiated under four distinct conditions: TGFβ alone; TGFβ + Vitamin C (VitC); TGFβ + retinoic acid (RA);

and TGFβ + RA + VitC (referred to as Th0; TGFβ; VitC; RA; and RA + VitC for simplicity). We assessed gene transcription and chromatin accessibility changes in these cells by RNA-seq and ATAC-seq; additionally, since RA alone has little or no effect on demethylation of *Foxp3 CNS1* and *CNS2* (Yue *et al.* 2016), we examined 5hmC and DNA methylation changes by CMS-IP and WGBS in iTreg cells differentiated with TGFβ versus TGFβ + RA + VitC.

Hierarchical clustering of gene expression data showed that the samples clustered by cell type (**Fig. S3.2A**). K-means clustering of 713 genes differentially expressed in selected pairwise comparisons (**Fig. 3.5A,** *arrows,* **Table S3.9**) identified clusters with distinct gene expression patterns (cluster 1, genes whose expression was largely suppressed in iTreg cells compared to Th0 cells; cluster 2, genes whose expression in iTreg cells was suppressed by RA; cluster 3, genes induced by TGFβ +RA and further induced by VitC; and cluster 4, genes – including the Treg signature genes *Foxp3*, *Nrp1*, *Ikzf4*, *Ccr6, Il2ra* and *Ctla4*, **Fig. S3.2B** – that showed the strongest induction of gene expression with VitC in the presence or absence of RA). Consistent with enrichment of the IL2/STAT5 signaling gene set and depletion of the E2F target gene set in WT compared to *Tet2/3* DKO Treg cells, the IL2/STAT5 signaling gene set was induced while the E2F target gene set was depleted by Vitamin C treatment (GSEA, **Fig. S3.2C**). Genes upregulated *(i)* in WT Treg cells compared either to conventional $CD4^+$ T cells (Hill *et al.* 2007) or to $CD4^+$ naïve T cells (our data) and *(ii)* in WT Treg cells compared to *Tet2/3* DKO Treg cells (our data), were highly represented in iTreg cells differentiated with TGFβ + VitC or TGFβ + RA + VitC, compared to cells differentiated with TGFβ alone or TGFβ + RA (GSEA, **Fig. 3.5B**). Conversely, genes downregulated in WT Treg cells compared either to conventional $CD4^+$ T cells (Hill *et al.* 2007) or to $CD4^+$ naïve T cells and genes downregulated in WT Treg cells compared to *Tet2/3* DKO Treg cells, were underrepresented in iTregs generated with TGFβ + VitC or TGFβ + RA +

VitC compared to iTregs differentiated with TGFβ alone or TGFβ + RA (GSEA, **Fig. S3.2D**). Thus, Vitamin C acts via TET proteins to promote the transcription of Treg signature genes during iTreg differentiation *in vitro*.

In parallel, we analyzed the chromatin accessibility changes in these cell types by ATAC-seq. K-means clustering of the 892 regions differentially accessible in selected pairwise comparisons (**Fig. 3.5C**, *arrows*, **Table S3.10**) in Th0 and iTreg cells identified clusters with distinct patterns of chromatin accessibility changes: regions that lost chromatin accessibility in all iTreg cells compared to Th0 cells (cluster 1); regions that lost accessibility in iTreg cells differentiated in the presence of Vitamin C compared to Th0 cells or to iTreg cells differentiated in the absence of Vitamin C (cluster 2); regions that were more accessible in iTreg compared to Th0 cells (cluster 3); and regions that were most accessible in iTreg cells differentiated in the presence of Vitamin C (cluster 4) (**Fig. 3.5C**). Vitamin C increased the number of differentially accessible regions in all pairwise comparisons (**Fig. 3.5D**), but RA had minimal effects (**Fig. S3.2E**). The conserved non-coding sequence *CNS2* was the major region in the *Foxp3* locus whose accessibility was increased by treatment with Vitamin C; it was barely accessible, if at all, in Th0 cells or in TGFβ or TGFβ + RA iTregs (**Fig. 3.5E**). Differentially accessible regions (DARs) that were more accessible in iTregs generated with TGFβ + VitC compared to TGFβ alone (**Fig. 3.5F**, *left*), or in TGFβ + RA + VitC iTregs compared to TGFβ + RA iTregs (**Fig. 3.5F**, *right*), were also more accessible in endogenous WT Treg cells compared to either $CD4^+$ naïve T cells or endogenous Treg cells isolated from *Tet2/3* DKO mice. Overall, Vitamin C not only skewed the gene expression profiles of iTreg cells toward those of endogenous Treg cells, but also enhanced the chromatin accessibility changes occurring during iTreg differentiation *in vitro*, to resemble the changes induced by TET enzymes during endogenous Treg differentiation in vivo.

As in all proliferating immune cell types that we have examined – Th1 and Th2 cells, iTreg cells, and B cells, (Lio *et al.* 2019a; Tsagaratou *et al.* 2014; Yue *et al.* 2016) – iTreg cells activated and differentiated with TGFβ alone showed a sharp decrease in 5hmC levels by 3 days and a further drop by 6 days of differentiation (**Fig. S3.2F**). In both cases, this drop was partially prevented by the addition of Vitamin C. In line with this, nearly all of DhmRs (99.9%) were more hydroxymethylated in TGFβ + RA + VitC iTregs compared to iTregs generated with TGFβ alone (**Fig. S3.2G, Table S3.11**). As examples, genome browser tracks and quantified 5hmC signal for the *Foxp3* and *Il2ra* genes are shown in **Fig. S3.2H,I**; again, 5hmC levels in the latter part of gene bodies correlated well with the levels of gene expression (**Fig. S3.2J**). Overall, we conclude that despite the strong global decrease of 5hmC in T cells activated *in vitro*, addition of Vitamin C during TGFβ-induced iTreg differentiation substantially increased genome-wide 5hmC levels.

WGBS analysis of DNA methylation patterns showed that 4,212 CpGs were differentially methylated between iTreg cells differentiated with TGFβ alone and iTreg cells generated with TGFβ + RA + VitC (**Table S3.12**); cells differentiated with TGFβ alone contained more highly-methylated DM-CpGs (3281) than TGFβ + RA + VitC iTreg cells (931, **Fig. 3.5G**). DM-CpGs demethylated in iTreg cells generated with TGFβ + RA + VitC vs TGFβ alone (**Fig. 3.5H**, *left*) were also demethylated in endogenous WT Treg cells, compared to either naïve T cells or endogenous Treg cells from *Tet2/3* DKO mice (**Fig. 3.5H**, *right*). Similarly, DM-CpGs that were demethylated in WT Treg cells compared to either naïve T cells or *Tet2/3* DKO Treg cells (**Fig. 3.5I,J**, *left*) were also demethylated in iTreg cells generated with TGFβ + RA + VitC compared to TGFβ alone (**Fig. 3. 5I,J**, *right*). On the other hand, DM-CpGs that were more methylated in TGFβ + RA + VitC versus TGFβ alone iTreg cells showed no difference in methylation levels when

naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells were compared (**Fig. S3.2K**); and DM-CpGs that were more methylated in WT Tregs compared to naïve T cells or *Tet2/3* DKO Treg cells displayed similar methylation levels in iTreg cells generated with TGFβ + RA + VitC compared to TGFβ alone (**Fig. S3.2L-M**). Together these data emphasize the clear correspondence between TET-dependent genome-wide demethylation of endogenous Tregs *in vivo*, and Vitamin C-dependent genome-wide demethylation of iTregs induced with TGFβ + RA *in vitro*.

Overall, our data demonstrate that Vitamin C potentiates Treg signatures during iTreg differentiation in vitro at the transcriptional, chromatin accessibility, 5hmC distribution and DNA methylation levels.

Increased chromatin accessibility correlates with increased 5hmC and decreased DNA methylation. We examined the relation between 5hmC levels and DNA methylation status in differentially accessible regions (DARs) identified by ATAC-seq (**Fig. 3.6**). The 2,270 regions with higher accessibility in WT Treg cells compared to *Tet2/3* DKO Treg cells showed more 5hmC deposition at the center of the accessible regions in WT Treg cells; correspondingly, DNA methylation (5mC + 5hmC) was lower in WT compared to *Tet2/3* DKO Treg cells in the 11,303 CpGs covered by these DARs (**Fig. 3.6A,** *top*). In contrast, the 1,201 regions with decreased accessibility in WT Treg cells compared to *Tet2/3* DKO Treg cells were depleted of 5hmC in WT Treg cells, and the 4,785 CpGs covered by these DARs showed less pronounced DNA demethylation in WT Treg compared to *Tet2/3* DKO Treg cells in (**Fig. 3.6A,** *bottom*). Moreover, regions located ±1kb of DM-CpGs that were demethylated in WT Tregs compared to either naïve T cells or *Tet2/3* DKO Tregs (**Fig. A3.6A**) were also considerably more accessible in WT Tregs compared to naïve T cells or *Tet2/3* DKO Tregs (**Fig. A3.6B**).

Similarly, the 174 regions with increased accessibility in iTregs differentiated with TGFβ + RA + VitC had increased 5hmC and decreased DNA methylation compared to the same regions in iTregs differentiated with TGFβ alone (**Fig. A3.6B**, *top*). In contrast, the 144 regions with decreased accessibility in the same comparison showed only a minor increase in 5hmC and no changes in DNA methylation (**Fig. A3.6B**, *bottom*). These findings are consistent with our previous data in NKT cells, pro-B cells and mature B cells (Lio *et al.* 2019a; Lio *et al.* 2016; Tsagaratou *et al.* 2017a). The lower magnitude of the change observed in iTregs differentiated *in vitro* likely reflects the sharp decrease in 5hmC levels in T cells upon activation and/or differentiation in cell culture.

Taken together, there is a strong positive correlation of increased 5hmC levels and decreased DNA methylation with increased chromatin accessibility in WT Tregs versus *Tet2/3* DKO Tregs *in vivo* and in iTregs differentiated with TGFβ + RA + VitC versus TGFβ *in vitro*.

Enrichment of STAT motifs and increased STAT5 occupancy in accessible regions of Vitamin C-treated iTreg cells. We used de novo and known motif enrichment analyses to ask whether Vitamin C preferentially increased the genome-wide occupancy of specific transcription factors in iTreg cells. Comparing three sets of differentially accessible regions (DARs) induced by Vitamin C, we observed a striking enrichment for consensus binding sequences (TTCnnnGAA) for STAT transcription factors in DARs that were more accessible in TGFβ + RA + VitC vs TGFβ iTregs; TGFβ + VitC vs TGFβ iTregs; and TGFβ + RA + VitC vs TGFβ + RA iTregs (**Fig. 3.7A**; **Fig. S3.3A**). The key Treg cytokine IL-2, known to be essential for Treg differentiation *in vivo*, is typically added at high concentrations (100 U/ml) to cell culture media during iTreg differentiation *in vitro*, and induces STAT5 phosphorylation and activation through JAK-STAT pathway

activation (Abbas *et al.* 2018; Ross & Cantrell, 2018). We therefore examined the features of iTreg differentiation in response to varying IL-2 concentrations in the presence or absence of RA and Vitamin C (**Fig. 3.7B, Fig. S3.3B-C**). We labeled naïve $CD4^+$ T cells from Foxp3-IRES-eGFP reporter mice with Cell Trace Violet (CTV), a dye that is diluted by half with each cell division; differentiated them into iTregs in media containing TGFβ or TGFβ + RA + VitC with varying amounts of IL-2 (0, 1, 5, 10, 25, 50, 75 and 100 U/ml); and monitored the percentage of Foxp3 (eGFP)-expressing cells, cell proliferation and cell survival on a daily basis. iTregs differentiated with TGFβ + RA + VitC showed a striking increase in the level and stability of Foxp3 expression, compared to iTregs differentiating in the presence of TGFβ alone (**Fig. 3.7B, Fig. S3.3B**). On day 2 of differentiation, over 80% of cells activated and cultured in TGFβ + RA + VitC media expressed Foxp3, compared to only ~40% of cells activated and cultured in media containing TGFβ alone (**Fig. S3.3B**). The increased Foxp3 stability was apparent in TGFβ + RA + VitC iTregs compared to TGFβ iTregs as a function of the number of days of differentiation as well as the number of cell divisions, and was observed in both the absence and presence (100 U/ml) of IL-2 (**Fig. 3.7B-C**). Notably, during the first 5 days of differentiation, the presence or absence of IL2 in culture media had no effect on cell proliferation as assessed by CTV dilution (**Fig. 3.7B**) or the percentage of Foxp3-expressing cells (**Fig. S3.3B**), likely because of endogenous IL2 secretion by the activated $CD4^+$ T cells. The decrease in the percentage of Foxp3-expressing cells at high levels of IL-2 (**Fig. S3.3B**) is consistent with the higher rates of cell proliferation under these conditions. Thus, as previously noted (Yue *et al.* 2016), Vitamin C is a major regulator of Foxp3 levels and stability in differentiating iTregs.

Addition of RA and Vitamin C also led to a moderate increase in cell viability (**Fig. 3.7D; Fig. S3.3C-D**). Starting on day 6, the percentage of viable iTreg cells correlated with the amount

of IL-2 in culture for both TGFβ alone and TGFβ + RA + VitC conditions: in particular, survival at low IL-2 concentrations (1 and 5 U/ml) was enhanced in the presence of RA and VitC (**Fig. 3.7D**, **Fig. S3.3C-D**). In addition, iTregs differentiated with TGFβ + RA + VitC showed increased phospho-STAT5 levels compared to iTregs differentiated with TGFβ alone; the increase was most obvious at low levels of IL-2 (3-10 U/ml), and was observed in iTreg cells differentiated in the presence of either 0 or 100 U/ml IL-2 (**Fig. 3.8A-B**; **Fig. S3.4A-B**). Consistent with this increase, we also observed increased binding of STAT5 to the *Foxp3 CNS2* enhancer and to a known STAT5-binding region in an intronic region of the *Il2ra* gene (Li *et al.* 2017) (**Fig. 3.8C**). Concomitantly, on day 6 after differentiation with 100 U/ml IL-2, we observed an increase in the expression of *Il2ra* mRNA by RNA-seq in iTregs differentiated with TGFβ + RA + VitC compared to iTregs differentiated with TGFβ alone (**Fig. S3.4C**, *left panel*), as well as increased expression of IL2Rα (CD25) protein by flow cytometry on day 6 after differentiation with either 0 or 100 U/ml IL-2 (**Fig. 3.8D-E**; **Fig. S3.4D-E**). The effect on *Il2rb* mRNA and IL2Rβ (CD122) protein was less pronounced (**Fig. S3.4C,F**), while expression of *Il2rg* chain mRNA and the common gamma chain (γc) protein decreased after differentiation (**Fig. S3.4C,G**). We observed similar effects when iTreg cells were differentiated in the absence of RA. Compared to iTregs differentiated with TGFβ alone, iTreg cells differentiated with TGFβ + VitC also showed an increased expression of both *Il2ra* mRNA by RNA-seq (**Fig. 3.5A**) and IL2Rα (CD25) protein by flow cytometry (**Fig. S3.5B**). iTregs differentiated with TGFβ + VitC also showed increased phospho-STAT5 levels especially at low level of IL-2 (10 U/ml) (**Fig. S3.5C**).

Overall, these data show that the presence of RA and Vitamin C during iTreg differentiation sensitizes the response to IL-2, by increasing both STAT5 phosphorylation and

STAT5 occupancy of two key enhancers in the *Foxp3* and *Il2ra* loci, and also by maintaining the levels of IL2Rα in response to limiting IL-2 stimulation.

3.4 Discussion

We showed previously that the two TET enzymes present at highest levels in mouse Treg cells, Tet2 and Tet3, control DNA demethylation of the *Foxp3 CNS2* enhancer and the stability of Foxp3 expression. To demonstrate these points, we used TET loss-of-function in endogenous Treg cells and TET gain-of-activity in TET activator Vitamin C treated iTreg cells in vitro. Vitamin C functions through TET proteins to mediate *CNS2* demethylation and increase of Foxp3 MFI, as these effects are abolished in TET-deficient iTreg cells treated with Vitamin C (Yue *et al.* 2016) **Fig. A3.5**). In this study, we have extended these results to a genome-wide level, by investigating for the first time the transcriptional profiles, chromatin accessibility landscapes, and DNA methylation/ hydroxymethylation changes in endogenous WT and TET-deficient Treg cells and in TGFβ-induced iTreg cells after addition of Vitamin C. We show that TET enzymes promote the acquisition of the molecular programs of gene expression, chromatin accessibility and DNA methylation that are characteristic of Treg cells in vivo; moreover, treatment of differentiating iTreg cells with the TET activator Vitamin C in vitro mimics these effects, albeit to a lesser extent.

Motif enrichment analysis of differentially accessible and differentially hydroxymethylated regions pointed to a correlation between Treg-specific and TET-regulated epigenetic alterations, identified by comparing WT $CD4^+$ naïve T vs WT Treg cells and WT Treg vs *Tet2/3* DKO Treg cells respectively. Consensus binding sequences of bZIP, RHD and NR transcription factor families were enriched in differentially accessible regions located in Treg-specific/TET-regulated cluster (cluster 2 in **Fig. 3.4C-D**) and in Treg-specific DhmRs (**Fig. S3.1C**). The DNA methylation levels surrounding consensus-binding motifs of these transcription factor families were also Treg-specific and TET-regulated. Specific members of each of these transcription factor families are known essential regulators of Treg development and function. The

consensus binding motifs for most of these transcription factors typically do not contain CpGs (**Fig. 3.3D**), implying that direct modulation of transcription factor binding by TET-dependent demethylation of the binding site is not likely to be a widespread mechanism for TET-dependent regulation of gene expression. It is more likely that the relevant transcription factors recruit Tet2 and Tet3 proteins to promoters and enhancers to maintain their chromatin accessibility and demethylated status (Lio *et al.* 2019a; Lio *et al.* 2016; Sardina *et al.* 2018; Suzuki *et al.* 2017; Wang *et al.* 2015).

Our data also reveal an intriguing connection between TET activity, Vitamin C and responses to IL-2, a key cytokine critical for Treg development, differentiation and function (Abbas *et al.* 2018; Ross & Cantrell, 2018). IL-2 is well-established as essential for Treg function: mice with germline deletion of *IL2*, *IL2ra* and *IL2rb* develop severe autoimmune diseases (Fontenot *et al.* 2005) and mice with Treg-specific ablation of *IL2ra* and *IL2rb* exhibit more severe lethal autoimmunity than mice with germline deletion of the same genes, exhibiting phenotypes similar to those of *Foxp3*-deficient *scurfy* mice (Chinen *et al.* 2016; Toomer *et al.* 2019). Likewise, *STAT5a/b*-deficient mice fail to generate Treg cells and also develop fatal autoimmunity (Snow *et al.* 2003). Thus the IL2-IL2R-STAT5 signaling axis is essential for the functional programming of Treg cells and maintenance of immune homeostasis (Ross & Cantrell, 2018). We observed that TET deletion in vivo resulted in impaired IL2/STAT5 signaling, with the IL2/STAT5 gene set being underrepresented in *Tet2/3* DKO Treg cells compared to WT Treg cells. Conversely, we found that STAT motifs were enriched in differentially-accessible regions induced by Vitamin C, and that iTregs induced in the combined presence of TGFβ, RA and Vitamin C survived better in low IL-2 media and were more responsive to stimulation with low IL-2 concentrations than iTregs induced in the presence of TGFβ alone. This feature was associated with increased levels of

phospho-STAT5, increased occupancy of STAT5 at key *Foxp3* and *Il2ra* enhancers and increased IL2Rα mRNA and protein levels (**Fig. 3.8**).

Treg cells lacking a single genomic region, *Foxp3 CNS2*, show cell cycle-dependent loss of Foxp3 upon *in vitro* stimulation through the TCR (Feng *et al.* 2014b; Li *et al.* 2014). Notably, this effect can be rescued by culturing the *CNS2*-deficient Treg cells in the presence of high concentrations of IL-2 (Feng *et al.* 2014b). Thus the CpG-rich intronic enhancer, *Foxp3 CNS2*, functions as a sensor of the cytokine IL-2 and its downstream target STAT5 to prevent cell cycle-dependent loss of Foxp3 expression in Treg cells *in vivo (Feng et al. 2014b)*. Here we show that *CNS2* is also a very specific sensor of TET activity and the TET activator Vitamin C during *in vitro* iTreg differentiation, based on Vitamin C-induced, TET-dependent and Treg-specific demethylation patterns, Vitamin C-induced and Treg-specific chromatin accessibility landscapes, and Vitamin C-induced STAT5 occupancy at low IL-2 concentrations (*CNS2* contains several STAT5 binding motifs).

TET proteins are increasingly recognized to be responsive to metabolic influences including those arising from diet and nutrition, and TET and Vitamin C deficiencies are both well-established as a risk factor for blood cancers (Huijskens *et al.* 2016; Khaw *et al.* 2001; Lio *et al.* 2019b; Loria *et al.* 2000). Our data emphasize that, in addition, low plasma Vitamin C levels may also predispose individuals to autoimmune disease secondary to decreased function of both thymically-derived and peripherally-generated Tregs. Vitamin C deficiency is expected to reduce the function of all Fe(II) and 2-oxoglutarate-dependent dioxygenases including TET enzymes, and to result in suboptimal function of both thymically-derived and peripherally-generated Tregs. It would be interesting to conduct an epidemiological study of the relation between plasma Vitamin

C levels and autoimmune disease in humans, similar to previous studies relating plasma Vitamin C levels to overall mortality and cancer (Huijskens *et al.* 2016; Khaw *et al.* 2001; Loria *et al.* 2000).

3.5 Materials and Methods

Mice. Foxp3-eGFP reporter mice (strain 006769) were originally obtained from Jackson Laboratory and further backcrossed to B6/C57 background for more than 10 generations. *Tet2*$^{fl/fl}$*Tet3*$^{fl/f}$ Foxp3-eGFP mice and *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP CD4Cre mice were generated in our laboratory using B6/Taconic background "Artemis" ES cells. All mice were on the B6 background and maintained in a specific pathogen-free animal facility in the La Jolla Institute for Immunology. Age and sex of the mice used for each experiment was stated in the methods. Deletion of *Tet2* and *Tet3* was mediated by CD4Cre, therefore in order to minimize the potential cell-extrinsic effects on Treg cells in this animal model, we used young (3-weeks old) mice for all the genomic studies. All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee of the La Jolla Institute for Immunology and were conducted in accordance with institutional guidelines.

RNA-seq library preparation. For WT and *Tet2/3* DKO Treg samples, CD4^{+}eGFP^{+} Treg cells were sorted from 3-weeks old *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP CD4-Cre mice and *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP littermate control mice (males). For Th0 cells and iTreg cells, CD4^{+}eGFP^{-}CD25^{-} CD62L^{hi}CD44^{lo} naïve T cells were sorted from Foxp3-eGFP reporter mice (5-6 weeks old, female) and differentiated into Th0 cells (activated with anti-CD3 and anti-CD28 antibodies in the absence of polarizing cytokines or neutralizing anti-cytokine antibodies) and iTreg cells under four different conditions: TGFβ alone; TGFβ + Vitamin C (VitC); TGFβ + retinoic acid (RA); TGFβ + RA + VitC. eGFP^{-} Th0 cells and eGFP^{+} iTreg cells were sorted on day 6 after differentiation.

Total RNA was isolated from samples using RNeasy plus mini or micro kit (Qiagen). RNA-sequencing libraries were prepared using Truseq stranded mRNA kit (Illumina) according to the manufacture's protocol. Libraries were pooled in equal quantity and sequenced using Illumina

HiSeq 2500 (Illumina) as 50-base-pair paired end reads (for WT and *Tet2/3* DKO Treg samples), 125-base-pair paired end reads (for Th0 and iTreg samples).

RNA-seq analysis. Processing of RNA-seq data was performed similar to that previously described (Barwick *et al.* 2016). Specifically, RNA-seq FASTQ files were quality and adapter trimmed using Trim Galore (v0.4.2) prior to mapping to the UCSC mouse genome mm9 using STAR (v2.5.3a) (Dobin *et al.* 2013) with the following parameters "-quantMode GeneCounts --chimSegmentMin 20 --outSAMtype BAM Unsorted" and the UCSC mm9 KnownGenes gtf transcript file. BAM files from STAR where sorted and duplicates were marked Samtools (v1.7) (Li *et al.* 2009). Quantification of gene expression used exons defined by the 'TxDb.Mmusculus.UCSC.mm9.knownGene' transcript database (v3.2.2) and the summarizeOverlaps function in mode "IntersectionNotEmpty" from the GenomicAlignments package (v1.20.1) (Lawrence *et al.* 2013) in R (v3.6.1) (Gentleman *et al.* 2004). Fragments per kilobase per million reads (FPKM) normalization was performed in R based on the number of exonic mapped reads per gene relative to all mapped autosomal reads. Differentially expressed genes were determined using EdgeR (v3.12.1) (Robinson *et al.* 2010) where a maximum FDR of 0.05 and minimum fold-change of 2 were enforced to determine significance. Differentially expressed genes were organized by K-means clustering using the 'kmeans' function of the stats package in R (v3.6.1). Gene set enrichment analysis (GSEA v4.0.2) was performed using a pre-ranked list based on the $-\log_{10}$(P-value) x sign(log fold-change) for each comparison using the mSigDB curated database (v6.2) (Subramanian *et al.* 2005).

ATAC-seq library preparation. For WT and *Tet2/3* DKO Treg samples, $CD4^{+}eGFP^{+}$ Treg cells were sorted from 3-weeks old $Tet2^{fl/fl}Tet3^{fl/fl}$ Foxp3-eGFP CD4-Cre mice and $Tet2^{fl/fl}Tet3^{fl/fl}$ Foxp3-eGFP littermate control mice (females). For Th0 cells and iTreg cells,

$CD4^+eGFP^-CD25^-CD62L^{hi}CD44^{lo}$ naïve T cells were sorted from Foxp3-eGFP reporter mice (5-6 weeks old, male) and differentiated into Th0 cells and iTreg cells under four different conditions: TGF alone; TGF + Vitamin C (VitC); TGF + retinoic acid (RA); TGF + RA + VitC. $eGFP^-$ Th0 cells and $eGFP^+$ iTreg cells were sorted on day6 after differentiation.

Cells were counted after sorting and 50,000 cells were washed once with 120 µl ice-cold 1 X PBS and spin down at 500g for 10mins at 4°C. Cell pellets were resuspended in 50 µl cold lysis buffer (10mM Tris-HCl pH7.4, 10mM NaCl, 3mM MgCl2, 0.1% NP-40) and immediately spin down at 500g for 15mins at 4°C. The supernatant was discarded and the nuclei pellets was resuspended in 25 µl of transposition mix (12.5 µl 2 X Nextera Tagment DNA Buffer, 1.25 µl Nextera Tagment DNA Enzyme, 11.25 µl H2O), incubated at 37°C for 30mins. DNA was purified using Zymo DNA Clean and Concentrator Kit (Zymo Research). Samples were then amplified using KAPA HiFi HotStart ReadyMix (KAPA Biosystems) for 11 cycles and purified using Ampure XP Beads (Beckman Coulter). The libraries were sequenced as 50-base-pair paired end reads using Illumina Hiseq 2500 (Illumina).

ATAC-seq analysis. ATAC-seq FASTQ files were quality and adapter trimmed using Trim Galore (v0.4.2) prior to mapping to the UCSC mouse genome mm9 using Bowtie2 (v2.2.6) (Langmead & Salzberg, 2012) with default parameters. Mapped SAM files were converted to BAM and putative PCR duplicates were marked with Samtools (v1.7) (Li *et al.* 2009). Regions of chromatin accessibility were determined using MACS2 (v2.1.0.20151222) with the following settings "-g mm -q 0.01" (Zhang *et al.* 2008). The union of accessible regions in all samples were determined using the 'GenomicRanges' package (v1.36.1) in R and data was normalized to reads in peaks per million (RPM) as previously described (Barwick *et al.* 2018). Differentially accessible regions were determined using EdgeR (v3.26.6) with a maximum FDR of 0.05 and a minimum

fold-change of 2. K-means clustering was performed using the 'kmeans' function of the stats package in R (v3.6.1) and plotted using the custom code similar to the heatmap2 function in R. Genome plots were made with custom R code using the 'rtracklayer' package (v1.44.4) (Lawrence *et al.* 2009). ATAC-seq data for Naïve CD4 T cells made use of ImmGen samples GSM2692322 and GSM2692323, which were used for comparison but not included in differential analyses. Transcription factor motifs were determined using the scanMotifGenomeWide.pl script of HOMER software (Heinz *et al.* 2010) and motifs enriched in differentially accessible regions were determined by Fisher's exact test with FDR correction.

CMS-IP library preparation. $CD4^+eGFP^+CD25^+$ WT Treg cells and $CD4^+eGFP^-CD25^-CD62L^{hi}CD44^{lo}$ naïve T cells were sorted from Foxp3-eGFP reporter mice (5-6 weeks old, male). For iTreg cells, $CD4^+eGFP^-CD25^-CD62L^{hi}CD44^{lo}$ naïve T cells were sorted from Foxp3-eGFP reporter mice (5-6 weeks old, male) and differentiated into iTreg cells under TGFβ alone and TGFβ + RA + VitC conditions. $eGFP^+$ iTreg cells were sorted on day6 after differentiation.

Sodium bisulfite treatment converts 5hmC into CMS (cytosine-5-methylenesulfonate), which is then recognized by anti-CMS serum with high sensitivity and selectivity (Pastor *et al.* 2011). CMS-IP was performed as previously described (Huang *et al.* 2012). Briefly, genomic DNA was sheared using Covaris S2 and was spiked with two 210-base-pair PCR amplicons (C amplicon and 5hmC amplicon) at a ratio of around 1:20,000. DNA was purified with Ampure XP beads (Beckman Coulter) and processed with NEBNext End Repair and dA-Tailing Modules (NEB), and ligated to methylated Illumina Adaptors using NEBNext Quick Ligation Module (NEB). DNA with ligated adaptors was then treated with sodium bisulfite (MethylCode, Thermo Fisher Scientific) for 4hrs, denatured and immunoprecipitated with anti-CMS serum (input samples were reserved as 1% of total DNA before immunoprecipitation). Samples for immunoprecipitated DNA

and input DNA were then purified using Phenol/Chloroform and amplified with index primers using KAPA HiFi HotStart Uracil+ Ready Mix (KAPA Biosystems). The samples after PCR amplification were then purified using Ampure XP Beads (Beckman Coulter) and sequenced as 50- or 100-base-pair paired end reads using Illumina Hiseq 2500 (Illumina).

The spike-in amplicons were generated using lambda phage DNA as template and with dNTP mix or 5-Hydroxymethylcytosine dNTP mix (Zymo Research), respectively. Therefore, the cytosines in the C amplicon were unmethylated (used to monitor the bisulfite conversion efficiency), while the cytosines in the 5hmC amplicon were 5-hydroxymethylated (used for normalization of the number of reads to the global content of 5hmC levels in different samples). Reads that mapped to the 5hmC spike-in control in the CMS-IP samples reflected the global content of 5hmC in each sample: for example, iTreg cells induced with TGFβ had the lowest level of 5hmC assessed by CMS dot blot, thus the DNA samples from these cells had the largest number of reads that mapped to the spike-in control.

CMS-IP analysis. CMS-IP FASTQ files were quality and adapter trimmed using Trim Galore (v0.4.2). Trimmed FASTQ files were mapped to the *in-silico* bisulfite converted mm9 genome with the unmethylated and 5hmC amplicons added as artificial chromosomes. Mapped SAM files were converted to BAM and putative PCR duplicates were marked with Samtools (v1.7) (Li *et al.* 2009). Normalization was performed using the 5hmC spike-ins. Specifically, the proportion (P_s) of 5hmC spike-in reads to non-duplicate autosomal reads were determined for each sample (S) as follows:

$$P_s = \frac{5hmC\ spikein\ reads}{autosomal\ reads}$$

This proportion of 5hmC spike-in reads was used to create a 5hmC normalization ratio for CMS-IP and input samples separately (N_s), using the following formula:

$$N_s = \frac{\bar{P}}{P_s}$$

Where $\bar{P}$ represents the average proportion of 5hmC reads of all CMS-IP or all input samples. A final 5hmC normalization ratio (N_f) was calculated by correcting the 5hmC normalization ratio of CMS-IP samples (N_c) by the 5hmC normalization ratio for their respective input samples (N_i) as follows:

$$N_f = \frac{N_c}{N_i}$$

Subsequently, CMS-IP reads were normalized for reads per million multiplied by the final 5hmC normalization ratio (N_f). A normalization ratio of 1 was used for input samples. The resultant normalization reflected the global content of 5hmC in each sample. Regions of enriched 5hmC were determined using MACS2 (v2.1.0.20151222) with the following settings "-g mm --broad -q 0.01" (Zhang *et al.* 2008). The union of 5hmC enriched regions were determined using the 'GenomicRanges' package (v1.36.1) in R and differentially accessible regions were determined using EdgeR (v3.26.6) with an FDR ≤0.05 and a fold-change ≥2.

WGBS library preparation. For WT Treg and $CD4^+$ naïve T cell samples, $CD4^+eGFP^+CD25^+$ Treg cells and $CD4^+eGFP^-CD25^-CD62L^{hi}CD44^{lo}$ naïve T cells were sorted from Foxp3-eGFP reporter mice (5-6 weeks old, male). For *Tet2/3* DKO Treg sample, $CD4^+eGFP^+CD25^+$ Treg cells were sorted from *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP CD4-Cre mice (3-weeks old, male). For iTreg cells, $CD4^+eGFP^-CD25^-CD62L^{hi}CD44^{lo}$ naïve T cells were sorted

from Foxp3-eGFP reporter mice (5-6 weeks old, male) and differentiated into iTreg cells under TGFβ alone and TGFβ + RA + VitC conditions. $eGFP^+$ iTreg cells were sorted on day6 after differentiation.

Genomic DNA was isolated using PureLink Genomic DNA Mini Kit (ThermoFisher Scientific, K182001) or FlexiGene DNA Kit (Qiagen, 51206). Unmethylated lambda DNA (Promega, D1521) was spiked into the genomic DNA samples at a ratio of 1:200 to monitor the bisulfite conversion efficiency. The samples were then sheared using Covaris S2, purified with Ampure XP beads (Beckman Coulter) and processed with NEBNext End Repair and dA-Tailing Modules (NEB), and ligated to methylated Illumina Adaptors using NEBNext Quick Ligation Module (NEB). DNA with ligated adaptors was then treated with sodium bisulfite (MethylCode, Thermo Fisher Scientific) for 4hrs and amplified with index primers using KAPA HiFi HotStart Uracil+ Ready Mix (KAPA Biosystems). The samples after PCR amplification were then purified using Ampure XP Beads (Beckman Coulter) and sequenced as 125- or 250-base-pair paired end reads using Illumina Hiseq 2500 (Illumina).

WGBS analysis. WGBS FASTQ files were quality and adapter trimmed using Trim Galore (v0.4.2) prior to mapping to the *in silico* bisulfite converted UCSC mouse genome mm9 using Bismark (v0.19.0) (Krueger & Andrews, 2011) to call Hisat2 (v2.1.0). Mapped BAM files were sorted with Samtools (v1.7) (Li *et al.* 2009) and CpG methylation calls were extracted using bespoke code as previously described (Barwick *et al.* 2016). CpG methylation calls were collapsed to one strand and DNA methylation coverage was determined as CpGs with ≥5x coverage per group resulting in DNA methylation data at 18,531,012 CpGs with an average coverage of 14.9x per CpG per sample (range 11.7x – 17.8x). Differentially methylated loci were determined using DSS (v2.32.0) (Feng *et al.* 2014a) where significance was determined by an FDR ≤0.05 and a

mean CpG methylation difference of 0.2. K-means clustering was performed in R (v3.6.1) and CpG loci were annotated using GREAT (v4.0.4) (McLean *et al.* 2010).

Transcription factor binding motif analyses. Enrichment of transcription factor consensus binding motifs was performed as previously described (Barwick *et al.* 2018). Briefly, consensus binding motifs were determined using HOMER software (http://homer.ucsd.edu/homer/; v4.10). Specifically, the 'scanMotifGenomeWide.pl' script was used with default parameters to catalogue motifs in the mm9 genome. Enrichment of motifs in DhmRs, DM-CpGs and DARs relative to assay coverage was determined using Fisher's exact test with a FDR correction.

***In vitro* iTreg differentiation.** $CD4^+CD25^-CD62L^{hi}CD44^{lo}eGFP^-$ naïve T cells were FACS sorted from spleen and lymph nodes of Foxp3-IRES-eGFP reporter mice (6- to 8-week-old) or *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP CD4Cre mice (2.5- to 3-week-old), and differentiated into iTregs with plate-bound anti-CD3 (clone 2C11) and anti-CD28 (clone 37.51) antibodies at 1 μg/ml in the presence of 2 ng/ml recombinant human TGFβ (Peprotech) and 100U/ml rhIL-2. For the conditions with retinoic acid (RA, sigma) and/or Vitamin C (VitC, sigma), 100 nM of RA, 100 μg/ml of VitC were added into the culture.

For iTreg cell differentiation in the presence of different IL-2 concentration, $CD4^+$ naïve T cells were sorted and labeled with the proliferation dye cell trace violet (Thermo Fisher Scientific) and differentiated into iTreg cells in the presence of TGFβ alone or TGFβ + RA + VitC at different concentration of rhIL-2 (0, 1, 5, 10, 25, 50, 75 and 100U/ml). The percentage of Foxp3 $(eGFP)^+$ cells, the percentage of viable cells (assessed using eBioscience fixable viability dye eFluor780) and cell proliferation (assessed using the cell proliferation dye cell trace violet) were monitored from day 2 to day 9 after differentiation.

Cell sorting and flow cytometry. Single-cell suspensions were prepared from spleen and lymph nodes for cell sorting. iTreg cells were collected from *in vitro* cultures for staining or cell sorting. The anti-mouse antibodies used for sorting of $CD4^+CD25^-CD62L^{hi}CD44^{lo}eGFP^-$ naïve T cells and $CD4^+$Foxp3 $(eGFP)^+$ or $CD4^+$Foxp3 $(eGFP)^+CD25^+$ Treg cells are as following (Clone name, conjugated fluorescence, dilution, manufacturer and catalog number shown in brackets): CD4 (RM4-5, PerCP-Cy5.5, 1:200, Biolegend, #100540); CD25 (PC61, APC, 1:200, Biolegend, #102012); CD62L (MEL-14, BV421, 1:400, Biolegend, #104436); CD44 (IM7, PE, 1:200, Biolegend, #103008). The gating strategy used for cells sorting was shown in **Fig. A3.7**.

For analysis of the expression levels of IL2R subunits from day 3 to day 6 after primary differentiation of iTregs in the presence of TGFβ alone or TGFβ + RA + VitC, single-cell suspensions were stained with anti-mouse antibodies against the following: CD4 (RM4-5, PerCP-Cy5.5, 1:200, Biolegend, #100540); IL2Rα/CD25 (PC61, Pacific Blue, 1:200, Biolegend, #102021); IL2Rβ/CD122 (TM-β1, PE, 1:75, Biolegend, #123209); IL2Rγ/CD132 (TUGm2, APC, 1:75, Biolegend, #132307). The same IL2Rα/CD25 antibody was used to assess the expression level of IL2Rα/CD25 in iTreg cells differentiated in the presence of TGFβ + VitC, or endogenous Treg cells from *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP CD4-Cre mice and littermate controls.

Analysis of 5hmC levels using anti-CMS dot blots. Anti-CMS dot blot assay was performed as described previously (Yue *et al.* 2016). Briefly, genomic DNA samples were treated with sodium bisulfite using the Methylcode kit (Invitrogen). Bisulfite-treated DNA samples were then denatured in 0.4M NaOH, 10 mM EDTA at 95°C for 10 min, and neutralized by adding an equal volume of cold 2M ammonium acetate (pH 7.0). 2-fold serial dilutions of the denatured DNA samples were spotted on a nitrocellulose membrane in an assembled Bio-Dot apparatus (Bio-

Rad) according to manufacturer's instructions. The membrane was washed with 2X SSC buffer, air-dried and vacuum-baked at 80°C for 2hrs, then blocked with 5% non-fat milk for 1hr and incubated with anti-CMS antibody (1:3,000) overnight at 4°C. After incubating with HRP-conjugated anti-rabbit IgG secondary antibody, the membrane was visualized by enhanced chemiluminescence.

STAT5 phosphorylation. To assess STAT5 phosphorylation, iTreg cells differentiated in the presence of TGFβ alone, TGFβ + VitC or TGFβ + RA + VitC were collected and washed on day 6 after differentiation. 1.5 to $2X10^6$ cells were plated in 100µL culture medium in a round-bottom 96-well plate. The cells were rested for 1 hour at 37°C, then mixed with 100µL of 2X IL-2 cytokine mix to get a final concentration of 0U/ml, 3U/ml, 5U/ml, 10U/ml and 100U/ml of IL-2. The cells were then restimulated with IL-2 at 37°C for 1 hour. After restimulation, cells were fixed with 4% Paraformaldehyde (Electron Microscopy Sciences) at room temperature for 10min. The fixed cells were washed twice with MACS buffer and permeabilized by adding 100µL of ice-cold True-Phos Perm Buffer (Biolegend) and incubated for at least 1 hour at -20°C. The cells were then washed twice with MACS buffer, stained with mouse anti-Stat5 (pY694) (Clone: 47/Stat5 (pY694); AlexaFluor 647; 10uL/assay; catalog #612599; BD Biosciences) and analyzed by flow cytometry on LSR-II. To examine STAT5 phosphorylation in endogenous Treg cells, $CD4^+$ T cells were isolated from *Tet2*$^{fl/fl}$*Tet3*$^{fl/fl}$ Foxp3-eGFP CD4-Cre mice and littermate controls using Dynabeads Untouched Mouse CD4 Cell Kit (Invitrogen) and used for IL-2 restimulation (10U/ml and 100U/ml of IL-2). $CD4^+Foxp3^+$ Treg cells were gated to examine the level of STAT5 phosphorylation.

STAT5 ChIP-qPCR. For ChIP assay, iTreg cells differentiated in the presence of TGFβ alone or TGFβ + RA + VitC were collected on day 7 after differentiation and restimulated without

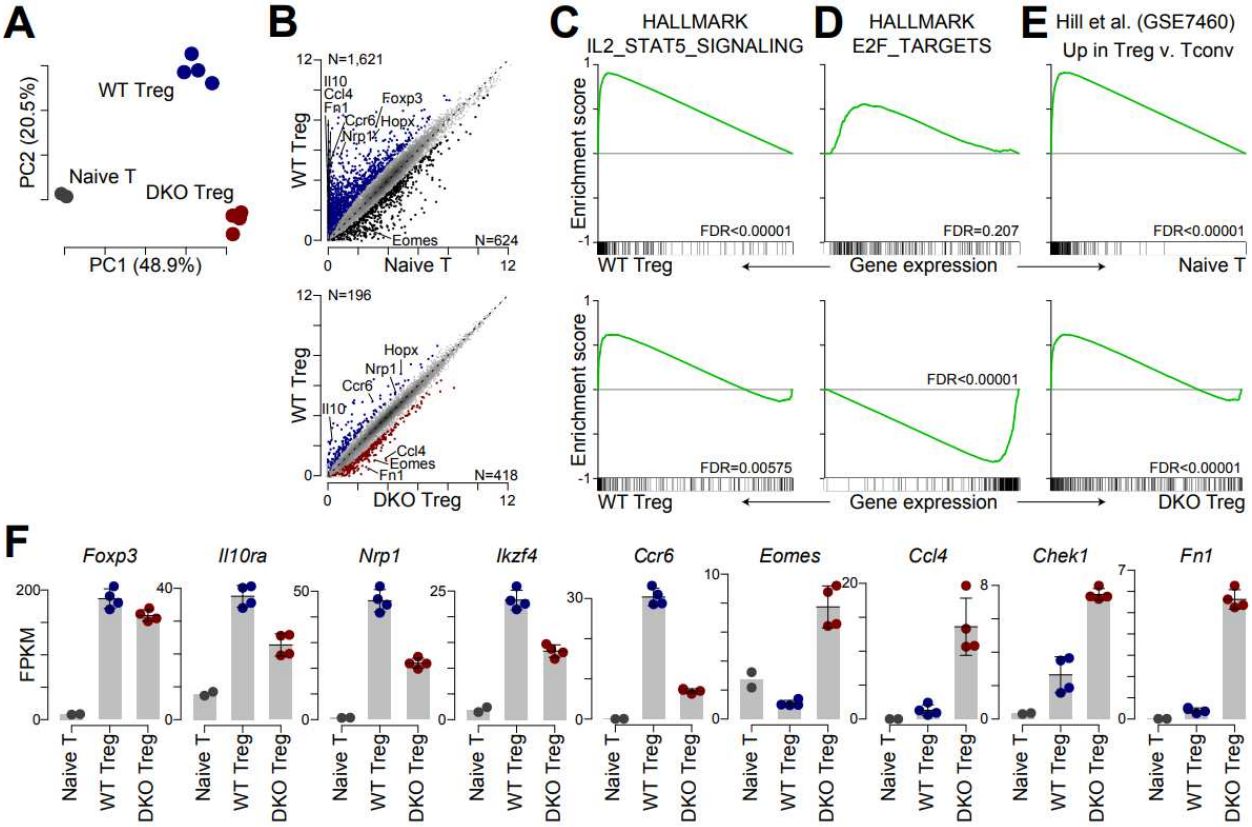
or with 3U/ml IL-2 for 1 hour at 37°C. Cells were then fixed with 1% formaldehyde (Thermo Fisher Scientific) at room temperature for 10 min at 1 X 10^6 cell/ml in culture medium. The reactions were quenched with 125mM glycine and washed twice with ice-cold 1XPBS. Cells were pelleted, snap-frozen with liquid nitrogen and stored at -80°C. All the buffers used in the following steps were supplemented with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific) and phosphatase inhibitor PhosStop (Roche). To isolate nuclei, cell pellets were thawed on ice, resuspended in lysis buffer (50mM HEPES pH7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) and incubated at 4°C for 10min with rotation. The cells were then washed twice with washing buffer (10mM Tris-HCl pH8.0, 200mM NaCl, 1mM EDTA and 0.5mM EGTA), once with shearing buffer (10mM Tris-HCl pH8.0, 1mM EDTA, 1% SDS). After the washing steps, nuclei were resuspended in shearing buffer in 1.5ml Bioruptor Pico Microtubes (5X10^6 cells/100µL) and sonicated using Bioruptor Pico sonication device (Diagenode) for 10 cycles (30 sec ON/30 sec OFF per cycle). After sonication, chromatin was diluted at a 1:10 ratio with RIPA buffer (10mM Tris-HCl ph7.5, 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% SDS, 0.1% sodium dexoycholate, 0.1% Triton X-100), insoluble debris was removed by centrifugation at 4°C. Protein G magnetic beads were pre-loaded with 5µg of anti-STAT5A/B antibody (AF2168; R&D Systems) and incubated with the chromatin (pre-cleared with Protein G magnetic beads) overnight at 4°C with rotation. Bead-bound chromatin was washed twice with RIPA buffer, once with high-salt wash buffer (50mM Tris-HCl pH8.0, 500mM NaCl, 1mM EDTA, 1% NP-40 and 0.1% SDS), once with low-salt wash buffer (10mM Tris-HCl pH8.0, 50mM NaCl, 1mM EDTA) and once with TE buffer (10mM Tris-HCl pH8.0, 1mM EDTA). Chromatin was eluted from beads with elution buffer (100mM $NaHCO_3$, 1% SDS) twice for 20min at room temperature with rotation. 1mg/mL RNaseA was added and incubated at 37°C for 30min. 0.5mg/mL Proteinase K (Ambion)

was then added. The chromatin was de-crosslinked at 65°C overnight with constant shaking. DNA was then purified with ChIP DNA Clean & Concentrator (Zymo Research). Quantitative real-time PCR was performed using FastStart Universal SYBR Green Master mix (Roche) on a StepOnePlus real-time PCR machine (Thermo Fisher Scientific). Primers used for Foxp3 CNS2 and Gmpr are from reference (Zheng *et al.* 2010). Primers used for Il2ra IN1b are as following and are designed based on the reference (Li *et al.* 2017).

Il2ra IN1b forward primer: CCA TTT GAA ACA AGC CCT CAC G
Il2ra IN1b reverse primer: GGG CCA TAT TTT TAG CGG AGC

3.6 Figures

Figure 3.1. Dysregulated transcriptional program in Treg cells in the absence of *Tet2* and *Tet3*.
(A) Principal component analysis of WT $CD4^+$ naïve T cell (*gray*), WT Treg cells (*blue*), and *Tet2/3* DKO Treg cells (*red*). The percent of variation explained by principle components 1 (PC1) and 2 (PC2) are shown in parenthesis. **(B)** Scatter plots showing the comparison of mean gene expression between WT Treg cells vs $CD4^+$ naïve T cells (*upper panel*) and WT Treg cells vs *Tet2/3* DKO Treg cells (*lower panel*). Differentially expressed genes (FDR $\leq$0.05 and fold-change $\geq$ 2) are shown in color (*blue*: WT Treg cells; *black*: $CD4^+$ naïve T cells; *red*: *Tet2/3* DKO Treg cells) and selected genes are labeled in the plots. The scale is $\log_2$(FPKM+1). **(C-E)** Gene set enrichment analysis of gene expression changes between WT Treg cells vs $CD4^+$ naïve T (*upper panels*) and WT Treg cells vs *Tet2/3* DKO Treg cells (*lower panels*) for the hallmark gene sets IL2/STAT5 signaling (**C**), E2F Target genes (**D**) and Treg signature gene set that are upregulated in Treg cells compared to conventional T cells (GSE7460) (**E**). The green line represents the enrichment score with enrichment in WT Treg cells shown to the left and enrichment in $CD4^+$ naïve T (*upper panels*) or Tet2/3 DKO Treg cells (*lower panels*) shown to the right. Overlap of genes in the given gene set is shown below and denoted by black ticks. **(F)** Examples of expression levels of Treg signature genes *Foxp3*, *Il10ra*, *Nrp1*, *Ikzf4* and *Ccr6*; and examples of expression levels of genes upregulated in *Tet2/3* DKO Treg cells *Eomes*, *Ccl4*, *Chek1* and *Fn1*. The Y-axis shows FPKM (Fragment Per Kilobase per Million reads), the X-axis shows the different cell types. Error bars show mean ± S.D. from two ($CD4^+$ naïve T cells) or four (WT and *Tet2/3* DKO Treg) biological replicates.


A
WT Treg
Naive T
DKO Treg
PC2 (20.5%)
PC1 (48.9%)
B
N=1,621
Il10
Ccl4
Fn1
Foxp3
Ccr6
Hopx
Nrp1
Eomes
N=624
WT Treg
Naive T
N=196
Hopx
Nrp1
Ccr6
Il10
Ccl4
Eomes
Fn1
N=418
DKO Treg
C
HALLMARK
IL2_STAT5_SIGNALING
Enrichment score
FDR<0.00001
FDR=0.00575
WT Treg
Gene expression
D
HALLMARK
E2F_TARGETS
FDR=0.207
FDR<0.00001
E
Hill et al. (GSE7460)
Up in Treg v. Tconv
FDR<0.00001
FDR<0.00001
Naive T
DKO Treg
F
Foxp3
Il10ra
Nrp1
Ikzf4
Ccr6
Eomes
Ccl4
Chek1
Fn1
FPKM
Naive T
WT Treg
DKO Treg

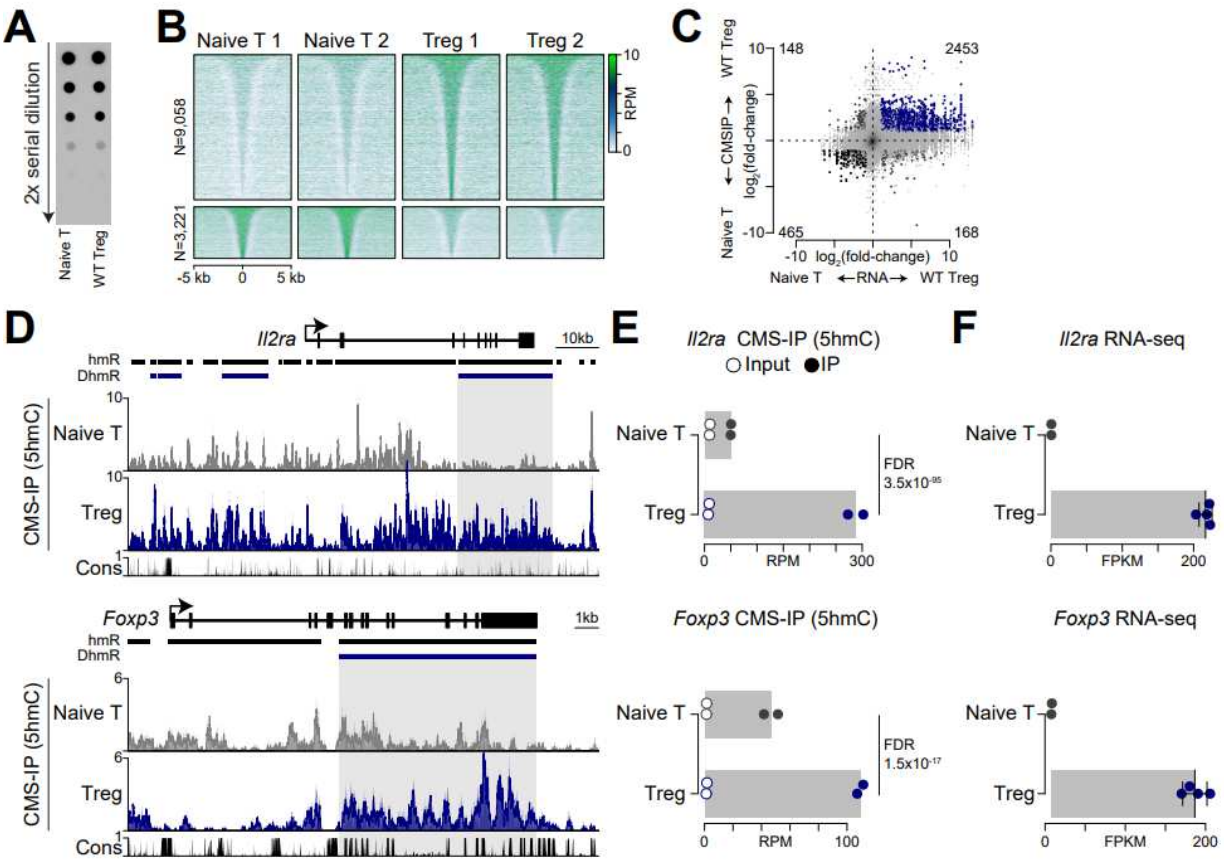


Figure 3.2. 5hmC correlates with gene expression.
(A) Anti-CMS dot blot assay assessing the global amount of 5hmC levels in CD4$^+$ naïve T cells and WT Treg cells. Data are representative of three biological replicates. **(B)** Heatmpas of differentially hydroxymethylated regions (DhmRs) between CD4$^+$ naïve T cells and WT Treg cells. **(C)** Scatter plot of 5hmC changes (Y-axis) and gene expression changes between CD4$^+$ naïve T and WT Treg cells (X-axis). DhmRs associated with a differentially expressed gene are denoted in color (*blue*: WT Treg cells; *black*: CD4$^+$ naïve T cells). **(D)** Genome browser view of CMS-IP data showing CMS (5hmC) distribution at *Il2ra* and *Foxp3* locus for CD4$^+$ naïve T cells and WT Treg cells. Hydroxymethylated regions (hmRs) and DhmRs are shown on the top; conservation (cons) is shown at the bottom. **(E)** The graphs show CMS (5hmC) signal (RPM) in DhmRs denoted by the shaded region in **D** for IP and Input in CD4$^+$ naïve T cells and WT Treg cells. FDRs are calculated based on the Benjamini-Hochberg correction of P-values determined using a negative binomial distribution (see methods). Error bars show mean ± S.D. from two biological replicates. **(F)** The graphs show the expression levels of *Il2ra* (*upper panel*) and *Foxp3* (*lower panel*) in CD4$^+$ naïve T cells and WT Treg cells. Error bars show mean ± S.D. from two (CD4$^+$ naïve T cells) or four (WT Treg) biological replicates.

Figure 3.3. Treg-specific DNA demethylation patterns are impaired in the absence of Tet2 and Tet3.
(A) Principal component analysis of DNA methylation data at 17,937,401 CpGs for $CD4^+$ naïve T cells (*gray*), WT Treg cells (*blue*) and *Tet2/3* DKO Treg cells (*red*). The percentage of variation explained by each component is shown in parentheses. **(B)** Scatter plots of DNA methylation data showing the comparisons between WT Treg vs $CD4^+$ naïve T cells (*left*) and WT Treg vs *Tet2/3* DKO Treg cells (*right*). The differentially methylated CpGs (DM-CpGs) are shown in color (*blue*: WT Treg; *black*: $CD4^+$ naïve T cells; *red*: *Tet2/3* DKO) and the numbers are shown on top of each graph. **(C)** K-means clustering of DNA methylation data for all the samples using CpGs differentially methylated between $CD4^+$ naïve T cells vs WT Treg cells and WT Treg vs *Tet2/3* DKO Treg cells (*left*); the number (N) of CpGs in each cluster is shown in parenthesis. *Middle:* CpG loci proximal to selected genes are annotated. *Right*: the average DNA hydroxymethylation levels at each CpGs as measured by CMS-IP. **(D)** The average DNA methylation levels at selected consensus transcription factor-binding motifs (*gray*: $CD4^+$ naïve T cells; *blue*: WT Treg cells; *red*: *Tet2/3* DKO Treg cells). Standard error is denoted in transparency of the corresponding colors. **(E)** Genome browser view of DNA methylation patterns of *Hopx* and *Il10* loci for $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells. Differentially hydroxymethylated regions (DhmRs) and differentially methylated CpGs (DM-CpGs) are shown on top of the tracks. Conservation track (Cons) and denotation of all CpGs are shown at the bottom. WGBS data are from two biological replicate samples.

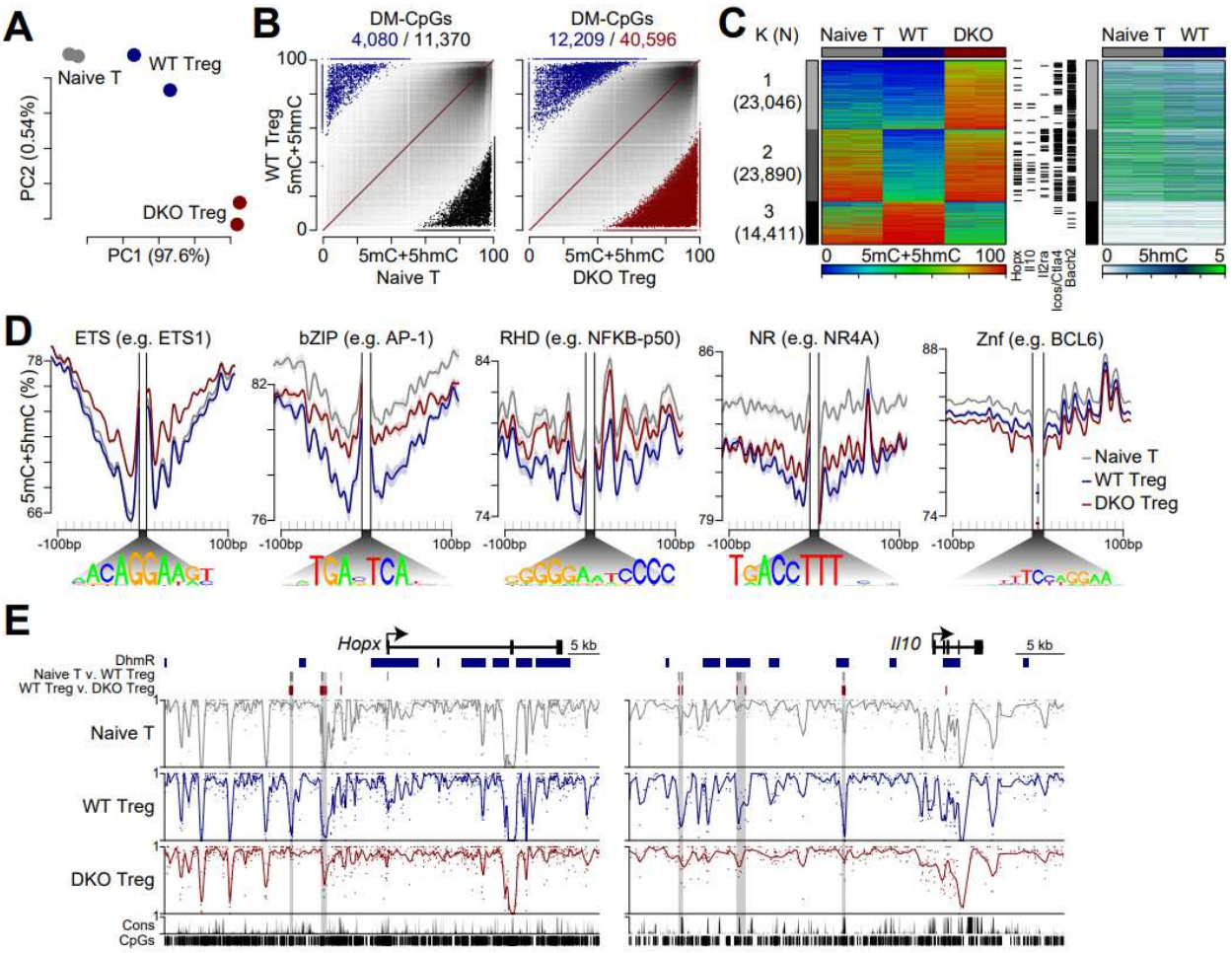

A
Naive T
WT Treg
DKO Treg
PC2 (0.54%)
PC1 (97.6%)
B
DM-CpGs
4,080 / 11,370
DM-CpGs
12,209 / 40,596
WT Treg 5mC+5hmC
5mC+5hmC Naive T
5mC+5hmC DKO Treg
C
K (N)
Naive T WT DKO
1 (23,046)
2 (23,890)
3 (14,411)
0 5mC+5hmC 100
Hopx
Il10
Il2ra
Icos/Ctla4
Bach2
Naive T WT
0 5hmC 5
D
ETS (e.g. ETS1)
bZIP (e.g. AP-1)
RHD (e.g. NFKB-p50)
NR (e.g. NR4A)
Znf (e.g. BCL6)
5mC+5hmC (%)
-100bp
100bp
Naive T
WT Treg
DKO Treg
E
Hopx
Il10
5 kb
DhmR
Naive T v. WT Treg
WT Treg v. DKO Treg
Naive T
WT Treg
DKO Treg
Cons
CpGs

Figure 3.4. Changes of chromatin accessibilities in the absence of Tet2 and Tet3.
(A) Principal component analysis of WT $CD4^+$ naïve T cell (*gray*), WT Treg cells (*blue*), and *Tet2/3* DKO Treg cells (*red*). The percent of variation explained by principle components 1 (PC1) and 2 (PC2) are shown in parenthesis. **(B)** Scatter plot of ATAC-seq data for WT Treg vs *Tet2/3* DKO Treg cells. Differentially accessible regions (DARs) are shown in color and the total number of DARs (N) is shown on the graph. **(C)** K-means clustering of ATAC-seq data for DARs between WT Tregs and *Tet2/3* DKO Tregs. ATAC-seq data for $CD4^+$ naïve T cells is from Yoshida *et al.* 2019. The number of regions in each cluster is denoted. **(D)** Heatmap of transcription factor binding motifs in each K-means cluster. The significance of enrichment (-log10(FDR) is denoted by color scale. Transcription factor families of particular interest are highlighted on top of the heatmap. **(E)** Genome browser view of chromatin accessibilities at *Hopx* and *Il10* loci for $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells. DhmRs, DM-CpGs and DARs (*shaded*) are shown on top of the tracks. Conservation track (Cons) is shown at the bottom. **(F)** Expression levels of *Hopx* and *Il10* in $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells. Error bars show mean ± S.D. from two biological replicate samples.

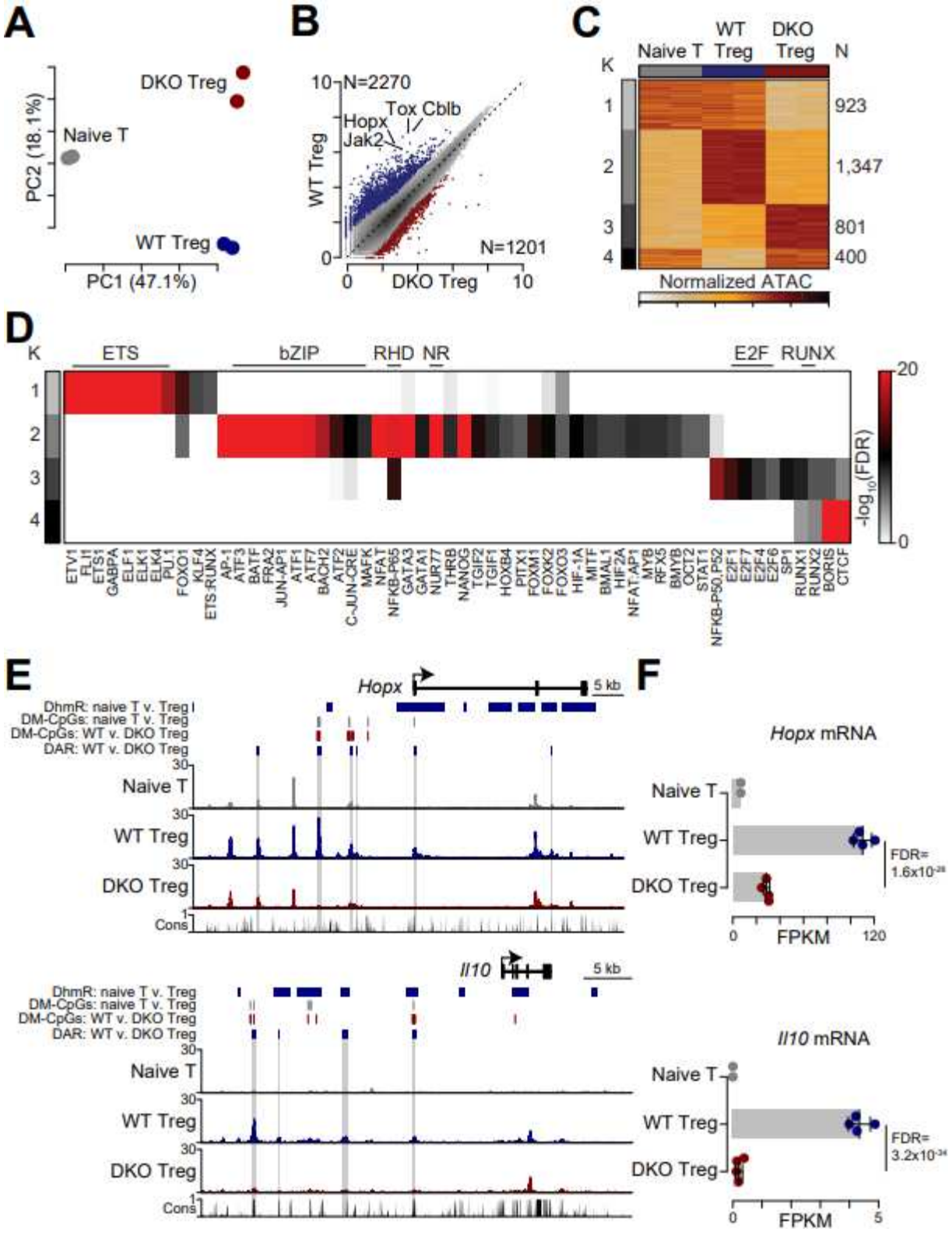

A
DKO Treg
Naive T
WT Treg
PC2 (18.1%)
PC1 (47.1%)
B
N=2270
Tox Cblb
Hopx
Jak2
WT Treg
DKO Treg
N=1201
C
WT Treg
DKO Treg
Naive T
N
K
923
1,347
801
400
Normalized ATAC
D
K
ETS
bZIP
RHD
NR
E2F
RUNX
-log10(FDR)
ETV1
FLI1
ETS1
GABPA
ELF1
ELK1
ELK4
PU.1
FOXO1
KLF4
ETS-RUNX
AP-1
ATF3
BATF
FRA2
JUN-AP1
ATF1
ATF7
BACH2
ATF2
C-JUN-CRE
MAFK
NFAT
NFKB-P65
GATA3
GATA1
NUR77
THRB
NANOG
TGIF2
TGIF1
HOXB4
PITX1
FOXM1
FOXK2
FOXO3
HIF-1A
MITF
BMAL1
HIF2A
NFAT-AP1
MYB
RFX5
BMYB
OCT2
STAT1
NFKB-P50,P52
E2F1
E2F7
E2F4
E2F6
SP1
RUNX1
RUNX2
BORIS
CTCF
E
Hopx
5 kb
DhmR: naive T v. Treg
DM-CpGs: naive T v. Treg
DM-CpGs: WT v. DKO Treg
DAR: WT v. DKO Treg
Naive T
WT Treg
DKO Treg
Cons
Il10
F
Hopx mRNA
FDR= 1.6x10-28
FPKM
Il10 mRNA
FDR= 3.2x10-34

Figure 3.5. Vitamin C promotes Treg molecular features in TGFβ-induced iTreg cells.
(A) K-means clustering of gene expression data for Th0 cells and iTreg cells differentiated *in vitro* under the conditions of TGFβ alone, TGFβ + VitC, TGFβ + RA; TGFβ + RA + VitC using genes differentially expressed between the pairwise comparisons indicated by the arrows on top of the heatmap. The number of genes in each cluster is shown on the left. **(B)** Gene set enrichment analysis (GSEA) of expression changes induced by Vitamin C treatment, for genes upregulated in Treg cells vs conventional T cells (GSE7460, *left panel*), genes upregulated in WT Treg cells vs $CD4^+$ naïve T cells (our data, *middle panel*) and genes upregulated in WT Treg cells vs *Tet2/3* DKO Treg cells (our data, *right panel*), presented as enrichment score (Y-axis) and overlap of genes from each set (X-axis, shown on the bottom) with genes expressed higher in TGFβ + VitC and TGFβ + RA + VitC compared to TGFβ alone and TGFβ + RA, respectively. Data are from two biological replicates. **(C)** K-means clustering of chromatin accessibility data (ATAC-seq) for Th0 cells and *in vitro* differentiated iTregs using regions differentially accessible between the pairwise comparisons indicated by the arrows on top of the heatmap. The number of regions is shown on the right and data are represented as normalized log2 (RPM+1). **(D)** Scatter plot of ATAC-seq data for TGFβ + VitC vs TGFβ iTregs (*left*), TGFβ + RA + VitC vs TGFβ + RA iTregs (*middle*) and TGFβ + RA + VitC vs TGFβ iTregs (*right*). Differentially-accessible regions (DARs) are shown in color and the total number of DARs (N) are shown on each graph. **(E)** Genome browser view of ATAC-seq data showing the accessibility profile at the *Foxp3* locus; the *CNS2* region is highlighted (*gray shading*). DARs are denoted above the cell type tracks with regions of more accessibility identified by color. Scale is reads per million (RPM) and conservation track (Cons) is shown at the bottom. **(F)** Beeswarm plots of mean chromatin accessibility measured as log2 (RPM+1) for regions more accessible in TGFβ + VitC vs TGFβ (*left*) and TGFβ + RA + VitC vs TGFβ + RA (*right*). DARs denote the regions in the respective cell types used for the comparison, and are indicated by the gray shading. P-values are calculated using an analysis of variances with Tukey's post-hoc test and DARs are determined using a negative binomial test with Benjamini-Hochberg correction FDR ≤ 0.05 (see methods). Data are from two biological replicates. **(G)** Scatter plots show the comparisons of whole genome bisulfite sequencing (WGBS) data between TGFβ + RA + VitC iTregs vs TGFβ iTregs. The differentially-methylated CpGs (DM-CpGs) are shown in color and the numbers are shown on top of each graph. **(H-J)** Box-and-whisker plots of DNA methylation levels for DM-CpGs less methylated in TGFβ + RA + VitC iTregs vs TGFβ iTregs (**H**), DM-CpGs less methylated in WT Treg cells vs $CD4^+$ naïve T cells (**I**) and DM-CpGs less methylated in WT Treg cells vs *Tet2/3* DKO Treg cells (**J**). DM-CpGs denotes the CpGs in the respective cell types used for the comparison. Boxplots show the median and quartiles with the whiskers extending to the most extreme data point within 1.5 times the interquartile range. P-values are shown above relevant comparisons and are calculated using an analysis of variances with Tukey's post-hoc test. Data are from two biological replicates.

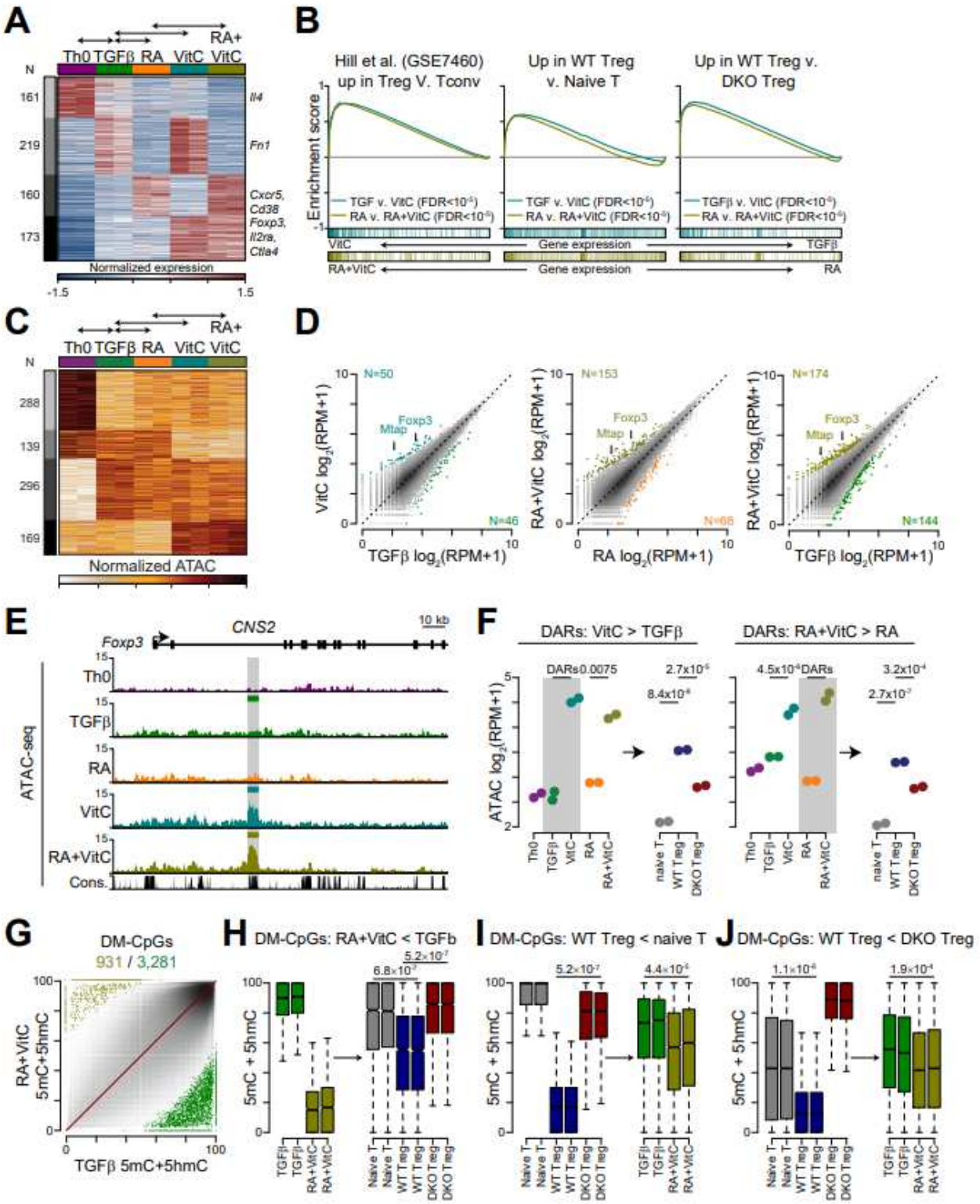

A
RA+
Th0 TGFβ RA VitC VitC
N
161
219
160
173
Il4
Fn1
Cxcr5, Cd38 Foxp3, Il2ra, Ctla4
Normalized expression
-1.5
1.5
B
Hill et al. (GSE7460) up in Treg V. Tconv
Up in WT Treg v. Naive T
Up in WT Treg v. DKO Treg
Enrichment score
TGF v. VitC (FDR<10-5)
RA v. RA+VitC (FDR<10-5)
TGFβ v. VitC (FDR<10-5)
VitC
Gene expression
TGFβ
RA+VitC
RA
C
RA+
Th0 TGFβ RA VitC VitC
N
288
139
296
169
Normalized ATAC
D
N=50
Foxp3
Mtap
N=46
VitC log2(RPM+1)
TGFβ log2(RPM+1)
N=153
N=66
RA+VitC log2(RPM+1)
RA log2(RPM+1)
N=174
N=144
E
CNS2
10 kb
Foxp3
Th0
TGFβ
RA
VitC
RA+VitC
Cons.
ATAC-seq
F
DARs: VitC > TGFβ
DARs: RA+VitC > RA
DARs
0.0075
2.7x10-6
8.4x10-6
4.5x10-6
3.2x10-4
2.7x10-7
ATAC log2(RPM+1)
Th0
TGFβ
VitC
RA
RA+VitC
naive T
WT Treg
DKO Treg
G
DM-CpGs
931 / 3,281
RA+VitC 5mC+5hmC
TGFβ 5mC+5hmC
H
DM-CpGs: RA+VitC < TGFb
5.2x10-7
6.8x10-7
5mC + 5hmC
TGFβ
RA+VitC
Naive T
WT Treg
DKO Treg
I
DM-CpGs: WT Treg < naive T
5.2x10-7
4.4x10-5
J
DM-CpGs: WT Treg < DKO Treg
1.1x10-6
1.9x10-4

Figure 3.6. Correlations among chromatin accessibility, 5hmC distribution and DNA methylation status.
(A) *Left*: differentially accessible regions (DARs) between WT Treg cells and *Tet2/3* DKO Treg cells. 2,270 regions are more accessible in WT Treg cells (*upper panel*); 1,201 regions are less accessible in WT Treg cells (*lower panel*). *Middle:* 5hmC distribution assessed by CMS-IP in the DARs shown on the left. *Right*: DNA methylation status assessed by WGBS in DARs shown on the left: 11,303 CpGs in 2,270 regions more accessible in WT Treg cells (*upper panel*); 4,785 CpGs in 1,201 regions less accessible in WT Treg cells (*lower panel*). $CD4^+$ naïve T cells were used as a reference. P-values are calculated using analysis of variance and boxplots show the median and quartiles with the whiskers extending to the most extreme data point within 1.5 times the interquartile range. **(B)** *Left*: differentially accessible regions (DARs) between TGFβ iTregs and TGFβ + RA + VitC iTregs. 174 regions are more accessible in TGFβ + RA + VitC iTregs (*upper panel*); 144 regions are less accessible in TGFβ + RA + VitC iTregs (*lower panel*). *Middle:* 5hmC distribution assessed by CMS-IP in the DARs shown on the left. *Right*: DNA methylation status assessed by WGBS in DARs shown on the left: 1,123 CpGs in 174 regions more accessible in TGFβ + RA + VitC iTregs (*upper panel*); 678 CpGs in 144 regions less accessible in TGFβ + RA + VitC iTregs (*lower panel*). P-values are calculated using analysis of variance and boxplots show the median and quartiles with the whiskers extending to the most extreme data point within 1.5 times the interquartile range.

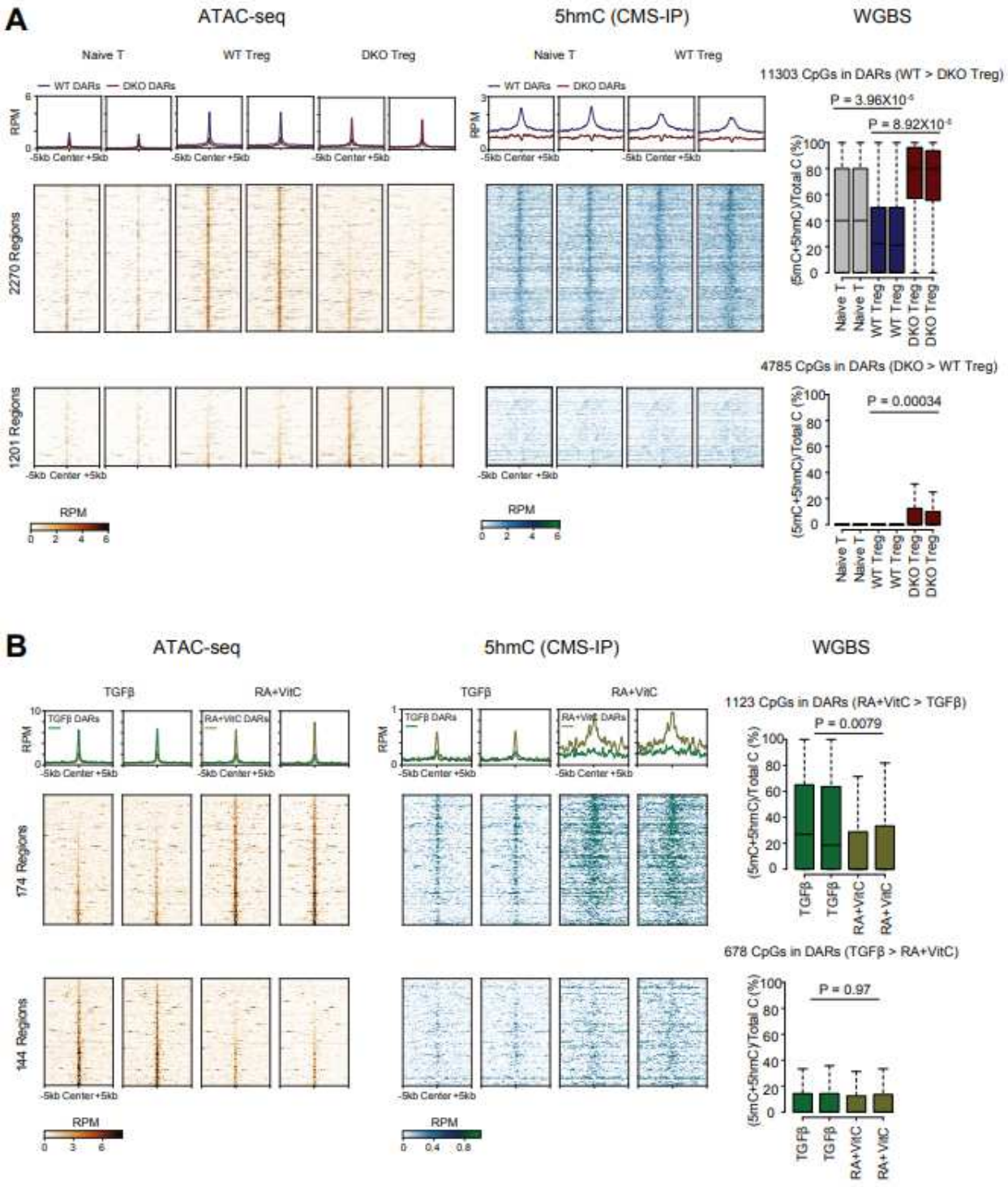

A
ATAC-seq
5hmC (CMS-IP)
WGBS
Naive T
WT Treg
DKO Treg
WT DARs
DKO DARs
RPM
-5kb Center +5kb
2270 Regions
1201 Regions
11303 CpGs in DARs (WT > DKO Treg)
P = 3.96X10-5
P = 8.92X10-5
(5mC+5hmC)/Total C (%)
4785 CpGs in DARs (DKO > WT Treg)
P = 0.00034
B
TGFβ
RA+VitC
TGFβ DARs
RA+VitC DARs
174 Regions
144 Regions
1123 CpGs in DARs (RA+VitC > TGFβ)
P = 0.0079
678 CpGs in DARs (TGFβ > RA+VitC)
P = 0.97

Figure 3.7. Enrichment of STAT motifs in regions that are more accessible with Vitamin C treatment.

(A) Scatter plot of pairwise comparison of chromatin accessibility changes between TGFβ + RA + VitC vs TGFβ iTregs showing the top identified motifs. **(B)** $CD4^+$ naïve T cells were sorted from Foxp3-IRES-eGFP reporter mice, labeled with cell proliferation dye cell trace violet and differentiated into iTreg cells with TGFβ alone or TGFβ + RA + VitC with different amounts of IL-2. Cells in the cultures were analyzed daily by flow cytometry. The two upper panels show representative dot plots of Foxp3 expression vs cell proliferation for TGFβ alone and TGFβ + RA + VitC iTregs with 100 U/ml IL-2 from day 2 to day 9. The two lower panels show representative dot plots of Foxp3 expression vs cell proliferation for TGFβ alone and TGFβ + RA + VitC iTregs with 0 U/ml IL-2 from day 2 to day 9. **(C)** Graph showing the percentage of $Foxp3^+$ cells at each day for TGFβ iTregs and TGFβ + RA + VitC iTregs with either 0 U/ml IL-2 or 100 U/ml IL-2. Error bars show mean ± S.D. from three biological replicates. **(D)** Graphs depicting the percentage of viable cells for TGFβ iTregs (*left*) and TGFβ+RA+VitC iTregs (*right*) differentiated with 0 U/ml, 1 U/ml, 5 U/ml and 10 U/ml IL-2. Error bars show mean ± S.D. from three biological replicates.

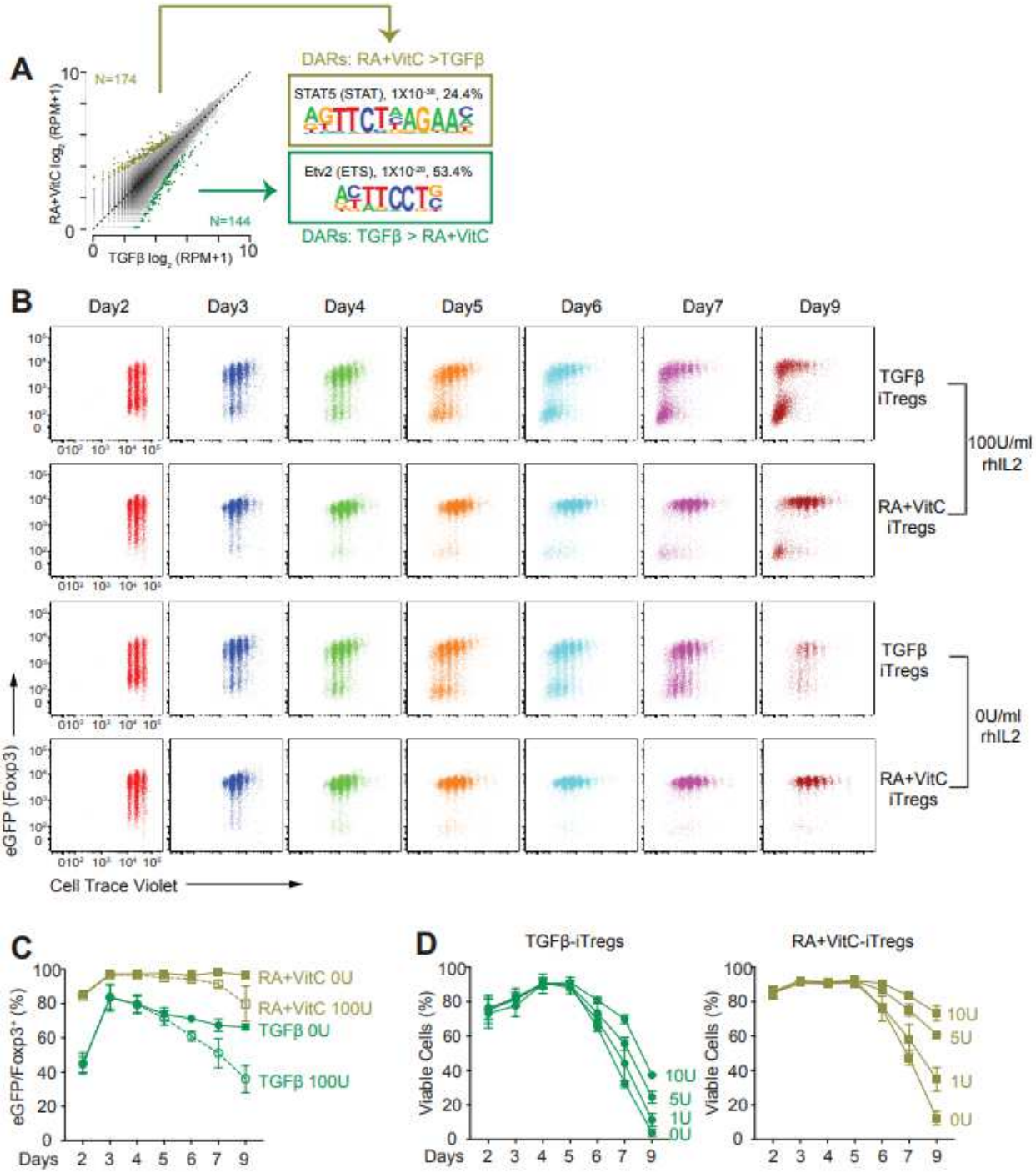
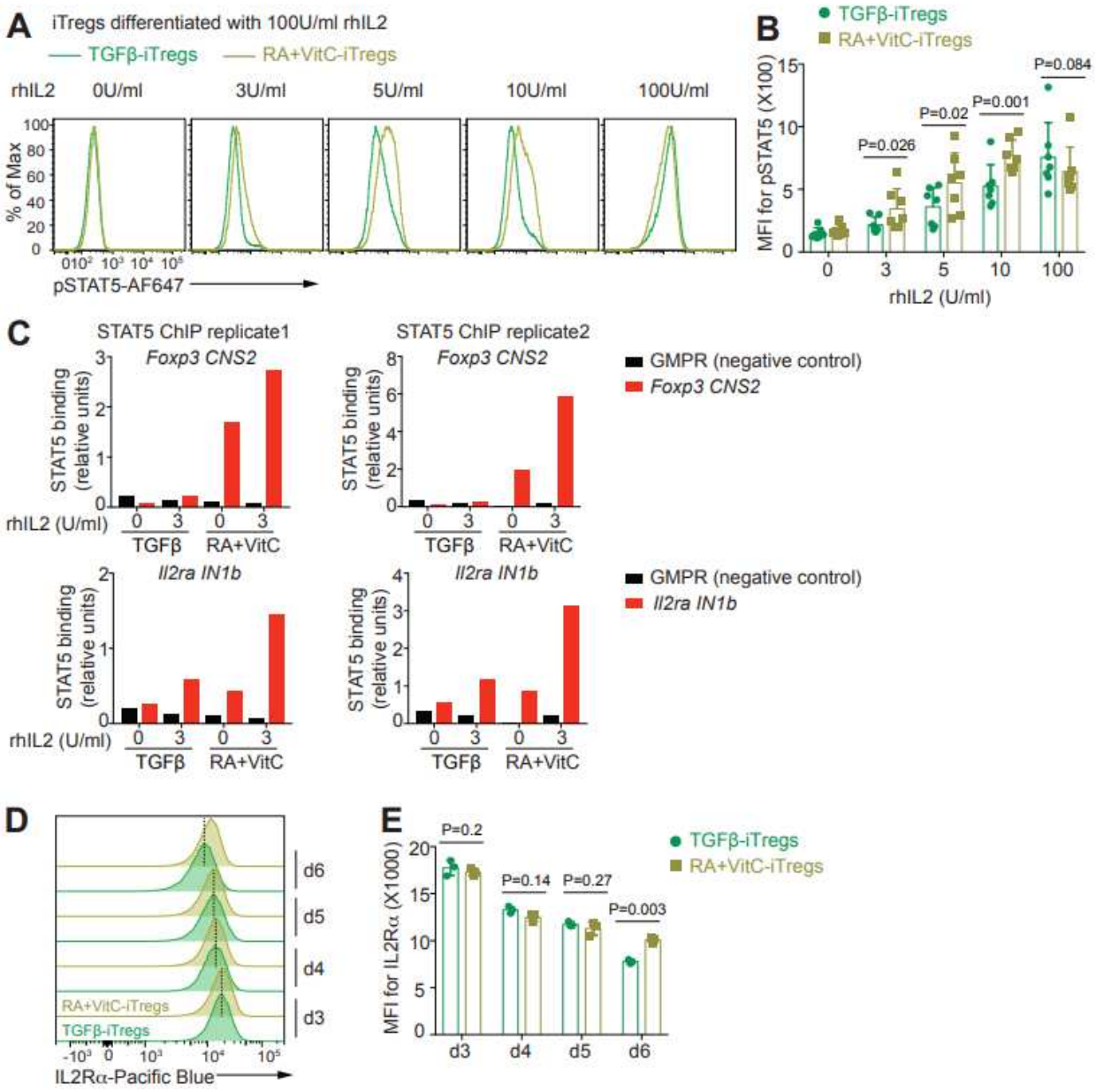

A
DARs: RA+VitC >TGFβ
STAT5 (STAT), 1X10-38, 24.4%
Etv2 (ETS), 1X10-20, 53.4%
DARs: TGFβ > RA+VitC
N=174
N=144
RA+VitC log2 (RPM+1)
TGFβ log2 (RPM+1)
B
Day2
Day3
Day4
Day5
Day6
Day7
Day9
TGFβ iTregs
RA+VitC iTregs
100U/ml rhIL2
0U/ml rhIL2
eGFP (Foxp3)
Cell Trace Violet
C
RA+VitC 0U
RA+VitC 100U
TGFβ 0U
TGFβ 100U
eGFP/Foxp3+ (%)
Days
D
TGFβ-iTregs
RA+VitC-iTregs
Viable Cells (%)
10U
5U
1U
0U
Days

Figure 3.8. Vitamin C facilitates the binding of STAT5 and maintains IL2Rα expression in iTregs.

(A-B) $CD4^+$ naïve T cells were sorted from Foxp3-IRES-eGFP reporter mice and differentiated into iTreg cells with TGFβ alone or TGFβ + RA + VitC in the presence of 100 U/ml IL-2 in cultures. On day 6, cells were restimulated without IL-2 or with 3, 5, 10 and 100 U/ml IL-2 at 37°C for 1hr, and stained for phospho-STAT5. Representative histogram overlay of phospho-STAT5 staining for TGFβ and TGFβ + RA + VitC iTregs after IL-2 restimulation (**A**). Graph depicting the geometric mean fluorescent intensity (MFI) of phospho-STAT5 in TGFβ and TGFβ + RA + VitC iTregs after IL-2 restimulation; each dot represents cells from an individual mouse (**B**). Data are from seven biological replicates. P-values are calculated using two-tailed paired Student's t test and error bars show mean ± S.D. **(C)** ChIP-qPCR showing STAT5 occupancy at *Foxp3 CNS2* and *IL2Rα IN1b* regions in TGFβ and TGFβ + RA + VitC iTregs without or with 3 U/ml IL-2 restimulation. GMPR is a negative control region. Data are from two biological replicates. **(D-E)** Representative histogram overlay (**D**) and geometric MFI (**E**) of IL2Rα staining for TGFβ and TGFβ + RA + VitC iTregs on day 3 to day 6 after differentiation in the presence of 100U/ml IL-2. Data are from three biological replicates. P-values are calculated using two-tailed paired Student's t test and error bars show mean ± S.D.


A
iTregs differentiated with 100U/ml rhIL2
TGFβ-iTregs
RA+VitC-iTregs
rhIL2
0U/ml
3U/ml
5U/ml
10U/ml
100U/ml
% of Max
pSTAT5-AF647
B
TGFβ-iTregs
RA+VitC-iTregs
MFI for pSTAT5 (X100)
P=0.026
P=0.02
P=0.001
P=0.084
rhIL2 (U/ml)
C
STAT5 ChIP replicate1
STAT5 ChIP replicate2
Foxp3 CNS2
Il2ra IN1b
STAT5 binding (relative units)
GMPR (negative control)
rhIL2 (U/ml)
TGFβ
RA+VitC
D
d6
d5
d4
d3
RA+VitC-iTregs
TGFβ-iTregs
IL2Rα-Pacific Blue
E
MFI for IL2Rα (X1000)
P=0.2
P=0.14
P=0.27
P=0.003

3.7 Supplemental and Appendix Figures

Figure S3.1. Related to Fig 3.3.
(A) Box-and-whisker plots showing the global levels of DNA methylation (5mC+5hmC) by WGBS data in all the samples: $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg. Boxplots show the median and quartiles with the whiskers extending to the most extreme data point within 1.5 times the interquartile range. **(B)** DNA methylation levels (5mC+5hmC) over the gene body, categorized on the basis of gene expression levels (RNA-seq analysis) in $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells. **(C)** Heatmap of motifs enriched in naïve T cell-specific DhmRs (*upper*) and Treg cell-specific DhmRs (*lower*). Only motifs of expressed transcription factors (FPKM $\geq$ 1) with FDR $\leq 10^{-10}$ are shown. **(D)** Genome browser view of DNA hydroxymethylation patterns of *Hopx* and *Il10* loci for $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells. Hydroxymethylated regions (hmRs) and differentially hydroxymethylated regions (DhmRs) are shown on top of the tracks. **(E)** Genome browser view of DNA methylation patterns of *Bach2*, *Cd28/Ctla4/Icos* and *Il2ra* loci for $CD4^+$ naïve T cells, WT Treg cells and *Tet2/3* DKO Treg cells. DhmRs and DM-CpGs are shown on top of the tracks. Conservation track (Cons) and denotation of all CpGs are shown at the bottom.

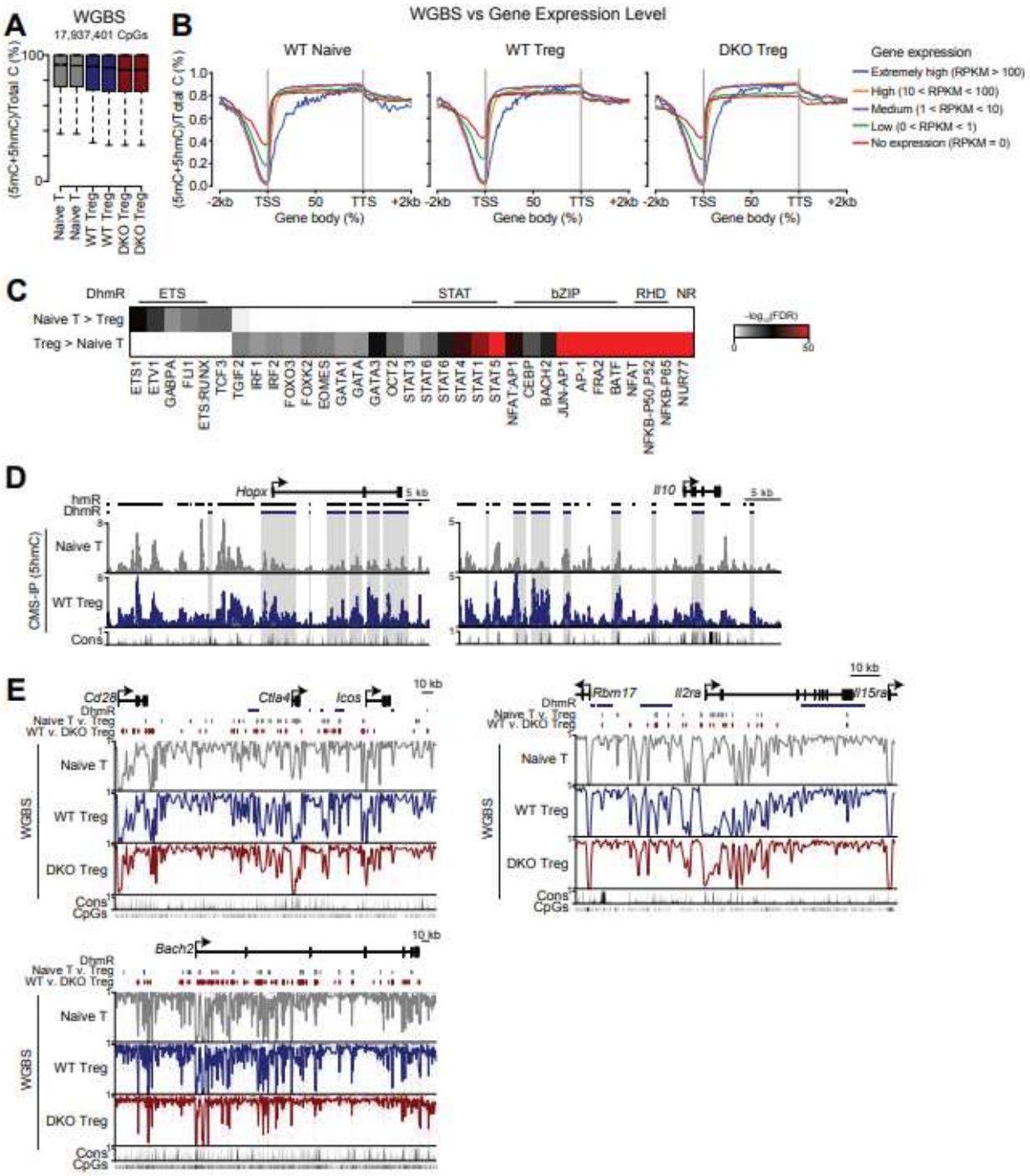
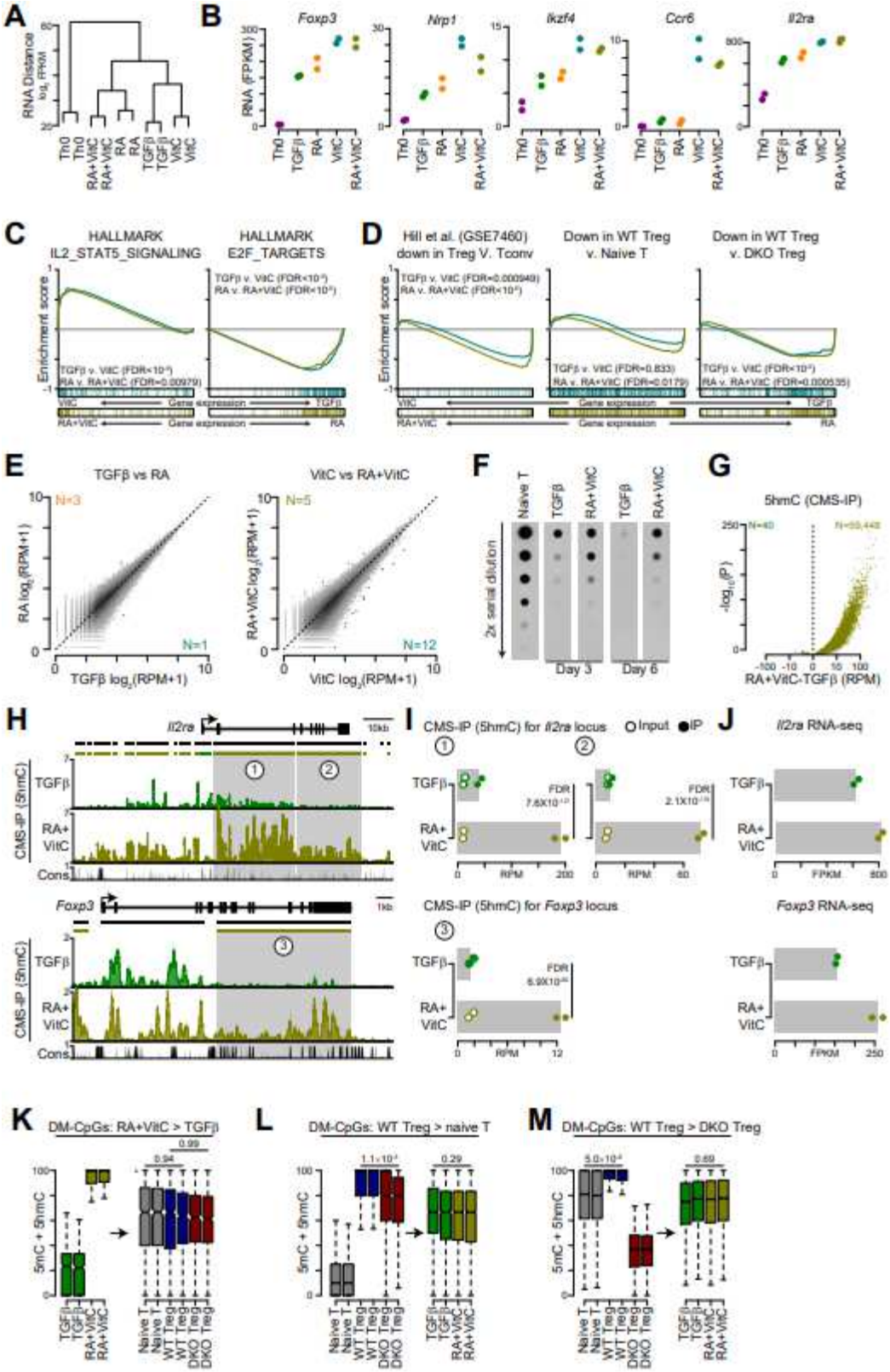

A
WGBS
17,937,401 CpGs
(5mC+5hmC)/Total C (%)
Naive T
Naive T
WT Treg
WT Treg
DKO Treg
DKO Treg
B
WGBS vs Gene Expression Level
WT Naive
WT Treg
DKO Treg
(5mC+5hmC)/Total C (%)
-2kb TSS 50 TTS +2kb
Gene body (%)
Gene expression
Extremely high (RPKM > 100)
High (10 < RPKM < 100)
Medium (1 < RPKM < 10)
Low (0 < RPKM < 1)
No expression (RPKM = 0)
C
DhmR
ETS
STAT
bZIP
RHD
NR
Naive T > Treg
Treg > Naive T
ETS1
ETV1
GABPA
FLI1
ETS-RUNX
TCF3
TGIF2
IRF1
IRF2
FOXO3
FOXK2
EOMES
GATA1
GATA
GATA3
OCT2
STAT3
STAT6
STAT4
STAT1
STAT5
NFAT-AP1
CEBP
BACH2
JUN-AP1
AP-1
FRA2
BATF
NFAT
NFKB-P50,P52
NFKB-P65
NUR77
D
Hopx
Il10
5 kb
hmR
DhmR
CMS-IP (5hmC)
Naive T
WT Treg
Cons
E
Cd28
Ctla4
Icos
10 kb
DhmR
Naive T v. Treg
WT v. DKO Treg
WGBS
Naive T
WT Treg
DKO Treg
Cons
CpGs
Rbm17
Il2ra
Il15ra
Bach2

Figure S3.2. Related to Fig 3.5.

(A) Hierarchical clustering of RNA-seq data for Th0 cells and iTreg cells differentiated under various conditions (two biological replicates for each cell type). **(B)** Examples of expression levels of Treg signature genes *Foxp3*, *Nrp1*, *Ikzf4*, *Ccr6* and *Il2ra* in Th0 and iTreg cells differentiated under various conditions. The Y-axis shows FPKM (Fragment Per Kilobase per Million reads), the X-axis shows the different cell types. Error bars show mean ± S.D. from two biological replicates. **(C)** GSEA of gene expression changes induced by Vitamin C treatment for the hallmark gene sets IL2/STAT5 signaling (*left panel*) and E2F target genes (*right panel*). **(D)** GSEA of gene expression changes induced by Vitamin C treatment, for genes downregulated in conventional T cells with Foxp3 overexpression vs mock (GSE7460, *left panel*), genes downregulated in WT Treg cells vs $CD4^+$ naïve T cells (our data, *middle panel*) and genes downregulated in WT Treg cells vs *Tet2/3* DKO Treg cells (our data, *right panel*), presented as enrichment score and overlap of genes from each set with ordering of the differences in expression. **(E)** Scatter plot of ATAC-seq data for TGFβ + RA vs TGFβ iTregs (*left*) and TGFβ + RA + VitC vs TGFβ + VitC iTregs (*right*). Differentially accessible regions (DARs) are shown in color and the total number of DARs (N) are shown on each graph. **(F)** Anti-CMS dot blot assay assessing the global amount of 5hmC levels in iTreg cells differentiated in the presence of TGFβ and TGFβ + RA + VitC for 3 days and 6 days. Data are representative of three biological replicates. **(G)** Volcano plot showing DhmRs between TGFβ iTregs vs TGFβ + RA + VitC iTregs; the number of DhmRs are shown on the plot. **(H)** Genome browser view of CMS-IP data showing CMS (5hmC) distribution at *Foxp3* and *Il2ra* locus for iTreg cells differentiated with TGFβ alone or TGFβ + RA + VitC. **(I)** The graphs show CMS (5hmC) signal (RPM) in regions identified in **(H)** for IP and Input in each cell type. FDRs are calculated based on the Benjamini-Hochberg correction of P-values determined using a negative binomial distribution (see methods). Error bars show mean ± S.D. from two biological replicates. **(J)** The graphs show the expression levels of *Il2ra* (*upper panel*) and *Foxp3* (*lower panel*) in each cell type. Error bars show mean ± S.D. from two biological replicates. **(K-M).** Box-and-whisker plots of DNA methylation levels for DM-CpGs more methylated in TGFβ + RA + VitC iTregs vs TGFβ iTregs (**K**); DM-CpGs more methylated in WT Treg cells vs naïve T cells (**L**) and DM-CpGs more methylated in WT Treg cells vs *Tet2/3* DKO Treg cells (**M**). P-values are shown above relevant comparisons and are calculated using an analysis of variances with Tukey's post-hoc test. DM-CpGs denotes the CpGs in the respective cell types used for the comparison. Boxplots show the median and quartiles with the whiskers extending to the most extreme data point within 1.5 times the interquartile range. Data are from two biological replicates.


A
RNA Distance log2 FPKM
Th0
Th0
RA+VitC
RA+VitC
RA
RA
TGFβ
TGFβ
VitC
VitC
B
Foxp3
Nrp1
Ikzf4
Ccr6
Il2ra
RNA (FPKM)
C
HALLMARK IL2_STAT5_SIGNALING
HALLMARK E2F_TARGETS
Enrichment score
TGFβ v. VitC (FDR<10-5)
RA v. RA+VitC (FDR=0.00979)
TGFβ v. VitC (FDR<10-5)
RA v. RA+VitC (FDR<10-5)
Gene expression
D
Hill et al. (GSE7460) down in Treg V. Tconv
Down in WT Treg v. Naive T
Down in WT Treg v. DKO Treg
TGFβ v. VitC (FDR=0.000949)
RA v. RA+VitC (FDR<10-5)
TGFβ v. VitC (FDR=0.833)
RA v. RA+VitC (FDR=0.0179)
TGFβ v. VitC (FDR<10-5)
RA v. RA+VitC (FDR=0.000535)
E
TGFβ vs RA
VitC vs RA+VitC
N=3
N=1
N=5
N=12
RA log2(RPM+1)
TGFβ log2(RPM+1)
RA+VitC log2(RPM+1)
VitC log2(RPM+1)
F
Naive T
TGFβ
RA+VitC
TGFβ
RA+VitC
2x serial dilution
Day 3
Day 6
G
5hmC (CMS-IP)
-log10(P)
RA+VitC-TGFβ (RPM)
H
Il2ra
Foxp3
CMS-IP (5hmC)
TGFβ
RA+ VitC
Cons
I
CMS-IP (5hmC) for Il2ra locus
Input
IP
CMS-IP (5hmC) for Foxp3 locus
RPM
J
Il2ra RNA-seq
Foxp3 RNA-seq
FPKM
K
DM-CpGs: RA+VitC > TGFβ
L
DM-CpGs: WT Treg > naive T
M
DM-CpGs: WT Treg > DKO Treg
5mC + 5hmC
Naive T
WT Treg
DKO Treg

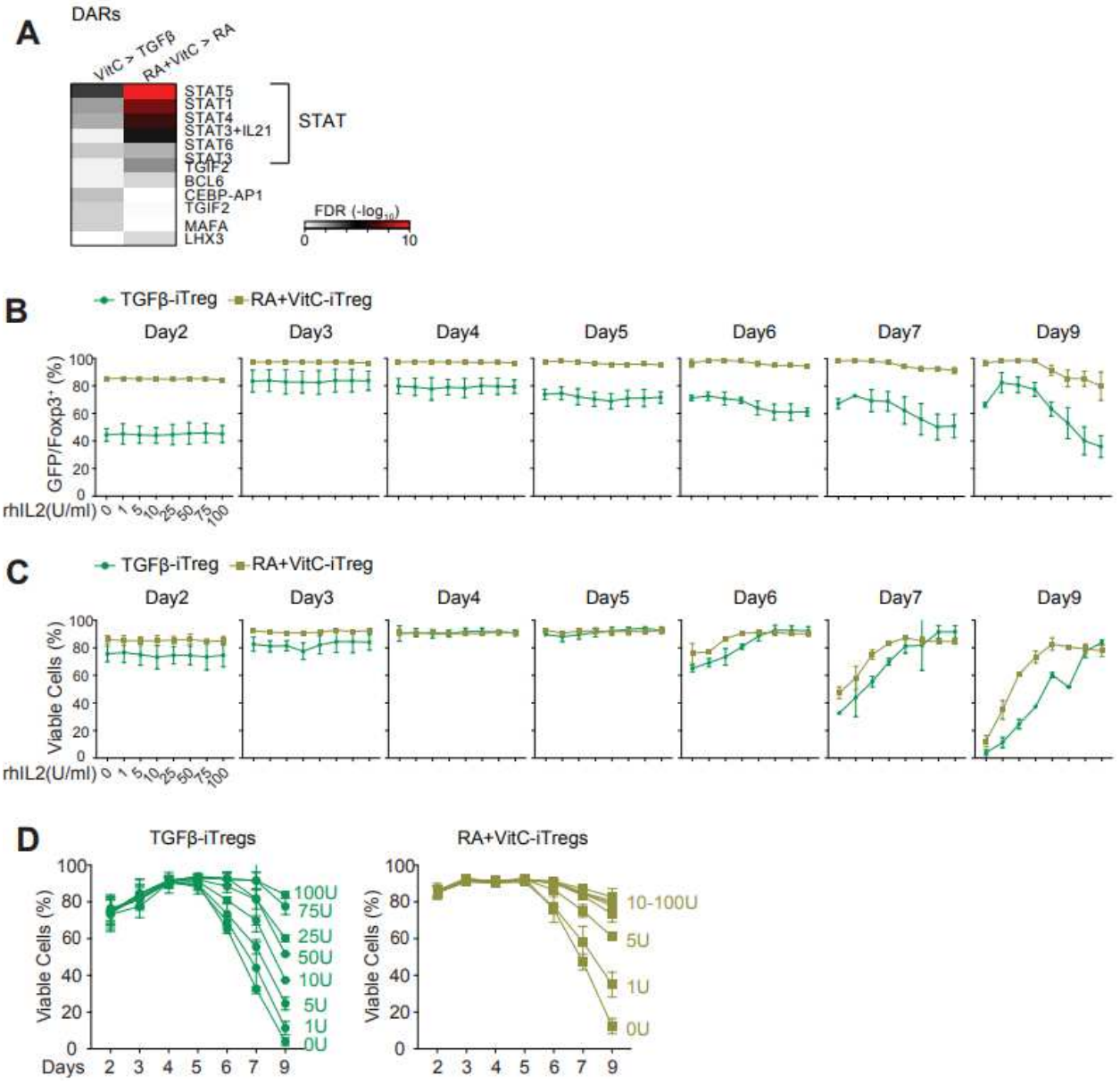
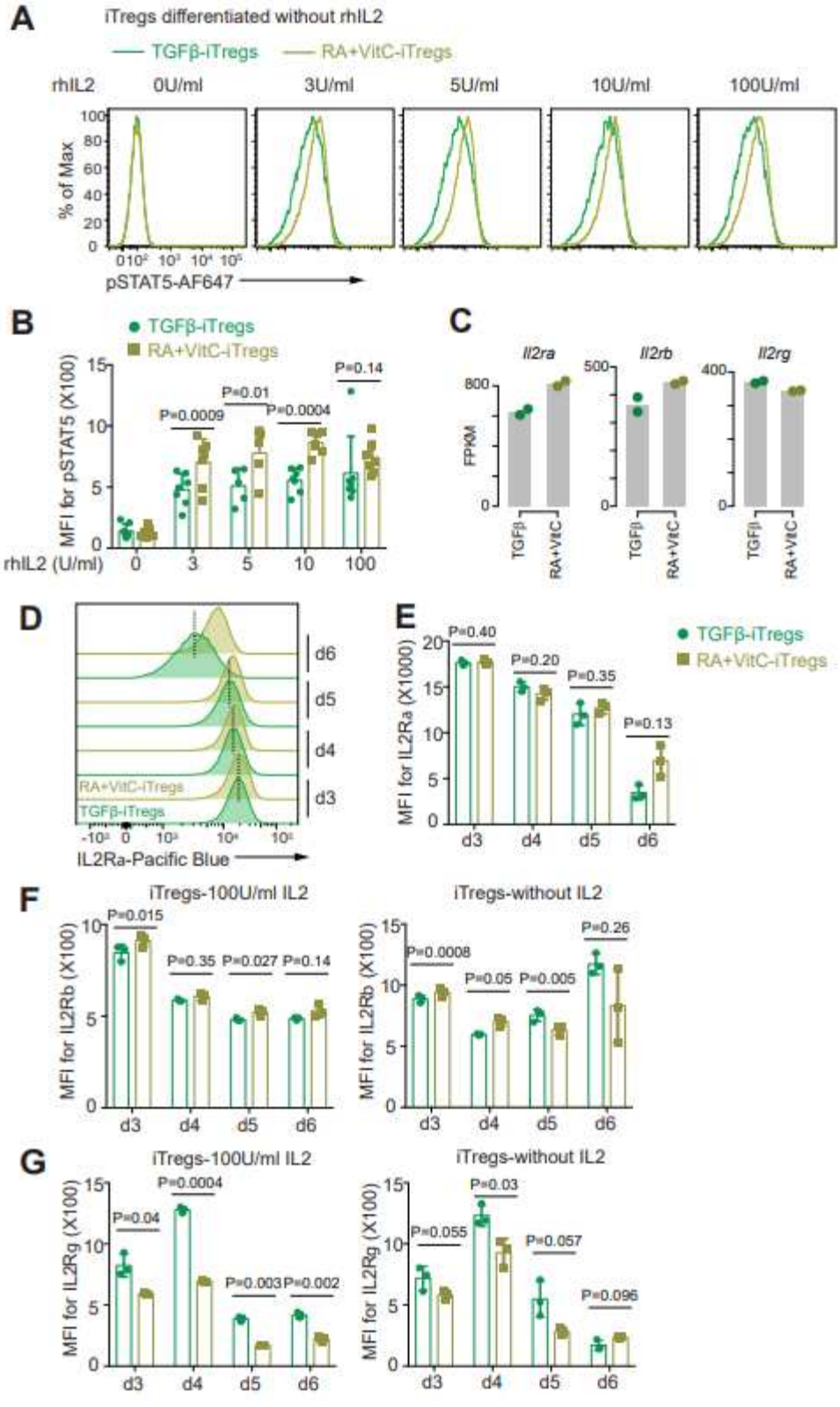


Figure S3.3. Related to Fig 3.7.

(**A**) Heatmap shows the identified motifs at regions more accessible with VitC treatment in comparison of TGFβ+VitC iTregs vs TGFβ iTregs and TGFβ+RA+VitC iTregs vs TGFβ+RA iTregs. (**B-C**) Graphs depicting the percentage of Foxp3$^+$ cells (**B**) and the percentage of viable cells (**C**) with different amount of IL-2 in cultures in TGFβ iTregs and TGFβ+RA+VitC iTregs from day 2 to day 9. Error bars show mean ± S.D. from three biological replicates. (**D**) Graphs depicting the percentage of viable cells for TGFβ iTregs (*left*) and TGFβ+RA+VitC iTregs (*right*) differentiated with different amount of IL-2. Error bars show mean ± S.D. from three biological replicates.

Figure S3.4. Related to Fig 3.8.

(A-B) $CD4^+$ naïve T cells were sorted from Foxp3-IRES-eGFP reporter mice and differentiated into iTreg cells with TGFβ alone or TGFβ + RA + VitC in the absence of IL-2 in cultures. On day 6, cells were restimulated with 0, 3, 5, 10 and 100 U/ml IL-2 at 37°C for 1 hour, and stained for phospho-STAT5. Representative histogram overlay of phospho-STAT5 staining for TGFβ and TGFβ + RA + VitC iTregs after IL-2 restimulation (**A**). Graph depicting the genometric mean fluorescent intensity (MFI) of phospho-STAT5 in TGFβ and TGFβ + RA + VitC iTregs after IL-2 restimulation; each dot represents cells from an individual mouse (**B**). Data are from at least five biological replicates. P-values are calculated using two-tailed paired Student's t test and error bars show mean ± S.D. **(C)** Expression levels of *Il2ra*, *Il2rb* and *Il2rg* in TGFβ iTregs and TGFβ + RA + VitC iTregs by RNA-seq. The Y-axis shows the FPKM (Fragment Per Kilobase per Million reads). Error bars show mean ± S.D. from two biological replicates. **(D-E)** Representative histogram overlay (**D**) and genometric MFI (**E**) of IL2Rα staining for TGFβ and TGFβ + RA + VitC iTregs (differentiated in the absence of IL-2) on day 3 to day 6 after differentiation. Data are from three biological replicates. P-values are calculated using two-tailed paired Student's t test and error bars show mean ± S.D.**(F)** Geometric MFI of IL2Rβ staining for TGFβ and TGFβ + RA + VitC iTregs differentiated for 6 days with 100 U/ml IL-2 (*left*) or without IL-2 (*right*). Data are from three biological replicates. P-values are calculated using two-tailed paired Student's t test and error bars show mean ± S.D. **(G)** Geometric MFI of IL2Rγ staining for TGFβ and TGFβ + RA + VitC iTregs differentiated for 6 days with 100 U/ml IL-2 (*left*) or without IL-2 (*right*). Data are from three biological replicates. P-values are calculated using two-tailed paired Student's t test and error bars show mean ± S.D.


A
iTregs differentiated without rhIL2
TGFβ-iTregs
RA+VitC-iTregs
rhIL2
0U/ml
3U/ml
5U/ml
10U/ml
100U/ml
% of Max
pSTAT5-AF647
B
TGFβ-iTregs
RA+VitC-iTregs
MFI for pSTAT5 (X100)
P=0.0009
P=0.01
P=0.0004
P=0.14
rhIL2 (U/ml)
C
Il2ra
Il2rb
Il2rg
FPKM
TGFβ
RA+VitC
D
d6
d5
d4
d3
RA+VitC-iTregs
TGFβ-iTregs
IL2Ra-Pacific Blue
E
TGFβ-iTregs
RA+VitC-iTregs
MFI for IL2Ra (X1000)
P=0.40
P=0.20
P=0.35
P=0.13
F
iTregs-100U/ml IL2
iTregs-without IL2
MFI for IL2Rb (X100)
P=0.015
P=0.35
P=0.027
P=0.14
P=0.0008
P=0.05
P=0.005
P=0.26
G
iTregs-100U/ml IL2
iTregs-without IL2
MFI for IL2Rg (X100)
P=0.04
P=0.0004
P=0.003
P=0.002
P=0.055
P=0.03
P=0.057
P=0.096

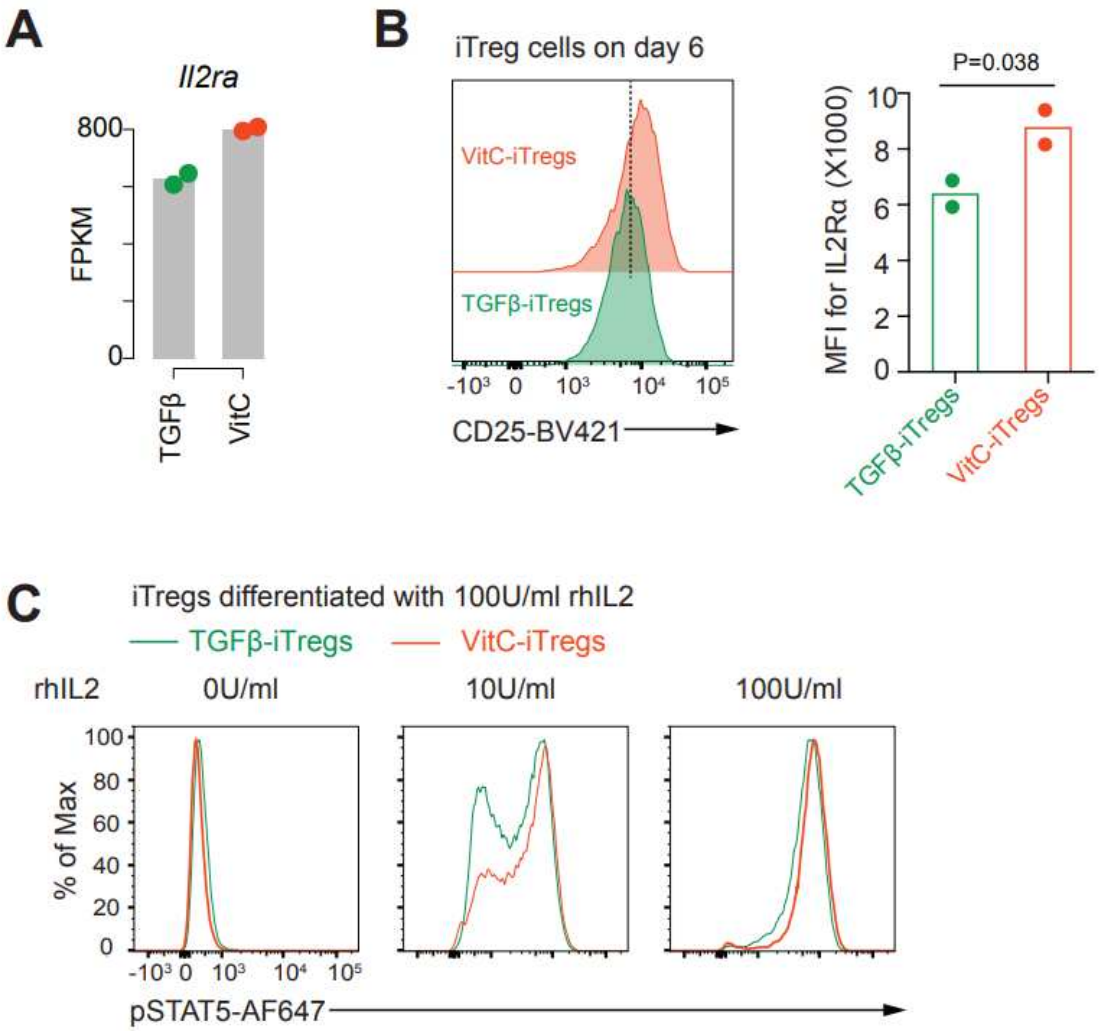


Figure S3.5. Related to Fig 3.8.

(**A**) Expression level of *Il2ra* in TGFβ and TGFβ + Vitamin C iTreg cells by RNA-seq. The Y-axis shows the FPKM (Fragment Per Kilobase per Million reads). (**B**) Representative histogram (*left*) and MFI (*right*) of IL2Rα staining for TGFβ and TGFβ + Vitamin C iTreg cells on day 6 after differentiation. P values were calculated by Student's *t* test. (**C**) iTreg cells differentiated with TGFβ and TGFβ + Vitamin C were rested and restimulated with 0, 10 and 100U/ml IL-2 at 37°C for 1 hour, and stained for phospho-STAT5. Histogram overlay of phospho-STAT5 staining for TGFβ and TGFβ + Vitamin C iTreg cells were shown. Data are representative from two biological replicates.

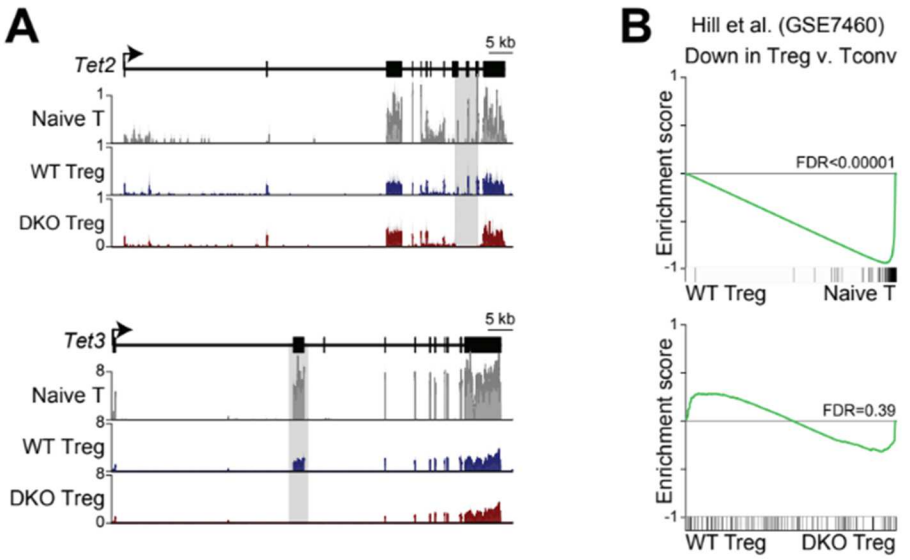


Figure A3.1. Deletion of exons 8-10 of *Tet2* and exon2 of *Tet3*.
(A) Genome browser view of RNA-seq data shows the expression profiles of *Tet2* and *Tet3* genes in CD4+naïve T cells, WT Treg and *Tet2/3* DKO Treg cells. Gray shades highlighting the deleted exons for *Tet2* (exon 8-10) and *Tet3* (exon 2). **(B)** Gene set enrichment analysis of gene expression changesbetween WT Treg cells vs CD4+naïve T (upper panels) and WT Treg cells vs *Tet2/3* DKO Treg cells (lowerpanels) for the Treg signature gene set that are downregulated in Treg cells compared to conventional T cells (GSE7460). The green line represents the enrichment score with enrichment in WT Treg cells shown to the leftand enrichment in CD4+naïve T (upper panels) or *Tet2/3* DKO Treg cells (lower panels) shown to the right. Overlap of genes in the given gene set is shown below and denoted by black ticks.

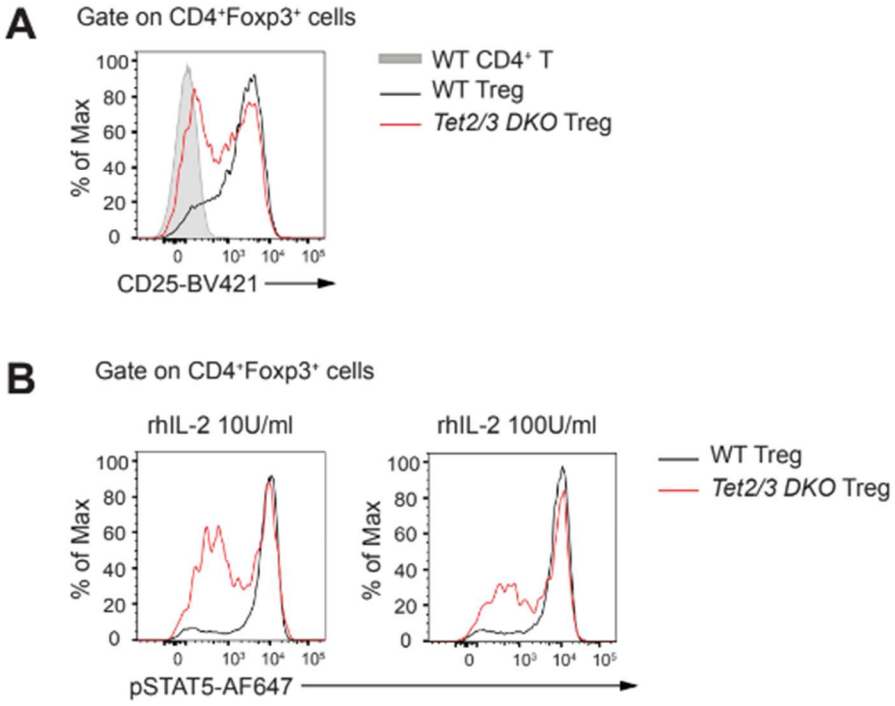


Figure A3.2. Reduccion of IL2Rα in *Tet2/3* DKO Treg cells, when compared to WT Treg cells.

(A) Flow cytometry analysis for the expression of CD25 in WT Treg cells (black line) and *Tet2/3* DKO Treg cells (red line) gated as CD4+ Foxp3+ cells. WT CD4+ Foxp3- T cells (shaded gray) was used asa control. **(B)** CD4+T cells were isolated using dynabeads from WT or *Tet2/3* DKO mice. Then the cells were rested and restimulated with 10 and 100U/ml IL-2 at 37°C for 1 hour, and stained for phospho-STAT5. Histogram overlay of phospho-STAT5 staining for WT Treg cells (black line) and *Tet2/3* DKOTreg cells (red line) gated asCD4+ Foxp3+ cells.

Figure A3.3. 5hmC enrichment profiles of $CD4^+$ naïve T cells and WT Treg cells.
(A) Total number of 5hmC enriched regions determined by CMS-IP data for two replicates of $CD4^+$ naïve T cells and WT Treg cells. **(B)** Regions enriched for 5hmC as determined by CMS-IP. 5hmC enriched regions are sorted by size for IP (top) and input (bottom). **(C)** Pie charts show the genomic distribution of CMS-IP peaks for $CD4^+$ naïve T cells (left) and WT Treg cells (middle). Representation of the annotated regions in the mouse mm9 genome is shown for comparison (right). **(D)** Enrichment of 5hmC containing regions in annotated genomic regions based on their relative abundance over the mouse genome mm9. Statistical analysis was done using Fisher's exact test. **(E)** Enrichment of 5hmC across genes stratified into three categories based on their expression levels for CD4+ naïve T cells (left) and WT Treg cells (right). A heat map of 5hmC is shown on the bottom where genes are sorted by expression level in each cell type independently. TSS: transcription start site; TTS: transcription termination site.

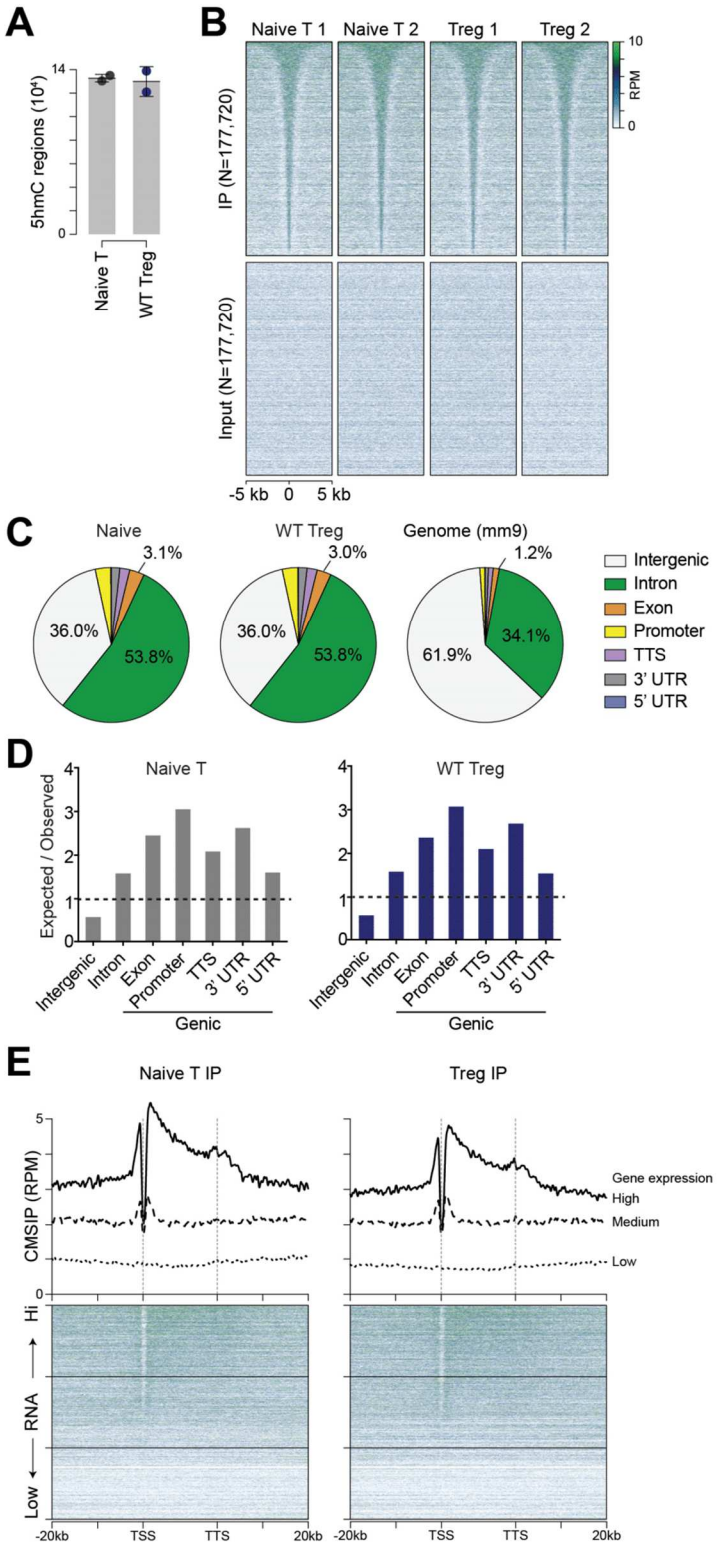
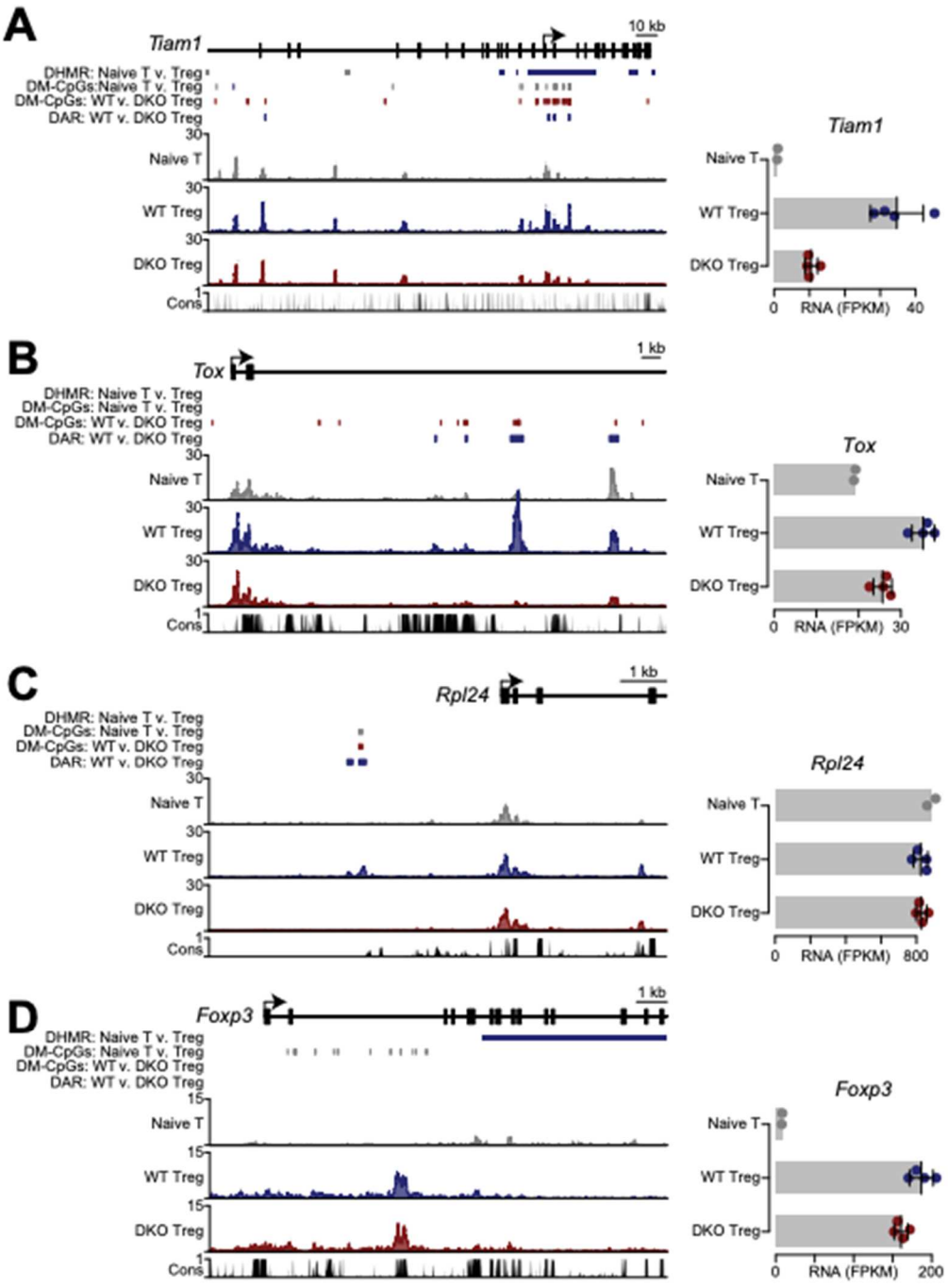

A
5hmC regions (10⁴)
14
0
Naive T
WT Treg
B
Naive T 1
Naive T 2
Treg 1
Treg 2
10
RPM
0
IP (N=177,720)
Input (N=177,720)
-5 kb
0
5 kb
C
Naive
WT Treg
Genome (mm9)
3.1%
3.0%
1.2%
36.0%
53.8%
36.0%
53.8%
61.9%
34.1%
Intergenic
Intron
Exon
Promoter
TTS
3' UTR
5' UTR
D
Naive T
WT Treg
Expected / Observed
Intergenic
Intron
Exon
Promoter
TTS
3' UTR
5' UTR
Genic
E
Naive T IP
Treg IP
CMSIP (RPM)
Gene expression
High
Medium
Low
Hi
RNA
Low
-20kb
TSS
TTS
20kb

Figure A3.4. $CD4^+$ naive T cells accessibility and expression patterns of *Tiam1*, *Tox, Rpl24* and *Foxp3* genes.
(A-D) Left panels: genome browser view of ATAC-seq data shows the accessibility profiles at or near *Tiam1* **(A)**, *Tox* **(B)**, *Rpl24* **(C)** and *Foxp3* **(D)** for $CD4^+$ naive T cells, WT Treg cells and *Tet2/3* DKO Treg cells. DhmRs, DM-CpGs and DARs are shown on top of the tracks. Conservation track (Cons) is shown at the bottom. Right panels: expression levels of *Tiam1* **(A)**, *Tox* **(B)**, *Rpl24* **(C)** and *Foxp3* **(D)** in $CD4^+$ naive T cells, WT Treg cells and *Tet2/3* DKO Treg cells.


A
Tiam1
10 kb
DHMR: Naive T v. Treg
DM-CpGs:Naive T v. Treg
DM-CpGs: WT v. DKO Treg
DAR: WT v. DKO Treg
Naive T
WT Treg
DKO Treg
Cons
Tiam1
RNA (FPKM)
B
Tox
1 kb
DM-CpGs: Naive T v. Treg
Tox
C
Rpl24
Rpl24
D
Foxp3
Foxp3
0
40
30
800
200

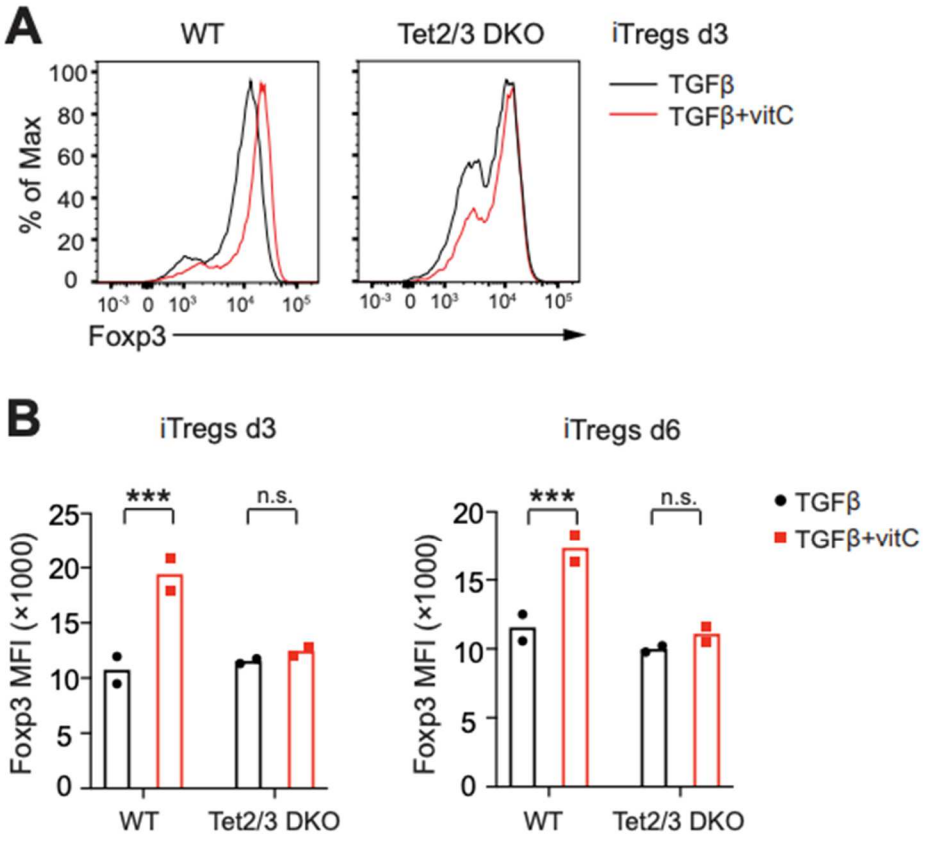


Figure A3.5. Vitamin C treatment increased the MFI of *Foxp3* expression in WT iTreg cells. **(A)** Representative histogram overlay of Foxp3 staining for iTreg cells differentiated for 3 days with TGFβ or TGFβ + Vitamin C from WT and Tet2/3-deficient mice. **(B)** Graphs for the MFI of Foxp3+ cells for iTreg cells differentiated for 3 days (left graph) and 6 days (right graph) with TGFβ or TGFβ + Vitamin C from WT and Tet2/3-deficient mice. *** $P < 0.001$ by Student's t test. n.s., not significant.

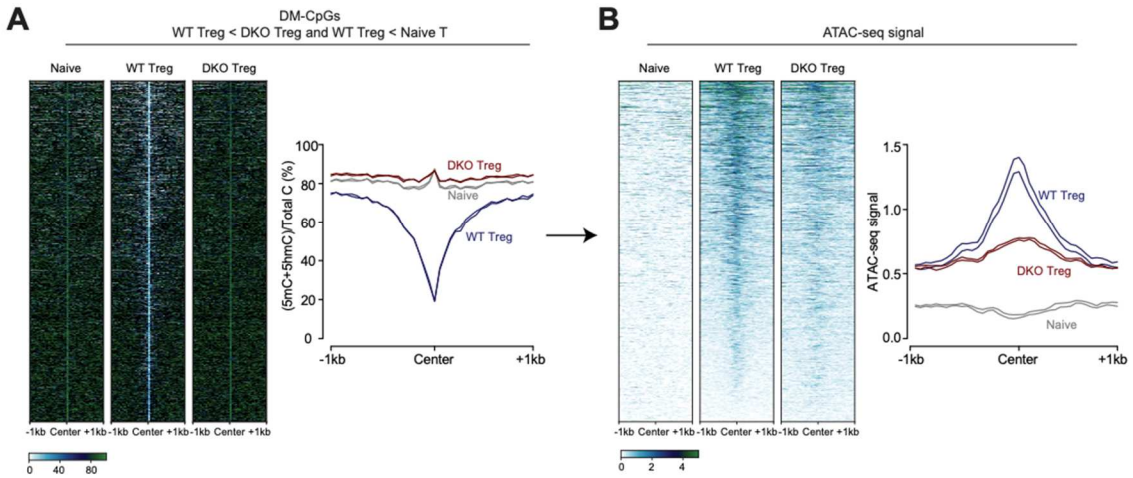


Figure A3.6. ***Tet2/3 DKO*** **iTreg cells regained methylation similar to the observed on naïve CD4$^+$, when compared to WT Treg cells.**

(A) Differentially methylated CpGs (DM-CpGs) that show more demethylation in WT Tregs compare to both CD4+ naïve T cells and Tet2/3 DKO Treg cells. **(B)** chromatin accessibility assessed by ATAC-seq in DM-CpGs shown in **(A)**.

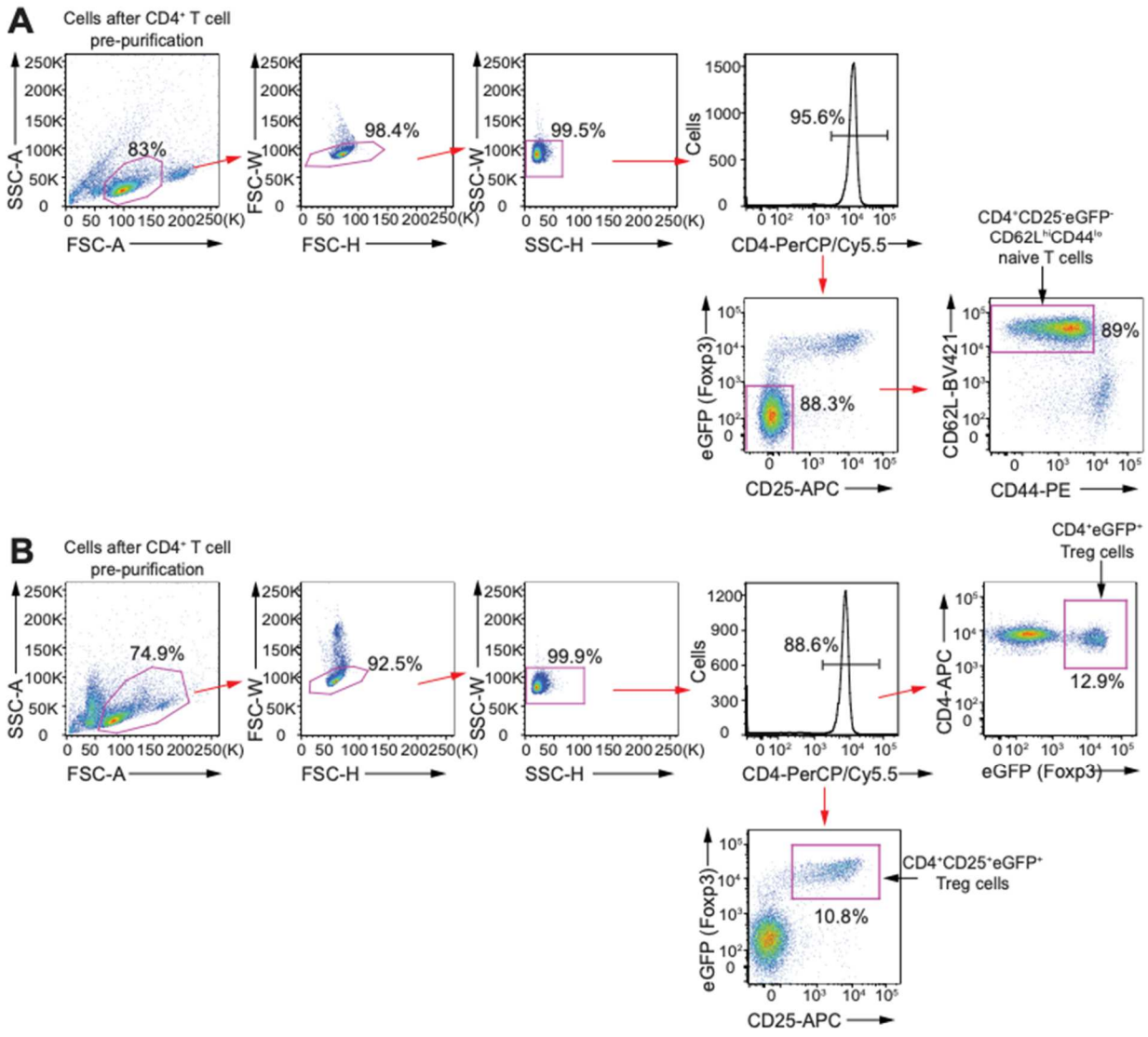


Figure A3.7. Gating strategies used for cell sorting.
(A) Gating strategy to sort CD4+CD25- eGFPCD62LhiCD44lo naive T cells from Foxp3-IRES-eGFP reporter mice. **(B)** Gating strategy to sort CD4+eGFP+ Treg cells for RNA-seq and ATAC-seq or CD4+CD25+eGFP+ Treg cells for WGBS and CMS-IP.

3.8 Author Contributions

A.R. and X.Y. conceived the project. X.Y. and X.L. performed experiments, analyzed and interpreted the results. D.S.C and B.G.B performed the bioinformatics and statistical analyses (RNA-seq, CMS-IP, WGBS and ATAC-seq). E.G.A performed the initial analysis for CMS-IP and WGBS data. B.G.B supervised the bioinformatics analyses. A.R. and X.Y. provided input for the biological aspects of the bioinformatics analyses. A.R. supervised project planning and execution. A.R., X.Y. and B.G.B wrote the manuscript, and all authors proofread the manuscript and provided editorial input.

3.9 Acknowledgements

We thank Dr. C.-W. J. Lio for critical reading of the manuscript and other members of the Rao laboratory for suggestions and discussions. We thank C. Kim, D. Hinz, L. Boggeman, C. Dillingham and M. Haynes at the La Jolla Institute Flow Cytometry facility for help with cell sorting experiments; J. Day of the La Jolla Institute Sequencing facility for help with next-generation sequencing. This work was supported by National Institutes of Health (NIH) R01 grants R35 CA210043 and AI 12858901 (to A.R.). FACSAria II Cell Sorter was acquired through the Shared Instrumentation Grant (SIG) Program S10 RR027366 and Hiseq 2500 was funded by S10OD016262. D.S.C. and E.G.A are supported by the CONACYT/UCMEXUS fellowship. X.L. was supported by a postdoctoral Fellowship from CIRM UCSD Interdisciplinary Stem Cell Research & Training Grant II (TG2-01154). B.G.B is supported by institutional funds, a postdoctoral Fellowship from American Cancer Society (PF-17-109-1-TBG), a Multiple Myeloma Research Foundation Fellowship and an American Society of Hematology Scholar Award.

Chapter 3, in full, is a reprint with modifications as it appears in "Whole-genome analysis of TET dioxygenase function in regulatory T cells", *EMBO reports* (2021). DOI: https://doi.org/10.15252/embr.202152716. The dissertation author was the co-primary investigator and co-first author of this paper. Other authors include Xiaojing Yue, Edahí González-Avalos, Xiang Li, Benjamin Barwick, and Anjana Rao.

3.10 References

Abbas, A.K., Trotta, E., R. Simeonov, D., Marson, A. and Bluestone, J.A., 2018. Revisiting IL-2: Biology and therapeutic prospects. Science immunology, 3(25), p.eaat1482.

Barwick, B.G., Scharer, C.D., Bally, A.P. and Boss, J.M., 2016. Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. Nature immunology, 17(10), pp.1216-1225.

Barwick, B.G., Scharer, C.D., Martinez, R.J., Price, M.J., Wein, A.N., Haines, R.R., Bally, A.P., Kohlmeier, J.E. and Boss, J.M., 2018. B cell activation and plasma cell differentiation are inhibited by de novo DNA methylation. Nature communications, 9(1), pp.1-14.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A. and Lorincz, M.C., 2013. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature, 500(7461), pp.222-226.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature methods, 10(12), pp.1213-1218.

Chinen, T., Kannan, A.K., Levine, A.G., Fan, X., Klein, U., Zheng, Y., Gasteiger, G., Feng, Y., Fontenot, J.D. and Rudensky, A.Y., 2016. An essential role for the IL-2 receptor in T reg cell function. Nature immunology, 17(11), pp.1322-1333.

Dang, L. and Su, S.S.M., 2017. Isocitrate dehydrogenase mutation and (R)-2-hydroxyglutarate: from basic discovery to therapeutics development. Annual review of biochemistry, 86, pp.305-331.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), pp.15-21.

Feng, H., Conneely, K.N. and Wu, H., 2014. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. Nucleic acids research, 42(8), pp.e69-e69.

Feng, Y., Arvey, A., Chinen, T., Van Der Veeken, J., Gasteiger, G. and Rudensky, A.Y., 2014. Control of the inheritance of regulatory T cell identity by a cis element in the Foxp3 locus. Cell, 158(4), pp.749-763.

Floess, S., Freyer, J., Siewert, C., Baron, U., Olek, S., Polansky, J., Schlawe, K., Chang, H.D., Bopp, T., Schmitt, E. and Klein-Hessling, S., 2007. Epigenetic control of the foxp3 locus in regulatory T cells. PLoS biology, 5(2), p.e38.

Fontenot, J.D., Rasmussen, J.P., Gavin, M.A. and Rudensky, A.Y., 2005. A function for interleukin 2 in Foxp3-expressing regulatory T cells. Nature immunology, 6(11), pp.1142-1151.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. and Hornik, K., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome biology, 5(10), pp.1-16.

Grinberg-Bleyer, Y., Oh, H., Desrichard, A., Bhatt, D.M., Caron, R., Chan, T.A., Schmid, R.M., Klein, U., Hayden, M.S. and Ghosh, S., 2017. NF-κB c-Rel is crucial for the regulatory T cell immune checkpoint in cancer. Cell, 170(6), pp.1096-1108.

Hausinger, R.P., 2004. Fe (II)/α-ketoglutarate-dependent hydroxylases and related enzymes. Critical reviews in biochemistry and molecular biology, 39(1), pp.21-68.

Hayatsu, N., Miyao, T., Tachibana, M., Murakami, R., Kimura, A., Kato, T., Kawakami, E., Endo, T.A., Setoguchi, R., Watarai, H. and Nishikawa, T., 2017. Analyses of a mutant Foxp3 allele reveal BATF as a critical transcription factor in the differentiation and accumulation of tissue regulatory T cells. Immunity, 47(2), pp.268-283.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. and Sun, Y., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science, 333(6047), pp.1303-1307.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell, 38(4), pp.576-589.

Hill, J.A., Feuerer, M., Tash, K., Haxhinasto, S., Perez, J., Melamed, R., Mathis, D. and Benoist, C., 2007. Foxp3 transcription-factor-dependent and-independent regulation of the regulatory T cell transcriptional signature. Immunity, 27(5), pp.786-800.

Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R. and Rao, A., 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PloS one, 5(1), p.e8888.

Huang, Y., Pastor, W.A., Zepeda-Martínez, J.A. and Rao, A., 2012. The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. Nature protocols, 7(10), pp.1897-1908.

Huehn, J., Polansky, J.K. and Hamann, A., 2009. Epigenetic control of FOXP3 expression: the key to a stable regulatory T-cell lineage?. Nature Reviews Immunology, 9(2), pp.83-89.

Huijskens, M.J., Wodzig, W.K., Walczak, M., Germeraad, W.T. and Bos, G.M., 2016. Ascorbic acid serum levels are reduced in patients with hematological malignancies. Results in immunology, 6, pp.8-10.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science, 333(6047), pp.1300-1303.

Iyer, L.M., Tahiliani, M., Rao, A. and Aravind, L., 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell cycle, 8(11), pp.1698-1710.

Josefowicz, S.Z., Lu, L.F. and Rudensky, A.Y., 2012. Regulatory T cells: mechanisms of differentiation and function. Annual review of immunology, 30, pp.531-564.

Khaw, K.T., Bingham, S., Welch, A., Luben, R., Wareham, N., Oakes, S. and Day, N., 2001. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. The lancet, 357(9257), pp.657-663.

Krueger, F. and Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. bioinformatics, 27(11), pp.1571-1572.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), pp.357-359.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Sung, K.W.K., Rigoutsos, I., Loring, J. and Wei, C.L., 2010. Dynamic changes in the human methylome during differentiation. Genome research, 20(3), pp.320-331.

Lawrence, M., Gentleman, R. and Carey, V., 2009. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics, 25(14), pp.1841-1842.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. and Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLoS computational biology, 9(8), p.e1003118.

Lee, D.U., Agarwal, S. and Rao, A., 2002. Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene. Immunity, 16(5), pp.649-660.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), pp.2078-2079.

Li, P., Mitra, S., Spolski, R., Oh, J., Liao, W., Tang, Z., Mo, F., Li, X., West, E.E., Gromer, D. and Lin, J.X., 2017. STAT5-mediated chromatin interactions in superenhancers activate IL-2 highly inducible genes: Functional dissection of the Il2ra gene locus. Proceedings of the National Academy of Sciences, 114(46), pp.12111-12119.

Li, X., Liang, Y., LeBlanc, M., Benner, C. and Zheng, Y., 2014. Function of a Foxp3 cis-element in protecting regulatory T cell identity. Cell, 158(4), pp.734-748.

Li, X. and Zheng, Y., 2015. Regulatory T cell identity: formation and maintenance. Trends in immunology, 36(6), pp.344-353.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P

Lio, C.W.J. and Rao, A., 2019. TET enzymes and 5hmC in adaptive and innate immune systems. Frontiers in Immunology, 10, p.210.

Lio, C.W.J., Shukla, V., Samaniego-Castruita, D., González-Avalos, E., Chakraborty, A., Yue, X., Schatz, D.G., Ay, F. and Rao, A., 2019a. TET enzymes augment activation-induced deaminase (AID) expression via 5-hydroxymethylcytosine modifications at the Aicda superenhancer. Science immunology, 4(34), p.eaau7523.

Lio, C.W.J., Yue, X., López-Moyado, I.F., Tahiliani, M., Aravind, L. and Rao, A., 2020. TET methylcytosine oxidases: new insights from a decade of research. Journal of biosciences, 45(1), pp.1-14.

Lio, C.W.J., Yuita, H. and Rao, A., 2019b. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. Blood, 134(18), pp.1487-1497.

Lio, C.W., Zhang, J., González-Avalos, E., Hogan, P.G., Chang, X. and Rao, A., 2016. Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. Elife, 5, p.e18290.

López-Moyado, I.F., Tsagaratou, A., Yuita, H., Seo, H., Delatte, B., Heinz, S., Benner, C. and Rao, A., 2019. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. Proceedings of the National Academy of Sciences, 116(34), pp.16933-16942.

Loria, C.M., Klag, M.J., Caulfield, L.E. and Whelton, P.K., 2000. Vitamin C status and mortality in US adults. The American journal of clinical nutrition, 72(1), pp.139-145.

Luo, C.T., Osmanbeyoglu, H.U., Do, M.H., Bivona, M.R., Toure, A., Kang, D., Xie, Y., Leslie, C.S. and Li, M.O., 2017. Ets transcription factor GABP controls T cell homeostasis and immunity. Nature communications, 8(1), pp.1-12.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M. and Bejerano, G., 2010. GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology, 28(5), pp.495-501.

Mouly, E., Chemin, K., Nguyen, H.V., Chopin, M., Mesnard, L., Leite-de-Moraes, M., Burlen-Defranoux, O., Bandeira, A. and Bories, J.C., 2010. The Ets-1 transcription factor controls the development and function of natural regulatory T cells. Journal of Experimental Medicine, 207(10), pp.2113-2125.

Muthusamy, N., Barton, K. and Leiden, J.M., 1995. Defective activation and survival of T cells lacking the Ets-1 transcription factor. Nature, 377(6550), pp.639-642.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C. and Oliviero, S., 2013. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. Genome biology, 14(8), pp.1-13.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P. and Tahiliani, M., 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature, 473(7347), pp.394-397.

Pastor, W.A., Aravind, L. and Rao, A., 2013. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nature reviews Molecular cell biology, 14(6), pp.341-356.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), pp.139-140.

Ronin, E., Lubrano di Ricco, M., Vallion, R., Divoux, J., Kwon, H.K., Grégoire, S., Collares, D., Rouers, A., Baud, V., Benoist, C. and Salomon, B.L., 2019. The NF-κB RelA transcription factor is critical for regulatory T cell activation and stability. Frontiers in immunology, p.2487.

Ross, S.H. and Cantrell, D.A., 2018. Signaling and function of interleukin-2 in T lymphocytes. Annual review of immunology, 36, pp.411-433.

Roychoudhuri, R., Hirahara, K., Mousavi, K., Clever, D., Klebanoff, C.A., Bonelli, M., Sciumè, G., Zare, H., Vahedi, G., Dema, B. and Yu, Z., 2013. BACH2 represses effector programs to stabilize Treg-mediated immune homeostasis. Nature, 498(7455), pp.506-510.

Sakaguchi, S., Yamaguchi, T., Nomura, T. and Ono, M., 2008. Regulatory T cells and immune tolerance. cell, 133(5), pp.775-787.

Samstein, R.M., Arvey, A., Josefowicz, S.Z., Peng, X., Reynolds, A., Sandstrom, R., Neph, S., Sabo, P., Kim, J.M., Liao, W. and Li, M.O., 2012. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell, 151(1), pp.153-166.

Sardina, J.L., Collombet, S., Tian, T.V., Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M. and Gut, I.G., 2018. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. Cell stem cell, 23(5), pp.727-741.

Nair, V.S., Song, M.H. and Oh, K.I., 2016. Vitamin C facilitates demethylation of the Foxp3 enhancer in a Tet-dependent manner. The journal of immunology, 196(5), pp.2119-2131.

Sekiya, T., Kashiwagi, I., Yoshida, R., Fukaya, T., Morita, R., Kimura, A., Ichinose, H., Metzger, D., Chambon, P. and Yoshimura, A., 2013. Nr4a receptors are essential for thymic regulatory T cell development and immune homeostasis. Nature immunology, 14(3), pp.230-237.

Snow, J.W., Abraham, N., Ma, M.C., Herndier, B.G., Pastuszak, A.W. and Goldsmith, M.A., 2003. Loss of tolerance and autoimmunity affecting multiple organs in STAT5A/5B-deficient mice. The Journal of Immunology, 171(10), pp.5042-5050.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43), pp.15545-15550.

Suzuki, T., Shimizu, Y., Furuhata, E., Maeda, S., Kishima, M., Nishimura, H., Enomoto, S., Hayashizaki, Y. and Suzuki, H., 2017. RUNX1 regulates site specificity of DNA demethylation by recruitment of DNA demethylation machineries in hematopoietic cells. Blood advances, 1(20), pp.1699-1711.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science, 324(5929), pp.930-935.

Toker, A. and Huehn, J., 2011. To be or not to be a Treg cell: lineage decisions controlled by epigenetic mechanisms. Science signaling, 4(158), pp.pe4-pe4.

Toomer, K.H., Lui, J.B., Altman, N.H., Ban, Y., Chen, X. and Malek, T.R., 2019. Essential and non-overlapping IL-2Rα-dependent processes for thymic development and peripheral homeostasis of regulatory T cells. Nature communications, 10(1), pp.1-16.

Tsagaratou, A., Äijö, T., Lio, C.W.J., Yue, X., Huang, Y., Jacobsen, S.E., Lähdesmäki, H. and Rao, A., 2014. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proceedings of the National Academy of Sciences, 111(32), pp.E3306-E3315.

Tsagaratou, A., González-Avalos, E., Rautio, S., Scott-Browne, J.P., Togher, S., Pastor, W.A., Rothenberg, E.V., Chavez, L., Lähdesmäki, H. and Rao, A., 2017a. TET proteins regulate the lineage specification and TCR-mediated expansion of i NKT cells. Nature immunology, 18(1), pp.45-53.

Tsagaratou, A., Lio, C.W.J., Yue, X. and Rao, A., 2017b. TET methylcytosine oxidases in T cell and B cell development and function. Frontiers in immunology, 8, p.220.

Tsagaratou, A. and Rao, A., 2013, January. TET proteins and 5-methylcytosine oxidation in the immune system. In Cold Spring Harbor symposia on quantitative biology (Vol. 78, pp. 1-10). Cold Spring Harbor Laboratory Press.

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H. and Jiao, B., 2015. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Molecular cell, 57(4), pp.662-673.

Wu, X. and Zhang, Y., 2017. TET-mediated active DNA demethylation: mechanism, function and beyond. Nature Reviews Genetics, 18(9), pp.517-534.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A.D., Stroud, J.C., Bates, D.L., Guo, L., Han, A., Ziegler, S.F. and Mathis, D., 2006. FOXP3 controls regulatory T cell function through cooperation with NFAT. Cell, 126(2), pp.375-387.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T. and Liu, L.X., 2011. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer cell, 19(1), pp.17-30.

Yang, R., Qu, C., Zhou, Y., Konkel, J.E., Shi, S., Liu, Y., Chen, C., Liu, S., Liu, D., Chen, Y. and Zandi, E., 2015. Hydrogen sulfide promotes Tet1-and Tet2-mediated Foxp3 demethylation to drive regulatory T cell differentiation and maintain immune homeostasis. Immunity, 43(2), pp.251-263.

Ye, D., Guan, K.L. and Xiong, Y., 2018. Metabolism, activity, and targeting of D-and L-2-hydroxyglutarates. Trends in cancer, 4(2), pp.151-165.

Yoshida, H., Lareau, C.A., Ramirez, R.N., Rose, S.A., Maier, B., Wroblewska, A., Desland, F., Chudnovskiy, A., Mortha, A., Dominguez, C. and Tellier, J., 2019. The cis-regulatory atlas of the mouse immune system. Cell, 176(4), pp.897-912.

Yue, X., Lio, C.W.J., Samaniego-Castruita, D., Li, X. and Rao, A., 2019. Loss of TET2 and TET3 in regulatory T cells unleashes effector function. Nature communications, 10(1), pp.1-14.

Yue, X. and Rao, A., 2020. TET family dioxygenases and the TET activator vitamin C in immune responses and cancer. Blood, 136(12), pp.1394-1401.

Yue, X., Trifari, S., Äijö, T., Tsagaratou, A., Pastor, W.A., Zepeda-Martínez, J.A., Lio, C.W.J., Li, X., Huang, Y., Vijayanand, P. and Lähdesmäki, H., 2016. Control of Foxp3 stability through modulation of TET activity. Journal of Experimental Medicine, 213(3), pp.377-397.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S., 2008. Model-based analysis of ChIP-Seq (MACS). Genome biology, 9(9), pp.1-9.

Zheng, Y., Josefowicz, S., Chaudhry, A., Peng, X.P., Forbush, K. and Rudensky, A.Y., 2010. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature, 463(7282), pp.808-812.

CHAPTER 4: Insights and future directions

In this dissertation we explore the role of TET proteins in B and Treg cells using in vivo and in vitro models. We found that Tet deletion in B or Treg cells is associated with altered gene expression, chromatin structure changes and changes in methylation and hydroximethylation levels. On one hand we analyze the phenotypic consequences of Tet2/3 deletion in mature B cells using a mouse CD19-Cre model, as well as documented relevant and interesting genomic and epigenomic features occurring on B cells. And on the other hand, we thoroughly analyze the genomic and epigenomic changes observed in Tet2/3-deleted Treg cells, using a CD4 Cre system, as well as the enhancement of the activity of Tet proteins in iTreg cells, after treating them with Vitamin C and retinoic acid.

In chapter2, in our CD19 Tet2/3 DKO mice model we found a causal relationship between TET deficiency and the development of mature B cell neoplasms. The fully penetrant progression of B cell lymphomas in CD19 Tet2/3 DKO mice is consistent with the frequent occurrence of TET gene mutations or dysregulation of TET activity in human DLBCL (Reddy *et al.* 2017; Schmitz *et al.* 2018; Chapuy *et al.* 2018). We show that Tet 2/3 deletion leads to spontaneous expansion of B cells with a GC phenotype. We also observe a clear association between TET deficiency, increased G-quadruplexes and R-loops, and increased DNA double strand breaks, particularly at Ig-switch regions in B cells. In this thesis we document for the first time an accumulation of G-quadruplex and R-loop structures is a novel feature of TET deficiency in B cells and cells of other hematopoietic lineages. TET-deficient B cells upregulated mRNAs encoding several proteins that recognize and regulate R-loops and G-quadruplex structures, including RNase H1, Rnaseh2b, DNMT1, Fancd2/FANCD2, ATRX, BLM, WRN, Recql4, and Pif1 (Crossley *et al.* 2019; Skourti-Stathaki *et al.* 2014; Hansel-Hertsch *et al.* 2017; Rhodes *et al.* 2015; Sauer *et al.* 2017). Additional

genetic deletion of one of the G4 binders, DNMT1, in CD19 Tet2/3 DKO B cells was associated with decreased levels of the precursor GC B cells, decreased G-quadruplex and R-loop structures and a notable increase in the survival of Tet2/3-deficient mice.

G-quadruplex structures have been implicated in initiation of the frequent IgH-BCL2 translocations in follicular lymphoma (Nambiar *et al.* 2011; Rabkin *et al.* 2008). We observed an association between increased G-quadruplex and R-loop structures and DNA breaks at the switch regions (primarily switch μ) of Ig isotypes in Tet2/3-deficient B cells. We used HTGTS to map DNA breaks relative to DNA junctions formed by break ligation (with possible errors due to indels), it is likely that we have underestimated the total number of DNA breaks. Nevertheless, we observed an increase in the absolute numbers of DNA DSBs in CD19 DKO B cells when compared with control B cells. The DNA breaks could arise either from conflicts of G-quadruplex and R-loop structures with transcription or DNA replication machineries, or from specific targeting of AID cytidine deaminase, the B cell mutator, to G4 structures and R-loops. In fact, AID possesses a G4-binding activity that is important for its genome-wide targeting (Qiao *et al.* 2017; Xu *et al.* 2020; Yewdell *et al.* 2020). A recent study reported an association between G4 structures and AID activity at commonly mutated genes in B cell lymphoma (Xu *et al.* 2020); consistent with these findings, we observed an enrichment of AID hotspot motifs within translocation sites with increased levels of G-quadruplexes and R-loops. The detailed mechanisms through which G-quadruplexes and R-loops recruit AID and/or promote genomic instability in B cells remain to be addressed. As an initial exploration, one could try to perform chromatin immunoprecitpitation for AID and explore the correlation with the G-quadruplex and R-loops peaks.

TET-dependent methylation changes do not necessarily play a direct biochemical role in the observed increase of R-loop and G quadruplex structures in Tet2/3 deficient B cells. We based

this conclusion on the fact that although we observed a slight but significant link between loss of TET activity and increased DNA methylation in the vicinity of R-loops and G-quadruplex structures, we observed no change in the levels of DNA methylation at the R-loop/G-quadruplex-containing regions (median size ~750 bp) themselves. Thus, further studies are needed to fully understand how TET deficiency in multiple cell types results in increased levels of G-quadruplex and R-loop structures. For example, it will be interesting to explore the possibility that TETs recruit other proteins (i.e. transcription factors) to modulate the formation or resolution of G-quadruplex and R-loops, and that TET-containing protein complexes, rather than TET catalytic activity, play an important role in regulating these structures.

DNMT1 has been previously shown to be induced in GC B cells (Shaknovich *et al.* 2011), where it may be needed to maintain the DNA methylation landscape during rapid GC B cell proliferation. Thus, the apparent upregulation of DNMT1 in TET-deficient B cells is likely due to their prominent GC phenotype compared to WT B cells. Conversely, Dnmt1 deletion results in a dramatic decrease of GC B cells, potentially the primary reason for the delayed oncogenesis observed in triple Tet2/3, Dnmt1-deficient B cells. The decrease in R-loop and G-quadruplex structures observed upon Dnmt1 deletion in total Tet2/3-deficient B cells suggests a possible role for DNMT1 activity in stabilization of these structures. DNMT1 is known to be overexpressed in several different hematological and solid cancers (Zhang & Xu *et al.* 2017). The potential interplay between TET and DNMT activities in regulating oncogenesis as well as R-loop and G-quadruplex levels in B cells and other cell types remains to be investigated.

Our studies suggest that G-quadruplexes and R-loops could be therapeutic vulnerabilities in cancers with TET loss-of-function. G4-stabilizing ligands were recently shown to decrease cell viability in ATRX-deficient gliomas and BRCA1/2-deficient tumor cells (Wang *et al.* 2019; Xu *et*

al. 2017). In our hands, the use of a G4-stabilizing ligand, or depletion of proteins known to regulate G4 or R-loop structures, was associated with increased DNA DSBs and a slight increase in apoptosis in TET-deficient B cells. Furthermore, deletion of DNMT1 in TET-deficient B cells prevented the accumulation of R-loop and G-quadruplex structures in splenic B cells and Peyer's patch GC B cells and rescued the survival of TET-deficient mice. Follow-up studies in pre-clinical models could test whether a combination of G4-stabilizing agents and DNA methyltransferase inhibitors might synergize to delay the onset and/or progression of B cell lymphomas and other malignancies with TET loss-of function.

In our second model, we used TET loss-of-function in endogenous Treg cells and TET gain-of-activity in TET activator Vitamin C treated iTreg cells in vitro. Vitamin C functions through TET proteins to mediate CNS2 demethylation and increase of Foxp3 MFI, as these effects are abolished in TET-deficient iTreg cells treated with Vitamin C (Yue *et al.* 2016). In this study, we have extended these results to a genome-wide level, by investigating for the first time the transcriptional profiles, chromatin accessibility landscapes, and DNA methylation/ hydroxymethylation changes in endogenous WT and TET-deficient Treg cells and in TGF-induced iTreg cells after addition of Vitamin C. We show that TET enzymes promote the acquisition of the molecular programs of gene expression, chromatin accessibility and DNA methylation that are characteristic of Treg cells in vivo; moreover, treatment of differentiating iTreg cells with the TET activator Vitamin C in vitro mimics these effects, albeit to a lesser extent.

Our data also reveal an intriguing connection between TET activity, Vitamin C and responses to IL-2, a key cytokine critical for Treg development, differentiation and function (Abbas *et al.* 2018; Ross & Cantrell, 2018). IL-2 is well-established as essential for Treg function: mice with germline deletion of IL2, IL2ra and IL2rb develop severe autoimmune diseases

(Fontenot *et al.* 2005) and mice with Treg-specific ablation of IL2ra and IL2rb exhibit more severe lethal autoimmunity than mice with germline deletion of the same genes, exhibiting phenotypes similar to those of Foxp3-deficient scurfy mice (Chinen *et al.* 2016; Toomer *et al.* 2019). Likewise, STAT5a/b-deficient mice fail to generate Treg cells and also develop fatal autoimmunity (Snow *et al.* 2003). Thus, the IL2-IL2R-STAT5 signaling axis is essential for the functional programming of Treg cells and maintenance of immune homeostasis (Ross & Cantrell, 2018). We observed that TET deletion in vivo resulted in impaired IL2/STAT5 signaling, with the IL2/STAT5 gene set being underrepresented in Tet2/3 DKO Treg cells compared to WT Treg cells. Conversely, we found that STAT motifs were enriched in differentially-accessible regions induced by Vitamin C, and that iTregs induced in the combined presence of TGF, RA and Vitamin C survived better in low IL-2 media and were more responsive to stimulation with low IL-2 concentrations than iTregs induced in the presence of TGF alone. This feature was associated with increased levels of phospho-STAT5, increased occupancy of STAT5 at key Foxp3 and Il2ra enhancers and increased IL2R mRNA and protein levels.

TET proteins are increasingly recognized to be responsive to metabolic influences including those arising from diet and nutrition, and TET and Vitamin C deficiencies are both well-established as a risk factor for blood cancers (Huijskens *et al.* 2016; Khaw *et al.* 2001; Lio *et al.* 2019b; Loria *et al.* 2000). Our data emphasize that, in addition, low plasma Vitamin C levels may also predispose individuals to autoimmune disease secondary to decreased function of both thymically-derived and peripherally-generated Tregs. Vitamin C deficiency is expected to reduce the function of all Fe(II) and 2-oxoglutarate-dependent dioxygenases including TET enzymes, and to result in suboptimal function of both thymically-derived and peripherally-generated Tregs. It would be interesting to conduct an epidemiological study of the relation between plasma Vitamin

C levels and autoimmune disease in humans, similar to previous studies relating plasma Vitamin C levels to overall mortality and cancer (Huijskens *et al.* 2016; Khaw *et al.* 2001; Loria *et al.* 2000).

In summary, here we demonstrate that TET loss of function alteres transcriptional regulation in immune cell development and activation. However, further studies will be necessary to elucidate the mechanisms involved of TET proteins and cellular transformation.

4.1 References

Abbas, A.K., Trotta, E., R. Simeonov, D., Marson, A. and Bluestone, J.A., 2018. Revisiting IL-2: Biology and therapeutic prospects. Science immunology, 3(25), p.eaat1482.

Chapuy, B., Stewart, C., Dunford, A.J., Kim, J., Kamburov, A., Redd, R.A., Lawrence, M.S., Roemer, M.G., Li, A.J., Ziepert, M. and Staiger, A.M., (2018) Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nature medicine, 24(5), pp.679-690.

Chinen, T., Kannan, A.K., Levine, A.G., Fan, X., Klein, U., Zheng, Y., Gasteiger, G., Feng, Y., Fontenot, J.D. and Rudensky, A.Y., 2016. An essential role for the IL-2 receptor in T reg cell function. Nature immunology, 17(11), pp.1322-1333.

Crossley, M.P., Bocek, M. and Cimprich, K.A. (2019) R-loops as cellular regulators and genomic threats. Molecular cell, 73(3), pp.398-411.

Fontenot, J.D., Rasmussen, J.P., Gavin, M.A. and Rudensky, A.Y., 2005. A function for interleukin 2 in Foxp3-expressing regulatory T cells. Nature immunology, 6(11), pp.1142-1151.

Hänsel-Hertsch, R., Di Antonio, M. and Balasubramanian, S., (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. Nature reviews Molecular cell biology, 18(5), pp.279-284.

Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W. and Frock, R.L., (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification–mediated high-throughput genome-wide translocation sequencing. Nature protocols, 11(5), pp.853-871.

Huijskens, M.J., Wodzig, W.K., Walczak, M., Germeraad, W.T. and Bos, G.M., 2016. Ascorbic acid serum levels are reduced in patients with hematological malignancies. Results in immunology, 6, pp.8-10.

Khaw, K.T., Bingham, S., Welch, A., Luben, R., Wareham, N., Oakes, S. and Day, N., 2001. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. The lancet, 357(9257), pp.657-663.

Lio, C.W.J., Yuita, H. and Rao, A., 2019. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. Blood, 134(18), pp.1487-1497.

Loria, C.M., Klag, M.J., Caulfield, L.E. and Whelton, P.K., 2000. Vitamin C status and mortality in US adults. The American journal of clinical nutrition, 72(1), pp.139-145.

Nambiar, M., Goldsmith, G., Moorthy, B.T., Lieber, M.R., Joshi, M.V., Choudhary, B., Hosur, R.V. and Raghavan, S.C., 2011. Formation of a G-quadruplex at the BCL2 major breakpoint region of the t (14; 18) translocation in follicular lymphoma. Nucleic acids research, 39(3), pp.936-948.

Qiao, Q., Wang, L., Meng, F.L., Hwang, J.K., Alt, F.W. and Wu, H., 2017. AID recognizes structured DNA for class switch recombination. Molecular cell, 67(3), pp.361-373.

Rabkin, C.S., Hirt, C., Janz, S. and Dölken, G., 2008. t (14; 18) Translocations and risk of follicular lymphoma. Journal of the National Cancer Institute Monographs, 2008(39), pp.48-51.

Reddy, A., Zhang, J., Davis, N.S., Moffitt, A.B., Love, C.L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L., Karjalainen-Lindsberg, M.L. and Nørgaard, P., 2017. Genetic and functional drivers of diffuse large B cell lymphoma. Cell, 171(2), pp.481-494.

Rhodes, D. and Lipps, H.J., 2015. G-quadruplexes and their regulatory roles in biology. Nucleic acids research, 43(18), pp.8627-8637.

Ross, S.H. and Cantrell, D.A., 2018. Signaling and function of interleukin-2 in T lymphocytes. Annual review of immunology, 36, pp.411-433.

Sauer, M. and Paeschke, K., 2017. G-quadruplex unwinding helicases and their function in vivo. Biochemical Society Transactions, 45(5), pp.1173-1182.

Schmitz, R., Wright, G.W., Huang, D.W., Johnson, C.A., Phelan, J.D., Wang, J.Q., Roulland, S., Kasbekar, M., Young, R.M., Shaffer, A.L. and Hodson, D.J., 2018. Genetics and pathogenesis of diffuse large B-cell lymphoma. New England Journal of Medicine, 378(15), pp.1396-1407.

Shaknovich, R., Cerchietti, L., Tsikitas, L., Kormaksson, M., De, S., Figueroa, M.E., Ballon, G., Yang, S.N., Weinhold, N., Reimers, M. and Clozel, T., 2011. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood, The Journal of the American Society of Hematology, 118(13), pp.3559-3569.

Skourti-Stathaki, K. and Proudfoot, N.J., 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & development, 28(13), pp.1384-1396.

Snow, J.W., Abraham, N., Ma, M.C., Herndier, B.G., Pastuszak, A.W. and Goldsmith, M.A., 2003. Loss of tolerance and autoimmunity affecting multiple organs in STAT5A/5B-deficient mice. The Journal of Immunology, 171(10), pp.5042-5050.

Tanaka, S., Ise, W., Inoue, T., Ito, A., Ono, C., Shima, Y., Sakakibara, S., Nakayama, M., Fujii, K., Miura, I. and Sharif, J., 2020. Tet2 and Tet3 in B cells are required to repress CD86 and prevent autoimmunity. Nature immunology, 21(8), pp.950-961.

Toomer, K.H., Lui, J.B., Altman, N.H., Ban, Y., Chen, X. and Malek, T.R., 2019. Essential and non-overlapping IL-2Rα-dependent processes for thymic development and peripheral homeostasis of regulatory T cells. Nature communications, 10(1), pp.1-16.

Wang, Y., Yang, J., Wild, A.T., Wu, W.H., Shah, R., Danussi, C., Riggins, G.J., Kannan, K., Sulman, E.P., Chan, T.A. and Huse, J.T., 2019. G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. Nature communications, 10(1), pp.1-14.

Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Dos Santos, N., Silvester, J., Wei, V., Garcia, J. and Kabeer, F., 2017. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. Nature communications, 8(1), pp.1-18.

Xu, Y.Z., Jenjaroenpun, P., Wongsurawat, T., Byrum, S.D., Shponka, V., Tannahill, D., Chavez, E.A., Hung, S.S., Steidl, C., Balasubramanian, S. and Rimsza, L.M., 2020. Activation-induced cytidine deaminase localizes to G-quadruplex motifs at mutation hotspots in lymphoma. NAR cancer, 2(4), p.zcaa029.

Yewdell, W.T., Kim, Y., Chowdhury, P., Lau, C.M., Smolkin, R.M., Belcheva, K.T., Fernandez, K.C., Cols, M., Yen, W.F., Vaidyanathan, B. and Angeletti, D., 2020. A Hyper-IgM Syndrome Mutation in Activation-Induced Cytidine Deaminase Disrupts G-Quadruplex Binding and Genome-wide Chromatin Localization. Immunity, 53(5), pp.952-970.

Young, R.M., Wu, T., Schmitz, R., Dawood, M., Xiao, W., Phelan, J.D., Xu, W., Menard, L., Meffre, E., Chan, W.C.C. and Jaffe, E.S., 2015. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proceedings of the National Academy of Sciences, 112(44), pp.13447-13454.

Yue, X., Trifari, S., Äijö, T., Tsagaratou, A., Pastor, W.A., Zepeda-Martínez, J.A., Lio, C.W.J., Li, X., Huang, Y., Vijayanand, P. and Lähdesmäki, H., 2016. Control of Foxp3 stability through modulation of TET activity. Journal of Experimental Medicine, 213(3), pp.377-397.

Zhang, W. and Xu, J., 2017. DNA methyltransferases and their roles in tumorigenesis. Biomarker research, 5(1), pp.1-8.